Auditory Plasticity –

Listening with the Brain

Illustration by Wet Designer Dog (www.wetdesignerdog.dk)

www.isaar.eu

The Danavox Jubilee Foundation

ISBN 978-87-990013-4-7 EAN 9788799001347

The Danavox Jubilee Foundation
c/o GN ReSound
Lautrupbjerg 7
DK-2750 Ballerup
Denmark

Layout: Graphic Mind
Printed by Centertryk A/S 2014

Preface

Organizing committee, ISAAR 2013

Scientific

Torsten Dau, Technical University of Denmark, Kgs. Lyngby, Denmark

Jakob Christensen-Dalsgaard, University of Southern Denmark, Odense, Denmark

Lisbeth Tranebjærg, University of Copenhagen, Copenhagen, Denmark

Ture Andersen, Odense University Hospital, Odense, Denmark

Administrative

Torben Poulsen, Technical University of Denmark, Kgs. Lyngby, Denmark

Caroline van Oosterhout, Technical University of Denmark, Kgs. Lyngby, Denmark

Abstract, programme, and manuscript coordinator – Webmaster

Sébastien Santurette, Technical University of Denmark, Kgs. Lyngby, Denmark

About ISAAR

The “International Symposium on Auditory and Audiological Research” is formerly known as the “Danavox Symposium”. The 2013 edition was the 25th symposium in the series and the 4th symposium under the ISAAR name, adopted in 2007. The Danavox Jubilee Foundation was established in 1968 on the occasion of the 25th anniversary of GN Danavox. The aim of the foundation is to support and encourage audiological research and development.

Funds are donated by GN ReSound (formerly GN Danavox) and are managed by a board consisting of hearing science specialists who are entirely independent of GN ReSound. Since its establishment in 1968, the resources of the foundation have been used to support a series of symposia, at which a large number of outstanding scientists from all over the world have given lectures, presented posters, and participated in discussions on various audiological topics.

A list of proceedings from previous symposia may be found at the ISAAR website: www.isaar.eu – ‘Previous Symposia’. At this page there is a link to the GN ReSound Audiological Library from where all contributions can be searched and downloaded.

Contents

I: Basic perceptual studies of training, learning, and generalization

II: Physiological correlates and modeling of auditory plasticity

III: Plasticity and auditory disorders

IV: Hearing rehabilitation with hearing aids and cochlear implants

V: Design and evaluation of hearing-aid signal processing

VI: New processing and fitting strategies in cochlear implants

VII: Hearing loss assessment and characterization

VIII: Factors influencing speech intelligibility

IX: Audiovisual interactions and music-listening habits

Addendum to ISAAR 2011: Speech Perception and Auditory Disorders

Induction of auditory perceptual learning

BEVERLY A. WRIGHT*

Department of Communication Sciences and Disorders, Knowles Hearing Center, Northwestern University, 2240 Campus Drive, Evanston, Illinois, 60208, USA

Performance on many perceptual tasks improves with practice even in adults, indicating that our sensory systems are not rigid but rather can be changed through experience. My co-workers and I have been investigating the factors that induce perceptual learning on auditory skills. We have evidence that two key requirements for perceptual improvement to occur across days are performance of the task to be learned and a sufficient amount of training per day. Beyond these core requirements, we also have documented that perceptual training can be made more efficient by not exceeding the required amount of daily training and by replacing a subset of the training trials with stimulus exposure alone. The elements of successful training regimens provide insights into perceptual-learning mechanisms. A greater knowledge of these mechanisms will lead to more effective training strategies to help restore perceptual skills in people with perceptual disorders as well as to enhance those skills in people with normal perception.

INTRODUCTION

Perceptual abilities improve with practice. This plasticity is of practical value because it provides an avenue for treating perceptual disorders as well as for enhancing normal perceptual skills. It is of scientific importance because it indicates that theories of perceptual processing must incorporate malleability.

My co-workers and I have been investigating the induction of perceptual learning in audition to gain a greater understanding of the kinetics and mechanisms of perceptual improvement. In these experiments we have focused on how a variety of multiple-day training regimens affect basic auditory skills in human adults. We chose to examine multiple-day as opposed to single-day regimens because improvement across days indicates that learning has moved to long-term memory (consolidated; McGaugh, 2000), performance across days is not necessarily predicted by performance within a day (Mednick *et al.*, 2002; Huyck and Wright, 2011), and the learning magnitude across days is typically greater than within a day. Our choice to evaluate learning on basic skills is based on the assumption that, at the physiological level, the general factors that trigger learning-related change are similar across a wide range of task and stimulus complexity. The particular neural circuits that are affected may differ, but the circumstances that lead them to change are largely the same. Though not described here, we have preliminary data

*Corresponding author: b-wright@northwestern.edu

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

suggesting that similar circumstances induce learning on both fine-grained auditory discrimination tasks and speech tasks, consistent with this assumption.

Here we summarize the results of these investigations, placed in the context of their implications for how best to elicit perceptual improvement. We suggest that effective and efficient perceptual training regimens include performance of the task to be learned and a sufficient, but not additional, number of training trials per day. We then show that a portion of the necessary daily training trials can be replaced with stimulus exposures without practice, providing a means to reduce the overall practice required to obtain improvement. We conclude with a brief discussion of what these training requirements suggest about learning mechanisms.

TASKS AND GENERAL PARADIGM

In all of the experiments featured in the following sections, the task was either frequency discrimination (Fig. 1A, left) or temporal-interval discrimination (Fig. 1A, right). In each two-presentation forced-choice trial a standard stimulus was presented in one randomly selected presentation and a signal stimulus in the other. The standard stimulus (filled horizontal bars) was the same for both tasks: two 15-ms, 1-kHz tones separated by a temporal interval of 100 ms. The signal stimulus (open horizontal bars) had a lower frequency than the standard in the frequency task and a longer temporal interval than the standard in the temporal-interval task. Discrimination thresholds were estimated using a three-down, one-up adaptive tracking procedure that yields the 79.4% correct point on the psychometric function (Levitt, 1971).

Each experiment consisted of a pre-training test, a training phase, and a post-training test. Trained listeners participated in all three segments. Controls participated only in the pre- and post-training tests, with no training in between. The time between the pre- and post-training tests was similar for the trained listeners and controls.

KEY TRAINING REQUIREMENTS

Practice on the task to be learned

One well-established requirement for learning on most perceptual tasks is practice on the task to be learned. The importance of active task performance is demonstrated primarily by two lines of evidence. First, learning resulting from performing one task rarely transfers to a different task even when both tasks are performed with the same standard stimulus. This lack of task transfer has been demonstrated repeatedly in the visual system. For just one example, observers who practiced either a local or a global visual orientation discrimination task with the same stimuli improved on the task on which they were trained, but did not transfer their learning to the other task (Ahissar and Hochstein, 1993). Figure 1 shows a similar outcome in the auditory domain. We trained two groups of adults 900 trials per day for 10-11 days on either a frequency-discrimination task or a temporal-interval discrimination task, using the same standard stimulus for both tasks (Fig. 1A), and then tested both groups on the

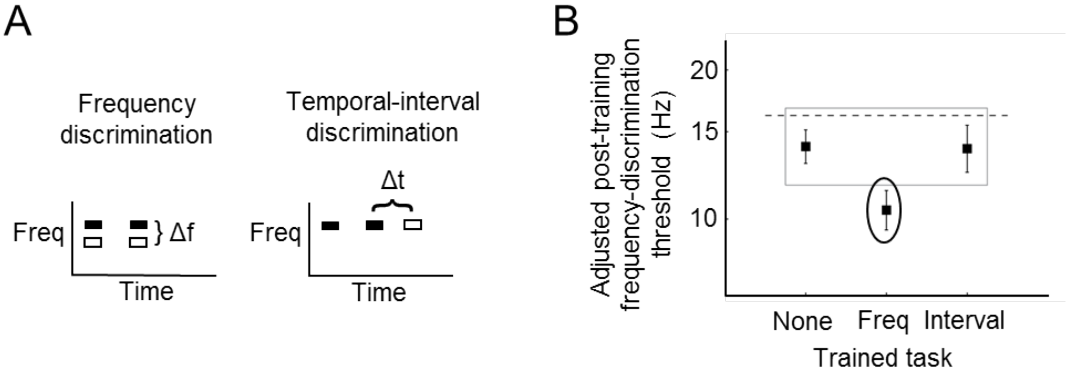


Fig. 1: Practice on the task to be learned. (A) Schematic diagrams of the frequency-discrimination (left) and temporal-interval discrimination (right) tasks. The standard stimulus was the same for both tasks (filled horizontal bars; two 15-ms 1-kHz tones separated by 100 ms), but the signal stimulus (open horizontal bars) had a lower frequency in the frequency task and a longer temporal interval in the temporal-interval task. The procedure was two-presentation forced-choice. (B) Mean post-training thresholds (filled squares; 79.4% correct detections) on the frequency-discrimination task following either no training (None), or 900 training trials per day for 10-11 days on frequency discrimination (Freq) or temporal-interval discrimination (Interval) (n = 6-10 per group). The post-training thresholds were adjusted to take into account individual differences in pre-training threshold (equation in Cohen, 1988). Also shown are the mean pre-training threshold across all listeners (dashed line), and the 95% confidence interval around the mean post-training threshold for the control group who participated in the pre- and post-training tests but received no training in between (None; gray box). Error bars indicate +/− 1 standard error of the mean. The post-training threshold that differed significantly from that of controls ($p < 0.05$) is marked (black circle). [Data from Wright and Sabin (2007) and Wright *et al.* (2010).]

frequency-discrimination task (Wright *et al.*, 2010). The frequency-trained group improved on the frequency task over the course of training and had lower frequency-discrimination thresholds at a post-training test than did the control group (Fig. 1B). In contrast, though the temporal-interval trained listeners improved on their trained task, their post-training thresholds on the frequency task did not differ from those of controls. Thus, learning did not transfer from temporal-interval discrimination to frequency discrimination. In another auditory case, learning did not transfer in either direction between the tasks of temporal-order discrimination and asynchrony detection at sound onset (Mossbridge *et al.*, 2006). If improvements were driven solely by stimulus exposure, learning should transfer between tasks. Second, cortical changes that have been observed to accompany perceptual learning either do not occur or are substantially reduced when the stimulus exposures are not linked with active performance of a task. For example, temporal resolution in primary auditory cortex improved in a group of rats trained to use an auditory temporal cue to locate food, but not in another group that were presented with the same sounds in a non-contingent manner (Bao *et al.*, 2004). Combining both lines of evidence, rats trained to perform one or another of two basic auditory tasks with the same stimuli showed behavioural improvement and corresponding cortical reorganization that was specific to the stimulus feature relevant to the task on which they were trained (Polley *et al.*, 2006).

Sufficient practice per day

Another apparent requirement for perceptual improvement across days is a sufficient, and sometimes substantial, amount of training per day. We observed this phenomenon on an auditory frequency-discrimination task, as shown in Fig. 2 (Wright and Sabin, 2007). We trained two groups of adults either 360 or 900 trials per day for 6 days on a frequency-discrimination task (Fig. 2A). The 900-trial-per-day group improved over the course of training and had lower frequency-discrimination thresholds at a post-training test than did the control group (Fig. 2B; data from Fig. 1B). In contrast, the 360-trial-per-day group showed no improvement on the frequency task over the course of training, and their post-training thresholds did not differ from those of controls. These conclusions held both when the total number of training days was held constant at 6 and when the total number of trials was held constant across the two groups. Thus, learning on this particular frequency-discrimination task required sufficient training per day. The need for sufficient training per day to yield learning across days has also been reported for a visual chevron-discrimination task (Aberg *et al.*, 2009) and a letter-enumeration task (Hauptmann and Karni, 2002).

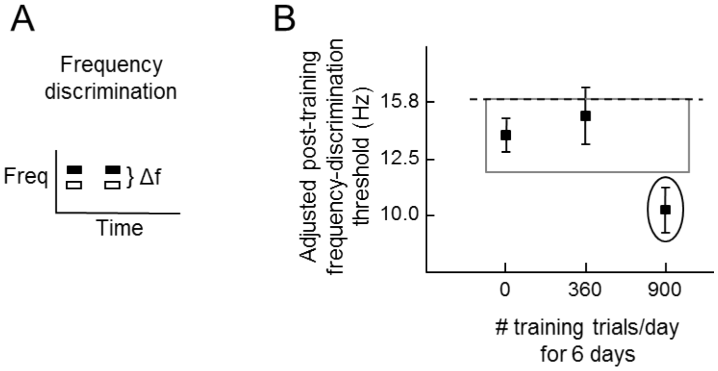


Fig. 2: Sufficient practice per day. (A) As in Fig. 1A. (B) Mean post-training thresholds on the frequency-discrimination task following 0 (no training), 360, or 900 training trials per day for 6 days on that task (n = 7-10 per group). Otherwise, as in Fig. 1B. [Data from Wright and Sabin (2007) and Wright *et al.* (2010).]

It is important to note that the sufficient amount of daily training required for learning can differ across tasks, and even across stimuli for the same task. We trained two groups of adults 360 trials per day for 6 days on either frequency discrimination or temporal-interval discrimination, using the same standard stimulus for both tasks (Wright and Sabin, 2007). The frequency-trained group did not improve on frequency discrimination, as illustrated in Fig. 2B (360 trials/day), but the temporal-interval trained group did improve on temporal-interval discrimination (not shown). However, sufficient daily training still seems necessary for learning on temporal-interval discrimination, because 50 training trials per day for 20 days yielded no improvement on that task (Rammsayer, 1994). Thus, the sufficient amount of daily training required for learning can differ across tasks, even when the standard stimulus is the same. Likewise, listeners who practiced ~360 training trials per day over multiple days on frequency discrimination improved on that task when the standard stimulus was a 300-ms, 1-kHz tone (Roth *et al.* 2003), but not when it was two brief 1-kHz tones separated by 100 ms (Fig. 2B). Thus, the sufficient amount of daily training required for learning can differ across stimuli, even when the task is the same.

Enough is enough

While perceptual learning across days appears to depend on sufficient training on each day, additional training beyond that amount can be superfluous. For example, we trained two groups of adults either 360 or 900 trials per day for 6 days on an auditory temporal-interval discrimination task (Wright and Sabin, 2007). Their learning curves essentially overlapped. Similar outcomes have been reported in investigations of learning on other tasks including auditory interaural-time-difference discrimination (Ortiz and Wright, 2010), visual chevron discrimination (Aberg *et al.*, 2009), and motor sequencing (Savion-Lemieux and Penhune, 2005). Thus, more daily training does not necessarily lead to greater improvement across days.

AN ALTERNATIVE ROUTE

Task practice plus additional stimulus exposure without practice

As described above, two core requirements for perceptual learning across days appear to be task performance and sufficient training trials per day. Here we show that these requirements can be met more efficiently through the combination of periods of practice and periods of additional stimulus exposure without practice (Wright *et al.*, 2010). This phenomenon is illustrated in Fig. 3, which shows thresholds on a frequency-discrimination task (Fig. 3A) at a post-training test for six different groups of adults (Fig. 3B): five groups who participated in different training regimens for 6-11 days, and a control group.

We trained two groups – Freq+Silence and All-Interval – using regimens that did not meet the requirements for learning established for this frequency-discrimination task. The Freq+Silence group practiced the frequency-discrimination task for 360 trials per day with the trials distributed in three bouts of 120 trials. The bouts were separated by ~6 minutes of silence during which the listeners completed a written symbol-to-number matching task. The post-training thresholds for this group were no better than those for controls, replicating the previous demonstration that 360 training trials per day are not sufficient to induce learning on this task (see Fig. 2B). The All-Interval group practiced 900 trials per day on a temporal-interval discrimination task using the same standard stimulus as in the frequency-discrimination task (Fig. 3A). As described above, this group did not transfer their learning from the temporal-interval to the frequency-discrimination task (data from Fig. 1B), demonstrating the need for performance of the task to be learned to induce improvement.

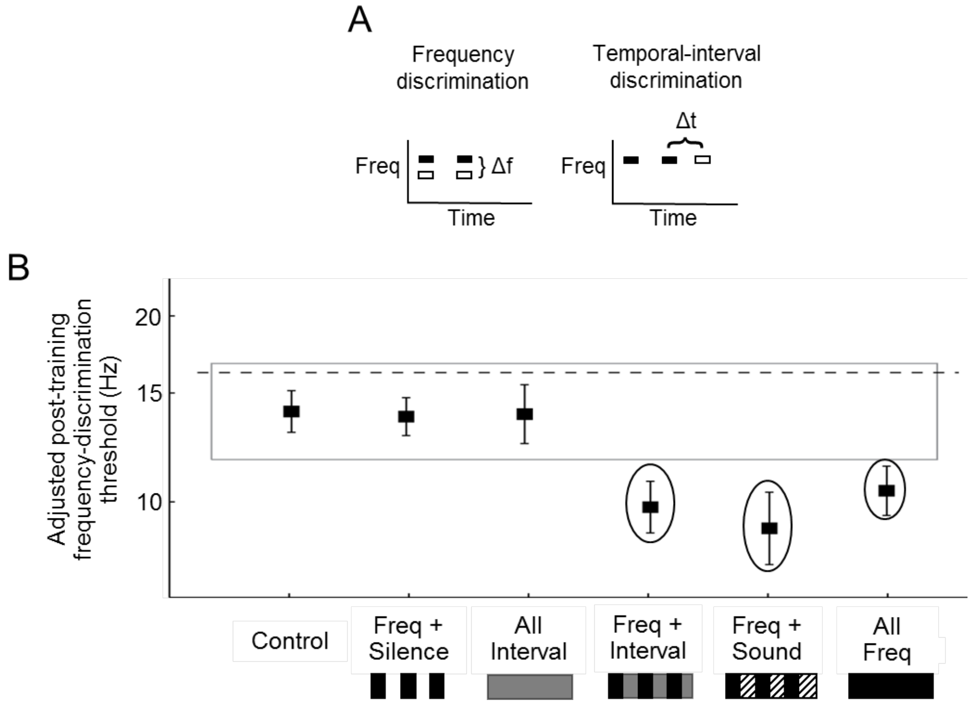


Fig. 3: Task practice plus additional stimulus exposure without practice. (A) As in Fig. 1A. (B) Mean post-training thresholds on the frequency-discrimination task following either no training (control), or one of five 6-11 day training regimens (n = 6-10 per group). See text for details. Otherwise, as in Fig. 1B. [Data from Wright and Sabin (2007) and Wright *et al.* (2010).]

We then combined variants of these two unsuccessful regimens to train two additional groups – Freq+Interval and Freq+Sound. Both combinations were successful. The Freq+Interval group practiced frequency discrimination for 360 trials per day and temporal-interval discrimination for 360 trials per day, alternating between the two tasks every 120 trials. The Freq+Sound group practiced frequency discrimination for 360 trials per day and also were exposed to, but did not perform, 360 trials per day of the temporal-interval discrimination task, alternating between the two tasks every 120 trials; the temporal-interval trials were presented in the background as the listeners completed a written symbol-to-number matching task. The post-training thresholds for these two groups were better than those for controls, and were similar to those for the All-Freq group (data from Fig. 1B) who practiced 900 trials per day on the frequency-discrimination task. Thus, though improvement on this task required task practice and sufficient daily training trials, task practice was not required throughout the entire training period. A portion of the practice trials could be replaced with additional stimulus exposures delivered either through performance of a different task or as background sounds.

In additional experiments we examined the influence of the temporal separation of the periods of task practice and additional stimulus exposure. Figure 4 shows the frequency-discrimination thresholds at a post-training test for four groups who participated in different 6-7 day training regimens and for a control group. One trained group practiced only frequency discrimination for 360 training trials per day (Short Freq). The other three trained groups practiced frequency discrimination followed by temporal-interval discrimination, each for 360 training trials per day, using the same standard stimulus for both tasks. The temporal separation between

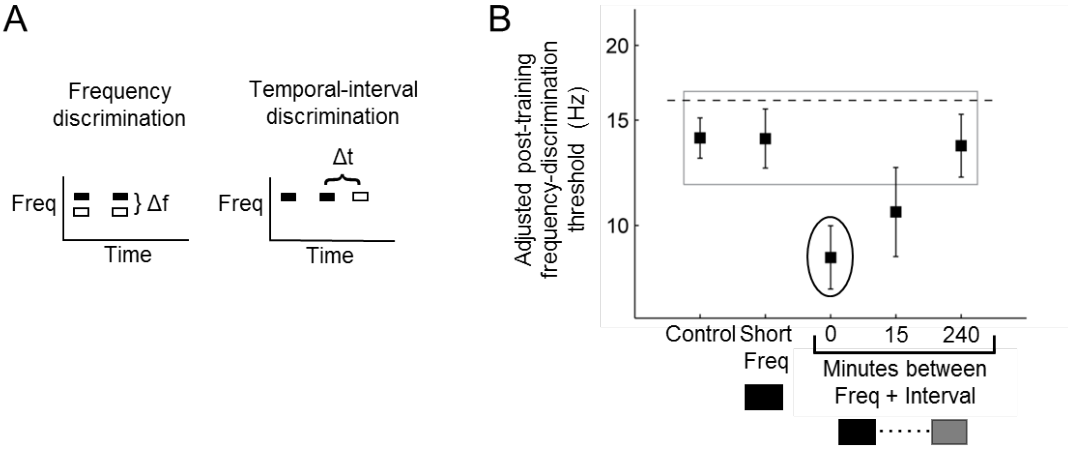


Fig. 4: Temporal separation between periods of task practice and periods of additional stimulus exposure. (A) As in Fig. 1A. (B) Mean post-training thresholds on the frequency-discrimination task following either no training (control), or one of four 6-7 day training regimens: 360 training trials per day on frequency discrimination alone (Short Freq), or 360 training trials per day on frequency discrimination and 360 on temporal-interval discrimination, using the same standard stimulus for both tasks, with the training trials on the two tasks separated by 0, 15, or 240 minutes (n = 8-10 per group). Otherwise, as in Fig. 1B. [Data from Wright and Sabin (2007) and Wright et al. (2010).]

the end of training on the frequency task and the beginning of training on the temporal-interval task was either 0, 15, or 240 minutes. Training on the frequency task alone yielded no improvement on that task, as described above (see Fig. 2B). Training on the temporal-interval task immediately after the frequency task did yield improvement on frequency discrimination, replicating the outcome obtained when the training alternated between these two tasks (see Fig. 3B). However, the effectiveness of the temporal-interval trials declined as the temporal separation between the two training periods increased to 15 minutes, and was gone when the periods were separated by 4 hours. Thus, the periods of task performance and of additional stimulus exposure need to occur within 15 minutes of each other.

We also examined the influence of the temporal order of the periods of task practice and additional stimulus exposure. We trained one group on the frequency task followed immediately by the temporal-interval task, as described above, and another group using the opposite order. Both groups improved on the frequency-discrimination task (data not shown). Thus, the temporal order of the two periods did not matter.

Finally, we examined the effect of stimulus differences between the practice and additional-stimulus-exposure periods. We trained two other groups on the frequency task followed immediately by the temporal-interval task, but varied the standard stimulus in the temporal-interval task. For one group, the temporal-interval standard had the same frequency as, but a different temporal-interval than, the frequency standard. This group improved on frequency discrimination. For the other group, the temporal-interval standard instead had the same interval as, but a different frequency than, the frequency standard. This group showed no improvement on the frequency task. Thus, the additional stimulus exposures had to share a key feature with the stimulus used during task practice, but the stimuli in the two periods did not need to be identical.

DISCUSSION

The elements of training regimens that yield perceptual improvement across days provide insights into perceptual-learning mechanisms, which, in turn, have implications for how to most effectively and efficiently train perceptual skills.

The need for task practice suggests that task performance provides an internal permissive signal that places the neural circuitry to be modified in a sensitized state. This permissive signal might arise from the attention required to perform the task or from rewards associated with performing the task, among other possibilities. The idea that top-down influences play a critical role in perceptual learning is well recognized (Ahissar and Hochstein, 2004; Seitz and Watanabe, 2005). The implication is that purely bottom-up exposure-based training regimens are unlikely to be successful.

Seemingly less appreciated is the apparent requirement for a sufficient amount, but no more, of practice per day. Most share the intuitive sense that training regimens should ‘provide enough training’, and accordingly design training plans that deliver the maximum amount of training allowed by time constraints. However, the observations that the actual number of daily training trials required for learning across days can be substantial, and that training beyond that amount can be superfluous, offer new insight into the learning process. The need for a sufficient amount of practice per day suggests that the neural circuitry to be modified must receive adequate stimulation to trigger consolidation (the transfer to long-term memory) (Wright *et al.*, 2010). That enough is enough suggests that consolidation may function as an all-or-none process (Wright and Sabin, 2007). By this view, the training (acquisition) and consolidation phases are functionally distinct. Additional support for this idea comes from reports in which the same intervening event (training on a non-target condition) disrupted learning on a target condition when presented during the acquisition stage, but not during the consolidation stage, of learning on that target condition (Banai *et al.*, 2010; Zach *et al.*, 2005). At a practical level, these observations suggest that training regimens could be made more effective and efficient by determining the amount of training that is necessary to generate improvement. Too little training will be ineffective, and too much inefficient.

Finally, the demonstration that the combination of task practice and additional stimulus exposure without practice can enhance perceptual improvement suggests that the influences of these two experiences on learning extend beyond the times in which they are elicited. The restriction of this temporal interaction to a period of minutes rather than hours implies that it is an aspect of the acquisition phase rather than the consolidation phase of learning. The lack of constraint on the presentation order of the two experiences raises the possibility that two different processes can create this beneficial interaction. The influence of task practice may extend into a following period of additional stimulus exposure, making those exposures function as if the task were still being performed, while a period of stimulus exposure may increase the effectiveness of subsequent task practice. It also appears that the neural circuitry engaged in the interaction is selective to stimulus features, not the whole stimulus, because, to be effective, the additional stimulus exposures needed to share a key feature with, but not necessarily be identical to, the stimulus used during task practice. Training regimens that take advantage of this interaction between task performance and additional stimulus exposures could reduce the amount of task practice necessary for learning on a given task by at least half. The saved practice trials could be replaced either with stimulus exposures without practice, to make the total regimen less work, or with training on a different task, to increase the regimen's overall impact.

In summary, we suggest that two key elements of successful multiple-day auditory perceptual training regimens are practice on the task to be learned, and a sufficient, and sometimes substantial, amount of training per day. Beyond these core requirements, perceptual training can be made more efficient by not exceeding the required amount of daily training and by replacing a subset of the training trials with stimulus exposure alone.

ACKNOWLEDGMENTS

Andy Sabin created preliminary versions of the figures. This work was supported by NIH/NIDCD.

REFERENCES

Aberg, K.C., Tartaglia, E.M., and Herzog, M.H. (**2009**). “Perceptual learning with chevrons requires a minimal number of trials, transfers to untrained directions, but does not require sleep,” Vis. Res., **49**, 2087-2094.

Ahissar, M., and Hochstein, S. (**1993**). “Attentional control of early perceptual learning,” Proc. Natl. Acad. Sci. USA, **90**, 5718-5722.

Ahissar, M., and Hochstein, S. (**2004**). “The reverse hierarchy theory of visual perceptual learning,” Trends Cogn. Sci., **8**, 457-464.

Banai, K., Ortiz, J.A., Oppenheimer, J.D., and Wright, B.A. (**2010**). “Learning two things at once: Constraints on the acquisition phase of perceptual learning,” Neurosci., **165**, 436-444.

Bao, S., Chang, E.F., Woods, J., and Merzenich, M.M. (**2004**). “Temporal plasticity in the primary auditory cortex induced by operant perceptual learning,” Nat. Neurosci., **7**, 974-981.

Cohen, J. (**1988**) “Statistical power analysis for the behavioural sciences,” in *The concepts of power analysis* (Lawrence Erlbaum Assoc, Hillsdale).

Hauptmann, B., and Karni A. (**2002**). “From primed to learned: the saturation of repetition priming and the induction of long-term memory,” Cogn. Brain Res., **13**, 313-322.

Huyck, J.J., and Wright, B.A. (**2011**). “Late maturation of auditory perceptual learning,” Develop. Sci., **14**, 614-621.

Levitt H **(1971)**. “Transformed up-down methods in psychoacoustics,” J. Acoust. Soc. Am., **49**, 467-477.

McGaugh, J.L. **(2000)** “Memory-A century of consolidation,” Science, **287**, 248-251.

Mednick, S.C., Nakayama, K., Cantero, J.L., Atienza, M., Levin, A.A., Pathak, N., and Stickgold, R. (**2002**). “The restorative effect of naps on perceptual deterioration,” Nat. Neurosci., **5**, 677–681.

Mossbridge, J.A., Fitzgerald, M.B., O’Connor, E.S., and Wright, B.A. (**2006**). “Perceptual-learning evidence for separate processing of asynchrony and order tasks,” J. Neurosci., **26**, 12708-12716.

Ortiz, J.A., and Wright, B.A. (**2010**). “Differential rates of consolidation of conceptual and stimulus learning following training on an auditory skill,” Exp. Brain Res., **201**, 441-451.

Polley, D.B., Steinberg, E.E., and Merzenich, M.M. (**2006**). “Perceptual learning directs auditory cortical map reorganization through top-down influences,” J. Neurosci., **26**, 4970-4982.

Rammsayer, T.H. (**1994**). “Effects of practice and signal energy on duration discrimination of brief auditory intervals,” Percept .Psychophys. **55**, 454-464.

Roth, D.A., Amir, O., Alaluf, L., Buchsenspanner, S., and Kishon-Rabin, L. (**2003**). "The effect of training on frequency discrimination: generalization to untrained frequencies and to the untrained ear," J. Basic Clin. Physiol. Pharmacol., **14**, 137-150.

Savion-Lemieux, T., and Penhune, V.B. (**2005**) "The effects of practice and delay on motor skill learning and retention," Exp. Brain. Res., **161**, 423-431.

Seitz, A.R., and Watanabe, T. (**2005**) "A unified model for perceptual learning," Trends. Cogn. Sci., **9**, 329-334.

Wright, B.A., and Sabin, A.T. (**2007**). "Perceptual learning: How much daily training is enough?" Exp. Brain Res., **180**, 727-736.

Wright, B.A., Sabin, A.T., Zhang, Y., Marrone, N., and Fitzgerald, M.B. (**2010**). "Enabling perceptual learning by alternating practice with sensory stimulation alone," J. Neurosci., **30**, 12868-12877.

Zach, N., Kanarek, N., Inbar, D., Grinvald, Y., Milestein, T., and Vaadia, E. (**2005**). "Segregation between acquisition and long-term memory in sensorimotor Learning," Eur. J. Neurosci., **22**, 2357–2362.

Studies of pitch mechanisms based on perceptual learning

BRIAN C. J. MOORE[1,*] AND HIROMITSU MIYAZONO[2]

[1] *Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 2EB, United Kingdom*

[2] *Department of Administration, Prefectural University of Kumamoto, 3-1-100 Tsukide, Kumamoto, Japan*

Mechanisms of pitch perception were studied using perceptual learning. In one set of studies, subjects discriminated the fundamental frequency (F0) of a target group of harmonics embedded in a background of harmonics with fixed F0. The results were potentially affected by pitch discrimination interference (PDI) and by cues related to pitch pulse asynchrony (PPA) between the target and background. Large learning effects occurred when PPA cues were available. Training was given using a stimulus with cosine-phase harmonics and high harmonics in the target, under conditions where PPA was not useful. Learning occurred, and it transferred to other cosine-phase, but not to random-phase, tones. The learning may reflect improvements in the ability to overcome PDI. In a second set of studies, F0 discrimination was measured for tones with cosine- or random-phase harmonics, bandpass filtered with five harmonics within the passband and presented in threshold-equalizing noise. Groups trained with LOW, MID, or MID-HIGH stimuli (harmonics 1-5, 11-15, or 14-18, respectively) showed learning effects that transferred to other stimuli except HIGH (28-32). A group trained with HIGH stimuli showed no learning effect, suggesting that a different mechanism was used for the HIGH stimuli than for the other stimuli. We propose that the LOW, MID, and MID-HIGH stimuli were discriminated using temporal fine structure information.

INTRODUCTION

This chapter describes a series of experiments in which perceptual learning was used to assess mechanisms of pitch perception for complex tones. It is widely believed that the pitch of complex tones containing low harmonics (below about the 8th), which are resolved in the peripheral auditory system (Plomp, 1964; Moore and Gockel, 2011), is derived from place and/or temporal information (patterns of phase locking) about the frequencies of the individual harmonics (Goldstein, 1973). Evidence for the involvement of phase locking comes from studies showing that the ability to 'hear out' individual components from complex sounds worsens at high frequencies, and even widely spaced components are difficult to hear out when their frequencies fall above 5 kHz (Moore *et al.*, 2006). Also, the pitch of a mistuned harmonic in a complex tone can be predicted using a model combining the effects of excitation pattern interaction and neural timing (Hartmann and Doty, 1996).

[*] Corresponding author: bcjm@cam.ac.uk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

For complex tones containing only very high harmonics (above about the 15th), the pitch is assumed to be based on the temporal envelope evoked on the basilar membrane by interfering harmonics (Moore and Moore, 2003a; de Cheveigné, 2005; Plack and Oxenham, 2005; Moore, 2012). There is less agreement about the mechanism that determines the pitch of complex tones with harmonics in the range 8-15. Some authors have argued that the pitch of such tones is derived from the temporal fine structure (TFS) of the waveform evoked on the basilar membrane by the interference of two or more harmonics (Schouten, 1940; Schouten *et al.*, 1962; Moore *et al.*, 2009). This idea is illustrated in Fig. 1. If this is the case, then the pitch mechanism might be similar for complex tones containing low, resolved, harmonics and for complex tones containing harmonics in the range 8-15.

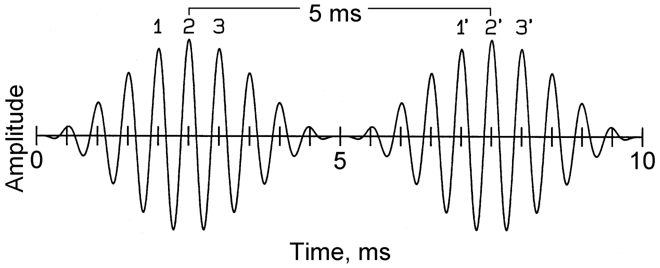


Fig. 1: Simulation of the waveform evoked on the basilar membrane at a place tuned to 2000 Hz by a complex tone with F0 = 200 Hz. Nerve spikes occur close to prominent peaks in the TFS (labelled 1, 2, 3 and 1′, 2′, and 3′). The pitch is assumed to be determined from the time interval between peaks close to adjacent envelope maxima (5 ms).

We have used perceptual learning to explore whether there are different pitch mechanisms for tones containing low, intermediate, and high harmonics. The rationale is that, if there are different pitch mechanisms, then training on fundamental frequency (F0) discrimination of tones with, for example, high harmonics will lead to improvements in performance only for tones with high harmonics; the training will not lead to better performance (i.e., transfer) to tones with low or intermediate harmonics because of the different mechanisms involved. However, if there is a single pitch mechanism for low, intermediate, and high harmonics, then training using tones with high harmonics might transfer to tones with low or intermediate harmonics, and vice versa. This rationale has been applied in a series of studies that are summarised below.

F0 DISCRIMINATION OF A GROUP OF HARMONICS EMBEDDED IN A BACKGROUND OF HARMONICS WITH FIXED F0

Several researchers have presented evidence suggesting that some harmonics are more important than others in determining the pitch of complex sounds (Plomp, 1967; Ritsma, 1967; Moore *et al.*, 1985). The harmonics that are most important are called the 'dominant' harmonics, and the frequency region in which they fall is called the 'dominant region'. In one series of studies (Miyazono and Moore, 2009; Miyazono *et al.*, 2010), we used stimuli similar to those that have been used to determine the dominant region. Thresholds for detecting a change in F0 (F0DLs) were measured for a group of harmonics (group B) embedded in a group of fixed non-overlapping harmonics (groups A and C) with the same mean F0. A schematic spectrum for one such stimulus is shown in Fig. 2.

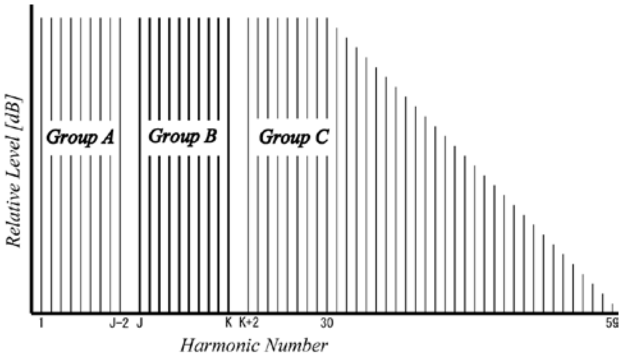


Fig. 2: Schematic spectrum showing components in groups A, B and C.

In the first experiment (Miyazono and Moore, 2009), a low F0 of 50 Hz was used. Group B contained harmonics 1-5, 1-25, 6-30, or 26-30. For the first two of these stimuli, there were no components in group A. The first two of these stimuli contained some resolved harmonics in group B, the third contained intermediate harmonics, and the fourth contained only completely unresolved harmonics. The components of the complex sound were added either starting with random phases or all starting in cosine phase (90°); the latter leads to a waveform with a high peak factor on the basilar membrane when the components are unresolved. In what follows, when 'group' forms part of a label referring to a stimulus it is spelled with a lower-case g, whereas when it refers to a group of subjects, it is written with an upper case G. One group of subjects was trained over multiple days using cosine-phase complex tones with harmonics 26-30 in group B (Group UC, unresolved-cosine). A second group was trained using random-phase complex tones with harmonics 1-5 in group B (Group RR, resolved-random). Group UC showed large improvements during training, which did not transfer to the other conditions tested (as assessed in the post-training session). Group RR did not show any clear improvement with training.

At first sight, these results might be taken as supporting the idea that there are different pitch mechanisms for low and high harmonics, as learning occurred only for the complex tones with high unresolved components in group B, and the learning did not transfer to complex tones with low harmonics in group B. However, Miyazono and Moore (2009) suggested an alternative explanation of the results. F0 discrimination of the cosine-phase tones with high harmonics might have been based on a cue called ‘pitch-pulse asynchrony’ (PPA) (Gockel *et al.*, 2005). Subjects may compare the timing of envelope peaks across different auditory filters. Consider, for example, an auditory filter centred on the 20th harmonic, within group A. For the cosine-phase stimuli, this would produce envelope peaks every 20 ms. For an auditory filter centred on the 28th harmonic, within group B, the envelope peaks would initially be synchronized to those of group A. However, in the interval where the F0 was shifted upwards, the period would be shorter, and towards the end of the stimulus the envelope peaks at the output of the filter centred in group B would occur earlier in time than those for the filter centred in group A; in other words, a PPA would occur. In the interval where the F0 was shifted downwards, a PPA in the opposite direction would occur. Thus, there would be a PPA across auditory filters, which would differ for the two intervals of a trial. The use of a cue based on PPA could account for the finding that, after training, thresholds for F0 discrimination of the cosine-phase complex tones with harmonics 26-30 in group B were very low, being below 0.1% of the F0 for several subjects.

Miyazono *et al.* (2010) confirmed that the learning effect found by Miyazono and Moore (2009) was indeed based on the use of a cue related to PPA. When PPA cues were disrupted by introducing a random temporal offset between the envelope peaks of the harmonics in group B and the remaining harmonics, F0DLs increased markedly.

Miyazono *et al.* (2010) examined perceptual learning using a training stimulus with cosine-phase harmonics, F0 = 50 Hz, and high harmonics in group B, under conditions where PPA cues were disrupted, as described above. Learning occurred, and it transferred to other cosine-phase tones, but not to random-phase tones. A similar experiment with F0 = 100 Hz showed a learning effect that transferred to a cosine-phase tone with mainly high unresolved harmonics, but not to cosine-phase tones with low harmonics, and not to random-phase tones. The learning found by Miyazono *et al.* (2010) appeared to be specific to tones for which F0 discrimination was based on distinct peaks in the temporal envelope.

A complication with the experiments described so far is that the results were almost certainly influenced by pitch discrimination interference (PDI), which is the phenomenon that F0 discrimination of a group of harmonics in one frequency region can be impaired by harmonics with a fixed (but nearby) F0 in a different region (Gockel *et al.*, 2004; 2009b). The learning effect found might partly reflect learning to overcome the interference produced by the harmonics in groups A and C. The experiments described next were intended to avoid this complication.

F0 DISCRIMINATION OF BANDPASS-FILTERED COMPLEX TONES IN BACKGROUND NOISE

Miyazono and Moore (2013) examined whether the pitch mechanism for tones with intermediate harmonics is similar to or different from the mechanisms for low and high harmonics. We studied perceptual learning for F0 discrimination using complex tones that were bandpass filtered so as to contain low resolved harmonics (stimulus LOW), high unresolved harmonics (stimulus HIGH), and intermediate harmonics (stimulus MID). All stimuli were presented in a background of threshold equalizing noise (TEN) (Moore *et al.*, 2000) to mask combination tones and to limit the audibility of components falling outside the passband.

Learning effects with harmonics 11-15 in the MID stimulus

In experiment 1, the filters were chosen to have relatively shallow slopes of 30 dB/oct so that, when the harmonics were unresolved, changes in F0 would result in minimal changes in the excitation pattern (Moore and Moore, 2003a; 2003b). Also, the use of shallow slopes meant that there were no ‘edge’ harmonics (harmonics with no adjacent harmonics above or below them), avoiding the possibility that edge harmonics might be unusually well resolved (Moore and Ohgushi, 1993).

Subjects were required to discriminate the F0 of two successive tones presented at 65 dB SPL. The nominal F0 was 100 Hz. Three fixed spectral envelopes were used, each with a flat bandpass region and slopes of 30 dB/oct. The passbands extended from 100 to 500, 1100 to 1500, and 2800 to 3200 Hz for cases LOW, MID, and HIGH, respectively. All components were added with cosine starting phase. The TEN spectrum ranged from 100 to 8000 Hz. The TEN level at 1 kHz, expressed as dB/ERB$_N$ (Moore, 2012), was set 20 dB below the level of the each component within the passband.

There were three groups of five young normal-hearing subjects, designated LOW, MID and HIGH, according to the stimuli used during training. Each subject was tested on 10 days, two for measurement of pre-training thresholds for all three conditions (LOW, MID, and HIGH), six for training with the stimulus allocated to that group (usually on successive days, but excluding weekends), and two for measurement of post-training thresholds.

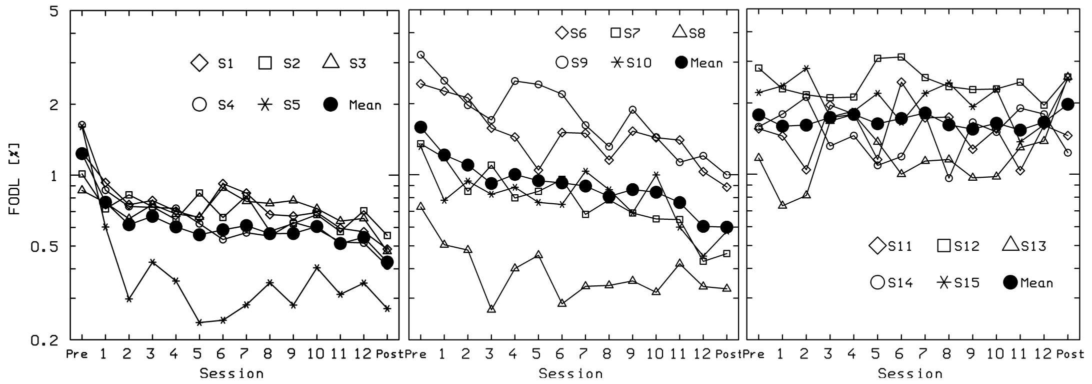


Fig. 3: Results obtained for the pre-training session (Pre), the training sessions, and the post-training session (Post), for Groups LOW (left), MID (middle) and HIGH (right). The F0DLs for the Pre and Post sessions are for the same stimuli as used during training.

Fig. 3 shows the results obtained for the pre-training session (Pre), the training sessions, and the post-training session (Post), for each group. Thin curves show geometric mean F0DLs for the individual subjects (based on at least three estimates for pre- and post-training sessions, and six for training sessions), and curves marked by large filled circles show the geometric mean across subjects. The F0DLs are expressed as relative values in % (100×ΔF/F0). Performance improved across days for Groups LOW and MID, but not for Group HIGH.

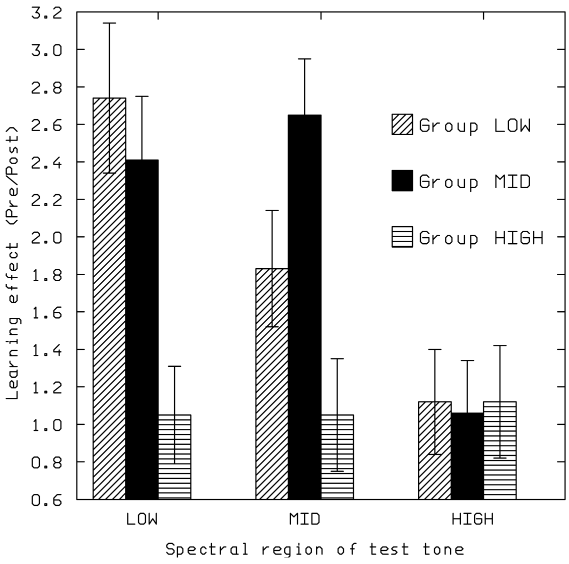


Fig. 4: Learning and transfer effects for each group. Each set of three bars denotes one stimulus type. Error bars indicate ± 1 standard error (SE).

Fig. 4 shows the overall learning effect for each group and each stimulus type, expressed as the mean F0DL for the pre-training session divided by the mean F0DL for the post-training session. The three sets of bars represent the three stimulus cases, and the three bars within each set represent the three groups. Group LOW showed a large learning effect for the LOW stimuli, with strong transfer to the MID stimuli, but no transfer to the HIGH stimuli. Group MID showed a large learning effect for the MID stimuli, with strong transfer to the LOW stimuli and no transfer to the HIGH stimuli. Group HIGH showed no learning effect for any stimuli. The fact that there was no learning effect for Group HIGH, while there was for Groups LOW and MID, suggests that the mechanism underlying F0 discrimination was different for the HIGH stimuli than for the LOW or MID stimuli.

The passband for stimulus MID contained harmonics 11 to 15. The harmonics in this stimulus were largely unresolved, and F0 discrimination was probably based on TFS information derived from unresolved harmonics. Hence, the similarity of the learning effect for cases LOW and MID, and the transfer of learning between these two cases, supports the idea that F0 discrimination was based on a common mechanism, probably using TFS information (from resolved harmonics for stimulus LOW and unresolved harmonics for stimulus MID). However, harmonics 7, 8, 9, and 10, which fell on the slope below the passband, would have been above the masked threshold in the TEN. Bernstein and Oxenham (2003) suggested that harmonics up to the 10th might be resolved. It is possible, therefore, that the lowest audible harmonics in stimulus MID were resolved to some extent. This could account for the similarity of the results for the LOW and MID stimuli, and the transfer of learning for these two types of stimuli. To assess this possibility, Miyazono and Moore (2013) measured learning and transfer effects using a new MID stimulus, which is described below.

Learning effects with harmonics 14-18 in the MID stimulus

In experiment 2, the new MID stimuli, denoted MID-HIGH, were filtered so that the passband contained harmonics 14-18. In addition, the spectral slope on the low side of the passband was made steeper, being 60 dB/oct rather than 30 dB/oct. This meant that fewer harmonics falling on the lower slope were above the masked threshold in the TEN. The lowest audible harmonic in the MID-HIGH stimuli was the 11th. This would not have been resolved, but the frequency region of the lowest audible harmonics might have been low enough for TFS information to be used. Seven new subjects were tested, denoted Group MID-HIGH. Training was performed only for the MID-HIGH stimuli, and transfer of learning to the LOW and HIGH stimuli was assessed. Only cosine-phase stimuli were used.

The left panel of Fig. 5 shows the learning curves. The group mean results improved significantly across days. Most individual subjects also showed improvements, but with marked variability. The right panel of Fig. 5 shows the learning and transfer effects. There was a large learning effect for the MID-HIGH stimuli, with strong transfer to the LOW stimuli and no transfer to the HIGH stimuli.

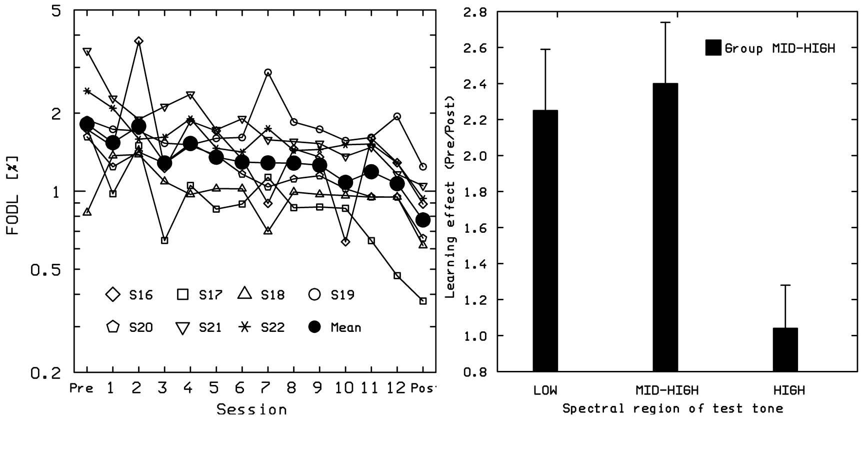


Fig. 5: The left panel shows F0DLs for the pre-training session (Pre), the training sessions, and the post-training session (Post), for each subject (open symbols) and for the mean (filled circles). The right panel shows learning and transfer effects for each stimulus type. Error bars indicate ± 1 SE.

The pattern of the results is the same as for experiment 1, despite the fact that all harmonics for stimulus MID-HIGH would have been unresolved. This supports the interpretation that the transfer of learning between stimuli MID and LOW and between MID-HIGH and LOW reflects a common underlying mechanism based on the use of TFS information.

Learning effects for random-phase tones

Experiments 1 and 2 were conducted using stimuli whose components were added in cosine starting phase, which leads to a waveform on the basilar membrane with a high peak factor when the components are not resolved. Experiment 3 was conducted to assess the importance of the peak factor. Components were added with random starting phase. This leads to a waveform on the basilar membrane with a lower peak factor than for cosine phase when the components are not resolved. Effects of component phase on F0 discrimination should only occur when the components on which discrimination is based are at least partly unresolved, so the results were intended to provide an additional check that the components for stimulus MID-HIGH were unresolved. The passbands extended from 100 to 500, 1400 to 1800, and 2800 to 3200 Hz, for cases LOW, MID-HIGH, and HIGH, respectively. Three groups of four (new) subjects were tested, designated LOW, MID-HIGH, and HIGH, according to the stimuli used during training.

The left panel of Fig. 6 shows the mean learning curves for each group. There was a clear improvement across sessions for Groups LOW and MID-HIGH, but not for Group HIGH. The mean F0DLs for Group HIGH were significantly higher than were obtained for Group HIGH in experiment 1, indicating that F0DLs based on temporal-envelope cues are affected by the peak factor of the waveform on the basilar membrane, which is consistent with previous work (Houtsma and Smurzynski, 1990; Wang *et al.*, 2012). Also, the mean F0DLs for Group MID-HIGH were significantly higher than the F0DLs for Group MID-HIGH in experiment 2, confirming that the components in stimulus MID-HIGH were at least partially unresolved.

The right panel of Fig. 6 shows the learning and transfer effects. Group LOW showed a large learning effect for the LOW stimuli, with some transfer to MID-HIGH and no transfer to the HIGH stimuli. Group MID-HIGH showed a large learning effect for the MID-HIGH stimuli, with some transfer to LOW and no transfer to the HIGH stimuli. Group HIGH showed no learning effect and no transfer to either of the other stimuli. The pattern of the learning and transfer effects was similar to that for experiments 1 and 2, indicating that the peak factor of the stimuli is not critical in determining whether or not learning and transfer of learning occur.

DISCUSSION

In the experiments with bandpass-filtered tones, F0DLs for tones with low harmonics improved with training, consistent with the results of Grimault *et al.* (2002). However, Grimault *et al.* also found a learning effect for tones with only

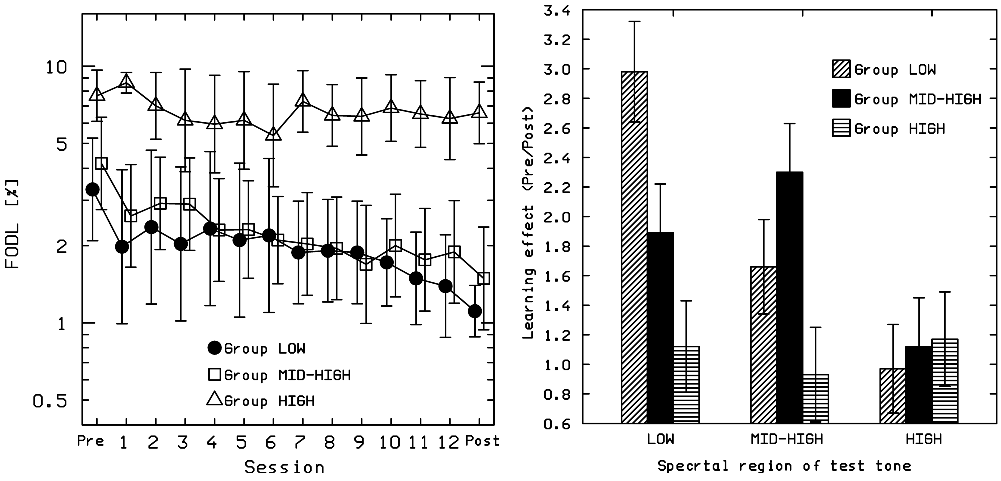


Fig. 6: The left panel shows mean results obtained for the pre-training session (Pre), the training sessions, and the post-training session (Post), for each group. The F0DLs for the Pre and Post sessions are for the same stimuli as used during training. The right panel shows learning and transfer effects for each group and stimulus type. Error bars indicate ± 1 SE.

high harmonics, while experiments 1 and 3 showed no such effect. The difference may have occurred because our subjects were tested using more trials during the pre-training sessions, which would have allowed fast perceptual learning (Hawkey *et al.*, 2004) and procedural learning. The learning effects found by Grimault *et al.* might have reflected fast perceptual and procedural learning. The results of experiments 1 and 3 also differ from those obtained for F0 discrimination of a group of harmonics embedded within harmonics whose F0 was fixed (Miyazono *et al.*, 2010), as described earlier in this chapter. For the earlier results, a learning effect was found when group B contained only high harmonics, but such an effect was not found in experiments 1 and 3. The difference probably reflects differences in the stimuli: discrimination of the F0 of a group of harmonics embedded within harmonics whose F0 was fixed in the earlier study, versus discrimination of a group of harmonics presented in TEN in the later study. In the earlier study, the learning may have involved reduction of PDI (Gockel *et al.*, 2004; 2009a). PDI seems to depend on the relative salience of the target and interfering sounds, and so PDI would have been strong when group B contained only high unresolved harmonics. It may be that effects of training on PDI are large when the PDI effect itself is large. When group B contained harmonics 1-5, the harmonics in group B would have had a higher pitch salience than those in groups A and C (Jackson and Moore, 2013), leading to a small PDI effect, and therefore to little scope for reducing PDI by training.

The experiments using bandpass-filtered tones included stimuli (MID and MID-HIGH) with intermediate harmonic numbers. In experiment 1, the lowest harmonic within the passband was the 11th, and the lowest component that was above threshold in the TEN was the 7th. In experiment 2, the lowest harmonic within the passband was the 14th, and the lowest harmonic that was above threshold in the TEN was the 11th. It seems likely that only harmonics up to the 8th are resolvable (Plomp, 1964; Plomp and Mimpen, 1968; Moore and Ohgushi, 1993; Moore *et al.*, 2006; Moore and Gockel, 2011), and the limit may be even lower for complex tones with low F0 (Jackson and Moore, 2013). Even if harmonics up to the 10th are resolvable (Bernstein and Oxenham, 2003), the audible harmonics in the MID-HIGH stimulus were almost certainly only unresolved. Consistent with this, F0 discrimination of the MID-HIGH stimuli was better when the components were added in cosine phase (experiment 2) than when they were added in random phase (experiment 3). The results showed clear learning effects for the LOW, MID, and MID-HIGH stimuli, and these effects transferred; training with LOW stimuli led to better F0 discrimination of MID stimuli, and training with MID or MID-HIGH stimuli led to better discrimination of LOW stimuli. This is consistent with the idea that F0 discrimination of the LOW, MID, and MID-HIGH stimuli was based on similar mechanisms, perhaps based on the use of TFS information. For the LOW stimuli, the TFS would have conveyed information about the frequencies of individual harmonics, whereas for the MID and MID-HIGH stimuli, the TFS would have conveyed information about the time intervals between prominent peaks in the waveform produced by the interaction of harmonics on the basilar membrane (Schouten *et al.*, 1962), as illustrated in Fig. 1. However, the two types of TFS information may be used in a similar way by the pitch processor (Meddis and O'Mard, 1997; Bernstein and Oxenham, 2005; Moore, 2012).

The results showed no learning effects for the HIGH stimuli, suggesting that the mechanism underlying discrimination of such stimuli is different from that for the LOW, MID, and MID-HIGH stimuli. It seems likely that F0 discrimination of the HIGH stimuli was based on envelope information only, not TFS information or information from resolved harmonics (Moore and Moore, 2003b).

CONCLUSIONS

F0 discrimination of a group of high harmonics embedded in harmonics with fixed F0 can be affected by cues related to PPA (for cosine-phase stimuli) and by PDI. The learning effects found for such stimuli may partly reflect learning to make effective use of PPA and learning to overcome PDI.

F0 discrimination of a group of bandpass-filtered harmonics presented in TEN showed learning effects for LOW, MID, and MID-HIGH stimuli (harmonics 1-5, 11-15, or 14-18), but not for HIGH stimuli. The learning effects obtained with LOW, MID, or MID-HIGH stimuli transferred to other stimuli except HIGH (28-32). These results suggest that the underlying pitch mechanisms are similar for LOW, MID, and MID-HIGH stimuli, but that a different pitch mechanism operates for HIGH stimuli. We propose that LOW, MID, and MID-HIGH stimuli are discriminated using TFS information, while HIGH stimuli are discriminated using temporal envelope information.

ACKNOWLEDGMENTS

Author BM was supported by the Medical Research Council UK (grant number G0701870). We thank Brian Glasberg for his collaboration in some of this work and Christian Füllgrabe for helpful comments on an earlier version of this chapter.

REFERENCES

Bernstein, J.G., and Oxenham, A.J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number?" J. Acoust. Soc. Am., **113**, 3323-3334.

Bernstein, J.G., and Oxenham, A.J. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am., **117**, 3816-3831.

de Cheveigné, A. (**2005**). "Pitch perception models," in *Pitch Perception*. Edited by C.J. Plack, A.J. Oxenham, R.R. Fay, and A.N. Popper (Springer, New York), pp. 169-233.

Gockel, H., Carlyon, R.P., and Plack, C.J. (**2004**). "Across-frequency interference effects in fundamental frequency discrimination: questioning evidence for two pitch mechanisms," J. Acoust. Soc. Am., **116**, 1092-1104.

Gockel, H., Carlyon, R.P., and Moore, B.C.J. (**2005**). "Pitch discrimination interference: The role of pitch pulse asynchrony," J. Acoust. Soc. Am., **117**, 3860-3866.

Gockel, H.E., Carlyon, R.P., and Plack, C.J. (**2009a**). "Further examination of pitch discrimination interference between complex tones containing resolved harmonics," J. Acoust. Soc. Am., **125**, 1059-1066.

Gockel, H.E., Hafter, E.R., and Moore, B.C.J. (**2009b**). "Pitch discrimination interference: The role of ear of entry and of octave similarity," J. Acoust. Soc. Am., **125**, 324-327.

Goldstein, J.L. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am., **54**, 1496-1516.

Grimault, N., Micheyl, C., Carlyon, R.P., and Collet, L. (**2002**). "Evidence for two pitch encoding mechanisms using a selective auditory training paradigm," Percept. Psychophys., **64**, 189-197.

Hartmann, W.M., and Doty, S.L. (**1996**). "On the pitches of the components of a complex tone," J. Acoust. Soc. Am., **99**, 567-578.

Hawkey, D.J., Amitay, S., and Moore, D.R. (**2004**). "Early and rapid perceptual learning," Nat. Neurosci., **7**, 1055-1056.

Houtsma, A.J.M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am., **87**, 304-310.

Jackson, H.M., and Moore, B.C.J. (**2013**). "The dominant region for the pitch of complex tones with low fundamental frequencies," J. Acoust. Soc. Am., **134**, 1193-1204.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am., **102**, 1811-1820.

Miyazono, H., Glasberg, B.R., and Moore, B.C.J. (**2010**). "Perceptual learning of fundamental frequency (F0) discrimination: Effects of F0, harmonic number, and component phase," J. Acoust. Soc. Am., **128**, 3649-3657.

Miyazono, H., and Moore, B.C.J. (**2009**). "Perceptual learning of frequency discrimination for tones with low fundamental frequency: Learning for high but not for low harmonics," Acoust. Sci. Tech., **30**, 383-386.

Miyazono, H., and Moore, B.C.J. (**2013**). "Implications for pitch mechanisms of perceptual learning of fundamental frequency discrimination: Effects of spectral region and phase," Acoust. Sci. Tech. (in press).

Moore, B.C.J., Glasberg, B.R., and Peters, R.W. (**1985**). "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am., **77**, 1853-1860.

Moore, B.C.J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am., **93**, 452-461.

Moore, B.C.J., Huss, M., Vickers, D.A., Glasberg, B.R., and Alcántara, J.I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol., **34**, 205-224.

Moore, B.C.J., and Moore, G.A. (**2003a**). "Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects," Hear. Res., **182**, 153-163.

Moore, B.C.J., Glasberg, B.R., Low, K.-E., Cope, T., and Cope, W. (**2006**). "Effects of level and frequency on the audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am., **120**, 934-944.

Moore, B.C.J., Hopkins, K., and Cuthbertson, S.J. (**2009**). "Discrimination of complex tones with unresolved components using temporal fine structure information," J. Acoust. Soc. Am., **125**, 3214-3222.

Moore, B.C.J., and Gockel, H. (**2011**). "Resolvability of components in complex tones and implications for theories of pitch perception," Hear. Res., **276**, 88-97.

Moore, B.C.J. (**2012**). *An Introduction to the Psychology of Hearing, 6th Ed.* (Brill, Leiden, The Netherlands), pp. 1-441.

Moore, G.A., and Moore, B.C.J. (**2003b**). "Perception of the low pitch of frequency-shifted complexes," J. Acoust. Soc. Am., **113**, 977-985.

Plack, C.J., and Oxenham, A.J. (**2005**). "The psychophysics of pitch," in *Pitch Perception*. Edited by C.J. Plack, A.J. Oxenham, R.R. Fay, and A.N. Popper (Springer, New York), pp. 7-55.

Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am., **36**, 1628-1636.

Plomp, R. (**1967**). "Pitch of complex tones," J. Acoust. Soc. Am., **41**, 1526-1533.

Plomp, R., and Mimpen, A.M. (**1968**). "The ear as a frequency analyzer II," J. Acoust. Soc. Am., **43**, 764-767.

Ritsma, R.J. (**1967**). "Frequencies dominant in the perception of the pitch of complex sounds," J. Acoust. Soc. Am., **42**, 191-198.

Schouten, J.F. (**1940**). "The residue and the mechanism of hearing," Proc. Kon. Ned. Akad. Wetenschap., **43**, 991-999.

Schouten, J.F., Ritsma, R.J., and Cardozo, B.L. (**1962**). "Pitch of the residue," J. Acoust. Soc. Am., **34**, 1418-1424.

Wang, J., Baer, T., Glasberg, B.R., Stone, M.A., Ye, D., and Moore, B.C.J. (**2012**). "Pitch perception of concurrent harmonic tones with overlapping spectra," J. Acoust. Soc. Am., **132**, 339-356.

Auditory learning: Uncorking performance bottlenecks

SYGAL AMITAY[1,*], PETE R. JONES[1,2], YU-XUAN ZHANG[1,3],
LORNA F. HALLIDAY[1,4], AND DAVID R. MOORE[1,5]

[1] *Medical Research Council Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom*

[2] *Current address: Institute of Ophthalmology, University College London, London EC1V 9EL, United Kingdom*

[3] *Current address: National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China*

[4] *Current address: Division of Psychology and Language Sciences, University College London, London WC1N 1PF, United Kingdom*

[5] *Current address: Communication Sciences Research Center, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229-3026, USA*

Internal noise is ubiquitous to information processing systems in the brain. It can originate in low-level, sensory systems (e.g., stochastic neural firing) or high-level cognitive functions (e.g., fluctuations in attention). Added to inefficiencies associated with the decision making process, it compromises our ability to make perceptual judgements even under ideal conditions (i.e., in the absence of external noise). We present evidence herein that performance-limiting internal noise and inefficiency of various origins can be reduced through training, resulting in improved behavioural performance. We promote the view that reducing or even removing these limiting processes is what defines perceptual learning, and that transfer of learning to untrained tasks critically depends on those tasks having a limiting process in common with the trained task. We present implications of this view for our understanding of perceptual learning during development and in atypical populations, as well as to the more practical aspects of designing perceptual and cognitive training programmes that will demonstrate benefits beyond the training tasks themselves.

INTRODUCTION

In detecting, discriminating, and identifying sounds, the accuracy of perceptual judgements critically depends on the fidelity with which the information arriving at the ears is encoded and subsequently processed. To make the perceptual decisions required by a psychophysical task, listeners must implicitly (or explicitly) deduce the structure of, and be able to extract the task-relevant information from, the physical stimulus. However, the fidelity with which information is encoded by the nervous system is subject to degradation by random effects such as transmission

*Corresponding author: sygal@ihr.mrc.ac.uk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

through physiologically noisy pathways (e.g., stochastic neural encoding both peripherally and more centrally; Vogels *et al.*, 1989; Javel and Viemeister, 2000) and fluctuations in arousal and attention (Fox *et al.*, 2006; 2007), as well as deterministic effects such as erroneous assumptions about the structure and statistics of the task (e.g., Tanner *et al.*, 1967; Maddox and Bohil, 2001).

The purpose of this paper is to promote the view that perceptual learning – the improvement in performance due to experience and practice – results from lifting the processing limitations that act as bottlenecks to performance (coined "learning the limiting process" by Dosher and Lu, 2005). Because processing limitations can occur at multiple levels, perceptual learning is not confined to its traditional bottom-up description; changes occur at the level of the bottleneck, not restricted to low-level stimulus encoding or decoding. Although we are drawing upon evidence primarily from auditory learning, we conjecture that these are general principles that would apply equally in other modalities.

The concept of internal noise is central to psychophysics. According to signal detection theory (Green and Swets, 1966; Macmillan and Creelman, 2005), making a decision involves comparing a decision variable derived from the representation of the sensory input with a subjective decision criterion (see Fig. 1). Both internal representation and decision criterion are subject to variations that are intrinsic to the listener and limit the accuracy of the decision. We have recently shown that, even in the absence of differences in the physical stimuli, early variations in electrophysiological brain activity (event-related potentials occurring less than 100 ms after stimulus onset and associated with stimulus encoding) can predict discrimination decisions in a multiple-interval, forced-choice procedure (Amitay *et al.*, 2013). This study demonstrated that the magnitude of the variation in the internal representations of the stimulus engendered by the internal noise is comparable to the behaviourally just-noticeable physical differences introduced when measuring discrimination thresholds.

Here we present evidence that auditory training reduces internal noise originating at various levels of the perceptual processing hierarchy, as well as inefficiencies resulting from suboptimal placement of the decision criterion. We show that by varying aspects of the stimulus, training task and procedure, we vary what is being learned by creating limitations on performance at different levels of processing.

REDUCING INTERNAL NOISE THROUGH TRAINING

Jones *et al.* (2013) demonstrated that internal noise was reduced over several practice sessions on a pure-tone frequency discrimination task. By adding external noise along the task-relevant dimension (jittering the frequency difference), we were able to show a significant reduction in internal noise, although the methods used precluded pinpointing the source(s) of the noise. Simulations based on the behavioural data suggested noise reduction was achieved through reweighting of frequency-specific channels, i.e., change in $[\omega_1,\omega_2,\ldots\omega_n]$ (Fig. 1A).

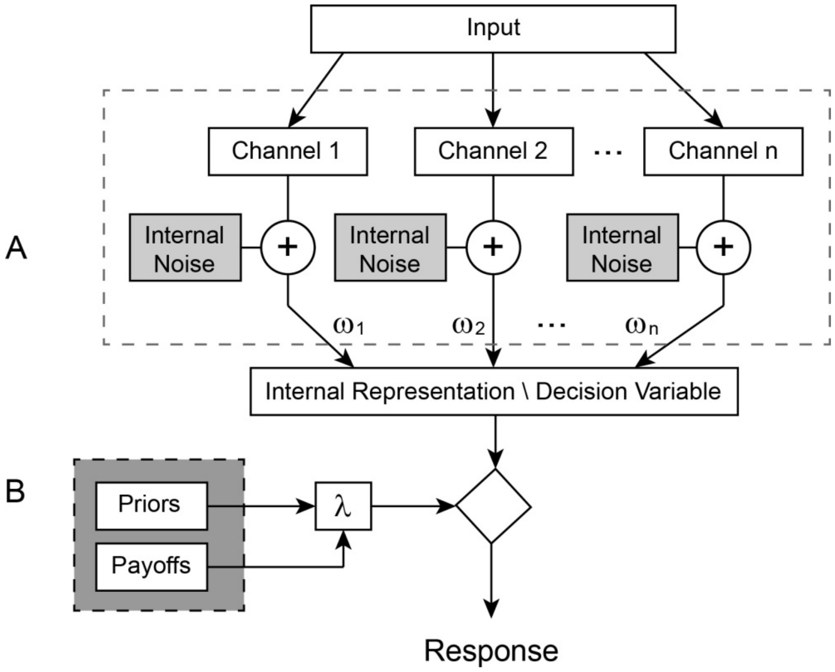


Fig. 1: A simple perceptual decision model. The incoming physical stimulus is transformed into an internal representation by summing over *n* independent information channels, each subject to internal noise (A). For simplicity we do not distinguish here between the internal representation and the decision variable computed from it, although this process may be subject to further internal noise (not specified in the text). A decision is made by comparing the decision variable to a criterion, λ, which may or may not be optimally placed (B).

Reducing low-level sensory noise

By varying tone duration in a frequency discrimination task, Amitay *et al.* (2012) observed that, while frequency discrimination thresholds improved on the trained task regardless of whether the tone was long (100 ms) or short (15 ms), training on short tones also improved discrimination of long tones, while training on long tones did not improve discrimination of short tones (Fig. 2). Our hypothesis, supported by simulations, was as follows: Frequency discrimination depends on the representations of each signal's frequency, and the accuracy of these representations is limited by phase locking noise due to the jitter in neural firing in the auditory nerve. One way of reducing this noise is increasing the integration time window (averaging over more cycles of the stimulus). We estimated the naïve (untrained) integration time for the 100-ms tones to be ~17 ms, while trained integration times are reportedly ~50 ms (Moore, 1973). Extending the integration window could not benefit the short tones, the duration of which was shorter than even the 17-ms naïve integration time window. Since extending the integration window was not possible for the short tones, we simulated the learning in this condition as reduced spike jitter in the auditory nerve. The simulation accurately predicted the transfer of learning from short to long tones (Fig. 2). The short tone duration imposed a limitation on which mechanism could support learning, resulting in a different limiting process being learned, but one which could benefit frequency discrimination regardless of tone duration.

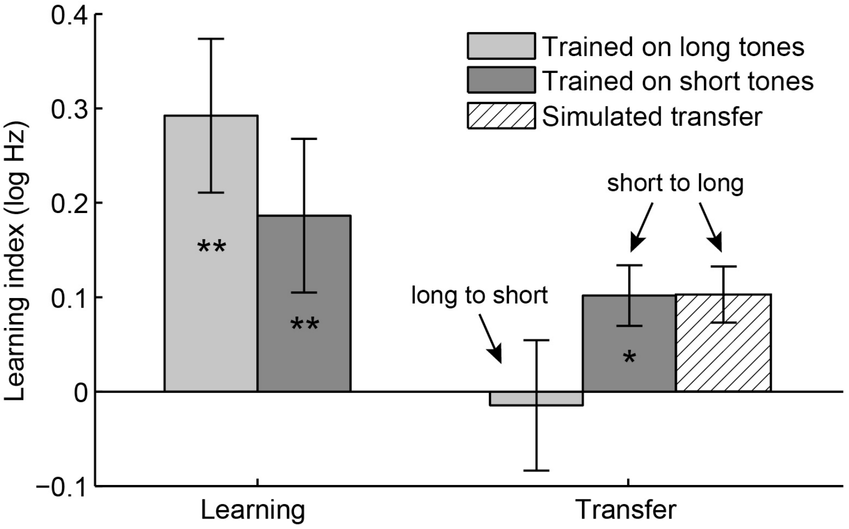


Fig. 2: Learning and transfer on a frequency discrimination task with long and short tones. The learning index is the difference between the pre-and post-training discrimination limens. Significant learning is marked by * p < 0.05; ** p < 0.01, corrected for multiple comparisons. Error bars denote s.e.m. Adapted from Amitay *et al.* (2012).

Reducing high-level cognitive limitations

Introducing uncertainty about the stimulus into a frequency discrimination task by roving the base value of the standard stimulus on a trial-by-trial basis impairs performance and slows learning down in good listeners (Amitay *et al.*, 2005). Since discrimination limens for a roving frequency exceed those observed when the individual frequencies are trained consecutively in a fixed frequency design, the limitation imposed on processing is unlikely to be due to bottom-up, sensory encoding of stimulus frequency. Despite the more protracted learning, training good listeners (those without exceedingly high naïve discrimination limens; see Amitay *et al.*, 2005) on a roving frequency discrimination task transferred fully to a fixed frequency discrimination task, while training on a fixed frequency resulted in naïve-like performance on the roving frequency task (Fig. 3).

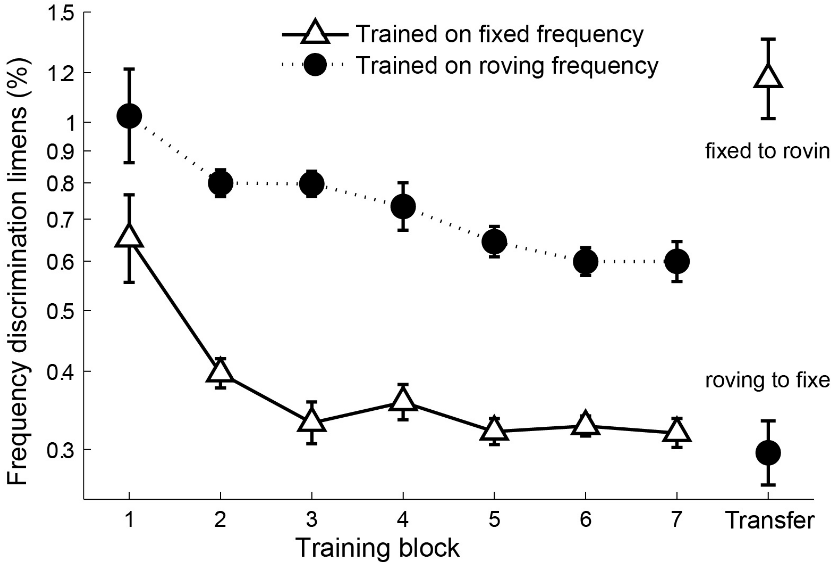


Fig. 3: Learning and transfer on a frequency discrimination task with a fixed- and roving frequency standard stimulus in good listeners (defined by thresholds on block 1). Discrimination limens (in percent of the standard frequency) are higher (poorer) and learning is slower in the roving frequency condition. However, training on roving frequency stimuli results in transfer to fixed frequency stimuli, but not *vice versa*. Each training block consisted of 500 trials. Limens are adjusted for initial (block 1) performance. Adapted from Amitay *et al.* (2005).

Two possible noise sources may affect processing under conditions of uncertainty. Firstly, uncertainty about the frequency of the incoming stimulus means that listeners cannot attend to a single frequency channel, but need instead to either shift their attention between channels in each trial or simultaneously monitor several different channels. However, learning to flexibly re-weight the channels (Fig. 1A) on every trial is an unnecessary skill when the frequency is not changing. Likewise, learning to weight them all equally should not benefit a discrimination that involves attending to just one channel. Therefore, learning a new weighting strategy does not explain the transfer results of Amitay *et al.* (2005).

A second alternative is that the constraint on processing is imposed by working memory. Unlike a fixed-stimulus discrimination for which a ‘perceptual anchor’ (Braida *et al.*, 1984) – a stable, long-term memory representation of the stimulus to which individual stimuli can be compared – can be formed, listeners trained with narrowly roving frequencies need to continually update working memory representations of the stimuli and compare them ‘online’ (see Banai and Amitay, 2012). Although reducing working memory limitations places greater processing demands on the system than forming perceptual anchors, this learning should benefit discrimination whether the stimulus is roved or not.

We have evidence that supports the suggestion that noise associated with working memory updating is at play in conditions involving stimulus uncertainty. Training on a roving frequency discrimination task differentially improved working memory capacity (compared to training on a fixed frequency discrimination) as measured using a tonal n-back task which required continual updating of tone representations but no fine discrimination of their frequencies (Zhang *et al.*, 2012). Moreover, training on the n-back task improved frequency discrimination for roving frequency stimuli. The learning was not specific to working memory for tones, and transferred to 3-back tasks with both visual shapes and auditory verbal stimuli (digits).

Taken together, these studies suggest that cognitive constraints such as working memory capacity may limit psychophysical performance, and that training on that psychophysical task, as well as training directed at the limiting process, serve to lift these constraints and improve performance on the trained task and other tasks constrained by the same limiting process (i.e., transfer). They also highlight the potential advantage of removing these types of processing limitations through training (i.e., transfer to very different tasks and between modalities). While perceptual learning is often very specific to the trained condition when the noise is of sensory origin, removing cognitive limitations appears to lead to transfer of learning to a much broader skill set (see also Green and Bavelier, 2003; Li *et al.*, 2009).

Reducing decision inefficiency due to response bias

Psychophysical thresholds are generally considered to measure perceptual sensitivity, but elevated thresholds can result from suboptimal placement of the criterion in decision making (Fig. 1B). Ideally, the decision criterion should be placed so as to maximise percent correct on the task. But an incorrect assessment of the utility (i.e., believing one response to be more beneficial; Maddox and Bohil, 2001), or erroneous *a priori* assumptions about the statistics of stimulus presentation (e.g., believing one response to be more likely to occur; Tanner *et al.*, 1967) that disregard the sensory evidence, can result in a systematic shift from the ideal criterion placement (bias), with an associated cost to performance.

Ratcliffe *et al.* (2012) have shown that bias in a yes/no amplitude-modulation detection task is reduced through training. Listeners were initially inclined to be liberal in their responses, responding ‘yes’ (‘signal present’) more often than ‘no’ (‘signal absent’). Training reduced this propensity. Even in multi-interval forced-choice procedures, considered to be bias-free, we have observed a response bias in naïve listeners that changed over the course of training (Halliday *et al.*, 2011).

Criterion placement can also be influenced by the responses to preceding trials. Jones *et al.* (2012) found this dynamic type of bias to be present in two-interval, forced-choice frequency discrimination tasks. Listeners were inclined to perseverate in their response choice after a correct response and alternate after an incorrect response. The bias was reduced, though not entirely eradicated, by training.

Simulations showed that this bias reduction could account for over one third of the shift in discrimination thresholds on psychoacoustic tasks.

Sources of decision inefficiency as well as sensory noise and cognitive constraints can therefore adversely affect performance, and play a part in perceptual learning.

AUDITORY LEARNING IN TYPICALLY DEVELOPING CHILDREN

In the previous section we have shown that perceptual learning can be described as a reduction in noise of sensory or cognitive origin, or inefficiencies associated with the decision. But learning in young adults, where both sensory and cognitive functions are largely mature, may be very different from learning in children. Children not only appear to have a greater degree of internal noise than adults (Buss *et al.*, 2006), but their perceptual performance may also be subject to constraints imposed by different sources of noise due to the different maturational trajectories of sensory and cognitive processes. While the ascending, sensory system is largely mature by 2 years of age (Moore, 2002), more central and cognitive functions continue to develop into adolescence and even adulthood (e.g., Bishop *et al.*, 2011; Moore and Linthicum, 2007). It is likely therefore that cognitive limitations will play a greater role than sensory limitations in children's difficulties in performing perceptual tasks (see Moore, 2012).

Indeed, Halliday *et al.* (2008) provided evidence in support of this suggestion by training 6-11 year old children on a frequency discrimination task with a fixed standard frequency. The children could be divided into subgroups based on their performance: Some were able to perform the task at the same level as naïve adults even without training ('adult-like'), some started with poorer performance but achieved adult-like performance levels with training ('trainable'), and some failed to achieve adult-like performance at any point ('non-adult-like'). Following training the children were tested on frequency discrimination with a roving standard frequency (Fig. 5 in Halliday *et al.*, 2008). Children in the non-adult-like and trainable subgroups had roving frequency difference limens that did not significantly differ from their pre-training fixed frequency difference limens. The adult-like subgroup, like adults (Fig. 3), had higher difference limens for roving- than fixed frequency discrimination. This transfer pattern suggests that the non-adult-like and trainable subgroups experience different limiting processes to the adult-like subgroup in their performance of the *fixed* frequency discrimination. It is possible these subgroups are unable to use the repetition in the stimuli to form perceptual anchors, resulting in similar discrimination limens for fixed and roving stimuli. This limiting process may have been learned by the trainable subgroup when training on fixed frequency discrimination, which would also explain why their learning failed to transfer to the roving condition.

The non-adult-like subgroup comprised of younger children (Table II in Halliday *et al.*, 2008) who had attentional lapses on 6.5% of trials (assessed as errors on trials in which the frequency difference was easily discriminable). In the two other subgroups the children were of a similar age and non-verbal IQ, but those who had adult-like performance from the start were distinguished by committing even fewer attentional lapses (1.1% in trainable, 0.1% in adult-like). Halliday *et al.* (2008) concluded that the inability to sustain attention was a limitation on frequency discrimination performance in the non-adult-like children. Until this bottleneck was removed, perceptual sensitivity could not increase through training.

Although inattention plays a large role in children's performance and ability to learn, in well motivated young adults it is unlikely to be an important factor in performance, or necessarily change significantly through training (see Jones *et al.*, 2013). This difference highlights the danger of applying learning rules derived from adults to children. Children need to overcome very different limitations to adults when training, so in effect they may be learning very different things. Moreover, it is likely that children will show a very different pattern of transfer to adults, because different tasks will be sharing the learned limiting process.

In addition to inattention, children may experience other performance bottlenecks different to those of adults. Although decisions and responses are identical in the model described in Fig. 1, this is not necessarily the case in every perceptual judgment task, and a distinction between these two processes may be more pronounced in children. For instance, motor errors may result in the response deviating from that intended, or the child may correctly identify the response but forget which key to press. Children may also be more susceptible to bias effects, though we are not aware of evidence in support of that.

A NOTE ON INDIVIDUAL VARIABILITY AND ATYPICAL LISTENERS

It is not only children who can be divided into subgroups with distinctly different learning and transfer patterns. Amitay *et al.* (2005) found a different learning and transfer pattern in 'good' and 'poor' listeners (distinguished by initially low or high discrimination limens, respectively). Unlike the good listeners described in Fig. 3, poor listeners had similar untrained thresholds for fixed- and roving frequency discrimination (Fig. 4). Surprisingly, it was the poor listeners in the group trained on fixed frequency that showed complete transfer to the roving frequency condition, while the poor listeners trained on roving showed only partial transfer to fixed-frequency (limens were lower than naïve, but higher than trained).

Although different sensory noise sources may contribute to the performance differences between good and poor listeners, it is more likely that the differences were rooted in cognitive limitations. Since both initial and transfer thresholds are similar for the fixed and roving conditions in the two groups of poor listeners, we could have concluded that the two training groups learned the same limiting process. However, the performance on the last training block is very different in the two groups, suggesting the learning is different. It is possible that both groups share learning of one limiting process but not another. For example, if the limitation imposed on poor listeners was of poor working memory affecting the forced-choice comparisons, both groups may have started by learning this process. If listeners in the fixed frequency training group then proceeded to improve thresholds through perceptual anchors in addition to working memory, this would explain transfer of learning to the roving condition. However, it is more difficult to reconcile the results for the poor listeners in the roving frequency training group with those from the good listeners, who showed full rather than partial transfer from roving- to the fixed frequency condition.

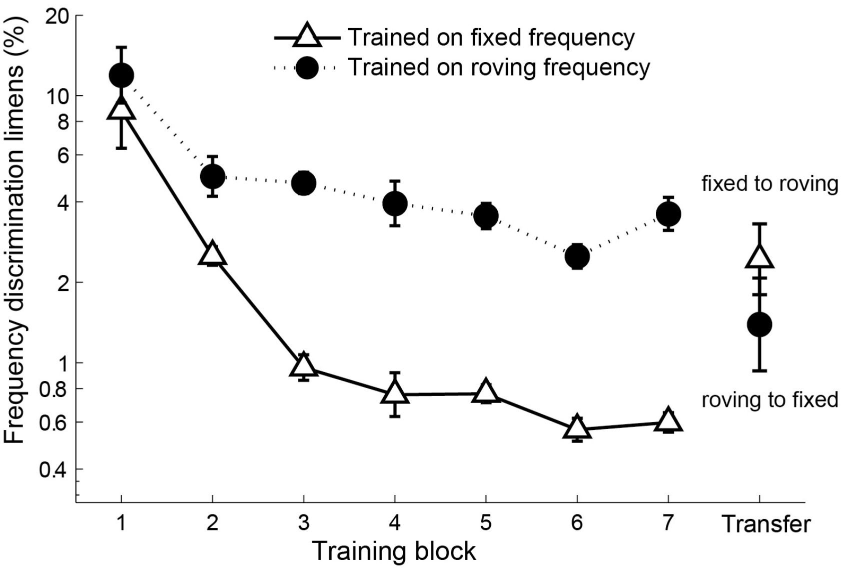


Fig. 4: Learning and transfer on a frequency discrimination task with a fixed- and roving frequency standard stimulus in poor listeners, defined based on naïve (block 1) limens. Initial discrimination limens (in percent of the standard frequency) are similar for the fixed- and roving frequency conditions both before and after training. Each training block consisted of 500 trials. Limens are adjusted for initial (block 1) performance. Adapted from Amitay *et al.* (2005).

Although we can only speculate on working memory as the performance-limiting noise that is removed through training in the Amitay *et al.* (2005) study, there is some evidence from other studies in atypical listeners that this may indeed be the case. For example, young adults with reading difficulties were shown to have similar limens for discrimination on fixed- and roving frequency tasks (Ahissar, 2007). This was interpreted as the inability to form perceptual anchors even when the task allowed for it. Training reading-disabled teenagers improved their ability in the fixed frequency condition, as well as showing transfer to an improvement in working memory function, suggesting cognitive-based constraints on performance are removed through training (Banai and Ahissar, 2009).

These results demonstrate that even within the adult population there is great variability in performance, some of which can be attributed to different limiting processes, or sources of noise, most likely of cognitive origin.

SUMMARY AND CONCLUSIONS

We have presented a view of perceptual learning as a process of removing performance-limiting constraints due to internal noise of both sensory (bottom-up) and cognitive (top-down) origin, as well as inefficiencies associated with the decision process. In this view, learning transfers from trained to untrained tasks when they share the limiting process that has been trained. We have shown that limiting processes may be different in adults and children (as well as other atypical populations), and that these differences affect not only what is learned but also how the learning transfers. Thus, the benefit of training a particular task may be specific to the trained population.

There are several implications to this view. First, from an applied perspective we cannot take a 'one-size fits all' approach to the learning process. Caution must be employed when applying results from young, motivated and well-rewarded adults to children, atypical learners or elderly populations (who may also suffer sensory and/or cognitive decline).

Second, from a more theoretical perspective, we should not assume that learning is a unitary and continuous process. The learning curve may actually result from a conglomeration of multiple effects, with different noise sources coming to the fore once the initially dominant noise source has been addressed. This also suggests that rather than just the length of practice (Jeter *et al.*, 2010), transfer of learning may depend on the cascade of limitations lifted through training. In support of this suggestion, Wright *et al.* (2010) have shown that transfer lags behind learning. Our interpretation of this would be that initial learning on the trained task addressed a limitation that was not shared by the transfer task, and that only once the shared limitation was learned did the learning transfer.

Finally, we offer a word of caution in interpreting the learning and transfer effects claimed for commercial training programs designed to address a variety of perceptual, cognitive and language difficulties through perceptual and/or cognitive training (e.g., Fast ForWord™: Tallal *et al.*, 1996). It is possible that the lack of conclusive evidence as to the efficacy of these programs lies in the choice of outcome measures, in terms of whether or not they share processing limitations with the trained tasks. In fact, these programs may be training something altogether different than the claims of their authors. For example, Fast ForWord™ may improve language by training the ability to selectively attend to sound (see Stevens *et al.*, 2008), or by lifting working memory constraints on updating rapidly presented stimuli rather than sensory-perceptual constraints on brief and rapidly presented stimuli, *per se*. A better understanding of the limiting processes in the target populations is imperative to help develop and optimise further training programs to address perceptual and cognitive processing limitations.

REFERENCES

Ahissar, M. (**2007**). "Dyslexia and the anchoring-deficit hypothesis," Trends Cogn. Sci., **11**, 458-465.

Amitay, S., Hawkey, D.J.C., and Moore, D.R. (**2005**). "Auditory frequency discrimination learning is affected by stimulus variability," Percept. Psychophys., **67**, 691-698.

Amitay, S., Zhang, Y.-X., and Moore, D.R. (**2012**). "Asymmetric transfer of auditory perceptual learning," Front. Psychol., **3**, 508.

Amitay, S., Guiraud, J., Sohoglu, E., Zobay, O., Edmonds, B.A., Zhang, Y.X., and Moore, D.R. (**2013**). "Human decision making based on variations in internal noise: an EEG study," PLoS ONE, **8**, e68928.

Banai, K., and Ahissar, M. (**2009**). "Perceptual learning as a tool for boosting working memory among individuals with reading and learning disability," Learn. Percept., **1**, 115-134.

Banai, K., and Amitay, S. (**2012**). "Stimulus uncertainty in auditory perceptual learning," Vision Res., **61**, 83-88.

Bishop, D.V.M., Anderson, M., Reid, C., and Fox, A.M. (**2011**). "Auditory development between 7 and 11 years: An event-related potential (ERP) study," PLoS ONE, **6**, e18993.

Braida, L.D., Lim, J.S., Berliner, J.E., Durlach, N.I., Rabinowitz, W.M., and Purks, S.R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context-coding," J. Acoust. Soc. Am., **76**, 722-731.

Buss, E., Hall, J.W., III, and Grose, J.H. (**2006**). "Development and the role of internal noise in detection and discrimination thresholds with narrow band stimuli," J. Acoust. Soc. Am., **120**, 2777-2788.

Dosher, B.A., and Lu, Z.L. (**2005**). "Perceptual learning in clear displays optimizes perceptual expertise: Learning the limiting process," Proc. Natl. Acad. Sci. USA, **102**, 5286-5290.

Fox, M.D., Snyder, A.Z., Zacks, J.M., and Raichle, M.E. (**2006**). "Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses," Nat. Neurosci., **9**, 23-25.

Fox, M.D., Snyder, A.Z., Vincent, J.L., and Raichle, M.E. (**2007**). "Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior," Neuron, **56**, 171-184.

Green, C.S., and Bavelier, D. (**2003**). "Action video game modifies visual selective attention," Nature, **423**, 534-537.

Green, D.M., and Swets, J.A. (**1966**). *Signal Detection Theory and Psychophysics* (John Wiley & Sons, New York).

Halliday, L.F., Taylor, J.L., Edmondson-Jones, A.M., and Moore, D.R. (**2008**). "Frequency discrimination learning in children," J. Acoust. Soc. Am., **123**, 4393-4402.

Halliday, L.F., Moore, D.R., Taylor, J.L., and Amitay, S. (**2011**). "Dimension-specific attention directs learning and listening on auditory training tasks," Atten. Percept. Psycho., **73**, 1329-1335.

Javel, E., and Viemeister, N.F. (**2000**). "Stochastic properties of cat auditory nerve responses to electric and acoustic stimuli and application to intensity discrimination," J. Acoust. Soc. Am., **107**, 908-921.

Jeter, P.E., Dosher, B.A., Liu, S.H., and Lu, Z.L. (**2010**). "Specificity of perceptual learning increases with increased training," Vision Res., **50**, 1928-1940.

Jones, P.R., Moore, D.R., and Amitay, S. (**2012**). "The role of response bias when learning a forced-choice task," Proceedings of the British Society of Audiology (BSA) Annual Conference, Nottingham, United Kingdom.

Jones, P.R., Shub, D.E., Moore, D.R., and Amitay, S. (**2013**). "Reduction of internal noise in auditory perceptual learning," J. Acoust. Soc. Am., **133**, 970-981.

Li, R., Polat, U., Makous, W., and Bavelier, D. (**2009**). "Enhancing the contrast sensitivity function through action video game training," Nat. Neurosci., **12**, 549-551.

Macmillan, N.A., and Creelman, C.D. (**2005**). *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates Inc., Mahwah, New Jersey).

Maddox, W.T., and Bohil, C.J. (**2001**). "Feedback effects on cost-benefit learning in perceptual categorization," Mem. Cognit., **29**, 598-615.

Moore, B.C.J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am., **54**, 610-619.

Moore, D.R. (**2002**). "Auditory development and the role of experience," Br. Med. Bull., **63**, 171-181.

Moore, D.R. (**2012**). "Listening difficulties in children: Bottom-up and top-down contributions," J. Comm. Disord., **45**, 411-418.

Moore, J.K., and Linthicum, F.H., Jr. (**2007**). "The human auditory system: A timeline of development," Int. J. Audiol., **46**, 460-478.

Ratcliffe, N., Jones, P.R., Moore, D.R., and Amitay, S. (**2012**). "The role of response bias when learning a yes/no task," Proceedings of the British Society of Audiology (BSA) Annual Conference, Nottingham, United Kingdom.

Stevens, C., Fanning, J., Coch, D., Sanders, L., and Neville, H. (**2008**). "Neural mechanisms of selective auditory attention are enhanced by computerized training: Electrophysiological evidence from language-impaired and typically developing children," Brain Res., **1205**, 55-69.

Tallal, P., Miller, S.L., Bedi, G., Byma, G., Wang, X., Nagarajan, S.S., Schreiner, C., Jenkins, W.M., and Merzenich, M.M. (**1996**). "Language comprehension in language-learning impaired children improved with acoustically modified speech," Science, **271**, 81-84.

Tanner, T.A., Haller, R.W., and Atkinson, R.C. (**1967**). "Signal recognition as influenced by presentation schedules," Percept. Psychophys., **2**, 349-358.

Vogels, R., Spileers, W., and Orban, G.A. (**1989**). "The response variability of striate cortical neurons in the behaving monkey," Exp. Brain. Res., **77**, 432-436.

Wright, B.A., Wilson, R.M., and Sabin, A.T. (**2010**). "Generalization lags behind learning on an auditory perceptual task," J. Neurosci., **30**, 11635-11639.

Zhang, Y.-X., Moore, D.R., Molloy, K., and Amitay, S. (**2012**). "Bidirectional transfer of working memory and perceptual learning," Proceedings of the British Association for Cognitive Neuroscience (BACN), Newcastle, United Kingdom.

More than adaptation – evidence for training-induced perceptual learning of time-compressed speech

KAREN BANAI[1,*] AND YIZHAR LAVNER[2]

[1] *Department of Communication Sciences and Disorders, University of Haifa, Haifa 31905, Israel*

[2] *Department of Computer Science, Tel Hai College, Tel Hai 12208, Israel*

The identification of time-compressed speech improves significantly following short-term exposure, but it is not clear whether additional practice yields additional learning. The goal of the experiment reported here was to determine whether 30-40 minutes of training, during which listeners practiced the identification of 100 different time-compressed sentences, yielded additional learning to that induced by a single brief exposure to 20 sentences. We also asked if this learning generalized to novel sentences and to a new speaker. Training resulted in more learning than a single brief exposure, and this learning generalized to a new speaker but not to new tokens. Brief exposure to 20 sentences did not result in any significant increases to performance when compared to naive listeners. We conclude that a prolonged learning phase exists for time-compressed speech, but that learning during this phase does not fully transfer to new, untrained tokens.

INTRODUCTION

The identification of time-compressed speech, an artificially created form of rapid speech, improves rapidly with exposure to a few time-compressed sentences, a phenomenon referred to as adaptation (e.g., Sebastian-Galles *et al.*, 2000; Peelle and Wingfield, 2005) or perceptual adjustment (e.g., Dupoux and Green, 1997; Pallier *et al.*, 1998). However, whether learning beyond this brief adaptation phase also occurs, and if so whether its characteristics are distinct from those of initial adaptation, remains unclear, because systematic training on more than 10-20 stimuli has been rare. Consistent with the finding that even highly experienced non-native speakers benefit from slower than normal presentation (Conrad, 1989; Zhao, 1997), we have previously observed a prolonged learning phase on a time-compressed speech identification task among non-native speakers of Hebrew (Banai and Lavner, 2012). The goal of the experiment presented here was to extend these findings to the learning of time-compressed speech in native speakers.

Relatively brief adaptation (10-20 sentences) to time-compressed speech substantially improves its perception, albeit not perfectly so. Previous reports suggest that after such exposure, performance improves from 20-76% correct to the range of 40-85% correct for the level of compression used during adaptation (Altmann and Young, 1993; Dupoux and Green, 1997; Pallier *et al.*, 1998;

*Corresponding author: kbanai@research.haifa.ac.il

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Sebastian-Galles *et al.*, 2000; Golomb *et al.*, 2007; Adank and Janse, 2009). In these studies, improvement appears quite general in the sense that transfer was observed across stimuli and even across languages. However, it is not clear whether performance does not reach ceiling (100% correct) due to inherent limitations of the learning process or due to other factors such as the duration of training. In a previous study we have shown that the latter might be one of the reasons (Banai and Lavner, 2012). In this study, we trained non-native speakers of Hebrew on the semantic verification of time-compressed sentences using an adaptive procedure. Listeners improved significantly over the course of a training program in which they had to verify 300 sentences per session for five sessions. Post training, the ability of trained listeners, but not of untrained controls who participated in pre- and post-test sessions only, to verify the trained sentences, became as good as that of naive native speakers. Both trained non-native listeners and naive native ones were able to consistently verify sentences compressed to less than 30% of their original length. The effects of learning generalized to the identification of time-compressed sentences produced by different talkers but not to untrained sentences or single words, leading us to hypothesize that prolonged learning might constitute a different, more specific form of learning than that observed after brief adaptation.

Although our previous study suggests that prolonged learning does occur on time-compressed speech, we could not document its presence among native speakers. During the pre-test phase native listeners could consistently verify sentences compressed to the maximum level of compression we allowed the adaptive procedure to reach during this phase (20%), making it impossible to uncover any further learning. It is thus still possible that the prolonged learning observed among non-native speakers simply reflect slower adaptation. Therefore, we now ask whether the identification of time-compressed speech continues to improve beyond the effects of adaptation to 20 sentences in native speakers of Hebrew. We also ask whether the pattern of generalization is similar or different from that observed among non-native speakers and after brief adaptation.

METHODS

Participants

Thirty native Hebrew speakers participated in this experiment. All were undergraduate University of Haifa students, with no history of hearing, learning, or language problems. Participants were paid for their participation. All aspects of the study were approved by the ethics committee of the Faculty of Social Welfare and Health Sciences at the University of Haifa. Participants were divided into three groups: a trained group (n = 10), an untrained control group (n=10), and a group of naive listeners (n = 10).

Organization of the experiment

The experiment had three phases. A pre-test on which trained listeners and untrained controls were exposed to 20 sentences (taken from the training set) presented by a male speaker and compressed to 30% of their naturally spoken duration; a training session during which trained listeners practiced a compressed speech verification task during which the degree of compression varied adaptively based on performance; and a post-test on which trained listeners, untrained controls (who attended the pre-test but not the training session) and naive listeners (who participated only in the post-test) were exposed to the 20 sentences from the pre-test as well as the same 20 sentences presented by a different speaker and 20 novel sentences spoken by the trained speaker, all compressed to 30%.

Stimuli

Stimuli were simple active subject-verb-object sentences in Hebrew, each 5-6 words long, taken from Prior and Bentin (2006). A total of 120 sentences were used. One hundred sentences formed the training set. The other 20 were used to assess the generalization of learning to untrained tokens. Sentences were recorded and then compressed with a WSOLA algorithm (Verhelst and Roelands, 1993) implemented in Matlab. For further details see Banai and Lavner (2012).

Tasks

Pre- and post-tests. Sentences were presented in blocks of 20 sentences. After hearing each sentence, listeners were asked to write it down as accurately as they could. No feedback was provided during this phase.

Training. Five blocks of 60 sentences selected at random from the training set were presented. On each trial listeners had to determine whether the sentence they heard was semantically plausible (e.g., ‘the grumpy waiter served the soup’) or implausible (e.g., ‘the grumpy potato served the soup’). Initial compression level was 65%. Subsequently, the level of compression changed based on listeners response using a 2-down/1-up staircase procedure. Feedback was provided for both correct and incorrect responses. For further details see Banai and Lavner (2012).

Experimental Conditions

The trained condition was comprised of 100 sentences presented by a male talker (designated the trained talker).

Three additional conditions were used:

1) Trained tokens: 20 sentences, randomly selected from the training set presented by the trained talker. These were presented to trained and control listeners during the pre- and post-test phases and to naive listeners during the post-test.
2) Untrained tokens: 20 sentences, not included in the training set, presented by the trained talker. These were administered during the post-test only to all listeners.
3) Untrained talker: The 20 trained tokens from above presented by a different speaker. These were administered during the post-test only to all listeners.

RESULTS

Training-induced learning

To determine whether additional learning to that induced by exposure to 20 sentences occurred, the learning curves from the 5 blocks of the training phase were analyzed. The mean group and the individual learning curves are shown in Fig. 1. The slopes of 9/10 individual curves were negative, indicating that in general, thresholds tended to improve with practice. The mean slope (−0.013 ± 0.01) was significantly negative with a 95% confidence interval of −0.019 to −0.006. A repeated measures ANOVA over the training blocks with contrasts comparing the mean of each block to the mean of all previous blocks suggests that learning was evident starting the third block and continued through the fourth block of training ($F(4,36) = 7.69$, $p < 0.001$; block 2 vs. block 1: $F(1,9) = 2.14$, $p = 0.18$; block 2 vs. previous blocks: $F(1,9) = 6.45$, $p = 0.032$; block 4 vs. previous blocks: $F(1,9) = 25.56$, $p = 0.001$; block 4 vs. previous: $F(1,9) = 4.49$, $p = 0.063$). Together it therefore appears that significant perceptual learning on a time-compressed speech task continues even beyond the initial adaptation phase.

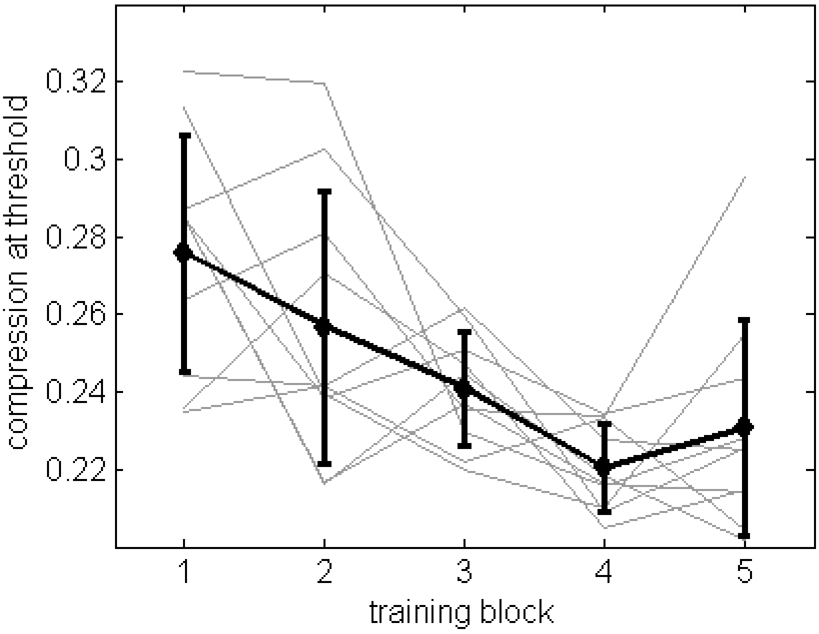


Fig. 1: Mean group learning curve (thick line) and individual curves (thin lines). Thresholds were determined as the mean compression over the last 5 reversals in a given block of trials.

Training versus rapid-learning

To determine whether the training-induced learning (discussed above) was significantly greater than the rapid learning induced by participating in the pre- and post-test only, performance on the trained tokens was compared between trained and control listeners. As shown in Fig. 2, pre- to post-test improvement was greater in the trained than in the control group. A significant group × test session interaction in an ANOVA with session as within subject factor and group as a between factor one ($F(1,17) = 10.69$, $p = 0.005$, partial $\eta^2 = 0.39$) suggested that the improvement was significantly greater among trained listeners. Likewise, an analysis of co-variance (ANCOVA, with pre-test scores as covariate) on the percentage of words correctly recognized during the post-test also suggests that mean post-training performance is significantly better among trained listeners even after taking into account putative pre-test differences ($F(1,18) = 31.84$, $p < 0.001$, partial $\eta^2 = 0.66$). Significant generalization of learning to the untrained talker condition ($F(1,18) = 8.47$, $p = 0.010$, partial $\eta^2 = 0.35$), but not to the untrained tokens condition ($F(1,18) = 0.38$, $p = 0.54$, partial $\eta^2 = 0.02$) was also observed.

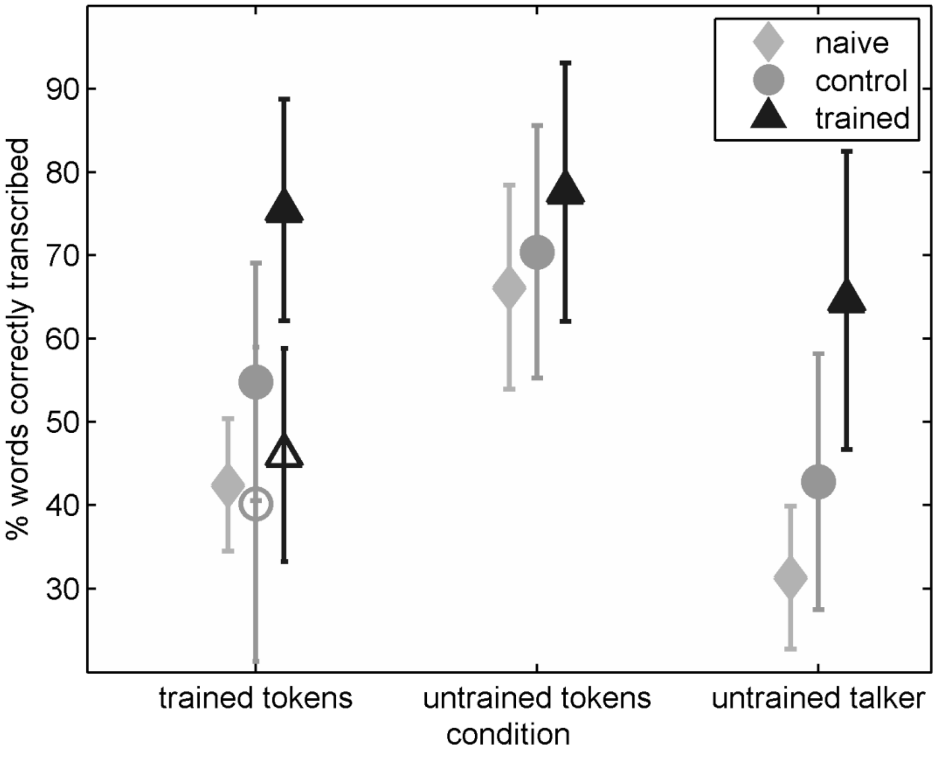


Fig. 2: Mean group performance across conditions and test sessions. Empty symbols denote pre-test performance; filled symbols denote post-test and naive performance. Error bars are ±1 standard deviation.

Training and pre-test participation versus naive performance

Finally, post-test performance was compared across the 3 group of listeners. ANOVAs suggest group differences on two of the three conditions (trained tokens: $F(2,26) = 31.46$, $p < 0.001$; untrained talker: $F(2,26) = 13.91$, $p < 0.001$), but not on the untrained tokens condition. To determine whether significant group differences were associated with the initial learning that occurred during the pre-test or with the effects of training, planned comparisons were carried out. Controls and naive listeners performed similarly on all conditions (see Fig. 2), suggesting that pre-test participation did not yield meaningful gains relative to naive performance. On the other hand, controls were significantly poorer than trained listeners on both the trained tokens condition ($t = 6.14$, $p < 0.001$) and the untrained talker condition ($t = 3.18$, $p = 0.004$). Together these suggest that group differences arose because training yielded more benefits than either pre-test or post-test participation, and pre-test induced learning had no lasting effects when compared with naive performance.

DISCUSSION

Similar to our previous findings in non-native speakers (Banai and Lavner, 2012), we now report that prolonged, training-induced learning of time-compressed speech occurs also among native speakers of Hebrew. Learning generalized to a novel talker of the opposite sex, but appeared specific to the sentences encountered during training. This pattern of learning and generalization suggests that training-induced learning might be more specific than adaptation-induced gains. Therefore, we conclude that adaptation-induced and training-induced learning of time-compressed speech might represent different types or phases of learning, such that learning starts with a rapid and broad phase during which learning generalizes quite widely and continues with a slower and more stimulus specific phase. This conclusion is consistent with the Reverse Hierarchy Theory (RHT) of perceptual learning (e.g., Ahissar *et al.*, 2009).

In contrast to the present findings, studies on the rapid adaptation to time-compressed speech suggest that learning during this phase is not stimulus-specific. For example, the ability of listeners to reproduce a specific time-compressed sentence was better if this sentence was encountered after 10 other sentences than if that same sentence was encountered after 5 other sentences (Dupoux and Green, 1997). Likewise, listeners who adapted to a set of 10 sentences in Catalan, reported a set of test sentences in Spanish more accurately than un-adapted controls (Pallier *et al.*, 1998). This was true even for listeners who spoke no Catalan, providing another demonstration of the generality of learning in this rapid phase. Generalization after training-induced learning in the current study was more limited. The lack of generalization to novel sentences in particular suggests that the learning we observed was different in nature than that reported in earlier studies. Otherwise trained listeners, who recognized a subset of the trained sentences more accurately than naive ones, would have also been more accurate on the set of untrained tokens.

ACKNOWLEDGEMENTS

This study was supported by the National Institute of Psychobiology in Israel.

We thank Yael Braver, Qamar Daher, Alex Frid, Dorit Nesher, and Neta Yuval-Weiss for help with various aspects of this project.

REFERENCES

Adank, P. and Janse, E. (**2009**). “Perceptual learning of time-compressed and natural fast speech,” J. Acoust. Soc. Am., **126**, 2649-2659.

Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (**2009**). “Reverse hierarchies and sensory learning,” Philos. Trans. R. Soc. Lond. B. Biol. Sci., **364**, 285-299.

Altmann, T.M., and Young, D. (**1993**). “Factors affecting adaptation to time-compressed speech,” Eurospeech '93, Berlin, 333-336.

Banai, K., and Lavner, Y. (**2012**). “Perceptual learning of time-compressed speech: more than rapid adaptation,” PLoS One, **7**, e47099.

Conrad, L. (**1989**). “The effects of time-compressed speech on native and EFL listening comprehension,” Stud. Second Lang. Acq., **11**, 1-16.

Dupoux, E., and Green, K. (**1997**). “Perceptual adjustment to highly compressed speech: Effects of talker and rate changes.” J. Exp. Psychol. Human, **23**, 914-927.

Golomb, J.D., Peelle, J.E., and Wingfield, A. (**2007**). “Effects of stimulus variability and adult aging on adaptation to time-compressed speech,” J. Acoust. Soc. Am., **121**, 1701-1708.

Pallier, C., Sebastian-Galles, N., Dupoux, E., Christophe, A., and Mehler, J. (**1998**). “Perceptual adjustment to time-compressed speech: a cross-linguistic study,” Mem. Cognit., **26**, 844-851.

Peelle, J.E., and Wingfield, A. (**2005**). “Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech,” J. Exp. Psychol. Human, **31**, 1315-1330.

Prior, A., and Bentin, S. (**2006**). “Differential integration efforts of mandatory and optional sentence constituents,” Psychophysiology, **43**, 440-449.

Sebastian-Galles, N., Dupoux, E., Costa, A., and Mehler, J. (**2000**). “Adaptation to time-compressed speech: Phonological determinants,” Percept. Psychophys., **62**, 834-842.

Verhelst, W., and Roelands, M. (**1993**). “An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech,” IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Minneapolis, MN, USA, 554-557.

Zhao, Y. (**1997**). “The effects of listeners' control of speech rate on second language comprehension,” Appl. Linguist., **18**, 49-68.

Assessing the benefits of auditory training to real-world listening: identifying appropriate and sensitive outcomes

HELEN HENSHAW[1,*] AND MELANIE FERGUSON[2]

[1] *NIHR Nottingham Hearing Biomedical Research Unit, School of Medicine, University of Nottingham, UK*

[2] *NIHR Nottingham Hearing Biomedical Research Unit, Nottingham University Hospitals NHS Trust, UK*

Auditory training is an intervention that aims to improve auditory performance and help alleviate the difficulties associated with hearing loss. To be an effective intervention, any task-specific learning needs to transfer to functional benefits in real-world listening. The present study aimed to identify optimal outcome measures to assess the benefits of auditory training for people with hearing loss. Thirty existing hearing-aid users with mild-moderate sensorineural hearing loss trained on a phoneme discrimination in noise task. Complex measures of listening and cognition were assessed pre- and post-training. Functional benefits to everyday listening were examined using a dual-task of listening and memory and an adaptive two-competing talker task. There was significant on-task learning for the trained task ($p < .001$), and significant transfer of learning to improvements in competing speech ($p < .05$) and dual-task performance ($p < .01$). For the dual-task, improvements were shown for a challenging listening condition (0 dB SNR), with no improvements where the task was either too easy (in quiet) or too difficult (-4 dB SNR). Findings suggest that for listening abilities, the development of complex cognitive skills may be more important than the refinement of sensory processing. Outcome measures should be sensitive to the functional benefits of auditory training and set at an appropriately challenging level.

INTRODUCTION

Accumulating evidence suggests that the challenges faced by older people with hearing loss cannot be explained by the audiogram alone (Kiessling *et al.*, 2003). Difficulties in hearing may be exacerbated by, or masquerade as, reductions in cognitive ability such as problems remembering or comprehending speech (Pichora-Fuller *et al.*, 1995).

Auditory training (AT) can be described as teaching the brain to listen through active engagement with sound (Henshaw and Ferguson, 2013). Typically, listeners learn to make perceptual distinctions between sounds (e.g., tones, phonemes, words) presented systematically. It is suggested that AT may lead to improvements in speech perception through the refinement of sensory processing (historically termed

*Corresponding author: helen.henshaw@nottingham.ac.uk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

analytic training), or the development of top-down repair strategies (synthetic training). A randomised controlled trial (RCT) of 50-74 year-old adults (n = 44) with mild sensorineural hearing loss (SNHL) who did not have hearing aids (Ferguson *et al.*, in press) showed significant improvements in a trained phoneme discrimination in quiet task (p < .001). Generalised improvements were shown for self-reported listening (particularly for a complex listening situation, p < .01, Cohen's d = .68), and complex cognitive tasks that engaged executive function (divided attention $p \leq$ .001, d = .53; working-memory updating p < .01, d = .50). No improvements were shown for simple cognitive tasks or perception of ASL sentences in modulated noise. These findings suggest that the development of complex cognition may be more important than the refinement of sensory processing to improve communication in everyday life.

The present study employed a short phoneme-discrimination-in-noise training task to identify appropriate outcomes that were sensitive to the functional benefits of AT for real-world listening in 30 adult hearing-aid (HA) users with mild-moderate SNHL, aged 50-74 years.

METHODS

Study design

A within-participant repeated measures design was used (Fig. 1). Participants attended two baseline outcome assessment sessions (T0 and T1) to help account for any procedural learning (test-retest) effects on outcome measure performance. This was followed by a 1-week no-contact control period and a second assessment session (T2). Participants then trained at home for one week before the final post-training assessment session (T3).

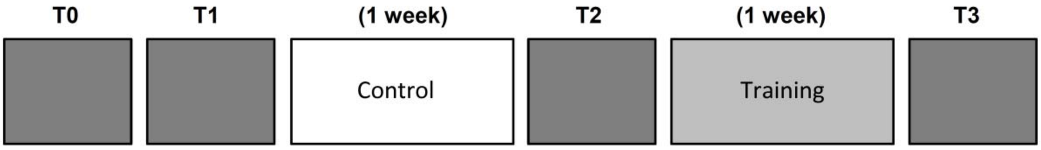


Fig. 1: Study design.

Participants

Thirty existing HA users (minimum HA experience = 3 months, mean = 10.3 years, SD = 10.7 years), aged 50-74 years (mean = 67.4 years, SD = 7.1 years) with mild or moderate SNHL (better-ear pure-tone thresholds averaged across 0.25, 0.5, 1, 2, and 4 kHz ranged between 21-69 dB HL, mean = 39.5 dB HL, SD = 12.7 dB), were recruited from the NIHR Nottingham Hearing Biomedical Research Unit research volunteer database.

Materials

Auditory training task: The phoneme-discrimination-in-noise task was delivered via computer game format (3I-3AFC oddball paradigm presented in ICRA multi-talker babble) using the IHR-STAR platform (for details, see Moore *et al.*, 2011). Participants trained using 11 different phoneme continua (/a/-/uh/, /b/-/d/, /d/-/g/, /e/-/a/, /er/-/or/, /i/-/e/, /l/-/r/, /m/-/n/, /s/-/sh/, /s/-/th/, and /v/-/w/). Each continuum transitioned from one phoneme to the other in 96 steps and was synthesised from end-points consisting of real voice recordings. Participants were presented with three discrete phonemes from one continuum per trial and were asked to identify the odd one out. Each phoneme continuum was presented for a block of 35 trials and the 11 continua were presented in sequential blocks on a rotational basis. A three-phase adaptive staircase procedure oddball response paradigm was used and threshold was the average of the last three trials in a block of 35 trials. Auditory and visual feedback (correct/incorrect response) was provided to participants after each trial. Participants completed two 15-minute training sessions each day, after which a graphical display showed the daily score for each continua plotted against their best score achieved. Visual rewards (on-screen fireworks) were shown when the participants improved on their previous best score.

Competing speech task: The Modified Coordinate Response Measure (MCRM) is a measure of speech intelligibility in the presence of a masker. The basic task, described by Hazan *et al.*, (2009), is based on the Coordinate Response Measure (Bolia *et al.*, 2000). For the present study, a single-talker masker was used. Participants were presented with sentences in the form of *'show the [animal] where the [colour] [number] is'*. There were six possible monosyllabic animals (cat, cow, dog, duck, pig, and sheep), six colours (black, blue, green, pink, red, and white) and eight numbers (1-9, excluding multisyllabic 7). Two sentences were presented concurrently, one by a female talker (target) and one by a male talker (distracter). Participants were asked to listen for the colour and number spoken by the female talker ('dog' was always the animal target) whilst ignoring the male talker, and to respond by pressing the corresponding target colour-number on a computer touchscreen. The test utilised an adaptive 1-up 1-down staircase method with an initial step size of 10 dB until reversal 1, reducing to 7 dB at reversal 2, and 4 dB at reversal 3 onwards. The test continued until eight reversals were achieved. Speech reception thresholds were calculated using the average of the last two reversals.

Letter-number sequencing task: A measure of working memory from the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 1997) was used. Participants were presented with a string of pre-recorded spoken numbers and letters and were asked to repeat them aloud, with the numbers in numerical order followed by the letters in alphabetical order. Sequences began at two items, with three trials at each sequence length. If the participant responded correctly for one out of the three sequence trials then the sequence length was increased by one item (up to a maximum sequence length of eight items), otherwise the test was discontinued. The task was scored as the total number of sequence trials correct.

Dual-task of listening and memory: The dual-task measured listening and memory, and was designed to assess listening effort (Howard *et al.*, 2010). Participants completed a five-digit memory task (secondary task) that flanked a speech-in-noise repetition task (primary task). A string of five digits was displayed visually on a computer screen for five seconds. Participants were asked to retain the digits in memory for later recall. Participants were then presented with a list of five AB Isophonemic Monosyllabic Words (Boothroyd, 1968) and asked to repeat each word immediately after presentation. After each word list, participants were asked to recall the five previously presented digits. Word lists were presented in three noise conditions (quiet, 0 dB, or −4 dB SNR using ICRA multi-talker babble). There were 12 word lists (four per condition), and the presentation order for noise conditions was counter-balanced across participants. This resulted in a maximum possible score of 20 correctly-repeated words and 20 correctly-recalled digits for each noise condition.

Procedure

Auditory training: Instructions and two initial (five-trial) phoneme-discrimination-in-noise training demonstration tasks were completed by participants alongside the researcher in the laboratory prior to commencing at-home training. Participants were asked to complete the training at home for 30 minutes a day (2 × 15 minute sessions with a minimum break of 15 minutes) for seven consecutive days (requested training duration = 3.5 hours), which equates to just over half the training provided in the previous RCT (6 hours; Ferguson *et al,* in press). Training was delivered, and responses logged, using a laptop computer (Toshiba A300), which was locked-down to run only the auditory training program. Auditory stimuli were delivered through Logitech LS11 speakers with a maximum signal level of 75 dB(A) at 30 cm.

Outcome assessment: Outcome measures were obtained at each outcome assessment session in the lab. Speech perception and cognitive tests took place in a quiet, purpose-designed test room. Auditory elements were delivered via a Logitech LS11 speaker placed directly in front of the participant at a distance of 1 m.

RESULTS

On-task learning

Participants trained at-home for an average of 197.8 minutes (SD = 28.7 minutes). A linear mixed model was used to assess any main effects of time (block) or phoneme continua (task) on phoneme discrimination thresholds and any task*block interaction. There was a highly-significant main effect of block ($F(1,1419.51) = 32.67$, $p < .001$) and phoneme-discrimination thresholds improved over time (Fig. 2). There was also a highly-significant main effect of task ($F(10,1414.43) = 22.33$, $p < .001$). A second linear mixed model with data divided by task showed a significant improvement by block for the majority of phoneme continua at either the $p < .001$ (/a/-/uh/, /i./-/e/), $p < .01$ (/er/-/or/, /m/-/n/, /s/-/th/, /v/-/w/), or $p < .05$ level (/e/-/a/, /l/-/r/). There was no significant improvement over time for three of the four phoneme continua that had the poorest initial thresholds, /s/-/sh/ ($p = .051$), /b/-/d/

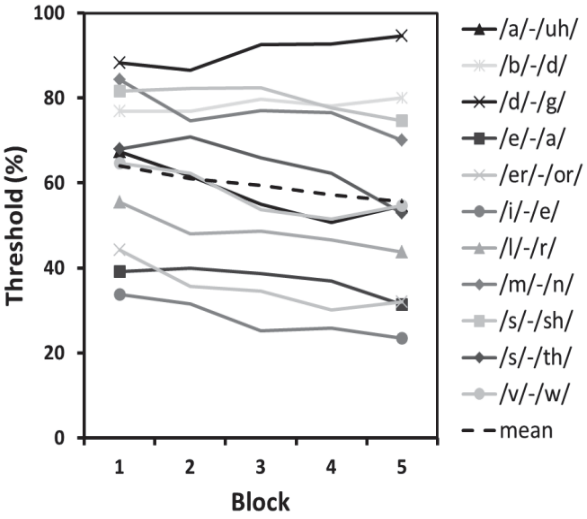


Fig. 2: Phoneme discrimination thresholds (across all participants) for each of the 11 phoneme continua over five training blocks; dashed line = group geometric mean.

(*p* = .855), and for /d/-/g/ performance got significantly worse (*p* < .001) over the course of training.

Transfer of learning to untrained measures

Identification of appropriate outcomes: competing speech

Analysis of performance for the competing-speech task across T1, T2, and T3 using a repeated measures ANOVA showed a significant main effect of time on speech reception thresholds (*F*(2,28) = 3.59, *p* < .05), see Fig. 3. Post-hoc comparisons showed no improvement for the control period (T1-T2), mean difference = −0.1, *p* = .89, and a significant improvement pre- to post-training (T2-T3), mean difference = 2.3 dB, *t*(29) = 2.55, p < .05, *d* = .47.

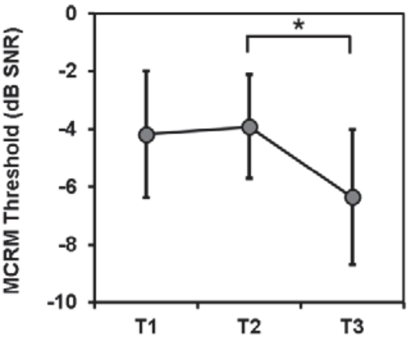


Fig. 3: Mean speech reception threshold (dB SNR) values for a two competing talker task (MCRM) with 95% confidence intervals at T1, T2, and T3, * *p* < .05.

Partial correlations controlling for age were used to explore the relationship between auditory and cognitive factors associated with performance on speech-perception tasks employed in either the present study (MCRM, two-competing-talker task), or in Ferguson *et al.,* in press (ASL sentences in 8-kHz modulated noise). Baseline pre-training measures at T1: better ear averaged hearing thresholds (BEA), self-reported listening (Initial Disability from the Glasgow Hearing Aid Benefit Profile), and working-memory (WM) scores (Digit Span forwards and backwards for Ferguson *et al.,* in press; Letter-Number Sequencing task for the present study), were correlated with baseline performance on the speech measures. Results are summarised in Table 1.

r =	BEA hearing thresholds	Self-reported listening	Working memory performance
Speech in noise (*n* = 44) (Ferguson *et al,* in press)	.38*	.08	.28
Competing speech (*n* = 30) (present study)	.49**	.45*	-.54**

Table 1: Partial correlations for baseline performance on speech-perception tasks (ASL sentences,) and (MCRM two competing talker task,), and baseline measures of better ear averaged hearing thresholds (BEA), self-reported listening, and working memory performance, * *p* < .05, ** *p* < .01.

Speech-perception performance on both tasks was significantly correlated with BEA hearing thresholds. Performance on the speech-in-noise task did not correlate significantly with self-reported listening or WM performance (Digit Span forwards and backwards). Performance on the competing speech task was significantly correlated with self-reported listening difficulties and with WM performance (Letter-Number Sequencing Task).

Identification of sensitive outcomes: dual-task of listening and memory

Individual task scores out of a possible 20 (number of digits correctly recalled and words correctly repeated) are plotted in Fig. 4.

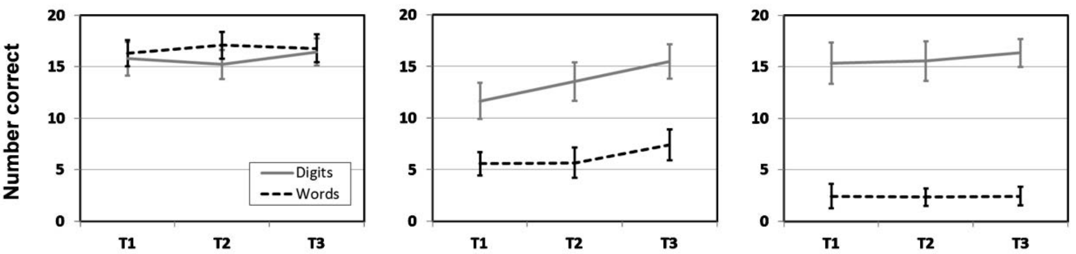


Fig. 4: Mean correct number of digits recalled and words repeated with 95% confidence intervals, across three noise conditions at T1, T2, and T3.

In quiet, performance was high for both the digit-recall and the word-repetition tasks. At 0 dB SNR, performance on the word-repetition task was reduced, with a reduction in performance for digit recall compared with the quiet condition. This may indicate an altered allocation of available resources to deal with the more difficult word-repetition demands. At −4 dB SNR, where participants were unable to identify the majority of words, digit-recall performance was once again comparable to that for the quiet condition.

Primary- and secondary-task scores were combined for each participant to give a dual-task score for each noise condition (maximum score = 40). A repeated-measures ANOVA showed no significant main effect of time on dual-task performance across the three noise conditions (*F*(2,87) = 1.75, *p* = .177), and no significant interaction between noise condition and time (*F*(2,87) = 0.33, *p* = .719). However, for the 0-dB SNR condition, where altered resource allocation was shown, there was a significant main effect of time on dual-task performance (*F*(2,28) = 7.72, *p* = .001). Post-hoc comparisons showed no improvement during the control period (T1-T2; mean difference = 0.2, *p* = 1.00), and a significant improvement pre- to post-training (T2-T3); mean difference = 3.6), *t*(29) = −4.24, *p* < .001, *d* = .77 (Fig. 5).

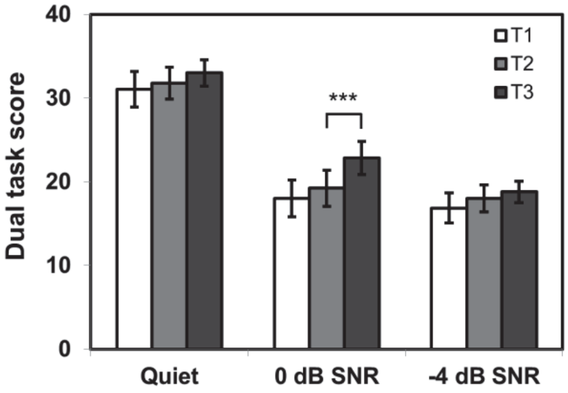


Fig. 5: Mean dual-task score for all participants with 95% confidence intervals across three noise conditions at T1, T2, and T3.

DISCUSSION

Results from the present study showed a significant improvement in phoneme-in-noise discrimination thresholds over time. The on-task learning effect was shown despite a substantially reduced AT schedule (just over half the training administered in Ferguson *et al.,* in press), and no significant improvements for three out of four of the trained phoneme continua with the poorest initial thresholds. As phoneme continua with the poorest initial thresholds improved the most during phoneme-discrimination-in-quiet training in the previous RCT (Ferguson *et al.,* in press), thus making the largest contribution to the on-task learning effect, it is likely that, these continua were too difficult for participants to discriminate when presented in a background of noise in the present study.

Despite a shorter auditory-training schedule and substantially less on-task learning than Ferguson *et al.,* (in press), generalised improvements were shown for a competing speech task that was associated with self-reported listening and cognitive abilities, and for a dual task of listening at a challenging SNR, but not where the task was too easy nor too difficult. These findings suggest that outcomes used to assess benefit of auditory training should be sensitive to the cognitive effects of training. Furthermore, benefits of training may be most evident when listening is challenging, and where resources need to be reallocated to meet listening demands. These results highlight a need for appropriate and sensitive outcomes to adequately assess the benefits of auditory training for people with hearing loss to ensure that those benefits are not overlooked.

ACKNOWLEDGMENTS

We would like to thank Dr. Daniel Clark and Miss Holly Thomas (NIHR Nottingham Hearing Biomedical Research Unit) for their assistance with study set up and data collection. Independent research presented in this article is funded by the National Institute for Health Research (NIHR) Biomedical Research Unit funding program. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

REFERENCES

Bolia, R.S., Nelson, W.T., Ericson, M.A., and Simpson, B.D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am., **107**, 1065-1066.

Boothroyd, A. (**1968**). "Developments in speech audiometry," Brit. J. Audiol, **2**, 3-10.

Ferguson, M.A., Henshaw, H., Clark, D.P.A., and Moore, D.R. (**in press**). "Benefits of phoneme discrimination training in a randomized controlled trial of 50-74 year olds with mild hearing loss," Ear Hearing.

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., and Shakespeare, B. (**2009**). "Speech perception abilities of adults with dyslexia: is there any evidence for a true defecit?" J. Sp. Lang. Hear. Res., **52**, 1510-1529.

Henshaw, H., and Ferguson, M.A. (**2013**). "Efficacy of individual computer-based auditory training for people with hearing loss: A systematic review of the evidence," PLoS ONE, **8**, e62836.

Howard, C.S., Munro, K.J., and Plack, C.J. (**2010**). "Listening effort at signal-to-noise ratios that are typical of the school classroom," Int. J. Audiol., **49**, 928-932.

Kiessling, J., Pichora-Fuller, M.K., Gatehouse, S., Stephens, D., Arlinger, S., Chisolm, T.H., Davis, A.C., Erber, N.P., Hickson, L., and Holmes, A.E. (**2003**). "Candidature for and delivery of audiological services: special needs of older people," Int. J. Audiol., **42**, S92-S101.

Moore, D.R., Cowan, J.A., Riley, A., Edmondson-Jones, A.M., and Ferguson, M.A. (**2011**). "Development of auditory processing in 6- to 11-yr-old children," Ear Hear, **32**, 269-285.

Pichora-Fuller, M.K., Schneider, B.A., Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise," J. Acoust. Soc. Am., **97**, 593-608.

Wechsler, D. (**1997**). *Wechsler Adult Intelligence Scale-3rd Edition (WAIS-3$^{®}$)* (San Antonio, TX: Harcourt Assessment).

Relationship of frequency-pattern training to speech perception

STANLEY SHEFT*, VALERIY SHAFIRO, AND KRISTEN CORTESE

Department of Communication Disorders and Sciences, Rush University Medical Center, 600 South Paulina Street, Suite 1012, Chicago, IL 60612, USA

Though discrimination of frequency patterns can relate to speech perception and the discrimination ability generally improves with training, the relationship between the training and speech perception is not known. Training regimens typically utilize simple repetition of the discrimination or identification trials. In the current work, the training protocol was based on interactive pattern reconstruction, increasing memory demands to accentuate learning. With either four- or five-tone patterns, the task was to assemble the constituent tones in the correct order. Tones were randomly selected from logarithmically scaled distributions (frequency: 400-1750 Hz, duration: 75-600 ms). In training but not test sessions, listeners were allowed multiple repetitions of the intact pattern to self-correct their interim response. To assess relationship to speech abilities, the same task was used in pre- and post-training measures with the tonal pattern replaced by samples of sinewave speech (SWS). Despite a high level of stimulus uncertainty, results showed a significant stimulus-specific benefit of training. Small but significant improvement in SWS intelligibility between pre- and posttest sessions was also obtained with greater relationship between results from intelligibility and pattern-reconstruction conditions post training.

INTRODUCTION

The frequency transitions and modulations of speech can serve multiple functions, not only conveying phonetic information but also enhancing signal coherence, segregation, and segmentation. Consistent with this involvement, Sheft *et al.* (2012a; 2012b) found significant age-mediated relationships in adults between speech perception in noise and the ability to discriminate random pitch patterns generated by frequency modulation (FM) of a tonal carrier with narrowband lowpass noise. Both behavioral and physiological measures show effect of auditory training on the processing of frequency patterns (Watson *et al.*, 1976; Tervaniemi *et al.*, 2001; Foxton *et al.*, 2004; Gaab *et al.*, 2006). Despite relationship between psychoacoustic and speech abilities, the effect of frequency-pattern training on speech perception is not known. The goal of the present study was to evaluate this effect as an initial step in determining the feasibility of a modified approach to frequency-pattern training in relationship to speech perception as a component of auditory rehabilitation.

*Corresponding author: stanley_sheft@rush.edu

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Training regimens in previous work have typically utilized simple repetition of the discrimination or identification trials used to assess pre- and post-training performance. The current experiment was designed to study the efficacy of a modified frequency-pattern training protocol. To make the task both more demanding and engaging for the listener, the basis of the protocol was interactive pattern reconstruction. Along with interactive participant involvement in the task through multiple self-corrected responses, the increased memory demands of the task were intended to accentuate learning.

Speech perception was primarily assessed using sinewave speech (SWS). SWS was chosen as a challenging and distorted stimulus set that would exhibit some level of direct relevance to the frequency patterns of the training protocol. In conjunction with measurement of the ability to reconstruct tonal frequency patterns, pattern reconstruction was also evaluated with segmented SWS tokens. The intent of theses SWS conditions was to use stimuli for which the intact pattern represented a form of speech in a task not requiring speech intelligibility.

METHOD

Participants were 26 normal-hearing English speakers (age: 20-28 yrs.), with 13 subjects in the experimental group which received training and 13 in the control group which did not. The experimental protocol consisted of a pretest, three training sessions, and a posttest. Six to seven days separated the pre- and posttest sessions. Speech intelligibility and psychoacoustic frequency-pattern reconstruction ability were assessed during the pre- and posttest sessions with training sessions only on the psychoacoustic task. All testing was with diotic stimulus presentation over headphones.

For the experimental group only, speech intelligibility was measured for AzBio sentences in comodulated noise (2.5-Hz lowpass noise modulator) at a −13 dB signal-to-noise ratio with speech at 70 dB SPL. The primary speech measures for both groups were intelligibility of SWS. SWS was generated with three component sinewaves continuously modulated in both frequency and amplitude with values estimated by linear prediction analysis. To set baseline intelligibility to allow for possible performance improvement and to strengthen relevance to the tonal frequency patterns of the training protocol (see below), component FM was lowpass filtered at 12 Hz with a 4th-order Butterworth filter, and amplitude variations were compressed by a factor of 0.7. SWS intelligibility was measured for vowel-consonant-vowel (VCV) tokens and consonant-nucleus-consonant (CNC) words. CNCs were scored both in terms of number of words and phonemes correct. Before scored testing, listeners were familiarized with SWS, first with sentences, and then with VCV and CNC tokens. Open-set SWS testing was at 75 dB SPL.

The initial condition of the pattern-reconstruction task used pure-tone stimuli. Working with four-tone frequency patterns, the listening task was to assemble the constituent tones in the correct order (Fig. 1). Subjects heard the target sequence only once, but could listen to both constituent tones and their interim reconstruction

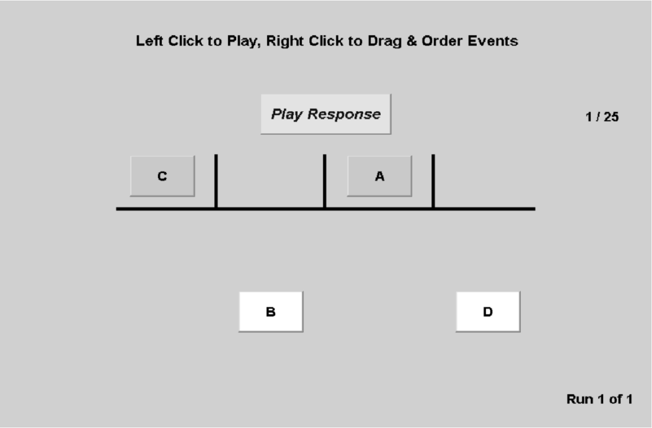


Fig. 1: Illustration of the pattern-reconstruction task with four elements. The pattern elements represented by the four boxes labeled A, B, C, and D are rearranged in the upper place holders to reconstruct the original pattern.

of the sequence as often as wanted. Correct-answer feedback for each sequence tone was provided after every trial.

Constituent tones of the patterns were randomly selected from logarithmically scaled distributions with frequency ranging from 400 to 1750 Hz and duration from 75 to 600 ms. These ranges were chosen to have a rough correspondence to the dominant regions of speech-element characteristics. To maintain discriminability of sequence components, any two component frequencies were separated by at least a factor of 1.2 with any two durations differing by at least a factor of 1.4. Constituent tones were shaped with a 50-ms rise/fall time or half the tone duration when less than 100 ms. The same task was also used in the pre- and posttest sessions with the tonal stimuli replaced by SWS VCVs and CNCs randomly segmented with exponential sampling (minimum duration: 75 ms) to create pattern elements. For each pattern-reconstruction condition (i.e., tone, VCV, CNC), data were collected from a single 25-trial block preceded by five practice trials. Stimulus level was 75 dB SPL.

Three training sessions between the pre- and posttest were completed by subjects on laptop computers at home. Each session lasted about 45 minutes. To document performance change, training sessions began with a repetition of the frequency-pattern task used in the pre- and posttest sessions. The actual training protocol required subjects to reconstruct five- instead of four-tone frequency patterns. Unlike the pre- and posttest sessions, during training, subjects were allowed multiple repetitions of the intact target sequence to self-correct their interim response. Subjects were encouraged to continue each training trial until they were confident that they had replicated the target sequence. Subjects completed two 25-trial training blocks in each session.

RESULTS

Figure 2 shows performance of the trained subject group on the five repetitions of the pattern-reconstruction task with four tonal elements. In this and all subsequent analysis, a Freeman-Tukey arcsine transform was applied to data before statistical analysis. A repeated-measures Analysis of Variance (ANOVA) showed a significant effect of repetition [$F(4,48) = 9.40$, $p < .001$, $\eta_p^2 = .44$], with Holm-Sidak *post-hoc* analysis indicating the contrasts of posttest performance with either the pretest or the first training session as the only significant differences. Across repetitions and subjects, error rate dropped by a factor of almost four with increasing stimulus duration within the 75-600 ms range. In contrast, no effect of stimulus frequency was observed within the range of frequencies used.

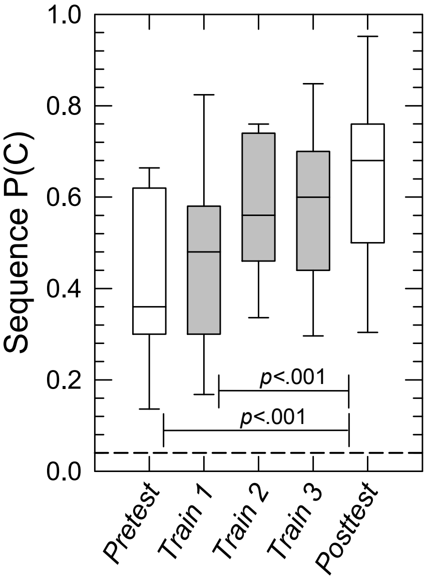


Fig. 2: For the trained subject group, box plots showing results in terms of proportion correct on the four-tone pattern-reconstruction task from pretest, the three training sessions, and posttest. The dashed line at the bottom indicates chance performance. Significant change re posttest is indicated by the horizontal line ending at the comparison condition.

Results from both subject groups comparing pre- to posttest performance with the three stimulus types in the pattern-reconstruction task are shown in Fig. 3. Across both groups, there was a significant effect of stimulus type in the pretest [$F(2,48) = 10.67$, $p < .001$, $\eta_p^2 = .31$] and posttest [$F(2,48) = 19.30$, $p < .001$, $\eta_p^2 = .45$], with *post-hoc* comparisons indicating significantly better performance ($p \leq .001$) in the VCV condition than with either CNC or tonal stimuli. Significant improvements from the pre- to posttest were found with each stimulus type for the trained group, and only in the VCV condition for the control group. For the trained subjects, the posttest change in performance obtained with tonal stimuli was significantly larger than that for VCV (t = 2.05, p = .05) or CNC (t = 2.74, p = .01) stimuli. Training effect was evaluated with a between-group Analysis of Covariance (ANCOVA) on posttest scores with the pretest as a covariate. In separate analyses for each stimulus type, a significant effect of training was obtained only with the tonal stimuli [$F(1,23) = 9.81, p = .005, \eta_p^2 = .30$].

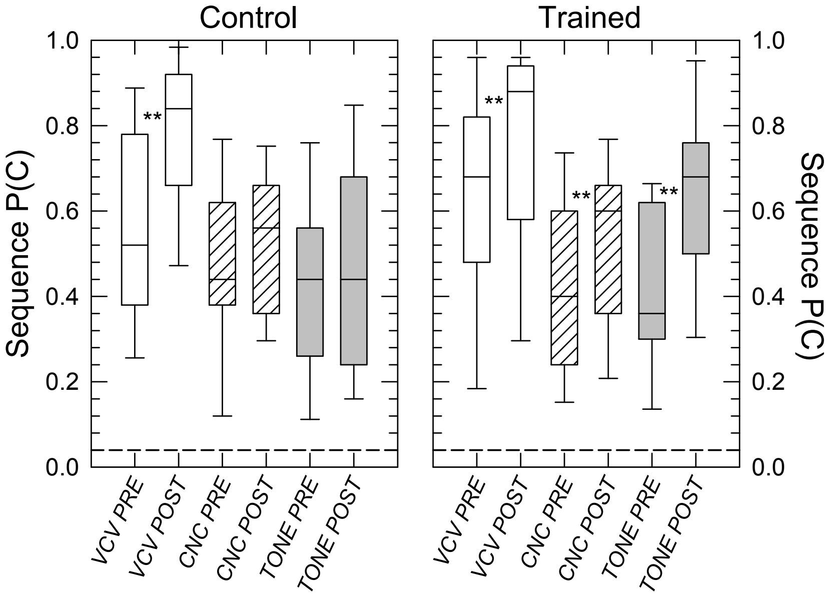


Fig. 3: For the control (left panel) and trained (right panel) subject groups, box plots showing pre- and posttest performance on the three pattern-reconstruction conditions using VCV, CNC, and tonal stimuli. Within group, significant change in performance from pre- to posttest at the p < .01 level is indicated by the double asterisk.

For the trained subjects, small but significant improvements in SWS intelligibility between the pre- and post-training sessions were obtained, with no effect on speech-in-noise performance as measured with AzBio sentences in modulated noise (Fig. 4). Controls showed no significant posttest improvement on any speech measure. Using an ANCOVA on posttest scores with pretest as a covariate, a significant effect of training was obtained only for VCV intelligibility [$F(1,23) = 7.95$, $p = .01$, $\eta_p^2 = .26$].

Pearson correlations are shown in Table 1 for both subject groups for pre- and posttest results from the three pattern-reconstruction conditions (i.e., VCV, CNC, and tone sequences) and two SWS measures. With the Bonferroni-Holm correction for multiple comparisons, only the two posttest relationships from the trained

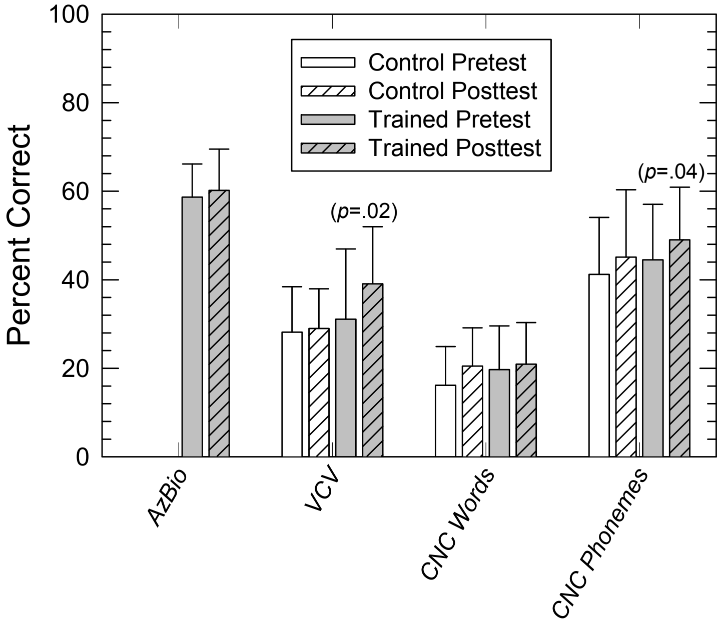


Fig. 4: For the control and trained subject groups, mean group performance on speech tests in terms of percent correct with error bars representing 1 SD. The control group was not tested with AzBio sentences. Within group, significant change from pre- to posttest is indicated by the level of significance in parentheses.

subjects involving the CNC phoneme measure and either the VCV- or CNC-sequence condition were significant in one-tailed testing. Thus, there was some indication of greater relationship post- than pre-training between pattern-reconstruction ability and speech perception as assessed by SWS intelligibility.

DISCUSSION

Watson *et al.* (1976) found that stimulus uncertainty diminished the benefit of training in learning to discriminate a local change to a single element of a tonal sequence. In contrast but consistent with results from Foxton *et al.* (2004), the current study demonstrated, despite a high level of stimulus uncertainty, significant improvement due to training on a task in which the signal was the pattern of the entire stimulus sequence. Though the current study intentionally employed an involved subject task, the effect of training was not solely procedural, as observed by dependence of both training effect and posttest correlation to speech intelligibility on stimulus type in the pattern-reconstruction conditions.

SWS stimuli were used in pattern-reconstruction task to incorporate speech structure into the signal pattern. Across all subjects, performance was best in the pattern task with the VCV stimuli in both the pre- and posttest measures, suggesting benefit from the added sequence structure. In separate pilot work, this benefit was lost when SWS modulation patterns were inverted to disrupt the speech-like structure of stimuli

Control

Pretest	VCV Sequence	CNC Sequence	Tone Sequence
VCV P(C)	-.243	-.057	-.418
CNC P(C)	.105	.541	.408

Posttest	VCV Sequence	CNC Sequence	Tone Sequence
VCV P(C)	-.279	.411	-.153
CNC P(C)	.283	.652	.428

Trained

Pretest	VCV Sequence	CNC Sequence	Tone Sequence
VCV P(C)	.341	.264	.038
CNC P(C)	.355	.665	.191

Posttest	VCV Sequence	CNC Sequence	Tone Sequence
VCV P(C)	.287	.467	-.045
CNC P(C)	**.709 (.04)**	**.703 (.04)**	.182

Table 1: For the control (top) and trained (bottom) subject groups, correlations between results from the three pattern-reconstruction conditions (VCV, CNC, and tone) and two SWS measures (VCV and CNC phonemes). Significant correlations are indicated in bold with the Bonferroni-Holm adjusted one-tailed *p* value in parentheses.

without altering the modulation statistics. In the present work, a role of sequence structure is also observed with performance from only the VCV and CNC pattern-reconstruction conditions showing significant correlation to a measure of speech intelligibility (Table 1). These relationships were obtained only posttest and only after training with a different stimulus type (i.e., tonal sequences). The benefit of training may then in part relate to change in the ability to utilize the manner by which auditory information is structured.

Despite indication of greater relationship post- than pretest between pattern-reconstruction ability and speech perception, the improvement in speech intelligibility due to training was small and limited to the SWS VCV stimulus set. Several factors may have contributed to this outcome. The absence of a training effect on sentence perception in noise may relate to speech redundancy so that young normal-hearing listeners were able to rely on cues other than frequency patterns and transitions, that is, stimulus aspects that were trained. This speculation suggests potentially greater training benefit for older and hearing-impaired listeners due to fewer available, or redundant, speech cues and possibly compromised but trainable frequency-pattern processing ability. Between the two SWS metrics studied, only VCV intelligibility showed a benefit due to training. This may have been due to VCV stimuli being more reliant than CNCs on low-level auditory processing. Correct response to VCVs requires identification of a single consonant while CNCs need processing of all sounds in the stimulus to form an open-set response. Presumably this results in less involvement of high-level linguistic factors in the VCV condition.

In the pattern-reconstruction conditions, effect of training was restricted to the trained stimulus type. The current protocol utilized only three training sessions. Wright *et al.* (2010) demonstrated that generalization lags behind stimulus-specific learning on auditory tasks and may not appear until after four days of training. In terms of practical applications of the current procedure, the concern is less with generalization to other stimulus types in the pattern-reconstruction task than with benefit for speech perception. Results showing strongest posttest relationships between speech intelligibility and the SWS pattern-reconstruction conditions suggest that these stimuli may lead to greater speech benefit from training than obtained with tonal sequences. Future work will investigate the effect of training length and stimulus type with study participants including older and hearing-impaired listeners, that is, populations that may benefit from auditory rehabilitation that includes the current approach to frequency-pattern training.

ACKNOWLEDGMENTS

Work was supported by Award Number R15DC011916 from the National Institute on Deafness and Other Communication Disorders (NIDCD). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIDCD or the National Institutes of Health.

REFERENCES

Foxton, J.M., Brown, A.C.B., Chambers, S., and Griffiths, T.D. (**2004**). “Training improves acoustic pattern perception,” Current Biol., **14**, 322-325.

Gaab, N., Gaser, C., and Schlaug, G. (**2006**). “Improvement-related functional plasticity following pitch memory training,” Neuroimage, **31**, 255-263.

Sheft, S., Risley, R., and Shafiro, V. (**2012a**). “Clinical measures of static and dynamic spectral-pattern discrimination in relationship to speech perception,” in *Speech Perception and Auditory Disorders*. Edited by T. Dau, M.L. Jepsen, T. Poulsen, and J.C. Dalsgaard (Danavox Jubilee Fndn., Ballerup), pp. 481-488.

Sheft, S., Shafiro, V., Lorenzi, C., McMullen, R., and Farrell, C. (**2012b**). “Effects of age and hearing loss on the relationship between discrimination of stochastic frequency modulation and speech perception,” Ear Hearing, **33**, 709-720.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R.J., and Näätänen, R. (**2001**). “Superior formation of cortical memory traces for melodic patterns in musicians,” Learn. Memory, **8**, 295-300.

Watson, C.S., Kelly, W.J., and Wroton, H.W. (**1976**). “Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty,” J. Acoust. Soc. Am., **60**, 1176-1186.

Wright, B.A., Wilson, R.M., and Sabin, A.T. (**2010**). “Generalization lags behind learning on an auditory perceptual task,” J. Neurosci., **30**, 11635-11639.

No generalization from training on a SAM-detection task to a SAM-rate discrimination task with different depths

LIPING ZHANG[1,*], FRIEDERIKE SCHLAGHECKEN[2], AND JAMES HARTE[1]

[1] *Institute of Digital Healthcare, WMG, University of Warwick, Coventry, CV4 7AL, UK*

[2] *Department of Psychology, University of Warwick, Coventry, CV4 7AL, UK*

Information is carried in speech and sounds both in subtle amplitude and frequency variations over time. Hearing-impaired people have a reduced ability to detect these cues, particularly in challenging auditory environments. Any improvements in these perceptual tasks, through for example auditory training, could help to alleviate some of these difficulties. Practice can improve the detection threshold for amplitude modulation (AM) in sound stimuli. A recent study (Fitzgerald and Wright, 2010) demonstrated that AM-detection learning generalizes from trained to untrained AM rates, but not to a new task (rate discrimination). The present study investigated whether this lack of generalization was due to the use of 100% AM depth in the rate-discrimination task. The present study aims to investigate if it is possible to improve the generalization of sinusoidal amplitude modulation (SAM) detection to rate discrimination by using lower AM depths, such as 70% and 40%, in the discrimination task. The results from this study do not show generalization from SAM-detection to SAM-rate discrimination with any of the lower modulation depths.

INTRODUCTION

Auditory learning is defined as an improvement in the skill to detect, discriminate, or group sounds and speech information (Goldstone, 1998; Halliday *et al.*, 2012). Training in the auditory system may lead to long-lasting changes to an organism’s perceptual system to improve its ability to receive environment sounds. There are two effects derived from auditory training, one is the learning effect, where a listeners’ ability to perform an auditory task could be improved through practice of the same task. The other is the generalization effect, where training in one task leads to improvement in another.

It is known that a normal-hearing person can make use of the context, rhythm, stress, and intonation in speech to understand another speaker. However, for the hearing impaired, it is difficult to use these cues, especially in noisy environments. Although speech recognition by cochlear-implant and hearing-aid users has improved significantly over the past years, most still have major difficulties in noisy environments (Dorman and Wilson, 2004; Ricketts and Hornsby, 2005). The ability of the brain to learn how to make use of an assistive device is as important as

*Corresponding author: liping.zhang@warwick.ac.uk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

developments in the technology (Plomp, 1978; Moore and Shannon, 2009). Therefore, rehabilitation and auditory-training programmes have the potential to optimise the performance of hearing-impaired users and help them get more benefit from their prosthetic device.

Amplitude and frequency fluctuations or modulations in sounds are important carriers of information for speech understanding (Plomp, 1983; Rosen, 1992). Sufficient auditory training could improve humans' perceptual skills to detect and discriminate sounds (Hall and Grose, 1994; Irvine *et al.*, 2000; Hawkey *et al.*, 2004). It is assumed that practise could lead to better performance to detect the changes in amplitude-modulated stimuli, especially for people with problems in detecting amplitude-modulated sounds. In theory, sinusoidal amplitude modulation (SAM) detection and SAM-rate discrimination tests have different perceptual cues that the auditory system uses during decision making (Fitzgerald and Wright, 2010). The SAM detection test mainly focuses on the differences of amplitude-modulated depths from the target to standard stimulus, while the modulation-rate difference between the target stimulus and the standard one is the critical cue for SAM-rate discrimination condition.

Wright and Zhang (2009) showed that auditory learning ability generalized across frequency, ear, stimulus duration, different presentation style, etc. However, Fitzgerald and Wright (2010) argued that the generalization effect could not transfer from SAM detection tasks to SAM-rate discrimination tasks. Fitzgerald and Wright (2010) used a 100% modulation depth for the SAM-rate discrimination tasks in their study. Patterson *et al.* (1978) indicated that 100% modulation depth for a discrimination test is too high to get the optimal rate-discrimination threshold. The present study hypothesises that a generalization effect may occur from SAM-detection to SAM-rate discrimination, if significantly lower modulation depths are used for the SAM-rate discrimination tasks. This project aims to see whether there will be a generalization effect from training on an SAM-detection test to an SAM-rate discrimination test with three different fixed modulation depths (100%, 70%, and 40%).

METHODS

Participants

Twenty normal-hearing volunteers (13 males and 7 females) participated in this experiment. All of the participants had no prior experience participating in psychoacoustic experiments, and their pure-tone thresholds were less than 20 dB HL. The age range was from 18 to 36 years old (with a mean age of 27 years). The participants were all volunteers recruited from the student and staff population of the University of Warwick.

Design

The twenty volunteers were randomly divided into a training group (n = 10) and control group (n = 10). Both groups were required to attend a pre-test and post-test session lasting approximately 2 hours. The pre- and post-test session included one SAM-detection condition and three SAM-rate discrimination conditions. The order of the four conditions was randomised in the pre- and post-tests but was the same across test participants. A three-interval three-alternative forced-choice procedure (3IFC/3AFC) was used to determine the thresholds for SAM-detection and SAM-rate-discrimination conditions. The modulation depth and rate were varied, targeting 79.4% correct performance on the psychometric curve (Levitt, 1971). Five SAM-detection and SAM-rate discrimination thresholds were obtained for each condition. The training group were required to attend 7 consecutive daily training sessions on SAM-detection tasks between the pre- and post-session. Twelve SAM detection thresholds were obtained in each training session. All experimental sessions were carried out within a single-walled sound-proofed room. Sound levels for the SAM detection and SAM-rate discrimination stimuli were calibrated using an IEC 711 acoustic coupler to 65 dB SPL (or at a spectrum level of 40 dB SPL). The experiment was approved by the biomedical and scientific research ethics committee of the University of Warwick.

Procedure

For the SAM-detection test, the target sound was a 3-4 kHz band-pass noise carrier modulated at 80Hz, while the reference sound was un-modulated. In this test condition, the modulation detection threshold was determined with an adaptive tracking procedure. There were three intervals, which include two reference signals and one target, randomly presented. The listener was instructed to decide which interval contains the target amplitude modulated stimuli. The starting modulation depth (m) was 100% modulation and the modulation index in decibels was 20Log_{10} (m). The initial step size was 4dB and then reduced to 2dB after three test reversals. The SAM-detection threshold was defined as the mean of the last 10 reversals in the adaptive tracking procedure.

For the SAM-rate-discrimination conditions test, a 3-4 kHz band-pass carrier-modulated at 80 Hz with three depths (high: 100%, mid: 70%, and low: 40%) was used as the reference sound and the target sound was the same carrier with a higher modulation rate. During this test, the modulation rate of target sound was measured to determine the modulation detection threshold by the 3IFC adaptive tracking procedure. Subjects were required to give a response about which interval was different from the other two. The initial rate difference between the standard and target stimulus was 15 Hz, then decreased to 3 Hz after the third interval, and 1 Hz thereafter, until the threshold was reached.

Data analysis

All participants produced pre-test threshold values within two standard deviations of the mean. No datasets were removed from the analysis, i.e., identified as outliers. The analysis of covariance (ANCOVA) with pre-test thresholds as the covariate was used to compare the test results between the trained and control group. Two way ANOVAs and *t*-tests were also used to confirm the test results.

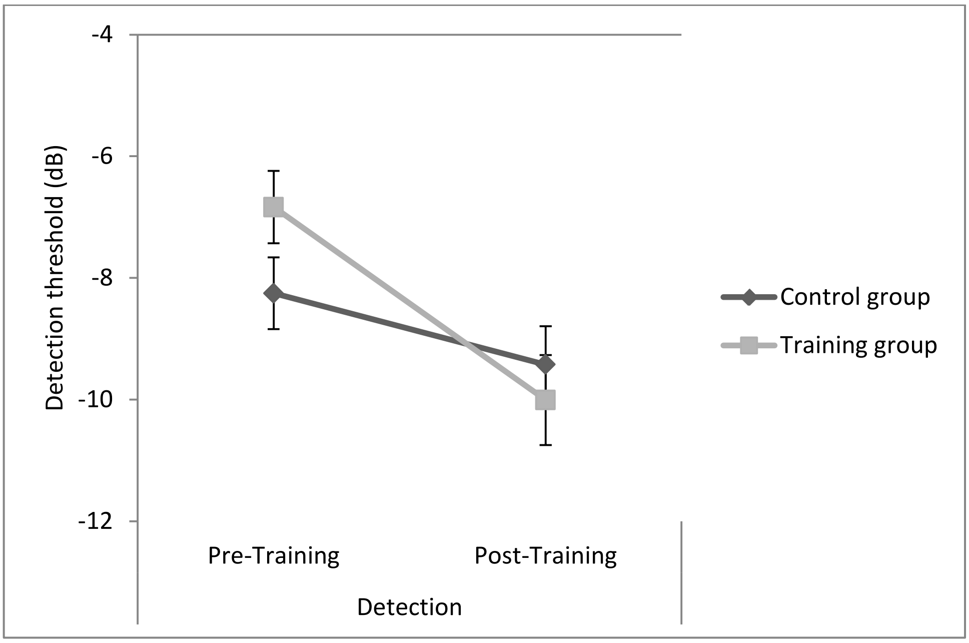


Fig. 1: Mean pre-test and post-test SAM detection thresholds for training (n = 10) and control group (n = 10).

RESULTS

As shown in Fig. 1, although the mean threshold for the trained listeners in the pre-test of the SAM detection condition (M = −6.84 dB, SD = 0.59) was higher than in the untrained listeners (M = −8.25 dB, SD = 0.59), the mean threshold for the trained listeners in the post-test of the SAM detection condition (M = −10.01 dB, SD = 0.74) was lower than the mean post-test SAM-detection threshold for the untrained group (M = −9.42 dB, SD = 0.63). Both two-way ANOVA and ANCOVA tests indicated that there was an overall learning difference between the pre- and post-test results for the trained group and control group (ANOVA: time, $F(1,18) = 100.73$, $p < 0.005$; group × time interaction, $F(1,18) = 21.33$, $p < 0.05$; ANCOVA: $F(1,17) = 18.51$, $p < 0.05$). The main effect comparing the two groups was not significant (ANOVA: $F(1,18) = 0.22$; $p > 0.05$).

Paired *t*-tests were conducted on threshold values from both the SAM-detection trained and SAM-detection untrained group. For the untrained group, there was a statistically-significant decrease in thresholds from the pre-test SAM-detection thresholds (M = −8.25 dB, SD = 1.87) to post-test SAM-detection thresholds (M = −9.42 dB, SD = 1.99), $t(9) = 4.34$, $p = 0.002$). For the trained group, there was also a statistically-significant decrease in thresholds from the pre-test SAM-detection thresholds (M = −6.84 dB, SD = 1.88) to the post-test SAM-detection thresholds (M = −10.01 dB, SD = 2.34), *t*(9) = 9.38, *p* < 0.0005). In order to find whether there was a significant difference in the improvement from pre- to post-test between the trained and untrained groups, an independent-samples *t*-test was carried out on the thresholds difference values from the pre- and post-test results between the two groups. It showed that there was a statistically-significant difference in improvement between the untrained group (M = 1.17 dB, SD = 0.85) and trained group (M = 3.17 dB, SD = 1.07), *t*(18) = −4.62, *p* < 0.0005).

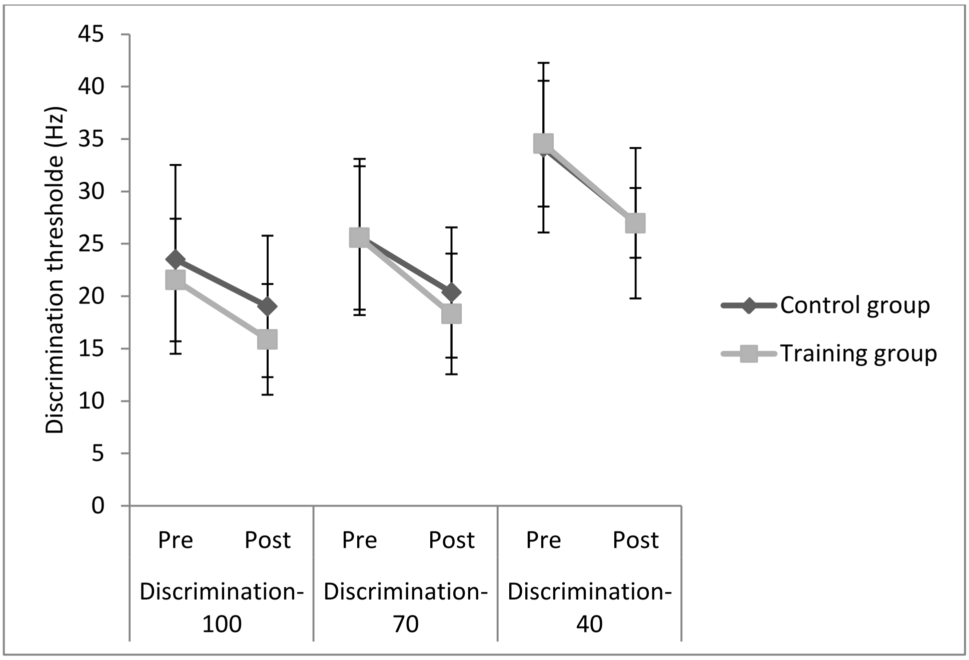


Fig. 2: Mean pre-test and post-test SAM-rate discrimination thresholds for the trained (n = 10) and untrained group (n = 10) under three conditions: 1: SAM-rate discrimination with modulation depth 100%; 2: SAM-rate discrimination with modulation depth 70%; 3: SAM-rate discrimination with modulation depth 40%.

According to Fig. 2, among all three different modulation depths (100%, 70%, and 40%) for SAM-rate discrimination conditions, participants had the largest improvement under the trained SAM-rate discrimination with modulation depth 40% (Pre-test: M = 34.55 Hz, SD = 1.90, Post-test: M = 26.96 Hz, SD = 2.27). The ANOVA test showed that there was a significant difference between the SAM-rate discrimination pre- and post- training sessions (time, *F*(1,18) = 49.00, *p* < 0.0005), but no significant different between the two groups (group × time interaction, *F*(1,18) < 1). Regarding the three different modulation depths, although there was a significant difference among these three depths (depth, $F(2,36) = 53.37$, $p < 0.0005$), no significant difference was observed from the trained and untrained groups with three different depths (time × depth, $F(1,18) = 2.29$, $p > 0.05$; group × time × depth interaction, $F(2,17) < 1$). The main effect comparing the two groups was also not significant (ANOVA: $F(1,18) = 0.23$, $p > 0.05$).The mean SAM-rate discrimination thresholds were 20.14, 22.47, and 30.68 Hz for modulation depths of 100%, 70%, and 40%, respectively. While the former two values were not significantly different from each other ($p > 0.05$), the latter was significantly higher than both (both $p < 0.001$).

DISCUSSION

This study confirmed that training improves abilities in the SAM detection task, as observed by Fitzgerald and Wright (2010). However, the results do not show generalization from SAM-detection to SAM-rate discrimination with any of the three modulation depths tested. Comparing the results from pre- and post- SAM detection thresholds and SAM-rate discrimination thresholds, both trained and untrained groups demonstrated significant improvement. Thus learning effects were observed for the SAM-detection and SAM-rate discrimination tests even after the initial pre-test session. When comparing the mean thresholds of SAM-rate discrimination tasks, no significant difference was observed between the trained and untrained group. So the study does not demonstrate a generalization effect from training on an SAM-detection task to an SAM-rate discrimination task.

Millward *et al.* (2005) presented evidence to suggest that the generalization effect between the trained auditory task and another task is more likely if both share a common stimulus dimension, i.e., the same masking noise or the same target stimulus is used. Further, they demonstrated an opposite effect to the desired synergistic generalization effect, where training in one task actually suppresses or reduces performance in another. This was more likely to occur if the two tasks did not share a common stimulus dimension. In the present study, the target sound in the SAM rate-discrimination test used an identical carrier to that used in the SAM-detection task. However, the stimulus feature of interest, namely modulation depth versus modulation frequency, differed between the two. It could be argued that the lack of generalization from training in SAM-detection to SAM-rate discrimination arose as a result of the auditory system processing these two tasks separately. Training on a range of different auditory stimuli may lead to a greater transfer learning effect (Halliday *et al.*, 2012), possibly because of improved attention and/or working memory. Further research should be carried out to explore whether there is better generalization when people are trained on more complex auditory stimuli, such as non-speech and speech sounds together.

REFERENCES

Dorman, M.F., and Wilson, B.S. (**2004**). “The design and function of cochlear implants,” Am. Scientist, **92**, 436-445.

Fitzgerald, M.B., and Wright, B.A. (**2010**). "Perceptual learning and generalization resulting from training on an auditory amplitude-modulation detection task," J. Acoust. Soc. Am., **129**, 898-906.

Goldstone, R.L. (**1998**). "Perceptual learning," Ann. Rev. Psychol., **49**, 585-612.

Hall, J.W. III, and Grose, J.H. (**1994**). "Development of temporal resolution in children as measured by the temporal modulation transfer function," J. Acoust. Soc. Am., **96**, 150-154.

Halliday, L.F., Taylor, J.L., Millward, K.E., and Moore, D.R. (**2012**). "Lack of generalization of auditory learning in typically developing children," J. Speech Lang. Hear. Res., **55**, 168-181.

Hawkey, D.J., Amitay, S., and Moore, D.R. (**2004**). "Early and rapid perceptual learning," Nat. Neurosci., **7**, 1055-1056.

Irvine, D., Martin, R., Klimkeit, E., and Smith, R. (**2000**). "Specificity of perceptual learning in a frequency discrimination task," J. Acoust. Soc. Am., **208**, 2964-2968.

Levitt, H. (**1971**). "Transformed up-down procedures in psychoacoustics," J. Acoust. Soc. Am., **49**, 467-477.

Millward, K.E., Hall, R.L., Ferguson, M.A., and Moore, D.R. (**2011**). "Training speech-in-noise perception in mainstream school children," Int. J. Pediatr. Otorhi., **75**, 1408-1417.

Moore, D., and Shannon R. (**2009**). "Beyond cochlear implants: awakening the deafened brain," Nat. Neurosci., **12**, 686-691.

Patterson, R.D., Johnson-Davies, D., and Milroy, R. (**1978**). "Amplitude modulated noise: The detection of modulation versus the detection of modulation rate," J. Acoust. Soc. Am., **63**, 1904-1911.

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am., **63**, 533-549.

Plomp, R. (**1983**). "The role of modulation in hearing," in *Hearing – Physiological Bases and Psychophysics*. Edited by R. Klinke and R. Hartman (Springer-Verlag, Berlin), pp. 270-276.

Ricketts, T.A., and Hornsby, B.W. (**2005**). "Sound quality measures for speech in noise through a commercial hearing aid implementing "digital noise reduction"," J. Am. Acad. Audiol., **16**, 270-277.

Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory, and linguistic aspects," Philos. T. Roy. Soc. A., **336**, 367-373.

Wright, B.A., and Zhang, Y. (**2009**). "A review of the generalization of auditory learning," Philos. T. Roy. Soc. A., **364**, 301-311.

HRTF adaptation and pattern learning

FLORIAN KLEIN* AND STEPHAN WERNER

Electronic Media Technology Lab, Institute for Media Technology, Technische Universität Ilmenau, D-98693 Ilmenau, Germany

The human ability of spatial hearing is based on the anthropometric characteristics of the pinnae, head, and torso. These characteristics are changing slowly over the years and therefore it is obvious that the hearing system must be adaptable to some degree. Researchers have already been able to measure this effect, but still there are many open questions like the influence of training time and stimuli, level of immersion, type of feedback, and inter-subject variances. With HRTF (head-related-transfer-function) adaptation it might also be possible to increase the plausibility of acoustical scenes over time. When measuring adaptation effects in a spatial hearing test it is important to distinguish between conscious pattern learning and perceptive adaptation. To increase the quality of virtual auditory display the amount of perceptive adaptation is of major interest. In an earlier spatial listening test high training effects could be observed within a short period of training. To investigate the different types of training a second listening test was conducted. The acoustic stimuli were altered between the test and training sessions to avoid pattern learning. The results are compared to the previous findings and give further insights into the topic of perceptive adaptation of HRTFs.

MOTIVATION AND STATE OF THE ART

In auditory research adapation effects of the auditory system are well known for example in the field of cochlear-implant (CI) treatment. In other research areas like in the development of spatial sound systems adaptation effects are mostly not evaluated. In recent publications the existence of spatial-hearing adaptation effects could be observed by Majdak (2012) and Parseihian and Katz (2012), and earlier by Hofman *et al.* (1998). The focus of research is the localisation performance of the listeners and the achievable accuracy gain by listening training. Researchers found better performances for quadrant errors (e.g., front-back confusions) and elevation perception after audio-visual or proprioceptive feedback. Training in virtual environments like in Majdak (2012) and Parseihian and Katz (2012) exhibit fast training effects after a short time of training. Under real conditions, by using ear molds to modify the head-related transfer functions instead of binaural synthesis, training effects are observed after many days of training.

Listening tests in our lab (referring to Klein and Werner (2013)) confirmed significant

*Corresponding author: florian.klein@tu-ilmenau.de

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

adaptation effects regarding the perception of elevation after audio-visual training in a virtual environment. Strong adaptation effects after short training periods are often understood as pattern learning (or known as procedural learning in Hawkey *et al.* (2004)) in contrast to perceptual adaptation because of the training on a specific task. A second listening test is compared to a previous adaptation test with the following questions:

1. Are adaptation effects persistent after a long time without training and could that be and indication for perceptual adaptation?
2. Can accuracy gains be achieved when test and training stimuli differ?

TEST ENVIRONMENT

Static binaural synthesis is used to create the spatialisation of different directions. A block diagram of the rendering system is shown in Fig. 1. Artificial HRTFs (CIPIC database by Algazi *et al.* (2001) and own KEMAR measurements) are used in combination with a least-squares-based headphone equalization according to Schärer (2008). For all tests STAX lambda pro headphones are used.

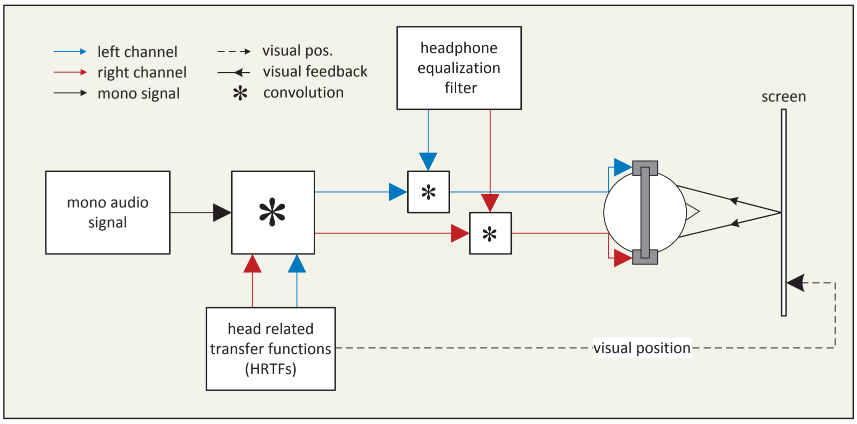


Fig. 1: Block diagram of the used binaural synthesis system combined with visual feedback over a screen; non-individual head-related transfer functions from the CIPIC database by Algazi *et al.* (2001) and headphone equalization according to Schärer (2008) are used.

For the audio-visual training participants are placed in front of a screen with loudspeaker symbols in direction of the virtual sound sources. Loudspeaker symbols for the virtually active loudspeaker are highlighted in green during the training sessions. A picture of a typical test situation is shown in Fig. 2. During the test sessions, the participants can simply use a computer mouse to select the loudspeaker symbol which is nearest to the perceived direction. Visual loudspeaker representations are placed at azimuth angles of $-30°$ to $30°$ in steps of $5°$ and at the vertical angles between $28.125°$ to $-16.875°$ in steps of $5.625°$.

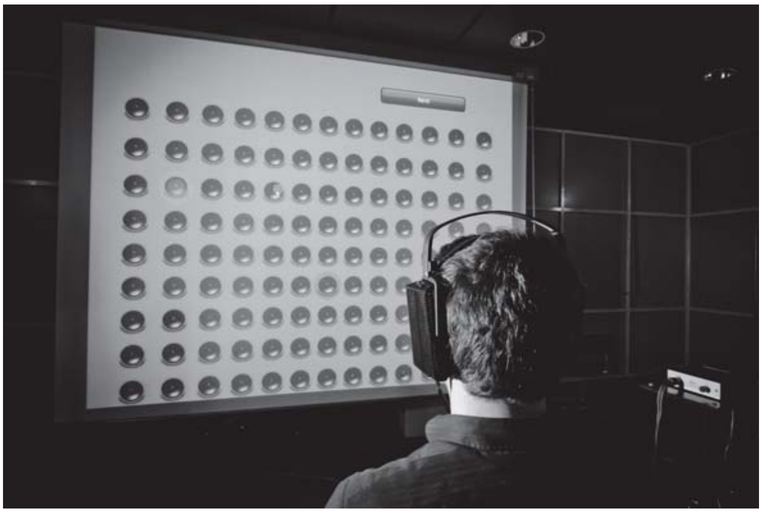


Fig. 2: Picture of the actual listening test setup in the listening lab of the TU Ilmenau.

Because a static binaural system is used and visual feedback is provided over screen, the positioning of the participants is crucial. Before each test the listeners are positioned at a defined distance and height in front of the screen. During the test the participant has to point his head towards the central loudspeaker symbol which is highlighted in red.

LISTENING TEST DESIGN

When using artificial HRTFs the localisation performance of the participants varies highly. Therefore initial pre-tests are conducted to measure the individual performance without any training and for each test stimulus. After the training sessions post-tests are conducted to measure the change in localisation performance. The different types of test and training sessions are described below.

Audio-visual training

For training, a sequence of virtual auditory sound sources is synthesized together with the spatially corresponding visual feedback. In one session 72 trials are presented which consist of four random azimuth directions for all nine elevation angles. Each direction gets repeated once.

Listening test with further audiovisual training

The audio stimuli are presented and the listener has to choose a perceived direction. After the response is given, the correct loudspeaker is highlighted in green. This session consists of 72 trials (six random azimuth angles at six different elevation angles and one repetition). In the test session virtual acoustic loudspeakers are synthesized only at the vertical angles of $28.125°, 16.875°, 5.625°, 0°, -5.625°$, and $-16.875°$. This kind of test can be understood as active training in contrast to the passive training by just watching and listening to a sequence of trials. Furthermore, this way test data can be acquired in the training phase.

Listening test without visual cues

During this test the participant has to rate 72 sound stimuli (equal to ‘Listening test with further audiovisual training’) and gets no visual feedback at all.

An overview of all test comparisons is shown in Fig. 3. The first test was done five months before the second test. The second test is divided into two parts to keep the test duration under 60 minutes. Overall 14 participants took part in the second test while nine of them already took part in the first test. For these nine listeners a comparison between test one and two can be done. The second test is aimed to compare the effect of different testing and training conditions by using different acoustic stimuli.

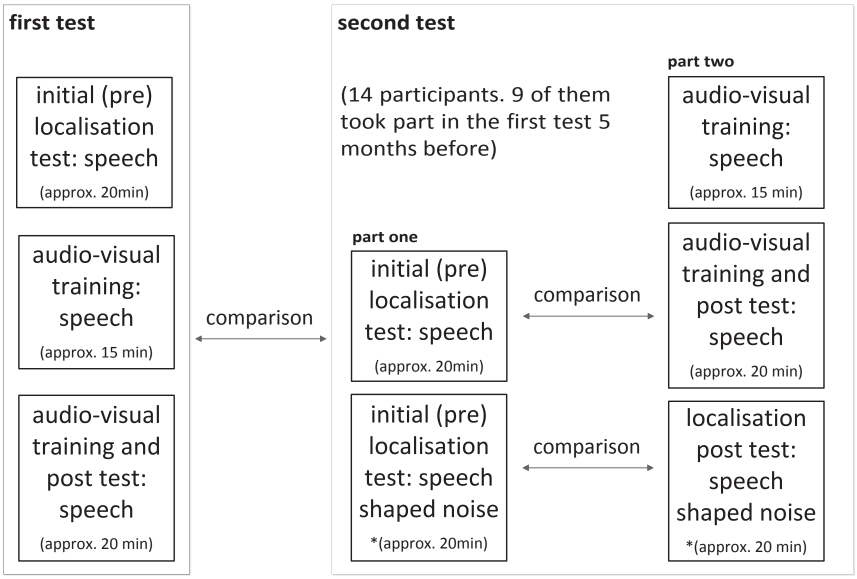

first test
initial (pre) localisation test: speech (approx. 20min)
audio-visual training: speech (approx. 15 min)
audio-visual training and post test: speech (approx. 20 min)
comparison
second test
(14 participants. 9 of them took part in the first test 5 months before)
part one
initial (pre) localisation test: speech (approx. 20min)
initial (pre) localisation test: speech shaped noise *(approx. 20min)
part two
audio-visual training: speech (approx. 15 min)
audio-visual training and post test: speech (approx. 20 min)
localisation post test: speech shaped noise *(approx. 20 min)
comparison
comparison

Fig. 3: Overview of the test procedures of the first and second test. The marked comparisons between tests are discussed in the results section.

The speech signal is about three seconds long and features a male foreign speaker. The other stimulus is CCITT coloured noise (according to ITU-T Rec. G227) with the same temporal envelope as the speech signal. The power spectrum of CCITT coloured noise is similar to the average power spectrum of typical speech.

RESULTS AND EVALUATION

Because the test setup only allows ratings in the frontal plane, training effects on front-back confusions can not be evaluated (in Majdak (2012) results about the reduction of quadrant angle errors like front-back confusion can be found). Therefore the focus of this publication is directed at changes in the perception of elevation.

Comparison between first and second test

In the first comparison the ability to discriminate different elevation angles is investigated. Elevation angles are ranked according to their perceived height and compared to the correct order of the elevation angles. If a participant orders all elevation angles correctly according to their height (for example -16.875° is perceived at -11.25°, 0° is perceived at -5.625°, and 16.875° is perceived at 5.625°), then the rank correlation equals one. Perceiving different target angles at the same angle or alternating the order of elevations results in a lower rank correlation. This approach gives no statement about the absolute height accuracy and is therefore more liberal than localisation error scores. Figure 4 shows box plots of the rank correlations for the different tests conducted with the nine participants from the first and second test.

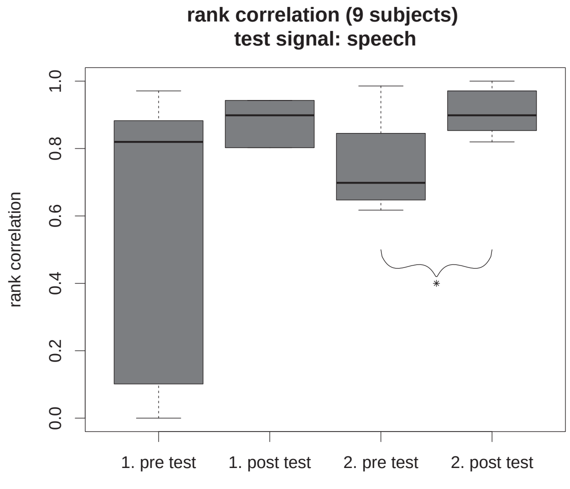


Fig. 4: Boxplot of rank correlations for each test. A rank correlation of one means that all participants were able to order the sound samples according to their height (no statement about absolute height accuracy). Significant differences are marked with an asterix symbol (Wilcoxon signed rank test with confidence interval < 0.05).

The comparison with the first test shows that the ability to discriminate elevation angles decreases without further training. There is no significant difference in rank correlation between the results of the first and the second pre-test. However the inter-individual differences are smaller in the second pre-test, which could be a sign for persistent learning effects. With a second training the same performance as in the first post-test can be achieved.

Second test: differences depending on test stimuli

In the following the results of the second test are presented. Figures 5 and 6 show the results for the pre- and post-tests for both test stimuli. The results of the pre-tests are always plotted on left. With both test stimuli participants show a compressed elevation perception in the pre-tests. The post-tests are plotted on the right and they show an increased ability in the absolute accuracy of elevation perception. When comparing the boxplots of the post-tests, the learning effect seems to be more prominent for the speech-shaped noise. In contrast to the speech signal, increased accuracy is observed for nearly all elevation angles. These results have to be compared carefully, because training time and conditions for both tests were not the same. For the speech stimuli only one session of audio-visual training was conducted, while in advance of the noise test two training sessions were conducted (compare with Fig. 3). On the other hand, all training sessions were conducted with the speech signal and no training with the noise stimuli was carried out.

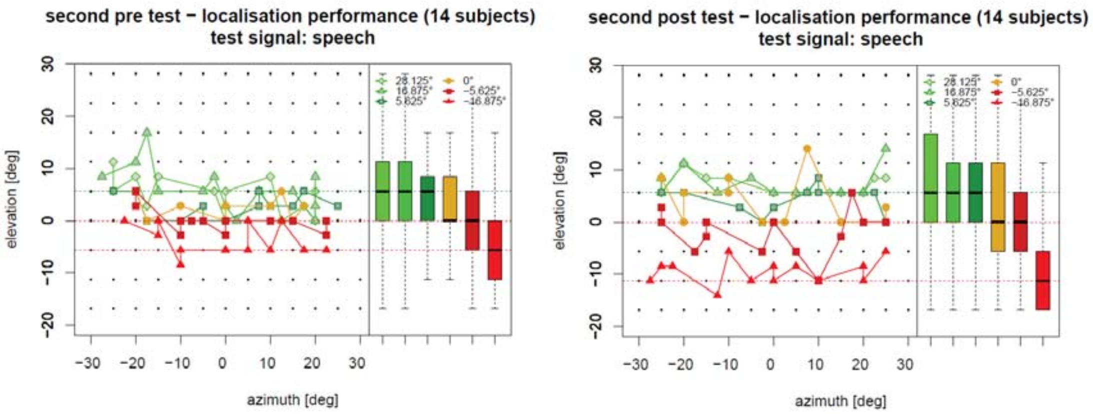


Fig. 5: Median localisation performance for the speech signal in the pre- and post-test; the left side of each plot shows the median ratings for the tested directions and the right side shows the perceived elevation as box plots for each target elevation angle.

All boxplots show a broad range of minimum and maximum values and at some angles high inter-quartile ranges. The reason is found when observing individual results. Depending on the subject the position of the virtual sources has a positive or negative offset (similar to the results of Hofman *et al.* (1998)). After training different learning

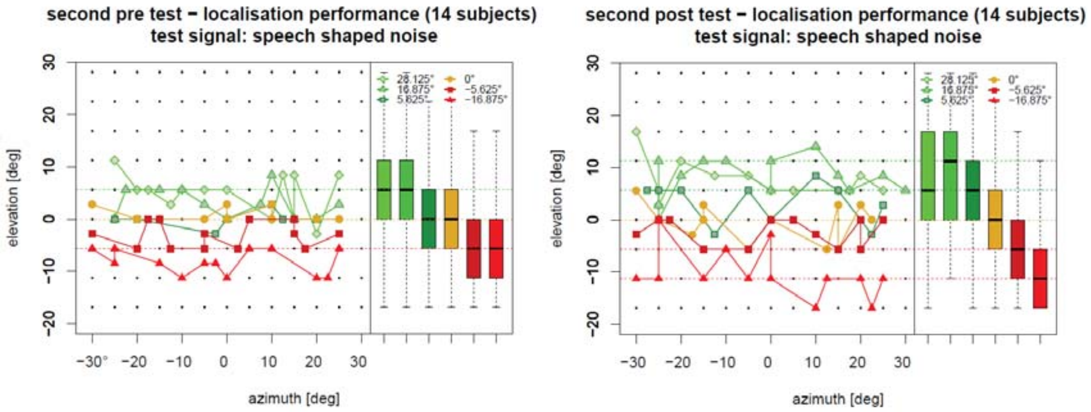


Fig. 6: Median localisation performance for the speech shaped noise signal in the pre- and post-test; the left side of each plot shows the median ratings for the tested directions and the right side shows the perceived elevation as box plots for each target elevation angle.

patterns can also be observed. One participant may increase his performance only in the lower hemisphere and another listener only in the upper hemisphere. Another interesting point are the ratings for the highest elevation angle in post-test for the speech-shaped noise signal (Fig. 6). The highest elevation angle was close to the maximum of the field of view and for some people this row of virtual loudspeaker symbols was barely visible. This might be an explanation for these elevation ratings being out of order.

SUMMARY AND DISCUSSION

As presented in Fig. 4 elevation perception decreases without further training to the initial performance. However smaller inter-individual differences are observable in the second test for people who also participated in the first test. This could mean that, at least for some people, a training effect remains over a longer period of time.

The second test showed that participants were able to increase their localisation performance despite a spectral difference between training signal (speech) and test signal (speech-shaped noise). This could be a sign for perceptual adaptation instead of pattern learning. A clear discrimination between those types of adaptation is still not possible while looking at these results. In the next test iteration, it could be useful to use different training tasks in combination with alternating stimuli. This way it might be possible to exclude pattern learning further. Another advancement would be a comparison to a control group (which gets no training but has to do the test as well) to distinguish between adaptation introduced by the audio-visual training and by the test procedure itself.

Further interesting effects can be observed when looking at the individual results:

Participants show different learning patterns. Performance increases are often observed in only one hemisphere (upper or lower) and the amount of accuracy gain varies highly.

ACKNOWLEDGEMENT

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (Grant BR 1333/14-1).

REFERENCES

Algazi, V.R., Duda, R.O., and Thompson, D.M. (**2001**). “The CIPIC HRTF Database”, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 99-102.

Hawkey, D.J.C., Amitay, S., and Moore, D.R. (**2004**). “Early and rapid perceptual learning,” Nature Neurosci., **7**, 1055-1056.

Hofman, P.M., Van-Riswick, J.G., and Van Opstal, A.J. (**1998**). “Relearning sound localization with new ears,” Nature Neurosci., **1**, 417-421.

Klein, F., and Werner, S. (**2013**). “HRTF adaption under decreased immersive conditions,” AIA-DAGA, Meran, Italy, 580-582.

Majdak, P. (**2012**). “Audio-visuelles Training der Schallquellenlokalisation mit manipulierten spektralen Merkmalen,” DAGA, Darmstadt.

Parseihian, G., and Katz, B.F.G. (**2012**). “Rapid head-related transfer function adaptation using a virtual auditory environment,” J. Acoust. Soc. Am., **131**, 2948-2957.

Schärer, Z. (**2008**). *Kompensation von Frequenzängen im Kontext der Binauraltechnik*. Master thesis, TU Berlin.

Formation of the mouse cochlea: roles of Sonic hedgehog

JINWOONG BOK[1], COLLEEN ZENCZAK[2], CHANHO HWANG[2], AND DORIS K. WU[2,*]

[1] *Department of Anatomy, Yonsei University College of Medicine, Seoul, 120-752, South Korea*

[2] *Laboratory of Molecular Biology, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD 20850, USA*

The formation of the mammalian cochlea is dependent on signalling from its surrounding tissues during embryogenesis. Using genetically engineered mutant mice and surgically manipulated chicken embryos, we demonstrated that Sonic hedgehog (Shh) secreted from the developing notochord and the floor plate is important for specification of ventral inner-ear structures that include the cochlear duct. Additionally, tissue-specific knockout of Shh in the developing spiral ganglion indicates that this source of Shh is required for mediating growth of the cochlear duct, timing of terminal mitosis of hair cell precursors, and subsequent differentiation of nascent hair cells along the cochlear duct.

INNER-EAR FORMATION

In the mouse, inner-ear formation initiates at embryonic day 8.5 (E8.5) from a region of the ectoderm next to the developing hindbrain. This specialized epithelial region, namely the otic placode, deepens to form a cup that separates from the rest of the ectoderm by pinching off to form a cyst (Fig. 1). Starting at the otic cup stage, some epithelial cells leave the antero-ventral region (defined as the neural-sensory competent domain) of the cup and these neuroblasts coalesce to form the cochleo-vestibular ganglion (CVG, VIII cranial ganglion). This neural delamination process in the mouse continues until at least E11.5, well after the otocyst is formed (Koundakjian *et al.*, 2007; Raft *et al.*, 2007). The neuroblasts within the CVG subsequently develop into bipolar neurons that send processes to innervate the sensory hair cells within the inner ear as well as nuclei in the brainstem and cerebellum.

Concomitant with the neuronal development, the otocyst undergoes a series of complex morphogenetic processes starting at E9.5 and reaching its adult pattern at approximately E16.5 (Fig. 2; Morsli *et al.*, 1998). The first structure that emerges from the rudimentary cyst is the endolymphatic duct, which subsequently develops into the endolymphatic sac and duct. This structure is essential for maintaining fluid homeostasis of the endolymph within the membranous labyrinth. Without a functional endolymphatic system for maintaining a proper balance of high potassium and low sodium ions in the endolymph, mechanotransduction of sensory hair cells will fail.

*Corresponding author: wud@nidcd.nih.gov

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

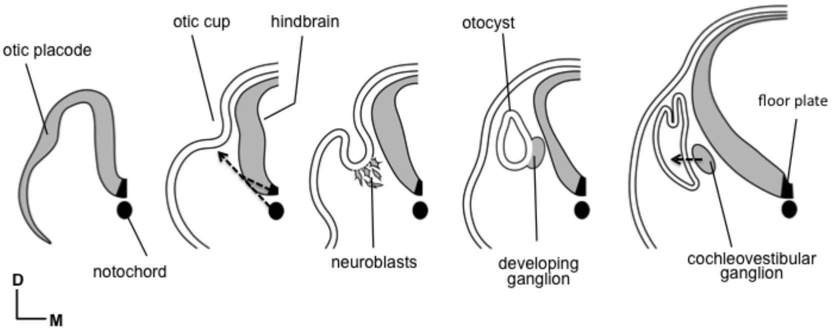


Fig. 1: Development of the inner ear from the otic placode to otocyst stage. Starting at the otic cup stage, neuroblasts delaminate from the antero-ventral region of the otic cup to form the cochleo-vestibular ganglion. Dotted arrows indicate the two sources of Shh that are important for cochlear development: floor plate and notochord required for specification of the cochlea at an early stage and the spiral ganglion required for growth of cochlear duct and timing of cell cycle exit and hair-cell differentiation at later stages. Orientations: D, dorsal; M, medial. Adapted from Chang *et al.* (2003).

The dorsal region of the otocyst proper develops into the vestibule consisting of three orthogonally arranged semicircular canals and the associated ampullae that are responsible for detecting angular acceleration. Additionally, the two macular organs, the utricle and saccule, develop within the mid-region of the otocyst and are important for detecting linear acceleration. The most ventral region of the otocyst develops into the auditory apparatus, the cochlear duct, which starts out in a postero-lateral position of the inner ear and then extends in a ventral-medial and anterior direction before coiling laterally to complete its 1.75 turns. While these series of morphogenetic events sculpture the otocyst into a labyrinth, they also coordinate the gradual separation of the neural-sensory competent domain into various sensory patches: one auditory and five vestibular sensory organs of the mouse.

MOLECULAR MECHANISMS REGULATING INNER-EAR DEVELOPMENT

What are the molecular mechanisms that underlie the formation of this intricate sensory organ? Similar to other organs of the body, the ear primordium first acquires its axial information from the surrounding tissues. Studies from chicken and mouse embryos suggest that the signals that confer this axial information may be quite conserved (Riccomagno *et al.*, 2002; Bok *et al.*, 2005; Riccomagno *et al.*, 2005; Bok *et al.*, 2011). For example, retinoic acid secreted by the somites caudal to the developing inner ear is important for patterning the anterior and posterior axes of the inner ear (Bok *et al.*, 2011). In the chicken, this posterior retinoic acid (RA) signal is opposed by the RA degradation enzyme, Cyp26c1, expressed in the ectoderm anterior to the ear primordium. As a result, the ear primordium receives a gradient of RA signalling: the anterior otic region that receives low levels of RA signalling develops into the neural-sensory competent domain. In contrast, the posterior otic region, closer to the source of RA emanating from the somites, receives higher levels of RA and develops into largely non-sensory structures (Bok *et al.*, 2011).

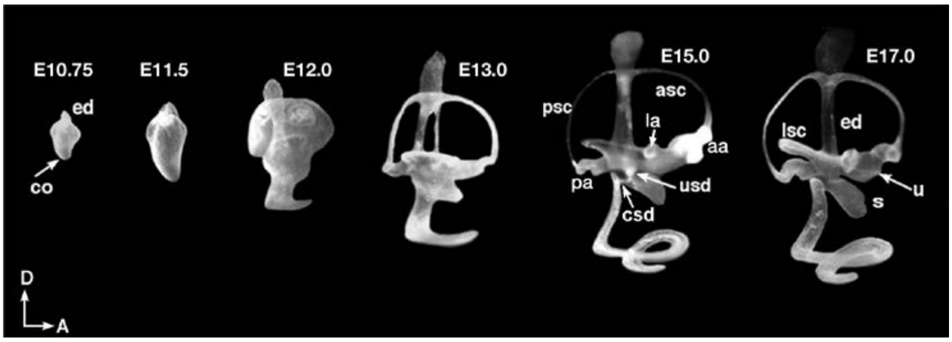


Fig. 2: Development of the inner ear from the otocyst stage to E17. Various stages of the developing mouse inner ear were injected with a paint solution in methyl salicylate. Abbreviations: aa, anterior ampulla; asc, anterior semicircular canal; co, cochlear duct; csd, cochlear-saccular duct; endolymphatic duct; la, lateral ampulla; lsc, lateral semicircular canal; pa, posterior ampulla; psc, posterior semicircular canal; s, saccule; u, utricle; usd, utricular-saccular duct. Adapted from Morsli *et al.* (1998).

The establishment of dorsal-ventral axis of the inner ear is dependent on secreted molecules from the developing hindbrain. For example, Wnts from the dorsal hindbrain are important for inner-ear development (Riccomagno *et al.*, 2005). On the other hand, Shh, secreted by the floor plate and notochord, is important for patterning ventral structures of the inner ear such as the cochlear duct and saccule (Riccomagno *et al.*, 2002). Thus, the developing hindbrain appears to be a key tissue in providing instructive signals for patterning the dorsal-ventral axis of the inner ear.

Does the inner ear develop autonomously once it acquires its axial identity? Studies on extirpating mouse or chicken inner ears at various developmental stages to a heterologous environment indicate that the three primary cell fates – neural, sensory and non-sensory – are established early during development (Li *et al.*, 1978; Swanson *et al.*, 1990). However, proper morphogenesis requires continuous instructions from the surrounding tissues and extirpated ears do not recapitulate well the anatomy of inner ears at equivalent stages in vivo. Thus, identifying the key molecules alone is insufficient to fully understand the molecular mechanisms underlying inner-ear formation. It is equally important to know when and where these key signalling molecules are required, especially since many of them are used repeatedly throughout inner ear development.

ROLES OF SONIC HEDGEHOG IN COCHLEAR FORMATION

Patterning of the cochlear duct

Many tissues such as the hindbrain, neural crest, mesenchyme, and spiral ganglion are important for the proper formation of the cochlear duct (Wu and Kelley, 2012; Bok *et al.*, 2013). Whether all of these tissues participate in conferring axial information to the cochlear anlage will require further investigation. Although it is clear that the cochlear duct is a ventral structure and will require ventral signalling, it is difficult to discern based on the location of the mature cochlea whether it is derived from the anterior or posterior compartment of the ear primordium, or both (Fig. 2). Chicken inner ears that received posteriorizing signals from both anterior and posterior sides of the ear rudiment developed a mirror duplication of the posterior half of the inner ear (Bok *et al.*, 2011). In these ears, a shortened cochlear duct (basilar papilla) without sensory tissue was found. Based on these results, we postulate that the cochlear duct is comprised of both anterior and posterior components: The organ of Corti, being part of the neural-sensory competent domain, belongs to the anterior compartment of the otic cup, whereas the rest of the cochlea, the non-sensory component, is derived from the posterior compartment. Thus, in chicken ears with duplicated posterior halves, the cochlear duct is comprised of non-sensory tissues only.

In chicken embryos, ablation or anterior-posterior inversion of a segment of the hindbrain adjacent to the developing inner ear affects the formation of the basilar papilla more readily than the vestibule, suggesting that cochlear formation is more sensitive to local signalling from the hindbrain (Liang *et al.*, 2010). While a majority of the cochlear phenotypes in the hindbrain manipulated embryos is attributed to the loss of Shh signalling (see below; Bok *et al.*, 2005), additional undetermined signal(s) are thought to be important for mediating the shape of the cochlear duct (Liang *et al.*, 2010). Additionally, mesenchyme surrounding the ear primordium also plays an important role in the proper coiling of the cochlear duct. Several genes expressed in the mesenchyme such as *Pou3f4* and *Tbx1* have been shown to affect the length and shape of the cochlear duct when deleted from the genome (Phippard *et al.*, 1999; Braunstein *et al.*, 2008).

Sonic hedgehog

Shh, one of the three vertebrate homologs of the *Drosophila hedgehog* gene, is by far the most important vertebrate *Hedgehog* gene during embryogenesis. It encodes a secreted molecule that is involved in formation of various organs such as the neural tube, retina, limbs, and gut (Ingham and McMahon, 2001). Shh functions by binding to the cell membrane receptor Patched, which allows the transducer of Shh signalling, Smoothened (also a trans-membrane protein), to be activated. A major role of Shh is to regulate the levels of the transcription factor Gli3 in target cells

(Litingtung *et al.*, 2002; te Welscher *et al.*, 2002). In the absence of Shh signalling, the Gli3 protein is cleaved and the N-terminus of the protein functions as a transcription repressor. In the presence of activated Smoothened, this enzymatic cleavage of Gli3 is inhibited and the full length of Gli3 protein functions as a transcription activator, which activates Shh target genes in the nucleus.

Contribution of Sonic hedgehog from the midline

Shh secreted by the floor plate and notochord in the midline is important for establishing dorsal-ventral patterning of the neural tube as well as for patterning paraxial structures such as the somites (Borycki *et al.*, 2000). The first indication that Shh from the midline is also important for patterning the inner ear stemmed from analyses of inner ears in *Shh-/-* mouse mutants (Riccomagno *et al.*, 2002). No ventral inner-ear structures are evident in these mouse embryos (Fig. 3A,B). To address the specific contribution of Shh secreted from the ventral midline in mediating the inner-ear phenotypes, we surgically ablated a segment of ventral neural tube and notochord near the developing chicken otic cup *in ovo*. As a result, the inner ear shows a normal vestibule but lacks the basilar papilla and saccule (Fig. 3C,D; Bok *et al.*, 2005). Remarkably similar phenotypes were also obtained when cells secreting antibodies blocking Shh bioactivities were implanted beside the ventral midline, suggesting that Shh is the main effector of ventral inner-ear patterning from the midline (Bok *et al.*, 2005). Taken together these results from chicken and mouse indicate that Shh secreted from the notochord and floor plate has a conserved role in specifying the ventral axis of the inner ear.

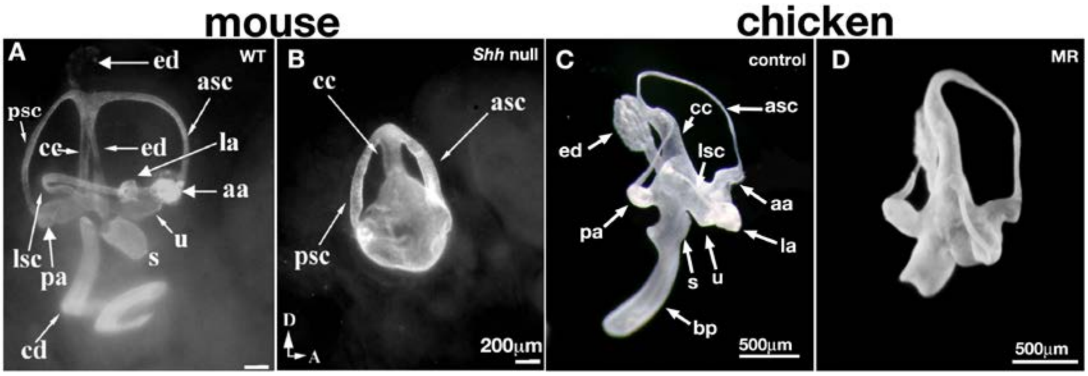


Fig. 3: Paint-filled mouse inner ears of (A) wildtype and (B) *Shh* null mutants at E15.5 as well as E7 chicken inner ears of (C) controls and (D) those with a segment of the notochord and floor plate beside the developing inner ear removed at E1.5 (midline removal, MR). This surgical operation affects cochlear development and results in an inner ear that resembles the *Shh*-/- mouse mutants. Abbreviations: bp, basilar papilla. Adapted from Riccomagno *et al.* (2002); Bok *et al.* (2005).

Since a major role of Shh is to remove the repressor function of Gli3 in other systems (Litingtung *et al.*, 2002; te Welscher *et al.*, 2002), we investigated the inner ear in *Shh* and *Gli3* double mutants. *Gli3* knockout ears are largely normal and only the lateral canal is absent (Bok *et al.*, 2007). If one of Shh's major roles in the inner ear is to remove the repressor function of Gli3, one would expect the inner ear phenotypes in *Shh*$^{-/-}$*; Gli3*$^{-/-}$ double mutants to be milder than that of *Shh*$^{-/-}$ alone. This is indeed the case. In *Shh*$^{-/-}$*; Gli3*$^{-/-}$ double mutant ears, the saccule and a shortened cochlear duct are present (Fig. 4B; Bok *et al.*, 2007) as opposed to the lack of all ventral structures in the *Shh*$^{-/-}$ mutants (Fig. 3B). The shortened cochlear duct in *Shh*$^{-/-}$*; Gli3*$^{-/-}$ double mutants also lacks *Msx1* expression (Fig. 4E), which is a marker for the apical region of the cochlear duct (Fig. 4D). This indicates that the duct is truncated, not simply smaller. While the absence of *Gli3* alleviated some of the phenotypes observed in *Shh* null mutants (presumably due to the loss of Gli3 repressor functions), the persistence of the missing apical region in the *Shh*$^{-/-}$*; Gli3*$^{-/-}$ double mutants suggests that the apical region of the cochlea requires the activator function of Gli3.

The notion that the apical cochlea requires higher levels of Shh signalling relative to the rest of the cochlea is also supported by the inner-ear phenotypes of *Δ*699/*Δ*699 mouse mutants, modelled after mutations observed in Pallister-Hall syndrome in humans. Both the mouse and human mutations resulted in a truncated Gli3 protein that has only repressor but no activator activity (Kang *et al.*, 1997; Bose *et al.*, 2002). In *Δ*699/*Δ*699 mutants, the cochlear duct is shortened and is missing *Msx1* expression (Fig. 4C,F). Thus, this mutant cochlea is presumably truncated and missing the apical region. The apical cochlear phenotypes in *Δ*699/*Δ*699 mouse mutants are consistent with patients with Pallister-Hall syndrome showing a prevalence of low-frequency hearing loss (Driver *et al.*, 2008).

Contribution of Sonic hedgehog from the spiral ganglion

Spiral ganglion Sonic hedgehog mediates growth of the cochlear duct:

In addition to the source of Shh from the ventral midline, *Shh* is also expressed in the developing spiral ganglion, first detectable at E11.75 (Bok *et al.*, 2013). What is the role of Shh secreted by the spiral ganglion? This question was addressed by generating tissue-specific knockout of *Shh* using the cre-lox approach. Three cre strains were used in the study: *Neurogenin1*^{cre} *(Ngn1*^{cre}*)*, *Neurogenin1*^{creER} *(Ngn1*^{creER}*)*, and *Foxg1*^{cre} (Bok *et al.*, 2013). All three promoters driving cre in these strains are active in the developing spiral ganglion and not in the floor plate and notochord (Hebert and McConnell, 2000; Koundakjian *et al.*, 2007; Quinones *et al.*, 2010). In the *Ngn1*^{creER} strain, the cre is fused to a mutated form of the estrogen receptor (ER), which provides a temporal control of cre activation pending tamoxifen administration (Hayashi and McMahon, 2002). The conditional knockout of *Shh* using each of the three cre strains generated inner ears with a shortened cochlear duct, the shortest being the *Foxg1*^{cre}*; Shh* $^{lox/-}$ mutants (Fig. 5). The length of the cochlear duct in the *Ngn1*^{creER}*; Shh* $^{lox/-}$ ears is also dependent on the timing of tamoxifen administration such that earlier administration leads to a shorter cochlea.

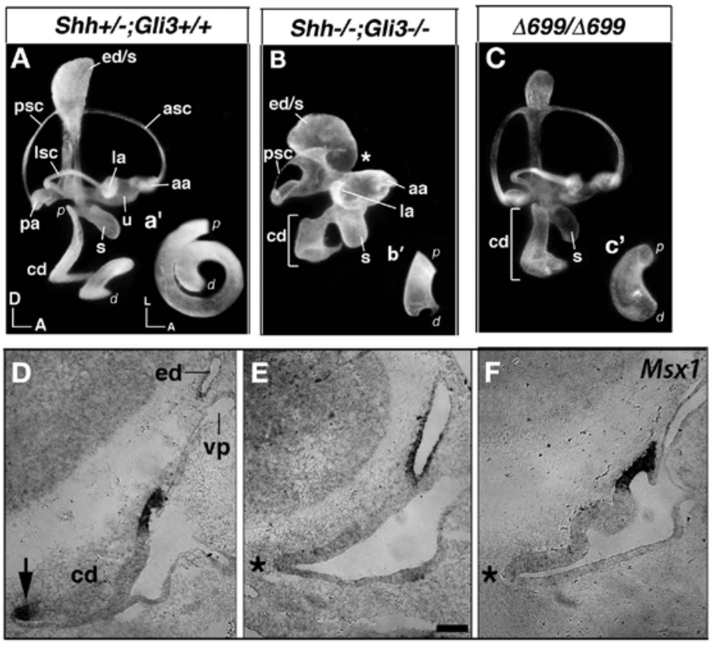


Fig. 4: Truncated cochlear ducts in *Shh*$^{-/-}$*; Gli3*$^{-/-}$ and *Δ*699/*Δ*699 mutants. (A-C) Paint-filled inner ears at E13.5 and (D-F) *Msx1* expression at E12.5 of *Shh*$^{+/-}$, *Gli3*$^{+/-}$ (A,D), *Shh*$^{-/-}$*; Gli3*$^{-/-}$ (B, E) and *Δ*699/*Δ*699 (C,F) mouse embryos. Cochlear ducts are shorter in *Shh*$^{-/-}$*; Gli3*$^{-/-}$ (B) and *Δ*699/*Δ*699 (C) ears than controls (A), and they are missing *Msx1* expression (E-F, asterisk), which marks the apical region of the cochlear duct (D, arrow). (a'-c') Ventral views of the cochlear duct shown in (A-C), respectively. Asterisk in (B) indicates the truncated anterior canal. Abbreviation: vp, vertical canal pouch. Adapted from Bok *et al.* (2007).

Does the shortened cochlear duct in these conditional mutants represent a truncation similar to the aforementioned *Shh*$^{-/-}$*; Gli3*$^{-/-}$ and *Δ*699/*Δ*699 mutant cochleae or a globally shortened duct? It is possible that specification of the apical cochlear duct requires higher or prolonged levels of Shh than other regions of the cochlear duct and that this extra Shh is supplied by the spiral ganglion acting in conjunction with the notochord and floor plate. Under such a scenario, the reduction of Shh signalling from the spiral ganglion in the *Shh* conditional mutants should affect apical cochlear development and abolish *Msx1* expression. In contrast, our analyses indicate that *Msx1* is expressed in the apical region of the shortened *Foxg1*^{cre}*; Shh*$^{lox/-}$ cochlea suggesting that this cochlear duct is only shortened and not truncated, unlike the *Shh*$^{-/-}$*; Gli3*$^{-/-}$ and *Δ*699/*Δ*699 cochleae. Taken together, these results suggest that Shh secreted by the spiral ganglion mediates only the growth of the cochlear duct, pre-patterned by Shh in the notochord and floor plate.

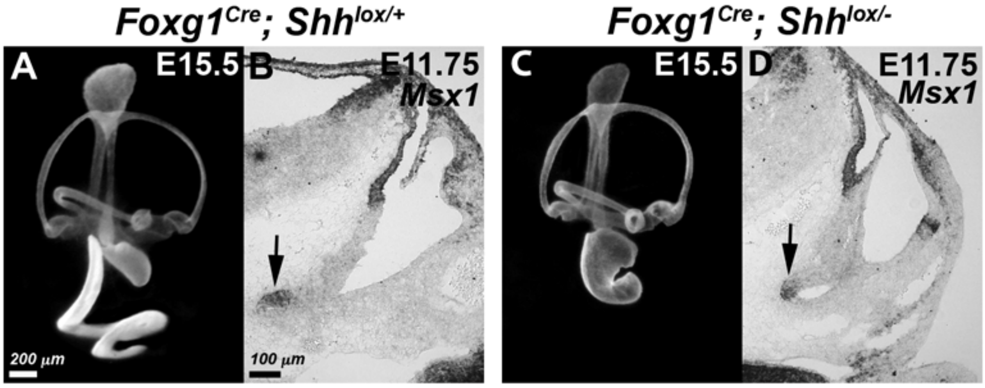


Fig. 5: The cochlear duct in the *Shh* conditional knockout mutant, *Foxg1*^{cre}*;* *Shh* $^{lox/-}$, is globally shortened. The cochlear duct in *Foxg1*^{cre}*; Shh* $^{lox/-}$ ears is shortened (C) but *Msx1* expression in the apex (D, arrow) is similar to controls (A, B). Adapted from Bok *et al.* (2013).

Spiral ganglion Sonic hedgehog mediates timing of cell cycle exit and hair-cell differentiation:

An unusual feature of hair-cell development in the organ of Corti is that hair-cell precursors exit from cell cycle in an apex-to-base direction along the cochlear duct, whereas hair-cell differentiation is initiated at the mid-basal region and progresses bi-directionally after terminal mitosis is completed (Ruben, 1967; Lee *et al.*, 2006). Thus, hair cells at the basal region exit from cell cycle promptly after cell cycle exit whereas their counterparts in the apex delay the differentiation process for several days. Previous in vitro studies indicate that Shh inhibits cochlear hair-cell formation (Driver *et al.*, 2008). Using a *Shh* reporter strain, it was shown that *Shh* expression in the spiral ganglion gets restricted towards the apical cochlear region over time (Liu *et al.*, 2010). Thus it was postulated that this restriction of *Shh* expression in the apex might be regulating the basal to apical wave of hair cell differentiation in the organ of Corti (Liu et al., 2010).

We reasoned that if Shh in the spiral ganglion is inhibiting hair-cell differentiation after terminal mitosis, then the lack of *Shh* in the spiral ganglion should cause hair-cell differentiation to proceed in the same direction as cell cycle exit (progressing from apex to base), provided that Shh has no effect on cell cycle exit of hair-cell precursors. We first determined the timing of cell cycle exit in mutants by injecting a thymidine analog, EdU, at different developmental times and determined the percentages of labelled hair cells present at E18.5, when hair cells can be unequivocally identified (Bok *et al.*, 2013). In principal, cells that undergo terminal mitosis shortly after EdU injection should retain the EdU and thus be heavily labelled, whereas cells that are post-mitotic or have undergone multiple rounds of cell division during this developmental period should not be labelled. The results from this cell-cycle-exit analysis indicate that hair-cell precursors exit from cell cycle prematurely in the *Foxg1*^{cre}*; Shh* $^{lox/-}$ cochlea but still in an apical to basal direction, similar to the wildtype (Bok *et al.*, 2013). In contrast, immunostaining of nascent hair cells indicates that hair-cell differentiation proceeds in the reverse apex-to-base direction predicted by our hypothesis (Fig. 6; Bok *et al.*, 2013). Together, these results indicate that Shh generated in the spiral ganglion promotes growth of the cochlear duct and proliferation of hair-cell precursors but inhibits hair-cell differentiation. Then, as *Shh* expression becomes restricted towards the apex of the cochlear duct, hair-cell differentiation is initiated starting at the basal cochlear region.

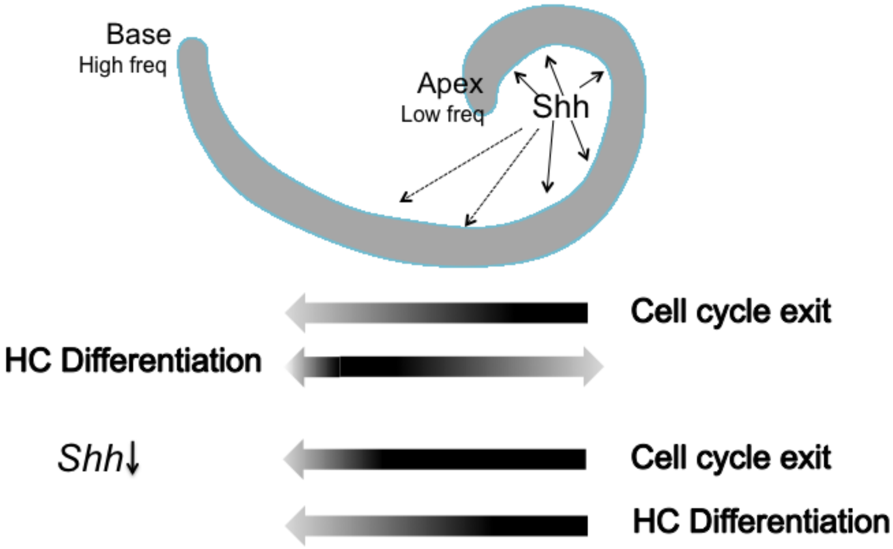


Fig. 6: Schematic summary of the role of Shh in regulating the timing of cell cycle exit and hair-cell differentiation in the mammalian cochlea. The mammalian cochlear duct is tonotopically organized such that hair cells at the base of the cochlea are tuned to high-frequency sound and hair cells at the apex to low-frequency sound. In the developing wildtype cochlea, hair-cell precursors exit from cell cycle in an apical to basal direction along the cochlear duct. Nevertheless, hair-cell differentiation is inhibited pending the restriction of Shh expression in the spiral ganglion towards the apical cochlear region (dotted arrows). Knocking out *Shh* expression in the spiral ganglion increases the timing of cell cycle exit and causes hair cells to differentiate promptly after cell cycle exit.

CONCLUSIONS

In summary, our results indicate that multiple sources of Shh are required for proper cochlear formation. Shh secreted by the notochord and floor plate is important for patterning the cochlear duct. At a slightly later time in development, Shh generated in the spiral ganglion mediates the growth of the cochlear duct. The dynamic relationship between the growing cochlear duct and the location of Shh expressing cells in the spiral ganglion dictate the timing of cell cycle exit of hair-cell precursors as well as differentiation of nascent hair cells in the cochlea. Finally, there are good evidence that suggest most of the aforementioned *Shh* functions are mediated by Shh acting directly on otic epithelial cells (Brown and Epstein, 2011; Tateya *et al.*, 2013).

REFERENCES

Bok, J., Bronner-Fraser, M., and Wu, D.K. (**2005**). "Role of the hindbrain in dorsoventral but not anteroposterior axial specification of the inner ear," Development, **132**, 2115-2124.

Bok, J., Dolson, D.K., Hill, P., Ruther, U., Epstein, D.J., and Wu, D.K. (**2007**). "Opposing gradients of Gli repressor and activators mediate Shh signaling along the dorsoventral axis of the inner ear," Development, **134**, 1713-1722.

Bok, J., Raft, S., Kong, K.A., Koo, S.K., Drager, U.C., and Wu, D.K. (**2011**). "Transient retinoic acid signaling confers anterior-posterior polarity to the inner ear," Proc. Natl. Acad. Sci. USA, **108**, 161-166.

Bok, J., Zenczak, C., Hwang, C.H., and Wu, D. K. (**2013**). "Auditory ganglion source of Sonic hedgehog regulates timing of cell cycle exit and differentiation of mammalian cochlear hair cells," Proc. Natl. Acad. Sci. USA, **110**, 13869-13874.

Borycki, A., Brown, A.M., and Emerson, C.P., Jr. (**2000**). "Shh and Wnt signaling pathways converge to control Gli gene activation in avian somites," Development, **127**, 2075-2087.

Bose, J., Grotewold, L., and Ruther, U. (**2002**). "Pallister-Hall syndrome phenotype in mice mutant for Gli3," Hum. Mol. Genet., **11**, 1129-1135.

Braunstein, E.M., Crenshaw Iii, E.B., Morrow, B.E., and Adams, J.C. (**2008**). "Cooperative function of Tbx1 and Brn4 in the periotic mesenchyme is necessary for cochlea formation," J. Assoc. Res. Otolaryngol., **9**, 33-43.

Brown, A.S., and Epstein, D.J. (**2011**). "Otic ablation of smoothened reveals direct and indirect requirements for Hedgehog signaling in inner ear development," Development, **138**, 3967-3976.

Chang, W., Cole, L.K., Cantos, R., and Wu, D.K. (**2003**). "Molecular genetics of vestibular organ development," in *Springer Handbook of Auditory Research: The vestibular system*, Vol. 19. Edited by S.M. Highstein, R.F. Fay, and A.N. Popper (New York: Springer-Verlag).

Driver, E.C., Pryor, S.P., Hill, P., Turner, J., Ruther, U., Biesecker, L.G., Griffith, A.J., and Kelley, M.W. (**2008**). "Hedgehog signaling regulates sensory cell formation and auditory function in mice and humans," J. Neurosci., **28**, 7350-7358.

Hayashi, S., and McMahon, A.P. (**2002**). "Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse," Dev. Biol., **244**, 305-318.

Hebert, J.M., and McConnell, S.K. (**2000**). "Targeting of cre to the Foxg1 (BF-1) locus mediates loxP recombination in the telencephalon and other developing head structures," Dev. Biol., **222**, 296-306.

Ingham, P.W., and McMahon, A.P. (**2001**). "Hedgehog signaling in animal development: paradigms and principles," Genes Dev., **15**, 3059-3087.

Kang, S., Graham, J.M., Jr., Olney, A.H., and Biesecker, L.G. (**1997**). "GLI3 frameshift mutations cause autosomal dominant Pallister-Hall syndrome," Nat. Genet., **15**, 266-268.

Koundakjian, E.J., Appler, J.L., and Goodrich, L.V. (**2007**). "Auditory neurons make stereotyped wiring decisions before maturation of their targets," J. Neurosci., **27**, 14078-14088.

Lee, Y.S., Liu, F., and Segil, N. (**2006**). "A morphogenetic wave of p27Kip1 transcription directs cell cycle exit during organ of Corti development," Development, **133**, 2817-2826.

Li, C.W., Van De Water, T.R., and Ruben, R.J. (**1978**). "The fate mapping of the eleventh and twelfth day mouse otocyst: an in vitro study of the sites of origin of the embryonic inner ear sensory structures," J. Morphol., **157**, 249-267.

Liang, J.K., Bok, J., and Wu, D.K. (**2010**). "Distinct contributions from the hindbrain and mesenchyme to inner ear morphogenesis," Dev. Biol., **337**, 324-334.

Litingtung, Y., Dahn, R.D., Li, Y., Fallon, J.F., and Chiang, C. (**2002**). "Shh and Gli3 are dispensable for limb skeleton formation but regulate digit number and identity," Nature, **418**, 979-983.

Liu, Z., Owen, T., Zhang, L., and Zuo, J. (**2010**). "Dynamic expression pattern of Sonic hedgehog in developing cochlear spiral ganglion neurons," Dev. Dyn. **239**, 1674-1683.

Morsli, H., Choo, D., Ryan, A., Johnson, R., and Wu, D.K. (**1998**). "Development of the mouse inner ear and origin of its sensory organs," J. Neurosci., **18**, 3327-3335.

Phippard, D., Lu, L., Lee, D., Saunders, J.C., and Crenshaw, E.B., 3rd (**1999**). "Targeted mutagenesis of the POU-domain gene Brn4/Pou3f4 causes developmental defects in the inner ear," J. Neurosci., **19**, 5980-5989.

Quinones, H.I., Savage, T.K., Battiste, J., and Johnson, J.E. (**2010**). "Neurogenin 1 (Neurog1) expression in the ventral neural tube is mediated by a distinct enhancer and preferentially marks ventral interneuron lineages," Dev. Biol., **340**, 283-292.

Raft, S., Koundakjian, E.J., Quinones, H., Jayasena, C.S., Goodrich, L.V., Johnson, J.E., Segil, N., and Groves, A.K. (**2007**) "Cross-regulation of Ngn1 and Math1 coordinates the production of neurons and sensory hair cells during inner ear development," Development, **134**, 4405-4415.

Riccomagno, M.M., Martinu, L., Mulheisen, M., Wu, D.K., and Epstein, D.J. (**2002**). "Specification of the mammalian cochlea is dependent on Sonic hedgehog," Genes Dev., **16**, 2365-2378.

Riccomagno, M.M., Takada, S., and Epstein, D.J. (**2005**). “Wnt-dependent regulation of inner ear morphogenesis is balanced by the opposing and supporting roles of Shh,” Genes Dev., **19**, 1612-1623.

Ruben, R.J. (**1967**). “Development of the inner ear of the mouse: a radioautographic study of terminal mitoses,” Acta Otolaryngol. Suppl., **220**, 1-44.

Swanson, G.J., Howard, M., and Lewis, J. (**1990**). “Epithelial autonomy in the development of the inner ear of a bird embryo,” Dev. Biol., **137**, 243-257.

Tateya, T., Imayoshi, I., Tateya, I., Hamaguchi, K., Torii, H., Ito, J., and Kageyama, R. (**2013**). “Hedgehog signaling regulates prosensory cell properties during the basal-to-apical wave of hair cell differentiation in the mammalian cochlea,” Development, **140**, 3848-3857.

te Welscher, P., Zuniga, A., Kuijper, S., Drenth, T., Goedemans, H.J., Meijlink, F., and Zeller, R. (**2002**). “Progression of vertebrate limb development through SHH-mediated counteraction of GLI3,” Science, **298**, 827-830.

Wu, D.K., and Kelley, M.W. (**2012**). “Molecular mechanisms of inner ear development,” Cold Spring Harb. Perspect. Biol., **4**, a008409.

Are receptive fields fixed or fluid?

JESSICA DE BOER[1,*], PAUL BRILEY[2], AND KATRIN KRUMBHOLZ[1]

[1] *MRC Institute of Hearing Research, Nottingham, NG7 2RD, United Kingdom*

[2] *Department of Psychology, University of York, York, YO10 5DD, United Kingdom*

Neural representations of sensory stimuli are affected by stimulus- and task context. These effects can be long term, such as observed after intensive training or sensory deprivation, or short term, for instance when stimuli are repeated or attended. Long-term effects are generally associated with changes in neural receptive fields, such as expanded representation of, and increased selectivity for, learned features after training, or cortical remapping after hearing loss. In contrast, short-term context effects are usually explained in terms of either suppressive (e.g., repetition suppression) or facilitatory (e.g., attentional facilitation) gain control, without any change in neural coding parameters. More recent models, however, propose that short-term effects, such as repetition suppression or attention, act not only through gain control of neuron populations, but also change the receptive fields of individual neurons. In this view, receptive fields are considered not as fixed, but rather as fluid and instantly adaptable. In this paper, new data are presented, based on non-invasive electrophysiological recordings in humans, which support the notion that short-term context effects cause rapid receptive-field plasticity.

INTRODUCTION

Neural receptive fields

The receptive field (RF) of a sensory neuron describes the selectivity with which that neuron responds to a particular stimulus feature. For example, the RF of an auditory neuron is characterised by the sound frequency that it is most responsive to, and by the steepness with which its responsiveness falls off with distance from this characteristic frequency (CF). This variation of responsiveness with frequency is also referred to as the RF 'tuning'. For primary auditory neurons, the RF is determined by the mechanical frequency tuning of the cochlea, and the neurons particular location along the tonotopic cochlear axis. At more central stages of the auditory pathway, the neural RF is determined by the synaptic input circuitry to the neuron, which receives converging afferent input from multiple units from more peripheral layers. Despite this convergence, the tonotopic arrangement originating from the cochlea is maintained along the ascending auditory pathway all the way to the auditory cortex, where CF varies gradually across the cortical surface, giving rise to a topographic representation of frequency. Such topographic maps of stimulus

[*] Corresponding author: jdb@ihr.mrc.ac.uk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

features are ubiquitous in the sensory cortices, and form the basis of the neural representations of sensory stimuli that underlie perception.

Experience-related receptive-field plasticity

Theoretically, neural receptive fields must be considered stable entities, in order to support deterministic neural models that provide perceptual constancy. However, in reality, neural receptive fields are known to be susceptible to modification by experience. In the auditory cortex, the cortical area that responds to a given sound frequency has been shown to be *expanded* after a period of intense identification training on that particular frequency (Polley *et al.*, 2006); vice versa, the cortical representation of a sound frequency has been found to *disappear* when peripheral sensitivity at that frequency is lost after noise trauma (Eggermont and Roberts, 2004). These cortical reorganisations are assumed to occur as a result of changes in the receptive fields of individual neurons, reflecting modifications to the neurons input circuitry. This receptive-field plasticity leads to changes in perception, which can be either beneficial, such as perceptual learning after training (Polley *et al.*, 2006), or detrimental, such as development of tinnitus after high-frequency hearing loss (Eggermont and Roberts, 2004). Experience-related receptive-field plasticity is generally assumed to develop over a relatively long time period, in the order of days or weeks. However, it is well-known that both neural and perceptual responses to sensory stimuli can be substantially affected by immediate experience on a much shorter time scale. For instance, attention can be switched between sensory streams within seconds, and is known to drastically and selectively alter the perceptual acuity for and neural responsiveness to sensory stimuli (Scharf *et al.*, 1987, Woldorff *et al.*, 1993). Another example is repetition suppression, or adaptation, which refers to the reduction in neural responsiveness after repeated stimulation. Adaptation is ubiquitous in the sensory cortex, where it acts on a time scale of 100s of milliseconds, and has been implicated in perceptual priming, the improved perceptual acuity for a repeated stimulus, as well as streaming and novelty detection (Grill-Spector *et al.*, 2006). Classically, both attention and adaptation have been considered to act through a gain mechanism, which either increases or decreases the input-output gain of selected neurons, without changing their receptive-field properties. More recently, alternative models have been put forward in which these types of short-term effects also affect the selectivity of neural responses to sensory stimuli. This suggests that neural receptive fields would be susceptible to modifications on a much more rapid time scale than has previously been assumed. Here, new results are presented that investigate this hypothesis by examining the short-term effects of immediate stimulus context and attention on neural receptive fields in the human auditory cortex. For this purpose, we recorded auditory evoked potentials (AEP) non-invasively using electro-encephalography (EEG). In order to infer neural receptive-field properties from the resulting AEPs, we used so-called adaptation paradigms that reveal feature selectivity of the neural population underlying the response. The principle of adaptation paradigms as a tool for measuring neural selectivity is explained below.

Measuring neural receptive fields non-invasively using adaptation paradigms

Adaptation has been observed at each spatial level of neural processing, ranging from single units in the auditory cortex (Wehr and Zador, 2005) to population responses captured in neuroimaging (Grill-Spector *et al.*, 2001; 2006). One particularly interesting property of adaptation is that it is stimulus-specific. This means that the reduction in neural responsiveness after repeated stimulation is greater when the repeated stimulus is identical than when one or more of its features is changed. The increased response elicited by a change in a repeated stimulus is referred to as the 'release from adaptation'. For population responses, a release from adaptation will be elicited only if the underlying neural population is selective for the changed stimulus feature. In this case, the release from adaptation is assumed to arise from activation of a 'fresh' subpopulation of neurons that had not been adapted, because the preceding stimulus fell outside these neurons' receptive fields (May and Tiitinen, 2009; Grill-Spector *et al.*, 2006). According to this 'fresh afferents' model, the release from adaptation will increase with increasing difference between two subsequent stimuli along the relevant feature dimension, as there will be increasingly less overlap between the neural populations responding to the first and second stimulus. In other words, the release from adaptation reflects the receptive-field tuning of the underlying neural population to that particular stimulus feature. A useful practical implication of this is that adaptation can be used to measure receptive-field properties of neural populations using non-invasive neuroimaging methods. Adaptation paradigms were first pioneered by Grill-Spector and colleagues, who applied them to functional magnetic resonance imaging (fMRI) to investigate feature selectivity in different areas of the human visual cortex (Grill-Spector *et al.*, 2001). They used a block design, in which they measured the average blood oxygenation level dependent (BOLD) response to blocked sequences of repeated stimuli, and compared responses between blocks in which the stimuli varied along different feature dimensions. Those feature changes for which the BOLD responses were largest were assumed to have elicited the greatest release from adaptation, and thus that feature was interpreted to be selectively represented by the underlying neural population. An alternative design is used to measure the neural selectivity to one particular stimulus. In this 'event-related' design, discrete trials are presented in which one stimulus (the adapter) is followed by another (the probe), with inter-trial intervals long enough to allow for recovery from adaptation between trials. Here, the response to the probe and the adapter are measured separately, and the amount of adaptation is measured by comparing the size of the (adapted) probe response to that of the (unadapted) adapter. By plotting the amount of adaptation as a function of the difference between adapter and probe along a particular feature dimension, an adaptation tuning curve is constructed that reflects the sharpness of neural tuning to that feature in the particular neural population that responds to the adapter.

METHODS

All experiments reported here recorded cortical AEPs in response to pure-tone stimuli of 100-150 ms duration presented with a stimulus onset asynchrony (SOA) of 500 ms. Stimuli were presented binaurally over headphones at 60 dB SPL to participants seated in a sound-attenuating and electrically-shielded booth. EEG signals were recorded from 33 electrodes placed according to the standard 10-20 arrangement. Data dimensionality was reduced either by fitting the data to a source model and extracting the average source waveform (experiment 1), or by calculating the global field power, which is the root-mean-square over all channels at each time point (experiments 2 and 3). The resulting AEPs showed the typical P1, N1, and P2 deflections, which are obligatory responses originating from the auditory cortex. Individual responses were quantified by the peak-to-peak amplitude difference between consecutive deflections. The difference between P1 and N1 is referred to as ‘N1’, and the difference between N1 and P2 as ‘P2’. The ‘N1’ component is thought to represent neural responses from the more peripheral input layer into the auditory cortex, whereas the ‘P2’ component is assumed to reflect more central, intra-cortical connections. All participants were normally-hearing young adults. 15 participants were tested in experiment 1, 24 in experiment 2, and 12 in experiment 3.

EXPERIMENT 1: DOES REPEATED EXPOSURE TO AN ADAPTER SHARPEN ADAPTATION TUNING?

Background

Both evoked potential studies (May and Tiitinen, 2009) and invasive recordings from the auditory cortex (Ulanovsky *et al.*, 2003) have reported that the neural response to the same stimulus increases as its occurrence in an oddball sequence becomes rarer. At first glance, this effect might be ascribed to stimulus-specific adaptation, as the response to a more often repeated stimulus would be more adapted and thus smaller. However, a study by Taaseh and colleagues reported that this ‘deviant’ response is elicited even when adaptation effects are controlled for (Taaseh *et al.*, 2011). Based on a modelling approach, the authors proposed that the deviant response results from a sharpening of adaptation tuning after repeated presentation of the adapting stimulus. This sharpening would decrease the overlap between the neurons activated by the probable and the rare stimuli in the oddball sequence, compared to a sequence in which the two stimuli are equally probable. This would cause a greater release from adaptation for the rare stimulus, resulting in the observed deviant response. While this hypothesis explained the results well, the sharpening hypothesis is not unequivocally supported by the findings. This is because the responses were measured in continuous sequences, in which the effect of single versus repeated adapters could not be evaluated separately. In order to test the sharpening hypothesis proposed by Taaseh and colleagues directly, we performed an experiment which compared the amount of adaptation after a single, two, or three identical adapters.

Design

An event-related paradigm was used in which discrete adapter-probe trials were presented with an inter-trial interval of 5 s. Based on current estimates of adaptation recovery time, this ensured that no adaptation effects from a preceding trial spilled over to a subsequent trial. The probe frequency was fixed at 1 kHz, and the adapter frequency ranged between 0 to 1.5 octaves above the probe frequency. The amount of adaptation was calculated as the ratio of the P2 amplitude of the probe to that of the first adapter in each trial, which represents the unadapted response.

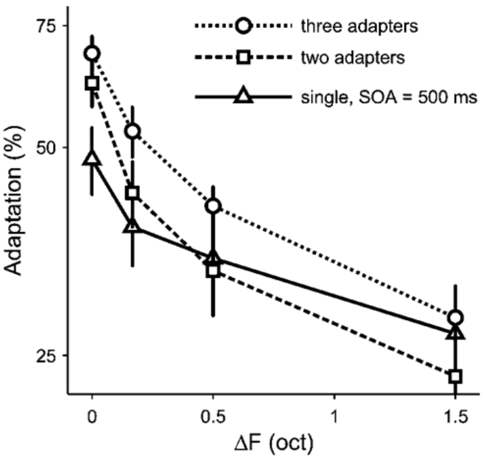


Fig. 1: Adaptation tuning curves compared for probes preceded by a single (triangles, solid line), two (squares, dashed line) and three (circles, dotted line) adapters presented in discrete trials. Plots show mean and standard error across participants of the percent adaptation of the P2 component of the AEP. [From Briley and Krumbholz, in revision.]

Results

The adaptation tuning curves measured for the single, two, and three adapter conditions are compared in Fig. 1. At zero frequency difference between the adapter and the probe (DF = 0), the amount of adaptation increases progressively with the number of adapters. This would be expected, as the effect of the successive adapters on the probe add up. However, as the frequency difference between adapter and probe increases, it can be seen that the amount of adaptation falls off more rapidly for multiple adapters than a single adapter. Notably, at the largest frequency difference, the multiple adapters are in fact no more effective than the single adapter. These effects were found to be significant, and suggest that multiple adapters are relatively less effective at adapting a deviant frequency than a single adapter. This supports the hypothesis put forward by Taaseh and colleagues that adaptation tuning is sharpened after repeated presentation of an adapter.

EXPERIMENT 2: DOES EXPOSURE LEAD TO LONGER-TERM SHARPENING OF ADAPTATION TUNING?

Background

The findings of experiment 1 suggest that the neural population that underlies the adaptation effect after multiple adapters is more sharply tuned than the neural population that is adapted after a single adapter. This could imply either that the adaptation effects observed for multiple versus single adapters involve different neural populations, or that the same neural population has become more sharply tuned after repeated exposure to the adapter. This latter effect could form a neural basis of perceptual priming, as a sharper representation of the repeated stimulus would be expected to improve the perceptual acuity for that stimulus. Priming has been proposed to be a short-term precursor to longer-term perceptual learning. By analogy, we hypothesized that the short-term sharpening observed here might form a precursor for longer-term receptive-field plasticity. In order to test this hypothesis, the next experiment investigated the longer-term effect of repeated exposure on adaptation tuning.

Design

AEPs were recorded in response to stimuli presented in random sequences in which one stimulus, here referred to as the adapter, was presented in 40% of trials, and six different stimuli, here referred to as the deviant probes, were presented in 10% of trials each. The deviant probes had frequencies spaced symmetrically within half an octave around the frequency of the adapter. Here, the average response to the adapter represents the amount of adaptation for a zero frequency difference, whereas the average response to the deviant probes reflects the release from adaptation as a function of frequency difference between adapter and probe. Two conditions were compared: in the ‘fixed’ condition, the adapter frequency was fixed throughout the recording at 1000 Hz; in the ‘roving’ condition, the frequency of the adapter was changed every two minutes, ranging within an octave around 1000 Hz. It was hypothesized that if any longer-term (i.e., > 2 minutes) sharpening effects occurred, this would lead to a difference in the adaptation tuning between the fixed and the roving condition. This is because in the roving condition, there would be no time for longer-term effects of exposure to the adapter to develop.

Results

First, the data were analyzed to estimate short-term sharpening effects. To this purpose, the responses to each stimulus were separately averaged depending on whether they were preceded by one, two, or three adapters in the random sequence. This analysis mimics the multiple and single adapter trials in experiment 1, but here the ‘trials’ were not discrete but embedded at random locations within the continuous random sequence. As these are very short-term effects, the data were averaged over the fixed and roving conditions, and over the entire duration of stimulus presentation and recording, which was 1.5 hours (interrupted by short breaks every 15 minutes). The resulting adaptation tuning curves are shown in Fig. 2A and 2B for the N1 and P2 components, respectively. It is immediately evident that the release from adaptation with frequency difference increases as the number of preceding adapters increases. This effect is significant at all deviant frequencies. Importantly, however, at zero frequency difference there is no significant change in response size for either the N1 or the P2 with increasing number of adapters. These results are analogous to the findings of experiment 1, and similarly suggest that the adaptation effect is more sharply tuned after multiple than after single presentations of the adapter.

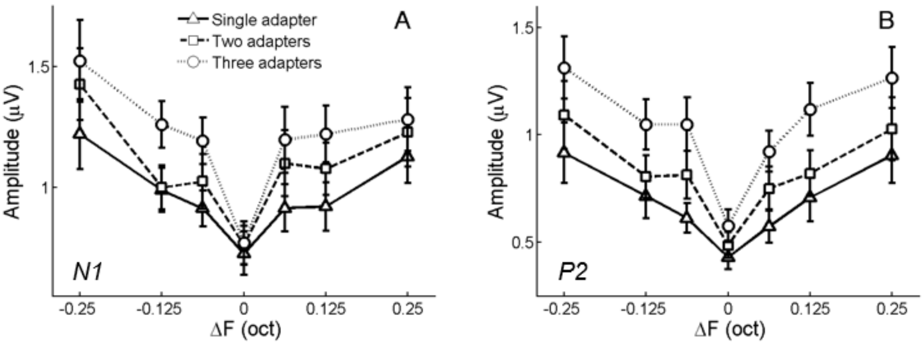


Fig. 2: Adaptation tuning curves compared for probes immediately preceded by one (triangles, solid line), two (squares, dashed line), and three (circles, dash-dot line) adapters in a random sequence. A: N1 B: P2.

Next, the data were analyzed for medium-term effects of exposure to the adapter. Responses to each stimulus were averaged separately over a consecutive time period of 15 minutes. If exposure to the adapter caused effects with memory spans of between 2 and 15 minutes, we would expect a difference between the fixed and the roving condition in the average tuning curve over the first 15 minutes. This comparison is shown in Fig. 3A and 3C for the N1 and P2 components, respectively. As can be observed, there was no significant difference between the two conditions for either component. Figures 3B and 3D show the same comparison for the average over the last fifteen minutes of the recording. Here, we would expect differences that might have developed over the preceding 45 minutes of exposure to the adapter. In fact, the only significant difference observed was a larger decrease in the N1 in the fixed versus the roving condition, which developed gradually during the recording session. This indicates that there is an exposure effect on the N1 with a memory span of up to 45 minutes, but this effect was not frequency-specific. The fact that the P2 did not show this effect may indicate that it is already maximally adapted after seconds of exposure, which is in line with previous findings. In summary, the results of experiment 1 provide further evidence that adaptation tuning is sharpened after repeated exposure to the adapter, but indicate that this is a purely short-term effect, with a memory span in the order of seconds.

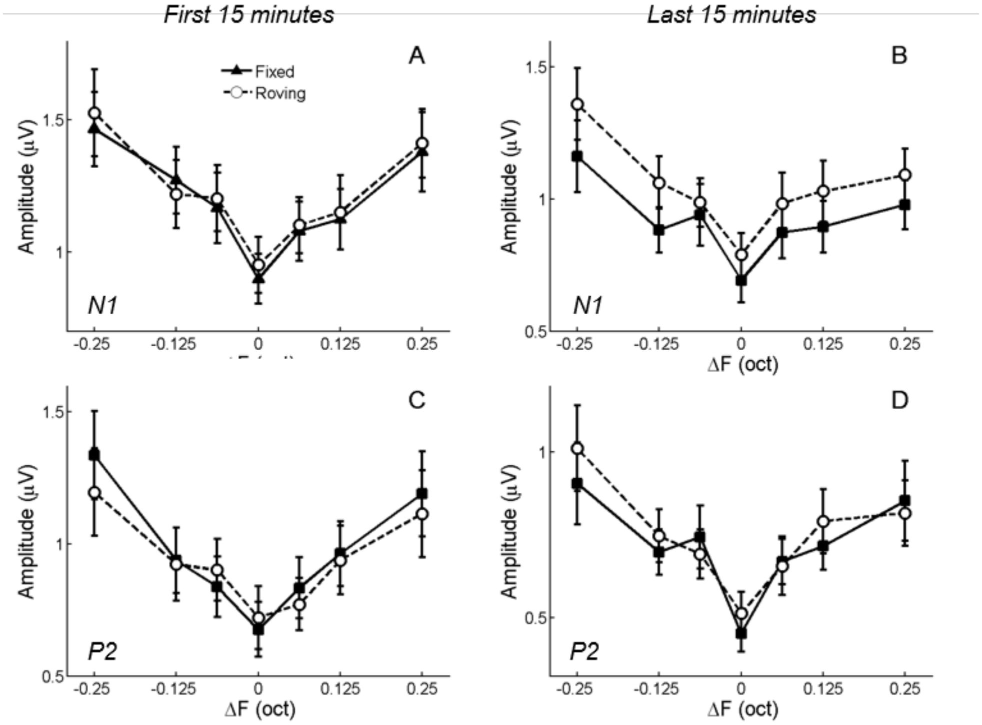


Fig. 3: Adaptation tuning curves obtained with a fixed (filled squares, solid lines) and a roving (open circles, dashed lines) adapter frequency. A and C: First 15 minutes of recording. B and D: Last 15 minutes of recording. A and B: N1; C and D: P2.

MECHANISMS UNDERLYING SHORT-TERM SHARPENING?

Two alternative mechanisms have been hypothesized to underlie the sharpening of adaptation tuning after repeated adapters. These different models explain the effect as arising from bottom-up and top-down processes, respectively.

Bottom-up explanation

Mill and colleagues developed a computational model in which sharpening of adaptation tuning is an emergent property of a convergent network of depressing synapses (Mill *et al.*, 2011). The model assumes that adaptation results from synaptic depression, which is supported by neurophysiological evidence (Wehr and Zador, 2005), and provides a good match to the time course of adaptation. An essential feature of the model is that, as a result of convergence, receptive-field tuning becomes broader from peripheral to central synaptic layers. It is then posited that more peripheral synapses are not depressed after a single adapter, but become depressed after repeated adapters. As a result, the peripheral synapses stop firing, which in turn allows more central synapses, which receive input from the peripheral layers, to recover and resume firing. In this situation, the adaptation tuning measured at the more central layer will actually reflect the tuning at the more peripheral, and thus more sharply tuned, neural layer. Thus, in this model the adaptation observed after multiple adapters reflects the tuning of a different neural population, rather than a change in tuning in the same population. Although this parsimonious model is compelling, it provides only a qualitative explanation of the findings, and is as yet not supported by any direct neurophysiological evidence.

Top-down explanation

An alternative hypothesis is that repeated stimulation elicits top-down feedback processes that modify neural receptive fields through efferent pathways that alter the synaptic input circuitry of individual neurons. Clearly, such a top-down feedback effect would have to act very rapidly to explain the short-term effects observed here. There is some evidence that rapid receptive-field sharpening can be elicited by top-down mechanisms from studies of selective attention. In an fMRI study, Murray and Wojciulik measured release from adaptation in the visual cortex in response to a change in the orientation of a visual stimulus (Murray and Wojciulik, 2004), and found that when the stimulus was selectively attended, the release from adaptation was increased. This indicates that attention caused an increased neural selectivity, i.e., a sharpening of receptive-field tuning, to stimulus orientation. As attention acts in a very immediate manner and can be switched rapidly, these findings suggest that top-down modulation of neural receptive fields can occur within a very short time.

EXPERIMENT 3: DOES ATTENTION SHARPEN NEURAL TUNING?

Background

In the auditory system, evidence of rapid task-related receptive-field plasticity has been reported from single neuron recordings in auditory cortex, which were suggested to result from top-down attentional modulation (Fritz *et al.*, 2003). However, results from human neuroimaging have been confounded by the use of paradigms in which apparent changes in selectivity could have resulted from changes in attentional load (e.g., Ahveninen *et al.*, 2011). In the final experiment presented here, we tested the hypothesis that attention sharpens neural tuning in the human auditory cortex directly, using a similar approach to Murray and Wojciulik.

Design

AEPs were recorded while participants performed a dichotic listening task. Pseudo-random tone sequences (Brimijoin and O’Neill, 2010) comprising four equally-probable frequencies were presented to one ear, while simultaneously a sequence of amplitude-modulated noises was presented to the other ear. The participants were instructed to attend to one ear at a time, which was changed every 2.5 minutes, and detect rare oddballs in the attended stream. In the tone sequences, the oddball was frequency modulated, whereas in the noise sequences, the oddball had a rising rather than falling amplitude profile. The modulation parameters of both types of oddballs were set to achieve an equal hit rate of ~75%. The noises were presented with an SOA of 666 ms plus a jitter ranging between 0 and 100 ms, to avoid synchronization to the tones. Only AEPs to the tones were recorded.

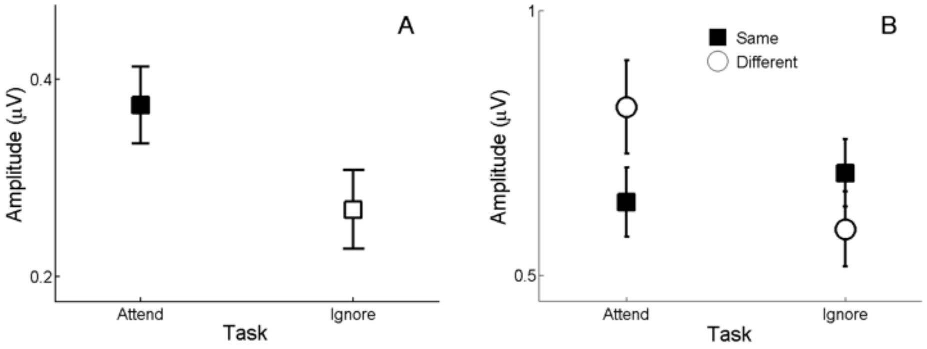


Fig. 4: Effect of attention on P2 amplitude of the AEP. A: Average response to all tones. B: Responses to tones preceded by the same frequency ('Same', filled squares) or by a different frequency ('Different', open circles).

Results

As expected, the response to the tones was significantly larger when participants attended to the tones ('attend') than when they ignored the tones and attended to the noises ('ignore'). This illustrated for the P2 amplitude in Fig. 4A. In order to evaluate whether attention caused a sharpening of neural tuning, the responses were separately averaged depending on whether they were immediately preceded by the same frequency ('same') or by a different frequency ('different'). The difference between these two conditions reflects the degree of frequency-dependent release from adaptation, with the 'different' response expected to be larger than the 'same' response. If attention sharpens frequency selectivity, we would expect a greater release from adaptation in the 'attend' versus the 'ignore' condition. In figure 4B, this comparison is shown for the P2 component. Note that in the 'attend' condition, the response to the 'different' tone is larger than the response to the 'same' tone, whereas in the 'ignore' condition, the 'same' response is slightly larger than the 'different' response. Statistical analysis revealed that the release from adaptation was significantly larger in the 'attend' than in the 'ignore' condition. This is similar to the findings of Murray and Wojciulik, and supports the hypothesis that attention sharpens neural tuning.

CONCLUSIONS

The close agreement between the results of experiment 1 and 2 and the findings of Taaseh and colleagues provides compelling evidence of sharpening of adaptation tuning after repeated adapters. However, the neural mechanism that underlies this effect has not yet been ascertained. Nevertheless, the results from experiment 3 suggest that receptive-field tuning is susceptible to rapid modulation via top-down pathways. It is plausible that the sharpening elicited by a repeated adapter results from a similar top-down mechanism. The stage of processing at which this top-down modification is effected is not necessarily at the cortex, but could be inherited from earlier stages of processing, via efferent connections that reach back towards the periphery. Efferent effects can even reach as far down as the cochlea, where they have been reported to mediate frequency-specific attentional modulation of cochlear gain (de Boer and Thornton, 2007; Maison *et al.*, 2001). Such peripheral effects could alter cortical receptive fields by changing the synaptic input into the cortex.

ACKNOWLEDGEMENTS

We thank Sarah Gibbs, Dominic Friston, and Margareta Szymczak for their assistance in data acquisition and processing; Owen Brimijoin for assistance with creation and analysis of tone sequences; Rob Mill, Cris Lanting, and Julien Besle for helpful discussions. This research was funded by the Medical Research Council (UK).

REFERENCES

Ahveninen, J., Hamalainen, M., Jaaskelainen, I.P., Ahlfors, S.P., Huang, S., and Lin, F.H. (**2011**). "Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise," Proc. Natl. Acad. Sci. USA, **108**, 4182-4187.

Brimijoin, W.O., and O'Neill, W.E. (**2010**). "Patterned tone sequences reveal non-linear interactions in auditory spectrotemporal receptive fields in the inferior colliculus," Hear Res., **267**, 96-110.

de Boer, J., and Thornton, A.R. (**2007**). "Effect of subject task on contralateral suppression of click evoked otoacoustic emissions," Hear. Res., **233**, 117-123.

Eggermont, J.J., and Roberts, L.E. (**2004**). "The neuroscience of tinnitus," Trends Neurosci., **27**, 676-682.

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (**2003**). "Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex," Nat. Neurosci., **6**, 1216-1223.

Grill-Spector, K., and Malach, R. (**2001**). "fMR-adaptation: a tool for studying the functional properties of human cortical neurons," Acta Psychol., **107**, 293-321.

Grill-Spector, K., Henson, R., and Martin, A. (**2006**). "Repetition and the brain: neural models of stimulus-specific effects," Trends. Cogn. Sci., **10**, 14-23.

Maison, S., Micheyl, C., and Collet, L. (**2001**). "Influence of focused auditory attention on cochlear activity in humans," Psychophysiology, **38**, 35-40.

May, P.J., and Tiitinen, H. (**2010**). “Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained,” Psychophysiology, **47**, 66-122.

Mill, R., Coath, M., Wennekers, T., and Denham, S.L. (**2011**). “A neuro-computational model of stimulus-specific adaptation to oddball and Markov sequences,” PLoS Comput. Biol., **7**, e1002117.

Murray, S.O., and Wojciulik, E. (**2004**). “Attention increases neural selectivity in the human lateral occipital complex,” Nat. Neurosci., **7**, 70-74.

Polley, D.B., Steinberg, E.E., and Merzenich, M.M. (2**006**). “Perceptual learning directs auditory cortical map reorganization through top-down influences,” J. Neurosci., **26**, 4970-4982.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., and Reeves, A. (**1987**). “Focused auditory attention and frequency selectivity,” Percept. Psychophys., **42**, 215-223.

Taaseh, N., Yaron, A., and Nelken, I. (2011). “Stimulus-specific adaptation and deviance detection in the rat auditory cortex,” PLoS One, **6**, e23369.

Ulanovsky, N., Las, L., and Nelken I. (**2003**). “Processing of low-probability sounds by cortical neurons,” Nat. Neurosci., **6**, 391-398.

Wehr, M., and Zador, A.M. (**2005**). “Synaptic mechanisms of forward suppression in rat auditory cortex,” Neuron, **47**, 437-445.

Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., and Sobel, D. (**1993**). “Modulation of early sensory processing in human auditory cortex during auditory selective attention,” Proc. Natl. Acad. Sci. USA, **90**, 8722-8726.

The severity of developmental hearing loss does not determine the magnitude of synapse dysfunction

TODD M. MOWERY, VIBHAKAR C. KOTAK, AND DAN H. SANES*

Center for Neural Science, New York University, New York, New York, USA

The loss of auditory experience can disrupt synapse function, particularly when it occurs during development. However, the extent of hearing loss can vary from mild to profound, and the duration of hearing loss can vary from days to years. Here, we asked whether the dysfunction of central auditory synapses scales with the severity of hearing loss. The manipulations range from mild sound attenuation to complete deafferentation at the time of hearing onset. Synapse function is measured in central auditory structures from the cochlear nucleus to cortex. The core finding is that even a ~25 dB attenuation in sensation level produces a quantitatively similar change to synaptic currents and membrane properties, as compared to deafferentation. Therefore, profound changes to central processing may occur even when developmental hearing loss is mild, provided it occurs when sound is first transduced.

INTRODUCTION

Many functional properties of central auditory neurons are use-dependent. They can be altered by passive exposure to sound, as well as active experiences such as learning. Central auditory plasticity is particularly evident throughout development during which it supports normal maturation of auditory processing (Sanes and Bao, 2009; Sanes and Woolley, 2011). However, this sensitivity to auditory experience can also introduce a risk: when sound-evoked activity is reduced due to the loss of hearing, both synaptic and membrane properties can assume a dysfunctional state (for a recent review, see Sanes, 2013). Much of primary evidence in support of this theory emerges from experiments in which the cochlea is damaged or removed. For example, our work on bilateral hearing loss has examined the synaptic consequences in the superior olive, inferior colliculus, and auditory cortex (e.g., Kotak and Sanes, 1997; Vale and Sanes, 2000; Vale and Sanes, 2002; Vale *et al.*, 2003; Kotak *et al.*, 2005). Therefore, the magnitude of functional changes in the auditory circuits during less severe forms of hearing impairment such as during middle ear damage has not been thoroughly explored.

In this review, we explore the issue of hearing loss severity by comparing cellular measures obtained following different experimental manipulations to the auditory periphery. For developmental hearing loss, these studies suggest that mild hearing loss can produce changes to cellular properties that are similar to those observed following cochlear damage. These results suggest that auditory deprivation during a

*Corresponding author: sanes@cns.nyu.edu

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

sensitive time window at hearing onset (postnatal day 11 in gerbil) leads to functional deficits in the brain independent of the severity of peripheral dysfunction. Further, they imply that behavioural delays reported for transient hearing loss in humans could be attributable to central nervous system alterations.

NORMAL DEVELOPMENT OF AUDITORY SYNAPSE FUNCTION

The functional properties of synapses have been measured as a function of age in several species. The measurements are commonly made from slice of tissue through a central auditory structure that is maintained in a warm, oxygenated saline solution for several hours. The synaptic responses are obtained with whole-cell current- or voltage clamp recordings, which allow one to make direct measurements of synaptic potentials or currents, respectively. Both evoked and spontaneous synaptic events can be quantified, with the most common parameters being amplitude and kinetics.

Both central excitatory and inhibitory synapses are functional well before the onset of hearing. In rodents, evoked synaptic responses can be observed in tissue obtained at birth, whereas sound transduction by the cochlea is first observed over one week later (Sanes and Walsh, 1997; Fitzgerald and Sanes, 2001). In fact, spontaneous action potentials are also observed in the auditory central nervous system (CNS) before sound first activates the cochlea. These action potentials may be evoked by hair cell activity (Jones *et al.*, 2007; Tritsch *et al.*, 2007; Johnson *et al.*, 2011), and by mechanisms that are intrinsic to the CNS (Kotak *et al.*, 2007b; Tritsch *et al.*, 2010; Kotak *et al.*, 2012). In either case, there is a great deal of spontaneous synaptic transmission in the auditory CNS prior to the onset of sound-evoked responses.

The amplitude and decay time of synaptic potentials mature rapidly beginning at about the time of ear canal opening in rodents (about 9-12 days postnatal, depending on the species). Figure 1 illustrates the developmental time course for excitatory and inhibitory postsynaptic potentials (EPSPs and IPSPs) recorded in the mouse auditory cortex (Oswald and Reyes, 2008; 2011). The measurements of amplitudes and decay times demonstrate that maturation continues after the onset of hearing and an adult-like state is reached within a few weeks. Many of the observed functional changes are highly significant, suggesting that mature auditory processing depends on sufficient development of synapse function.

Recordings obtained in auditory brainstem nuclei are generally consistent with this rate of development. Within about two weeks of hearing onset, EPSPs and IPSPs recorded in the lateral and medial superior olivary nuclei display adult-like kinetics (Sanes, 1993; Kandler and Friauf, 1995; Scott *et al.*, 2005; Magnusson *et al.*, 2005; Chirila *et al.*, 2007). Furthermore, these findings are consistent with observations from many other auditory brainstem nuclei (Chuhma and Ohmori, 1998; Taschenberger and von Gersdorff, 2000; Brenowitz and Trussell, 2001; Balakrishnan *et al.*, 2003; Nakamura and Takahashi, 2007; Gao and Lu, 2008; Sanchez *et al.*, 2010). Although exceptions to this pattern of development are seldom observed, IPSC decay time in cortical pyramidal neurons displays a relatively prolonged maturation, only reaching an adult value at about 3 postnatal months (Takesian *et al.*, 2012). A late maturation of synaptic inhibition is consistent with the prolonged transition of $GABA_A$ receptor subunit expression in human cortex (Pinto *et al.*, 2010).

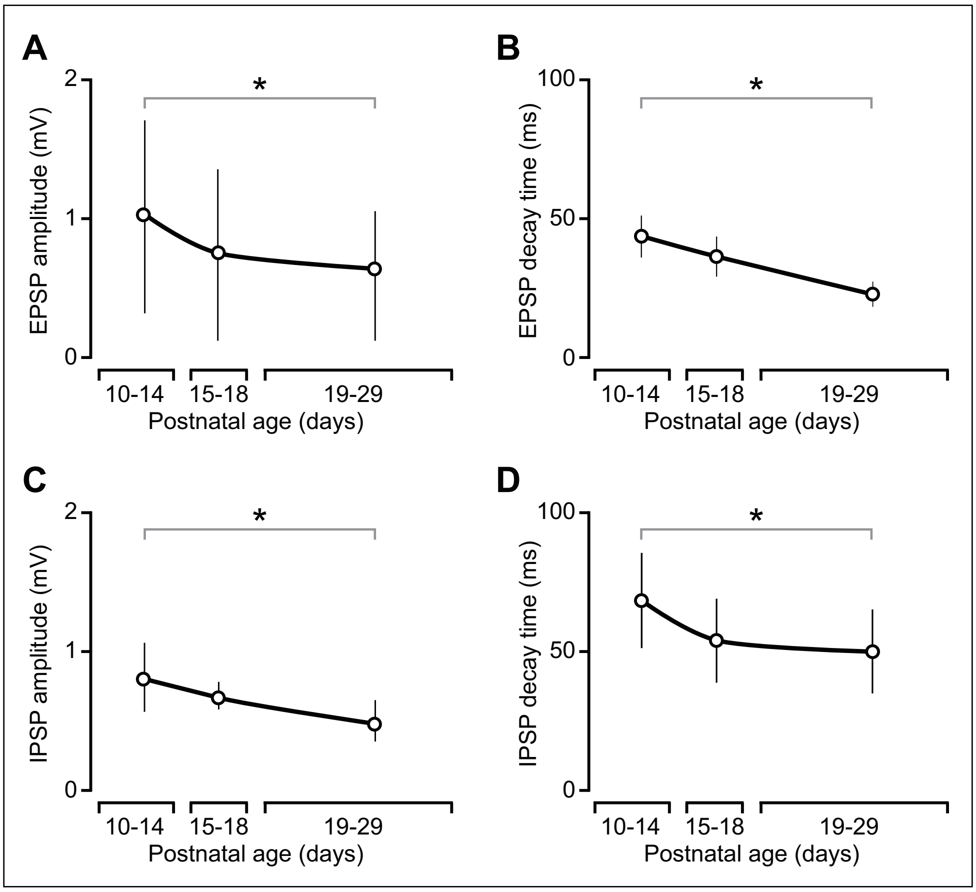


Fig. 1: Excitatory and inhibitory synaptic potential amplitudes and decay times measured from in vitro whole-cell recordings in mouse auditory cortex. Synaptic potentials are in response to stimulation of a single presynaptic neuron. In each graph, mature responses appear to emerge over about 14 days. There is a developmental decrease in (A) EPSP amplitude, (B) EPSP decay time, (C) IPSP amplitude, and (D) IPSP decay time. Asterisks indicate a statistically significant change. Adapted from Oswald and Reyes (2008; 2011).

The wealth of data obtained from recordings in brain slices is consistent with the few in vivo whole-cell studies that have been conducted on anesthetized animals. As shown in Fig. 2A, the decay times for EPSCs and IPSCs are relatively stable after about 25 days postnatal. Similarly, measures of sound-evoked plasticity mature at about the same rate. In animals younger than P21, sound stimulation leads to an increase in both EPSC and IPSC conductance (Fig. 2B), but this form of plasticity is absent after P25 (Dorrn *et al.*, 2010).

The mechanistic bases for changes in the amplitude of a synaptic response are manifold. For example, the number of neurotransmitter receptors at the synapse, the conductance of single receptor-coupled channels, the amount of transmitter released, and the distribution of ions across the membrane, can each determine the response amplitude. Furthermore, the specific molecular composition of a receptor will establish its mean open time when bound by neurotransmitter, and this will determine the decay time for a synaptic event. Therefore, measurement of synaptic amplitude and kinetics are a logical first step in determining the molecular and genetic basis for the effects associated with hearing loss.

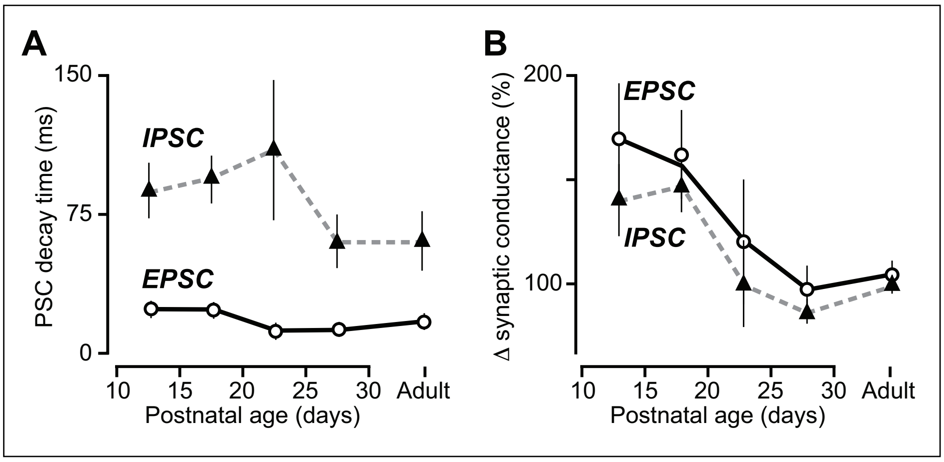


Fig. 2: Synaptic current decay times and sound-induced plasticity measured with in vivo whole-cell recordings from the rat auditory cortex. Mature responses are observed by postnatal day 25-30. (A) The decay times of sound-evoked EPSCs and IPSC decline after P20. (B) An increase in sound-evoked EPSC or IPSC conductance occurs following 3-5 min of sound stimulation, but this phenomenon fails to occur after P25. Adapted from Dorrn *et al.*, (2010).

SOUND ATTENTUATION VERSUS DEAFFERENTATION

Since synaptic responses display a well-characterized maturation in the rodent central auditory system, it is possible to study whether hearing loss induces an impairment. More importantly, it permits for the quantitative comparison of different forms of hearing loss. Conductive hearing loss (CHL), such as that which may occur during bouts of otitis media with effusion, leads to sound attenuation and a smaller neural response to a given SPL. Depending on its severity (e.g., ear canal atresia vs otosclerosis), the magnitude of sound attenuation can range from 10-50 dB. However, CHL is not associated with a direct injury to the cochlea, and the innervation density should not be altered. In contrast, sensorineural hearing loss (SNHL) involves a direct injury to the cochlea, and would include both a smaller neural response to a given SPL, as well as deafferentation of the CNS.

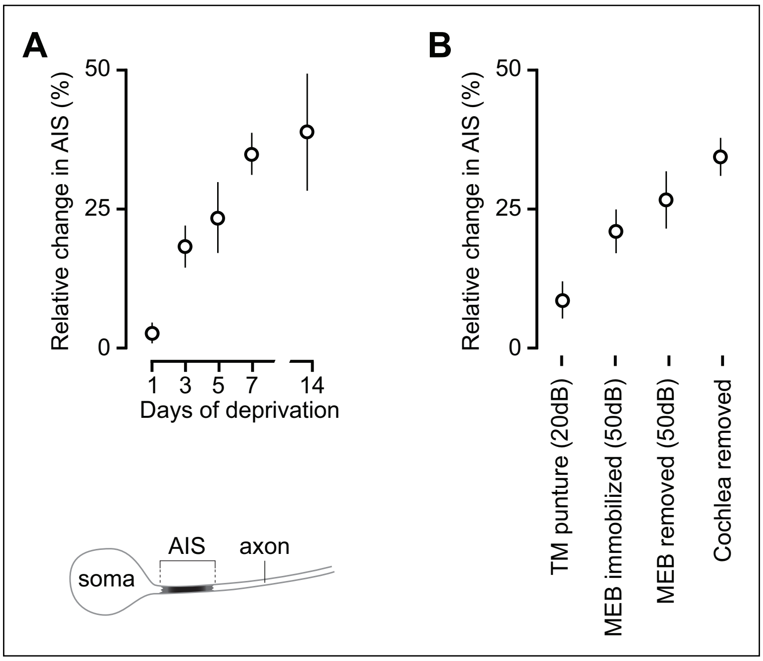


Fig. 3: Effect of hearing loss is correlated with severity in the chick cochlear nucleus. (A) Action potentials are initiated in a small region of membrane containing a high density of voltage-gated sodium channels, called the axon initial segment (AIS). The AIS became more extended along the axon following hearing loss, and the effect size was correlated with the severity of the manipulation. The SNHL-induced alteration of AIS length emerges over about 7 days. (B) The magnitude of AIS expansion is correlated with the severity of hearing loss. The first 3 conditions represent CHL (amount of attenuation shown in parentheses), and the final manipulation represents SNHL. (TM, tympanic membrane; MEB, middle ear bone). All animals were reared in the same acoustic environment. Adapted from Kuba *et al.* (2010).

There is evidence that hearing-loss-induced changes to cellular properties are correlated with the severity of deprivation. In the chick cochlear nucleus, developmental hearing loss causes a redistribution of sodium channels on the axon initial segment (AIS), a region of membrane responsible for action potential initiation. This effect begins to emerge after 1 day of hearing loss, and requires about 7 days to reach an asymptotic level (Fig. 3A). Furthermore, the magnitude of this change is correlated with the type of experimentally-induced hearing loss (Kuba *et al.*, 2010). As shown in Fig. 3B, there is only a modest change in AIS length in response to a mild CHL (puncture of the tympanic membrane) which induces approximately 20 dB of attenuation, as measured with auditory brainstem response (ABR). However, moderate CHL (middle-ear bone immobilization or removal; ~50 dB of attenuation) leads to a highly significant increase in AIS length. Finally, SNHL (cochlea removal) induces the largest effect. These results indicate that CHL can elicit significant changes to the CNS cellular function, but that the effects due to SNHL are quantitatively larger.

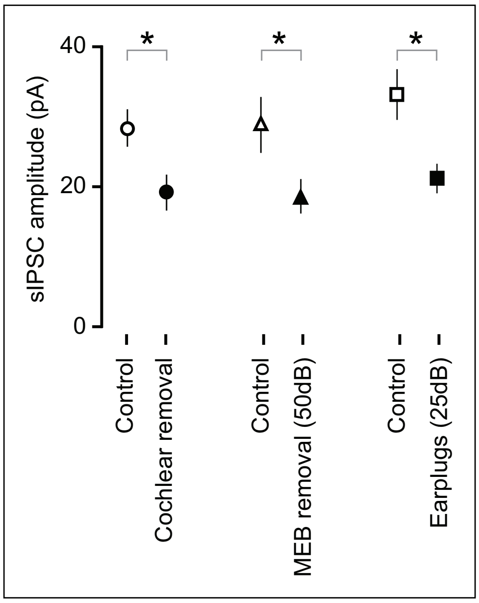


Fig. 4: A comparison of the impact of 3 forms of developmental hearing loss on auditory cortex inhibition. Bilateral cochlear removal, middle ear bone removal, or earplug insertion were induced at postnatal day 10-11, and spontaneous IPSCs were subsequently recorded from auditory cortex pyramidal neurons. IPSC amplitude declined to an equivalent degree in each form of hearing loss, as compared to age-matched control recordings (MEB, middle ear bone). Asterisks indicate a statistically significant change. Adapted from Kotak *et al.* (2008), Takesian *et al.* (2012), and Mowery *et al.* (2013).

Although no single study provides a similar comparison of hearing loss severity as it relates to synaptic function, we have performed identical measures of cortical inhibitory currents following 3 different forms of developmental deprivation (Kotak *et al.*, 2008; Takesian *et al.*, 2012; Mowery *et al.*, 2013). Figure 4 shows the mean amplitude of spontaneous IPSCs recorded in gerbil auditory cortex from control animals, in comparison to animals reared with SNHL (i.e., bilateral cochlea removal), moderate CHL (i.e, bilateral malleus removal), or mild CHL (i.e., bilateral earplugs). Each form of hearing loss induces a nearly identical decrease in IPSC amplitude. Furthermore, when CHL is induced in adult animals, it does not lead to a decrease in IPSC amplitude. The relative impact of developmental moderate CHL or SNHL on transmitter release has also been examined for both excitatory and inhibitory synapses in auditory cortex (Xu *et al.*, 2007; Takesian *et al.*, 2010). Again, there was little difference between bilateral CHL versus bilateral SNHL. Following either manipulation there was significantly greater synaptic depression in response to multiple stimuli. For excitatory synapses, the SNHL elicited effect was slightly larger than that observed with CHL (Xu *et al.*, 2007). Therefore, cortical synaptic function can be as sensitive to sound attenuation as it is to complete deafferentation.

Although very different manipulations can result in similar outcome measures, there are cases in which the SNHL elicits significantly larger effects. For example, SNHL leads to a greater reduction in spike frequency adaptation in response to trains of injected current pulses, as compared to CHL (Xu *et al.*, 2007). Finally, hearing loss can disrupt long-term synaptic plasticity (Kotak *et al.*, 2007a), a neuronal mechanism thought to be involved in learning. For example, inhibitory synapses in auditory cortex display long-term potentiation following trains of afferent stimulation, and this synaptic plasticity is diminished by hearing loss (Xu *et al.*, 2010). Furthermore, the reduction of plasticity is greater for SNHL than it is for CHL (Fig. 5). These cortical studies suggest that deafferentation can have a greater influence on the development of cellular properties, as compared to manipulations that result in sound attenuation.

SHORT- VERSUS LONG-TERM HEARING LOSS

Although developmental hearing loss has been shown to influence many neural properties, most of these results are obtained following a short survival time. These data demonstrate that the effects can appear within hours to days, but they do not address whether the changes are permanent. Our studies of the impact of hearing loss on synaptic inhibition in auditory cortex suggest that cellular deficits can persist into adulthood. Figure 6 plots the mean amplitudes of spontaneous IPSC recorded from auditory cortex neurons following developmental CHL, as a function of survival time. CHL causes a significant reduction in IPSC amplitude, and this effect is present at both short and long survival times (Takesian *et al.*, 2012). The long duration inhibitory currents that are observed after SNHL resemble IPSCs that are recorded in neurons from pre-hearing animals, suggesting that normal acoustic experience is essential for maturational progress of GABA_A receptor subunit function to occur (Kotak *et al.*, 2008). Specifically, the agonists of GABA_A receptor subunits α1 and β2/3 did not produce effects on IPSC kinetics, and this lack of an effect resembled that observed in neurons from pre-hearing animals.

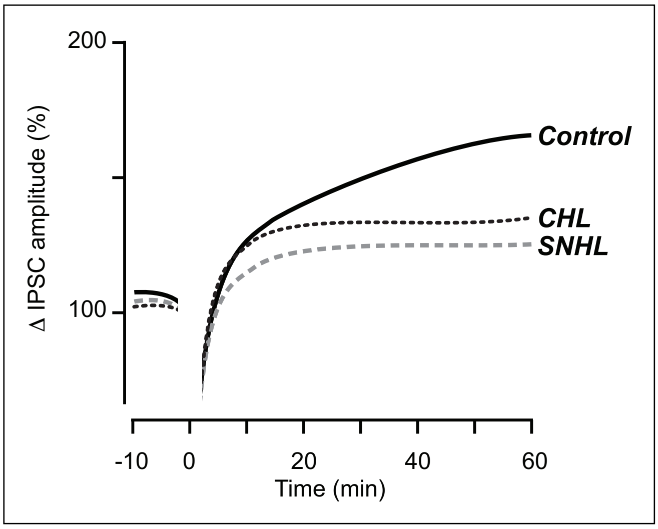


Fig. 5: The impact of hearing loss on inhibitory synaptic long-term potentiation recorded in the gerbil auditory cortex. Evoked IPSCs were recorded from pyramidal neurons for 10 minutes, followed by a series of stimulus trains that were designed to emulate the temporal discharge pattern of auditory cortex neurons in vivo. Following this treatment, the amplitude of evoked IPSCs was potentiated by 155%, as compared to the pre-treatment value (Control). The magnitude of this potentiation was much smaller for animals with developmental hearing loss. However, SNHL resulted in a greater effect, as compared to CHL. Adapted from Xu *et al.* (2010).

One mechanistic explanation for this effect is that adult GABA$_A$ receptor subunits are not properly trafficked to the synaptic membrane (Sarro *et al.*, 2008). Since CHL also causes IPSC decay times to remain longer, both at short and long survival times, the functional expression of GABA$_A$ receptors may remain compromised for the duration of hearing loss. In this regard, it is interesting to note that inhibitory maturation can be induced with pharmacological manipulations that boost GABAergic transmission, such that normal IPSC amplitudes and kinetics are observed in animals that remain deafened (Kotak *et al.*, 2013).

Although impairment of cellular function can persist during ongoing hearing loss, many cellular properties are normal after long survival times. Recordings obtained from adult cortical inhibitory interneurons following developmental CHL indicate that passive membrane properties are similar to those displayed by age-matched controls (Takesian *et al.*, 2012). A similar outcome has been observed following bilateral hearing loss in developing rats. Membrane excitability is altered at short-term survival intervals, but neurons no longer differ from controls at 1 month postnatal (Rao *et al.*, 2010). Interestingly, serotonin suppresses pyramidal cell discharge, but only at longer survival times. These findings suggest that perceptual deficits that are observed in adulthood are likely due to only a subset of the cellular alterations that have been described following a short survival time.

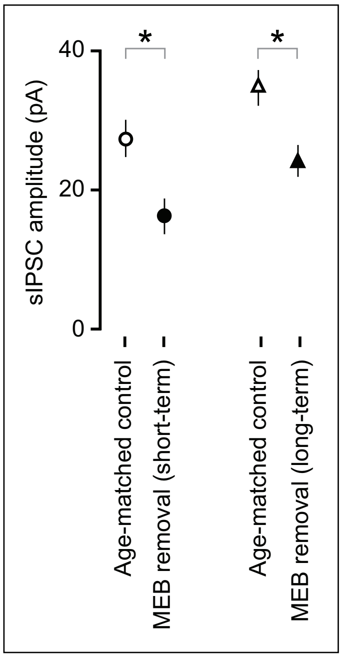


Fig. 6: Following developmental CHL at postnatal day 10, inhibitory synaptic currents remain depressed through adulthood. Spontaneous IPSCs were recorded in auditory cortex pyramidal neurons after 7-12 days of hearing loss (short-term), or 80-100 days of hearing loss (long-term). At both survival times, CHL resulted in smaller IPSCs, as compare to age-matched controls (MEB, middle ear bone). Asterisks indicate a statistically significant change. Adapted from Takesian *et al.* (2012).

SUMMARY

At the qualitative level, it is clear that the consequences of hearing loss on synaptic properties are similar for animals with experimentally induced moderate CHL or SNHL. However, there is not yet sufficient information on synaptic function following mild forms of developmental hearing loss to determine whether its consequences are comparable. Since mild unilateral hearing loss does induce significant changes to CNS coding properties, the likelihood is that mild hearing loss does induce substantive cellular changes. Certainly, our preliminary findings (bilateral earplugs, Fig. 4) are in accord with this conclusion. At the quantitative level, the effects of hearing loss are likely to be of larger magnitude when there is a loss of hair cells and/or spiral ganglion neurons. This is apparent in some (e.g., Fig. 5), but not all, of our measures from auditory cortex. Taken together, these findings suggest that profound changes to central processing may occur even when developmental hearing loss is moderate (CHL) and raises the question whether transient forms of CHL are equally detrimental to the cellular maturation of central auditory circuits.

REFERENCES

Balakrishnan, V., Becker, M., Lohrke, S., Nothwang, H.G., Guresir, E., and Friauf, E. (**2003**). "Expression and function of chloride transporters during development of inhibitory neurotransmission in the auditory brainstem," J. Neurosci., **23**, 4134-4145.

Brenowitz, S., and Trussell, L.O. (**2001**). "Maturation of synaptic transmission at end-bulb synapses of the cochlear nucleus," J. Neurosci., **21**, 9487-9498.

Chirila, F.V., Rowland, K.C., Thompson, J.M., and Spirou, G.A. (**2007**). "Development of gerbil medial superior olive: integration of temporally delayed excitation and inhibition at physiological temperature," J. Physiol., **584**, 167-190.

Chuhma, N., and Ohmori, H. (**1998**). "Postnatal development of phase-locked high-fidelity synaptic transmission in the medial nucleus of the trapezoid body of the rat," J. Neurosci., **18**, 512-520.

Dorrn, A.L., Yuan, K., Barker, A.J., Schreiner, C.E., and Froemke, R.C. (**2010**). "Developmental sensory experience balances cortical excitation and inhibition," Nature, **465**, 932-936.

Fitzgerald, K.K., and Sanes, D.H. (**2001**). "The development of stimulus coding in the auditory system," in *Physiology of the Ear, 2nd Edition*. Edited by E. Jahn and J. Santos-Sacchi (Singular Publishing, San Diego), pp. 215-240.

Gao, H., and Lu, Y. (**2008**). "Early development of intrinsic and synaptic properties of chicken nucleus laminaris neurons," Neurosci., **153**, 131-143.

Johnson, S.L., Eckrich, T., Kuhn, S., Zampini, V., Franz, C., Ranatunga, K.M., Roberts, T.P., Masetto, S., Knipper, M., Kros, C.J., and Marcotti, W. (**2011**). "Position-dependent patterning of spontaneous action potentials in immature cochlear inner hair cells," Nature Neurosci., **14**, 711-717.

Jones, T.A., Leake, P.A., Snyder, R.L., Stakhovskaya, O., and Bonham, B. (**2007**). "Spontaneous discharge patterns in cochlear spiral ganglion cells before the onset of hearing in cats," J. Neurophys., **98**, 1898-1908.

Kandler, K., and Friauf, E. (**1995**). "Development of glycinergic and glutamatergic synaptic transmission in the auditory brainstem of perinatal rats," J. Neurosci., **15**, 6890-6904.

Kotak, V.C., and Sanes, D.H. (**1997**). "Deafferentation of glutamatergic afferents weakens synaptic strength in the developing auditory system," Eur. J. Neurosci. **9**, 2340-2347.

Kotak, V.C., Fujisawa, S., Leem F.A., Karthikeyan, O., Aoki, C., and Sanes, D.H. (**2005**). "Hearing loss raises excitability in the auditory cortex," J. Neurosci., **25**, 3908-3918.

Kotak, V.C., Breithaupt, A.D., and Sanes, D.H. (**2007a**). "Developmental hearing loss eliminates long-term potentiation in the auditory cortex," Proc. Natl. Acad. Sci. USA, **104**, 3550-3555.

Kotak, V.C., Sadahiro, M., and Fall, C.P. (**2007b**) “Developmental expression of endogenous oscillations and waves in the auditory cortex involves calcium, gap junctions, and GABA,” Neurosci., **146**, 1629-1639.

Kotak, V.C., Takesian, A.E., and Sanes, D.H. (**2008**). “Hearing loss prevents the maturation of GABAergic transmission in the auditory cortex,” Cerebral Cortex, **18**, 2098-2108.

Kotak, V.C., Péndolam L.M., and Rodríguez-Contreras, A. (**2012**). “Spontaneous activity in the developing gerbil auditory cortex in vivo involves GABAergic transmission,” Neurosci., **226**, 130-144.

Kotak, V.C., Takesian, A.E., MacKenzie, P.C., and Sanes, D.H. (**2013**). “Rescue of inhibitory synapse function following developmental hearing loss,” PLoS One, **8**, e53438.

Kuba, H., Oichi, Y., and Ohmori, H. (**2010**). “Presynaptic activity regulates Na(+) channel distribution at the axon initial segment,” Nature, **465**, 1075-1078.

Magnusson, A.K., Kapfer, C., Grothe, B., and Koch, U. (**2005**). “Maturation of glycinergic inhibition in the gerbil medial superior olive after hearing onset,” J. Physiol., **568**, 497-512.

Mowery, T.M., Kotak, V.K., and Sanes, D.H. (**2013**), “Critical period for auditory cortex inhibitory maturation closes by postnatal day 19,” Soc. Neurosci. Abs. **43**.

Nakamura, Y., and Takahashi, T. (**2007**). “Developmental changes in potassium currents at the rat calyx of Held presynaptic terminal,” J. Physiol., **581**, 1101-1112.

Oswald, A.M., and Reyes, A.D. (**2008**). “Maturation of intrinsic and synaptic properties of layer 2/3 pyramidal neurons in mouse auditory cortex,” J. Neurophysiol., **99**, 2998-3008.

Oswald, A.M., and Reyes, A.D. (**2011**). “Development of inhibitory timescales in auditory cortex,” Cereb. Cortex, **21**, 1351-1361.

Pinto, J.G., Hornby, K.R., Jones, D.G., and Murphy, K.M. (**2010**). “Developmental changes in GABAergic mechanisms in human visual cortex across the lifespan,” Front. Cell. Neurosci., **4**, 16.

Rao, D., Basura, G.J., Roche, J., Daniels, S., Mancilla, J.G., and Manis, P.B. (**2010**). “Hearing loss alters serotonergic modulation of intrinsic excitability in auditory cortex,” J. Neurophys., **104**, 2693-2703.

Sanchez, J.T., Wang, Y., Rubel, E.W., and Barria, A. (**2010**). “Development of glutamatergic synaptic transmission in binaural auditory neurons,” J. Neurophysiol., **104**, 1774-1789.

Sanes, D.H. (**1993**). “The development of synaptic function and integration in the central auditory system,” J. Neurosci., **13**, 2627-2637.

Sanes, D.H., and Walsh, E.J. (**1997**). “Development of Auditory Processing,” in *Development of the Auditory System*. Edited by E.W Rubel, A.N. Popper, and R.R. Fay (Springer-Verlag, New York), pp. 271-314.

Sanes, D.H., and Bao, S. (**2009**). “Tuning up the developing auditory CNS,” Curr. Opin. Neurobiol., **19**, 188-199.

Sanes, D.H., and Woolley, S.M.N. (**2011**). “A behavioral framework to guide research on central auditory development and plasticity,” Neuron, **72**, 912-929.

Sanes, D.H. (**2013**). “Synaptic and cellular consequences of hearing loss,” in *Springer Handbook of Auditory Research: Deafness*. Edited by A. Kral, R.R. Fay, and A.N. Popper (Springer-Verlag: New York).

Sarro, E.C., Kotak, V.C., Sanes, D.H., and Aoki, C. (**2008**), “Hearing Loss Alters the Subcellular Distribution of Presynaptic GAD and Postsynaptic GABAA Receptors in the Auditory Cortex,” Cereb. Cortex **18**, 2855-2867.

Scott, L.L., Mathews, P.J., and Golding, N.L. (**2005**). “Posthearing developmental refinement of temporal processing in principal neurons of the medial superior olive,” J. Neurosci., **25**, 7887-7895.

Takesian, A.E., Kotak, V.C., and Sanes, D.H. (**2010**). “Presynaptic GABA(B) receptors regulate experience-dependent development of inhibitory short-term plasticity,” J. Neurosci., **30**, 2716-2727.

Takesian, A.E., Kotak, V.C., and Sanes, D.H. (**2012**). “Age-dependent effect of hearing loss on cortical inhibitory synapse function,” J. Neurophys., **107**, 937-947.

Taschenberger, H., and von Gersdorff, H. (**2000**). “Fine-tuning an auditory synapse for speed and fidelity: developmental changes in presynaptic waveform, EPSC kinetics, and synaptic plasticity,” J. Neurosci., **20**, 9162-9173.

Tritsch, N.X., Yi, E., Gale, J.E., Glowatzki, E., and Bergles, D.E. (**2007**). “The origin of spontaneous activity in the developing auditory system,” Nature, **450**, 50-55.

Tritsch, N.X., Rodríguez-Contreras, A., Crins, T.T., Wang, H.C., Borst, J.G., and Bergles, D.E. (**2010**). “Calcium action potentials in hair cells pattern auditory neuron activity before hearing onset,” Nature Neurosci., **13**, 1050-1052.

Vale, C., and Sanes, D.H. (**2000**). “Afferent regulation of inhibitory synaptic transmission in the developing auditory midbrain,” J. Neurosci., **20**, 1912-1921.

Vale, C., and Sanes, D.H. (**2002**). “The effect of bilateral deafness on excitatory synaptic strength in the auditory midbrain,” Eur. J. Neurosci., **16**, 2394-2404.

Vale, C., Schoorlemmer, J., and Sanes, D.H. (**2003**). “Deafness disrupts chloride transport and inhibitory synaptic transmission,” J. Neurosci., **23**, 7516-7524.

Xu, H., Kotak, V.C., and Sanes, D.H. (**2007**). “Conductive hearing loss disrupts synaptic and spike adaptation in developing auditory cortex,” J. Neurosci., **27**, 9417-9426.

Xu, H., Kotak, V.C., and Sanes, D.H. (**2010**). “Normal hearing is required for the emergence of long-lasting inhibitory potentiation in cortex,” J. Neurosci., **30**, 331-341.

A computational model of sound recognition used to analyze the capacity and adaptability in learning vowel classes

JEFFREY SPENCER[1,2,*], NEIL MCLACHLAN[3], AND DAVID B. GRAYDEN[1,2]

[1] *Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne, Australia*

[2] *Centre for Neural Engineering, University of Melbourne, Melbourne, Australia*

[3] *Centre for Music, Mind, and Wellbeing, Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia*

Sound recognition is likely to initiate early in auditory processing and use stored representations (spectrotemporal templates) to compare against spectral information from auditory brainstem responses over time. A computational model of sound recognition is developed using neurobiologically plausible operations. The adaptability and number of templates required for the computational model to correctly recognize 10 Klatt-synthesized vowels is determined to be around 1250 templates when trained with random fundamental frequencies from the male pitch range and randomized variation of the first three formants of each vowel. To investigate the ability to adapt to noise and other unheard vowel utterances, test sets with 1000 randomly generated Klatt vowels in babble at signal-to-noise ratios (SNRs) of 20 dB, 10 dB, 5 dB, 0 dB, and -5 dB are generated. The vowel recognition rates at each SNR are 99.7%, 99.6%, 97.0%, 77.6%, and 54.0%, respectively. Also, a test set of four vowel recordings from four speakers is tested with no noise, giving 100% recognition rate. These data suggest that storage of auditory representations for speech at the spectrotemporal resolution of the auditory nerve over a typical range of spoken pitch does not require excessive memory resources or computing to implement on parallel computer systems.

INTRODUCTION

Most research on sound recognition in computational systems has been on automatic speech recognition systems. Automatic Speech Recognition (ASR) has primarily used Hidden Markov Models (HMMs) to model the statistics of the acoustic features in human speech (Rabiner, 1989). The performance of ASR systems using HMMs is significantly worse in noisy compared to clean conditions especially in non-stationary noise such as babble noise. The recognition accuracy for vowel identification of current automatic speech recognition systems at -5 dB SNR is comparable to human vowel identification scores at -15 dB SNR (Kalinli *et al.*, 2010; Mi *et al.*, 2013). This performance gap increases even further if these ASR systems are only trained with clean data instead of trained with both clean and noisy data (Kalinli *et al.*, 2010;

*Corresponding author: jeffspencerd@gmail.com

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Pearce and Hirsch, 2000).

Template-based or exemplar-based approaches to sound recognition modeled on the neurobiology and psycholinguistics of the auditory system can provide more accurate modeling of auditory signals (Deng and Strik, 2007). A limitation often stated in the past for template-based systems is the computational power and memory resources required is too excessive (De Wachter *et al.*, 2007). This limitation has become less of a problem in recent years with availability of large increases in computing power and memory storage. Furthermore, template-based systems can be implemented in parallel in near real-time systems, such as field-programmable gate arrays (FPGAs) or graphics processing units (GPUs).

A recent neurobiological model of the auditory system, the Object-Attribute Model (OAM), postulates that long-term memory modulates neural spectrotemporal receptive fields through recognition mechanisms early in cortical processing (McLachlan and Wilson, 2010). Temporal information is not available at the onset of the sound, necessitating the use of sequential slices of spectral information to compare to long-term memory templates through time to recognize sounds. Hebbian learning enables the creation and adaptation of the long-term memory templates for commonly occurring sound timbres (McLachlan and Wilson, 2010). The stages of the computational sound recognition model described here are based on the mechanisms in the OAM and use neurobiologically plausible mechanisms for spectrotemporal processing of the auditory signals fed to the model.

Research on categorical vowel perception has shown that vowels are not perceived categorically but are rather perceived along a continuum (Schouten *et al.*, 2003). Human listeners do not make unanimous decisions about vowels when perceptual vowel boundaries overlap in the F1/F2 vowel space (Peterson and Barney, 1952; Hillenbrand and Gayvert, 1993; Neel, 2008). Although vowels have overlap in the perceptual boundaries, vowels presented in isolation do have a region in the F1/F2 vowel space where they are most consistently identified (Fairbanks and Grubb, 1961). This F1/F2 vowel-formant space is the region containing the points where at least 75% of the listeners correctly identified the produced vowel. This suggests that a centroid in the F1/F2 vowel space best represents a vowel.

The model is first trained until recognition accuracy proceeds to near 100% for ten Klatt-synthesized (Klatt, 1980) vowels from the most representative vowel region of the F1/F2/F3 vowel space in the seminal work of Peterson and Barney (1952). The training determines if the number of templates and memory storage required for correct recognition is feasible on current computer systems. Furthermore, the recognition accuracy of the model with different resolutions of fundamental frequency in the templates is compared to determine if fine pitch information is required for recognition. Then, the template database is used to explore the benefit and adaptability of the recognition system with speech babble noise added to Klatt-synthesized vowels at multiple SNRs. In addition, the template database is used for recognition of a small set of recorded vowels /ae, ɜ, i, u/ from two male speakers to see if the model can adapt from synthesized to real speech.

METHOD

The model is implemented in the programming language Python using the numpy and scipy packages (Oliphant, 2007). The beginning processing stages are similar to many other models involving the auditory periphery (Slaney, 1993). The stages include a Gammatone filter bank similar to Slaney (1993), half-wave rectification as an approximation of hair-cell transduction, and the formation of specific loudness by short-duration temporal integration at each filter channel (Viemeister and Wakefield, 1991). Next, lateral inhibition is performed to sharpen the spectral resolution by off-frequency inhibition (McLachlan, 2009). Following lateral inhibition is a nonlinear dynamic saturation stage that provides loudness invariance and noise robustness. The saturation stage calculates saturation thresholds across every filter channel independently. At each filter, a Gaussian function weights the neighboring filters, and the mean of the center filter and the weighted neighboring filters is taken as the saturation threshold for that filter channel. Therefore, the saturation threshold can increase in only specific regions of the spectrum to rise above the noise level. This saturation mechanism can not only rise above white noise but also non-stationary noise such as babble noise. More specific details of the processing stages of the computational model can be found in McLachlan (2011).

The model is based on normal-hearing listener classification of the vowels /i, ɪ, ɛ, æ, ʌ, a, ɔ, ʊ, u, ɜ/. The fundamental frequency (F0) and first three formants (F1/F2/F3) of these vowels are taken from the unanimously classified male spoken vowels in Peterson and Barney (1952). The unanimously classified vowels are the vowels that are heard as the intended vowel by all 70 listeners. The minimum and maximum F0 values (93-203 Hz) are used to define the range for choosing the fundamental frequency values. The F1(x_1), F2(x_2), and F3(x_3) values for each vowel are fitted with a three-dimensional multivariate Gaussian distribution,

$$f_{\mathbf{x}}(x_1,x_2,x_3) = \frac{1}{\sqrt{(2\pi)^3|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right), \qquad \text{(Eq. 1)}$$

where Σ is the covariance matrix, $|\Sigma|$ is the determinant of Σ, and μ is the vector of means. The tolerance region of the distribution is the region in which at least p percentage of the points are enclosed. The tolerance region is defined as

$$(x-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu) \leq c = \chi_3^2(p), \qquad \text{(Eq. 2)}$$

where c is the tolerance factor and $\chi_3^2(p)$ is the percent point function for probability p of the chi-squared distribution with three degrees of freedom (Krishnamoorthy and Mathew, 2009).

The surfaces in Fig. 1 are ellipsoids defined by a constant probability containing 30% ($p = 0.3$) of the points ($c = 1.42$). Enclosing 30% of the data points is roughly one

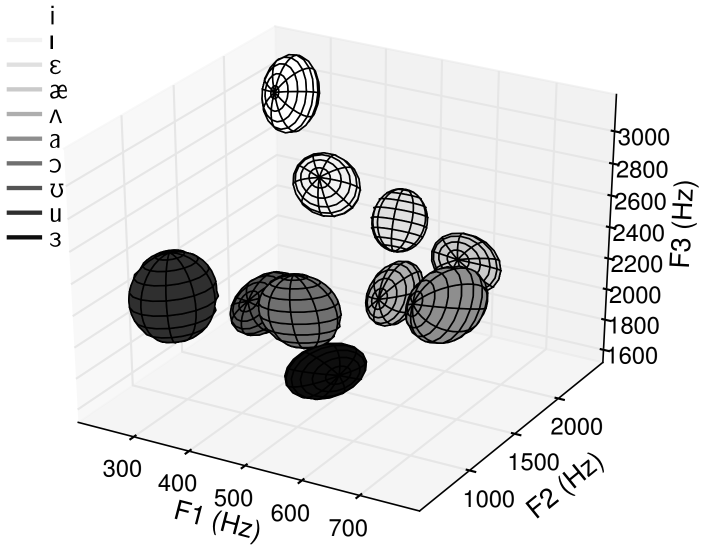


Fig. 1: Multivariate Gaussian distribution for each male spoken vowel that are unanimously classified (Peterson and Barney, 1952). The ellipsoids enclose at least 30% of the points in each distribution. This corresponds to a tolerance factor of $c = 1.42365$ and roughly one standard deviation in each direction F1/F2/F3.

standard deviation in each formant direction from the centroid of each vowel. These vowel ellipsoids are the F1/F2/F3 vowel spaces that best represent the spoken vowels from Peterson and Barney (1952) and recognition accuracy is expected to reach 100%.

Klatt-synthesized (Klatt, 1980) vowels are initially computed from the means of the formants of each vowel ellipsoid at the mean pitch from the male vowel recordings. The bandwidth for each formant frequency is determined by a fifth-order polynomial fit to measured closed-glottis bandwidths (Hawks and Miller, 1995). Separate fifth-order polynomials are fit to the data below and above 500 Hz. The other parameters fed into the Klatt-synthesizer, besides the default Klatt-synthesizer (Klatt and Klatt, 1990) values are F4 = 3400 Hz, F5 = 4000 Hz, sampling frequency = 12 kHz, and duration = 300 ms. This creates 10 Klatt-synthesized vowels at the mean values in F0, F1, F2, and F3 space.

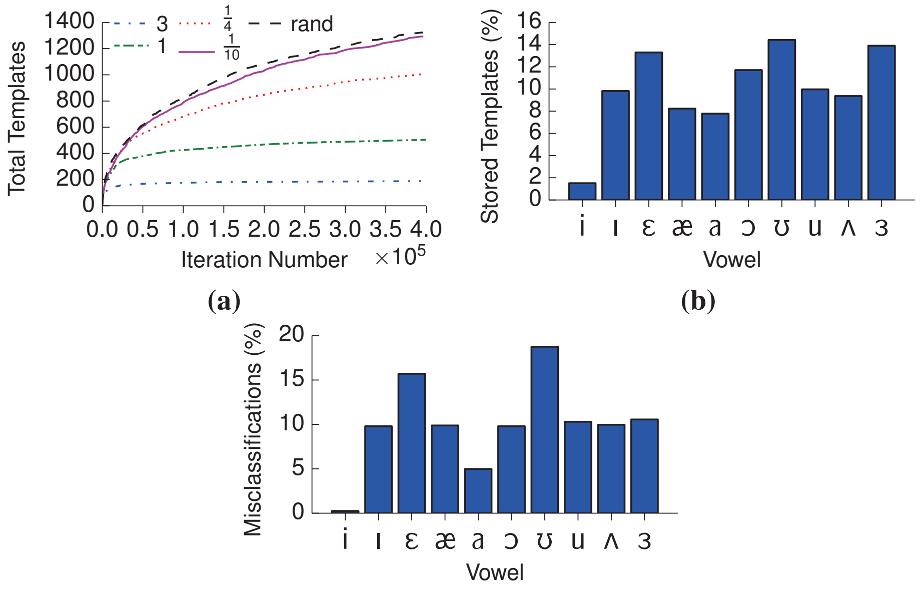
Spectral templates are computed from the 10 Klatt-synthesized vowels as the initial training set. Then, to determine the number of vowel templates required for the correct recognition of the 10 vowels, the program is iterated by choosing a random vowel, random fundamental frequency (93-203 Hz), and random formant parameters from the multivariate Gaussian distribution of the chosen vowel at each iteration. Multiple training sets are computed at different fundamental-frequency resolutions. The fundamental frequency in the training sets is chosen randomly at semitone intervals of 3, 1, 0.25, and 0.1 calculated from the lowest male F0 (93 Hz) or at random from the set of rational numbers. These parameters are used to Klatt-synthesize a new vowel, which is fed through the model. The output from the model is then compared to all vowels currently in the stored template database. The most activated template in the stored template database is checked to see if it comes from the same vowel as the computed spectral template. If the vowel does not match (a misclassification), the computed spectral template is added to the training database, and then the next iteration begins.

The procedure for selecting a random fundamental frequency from the rational numbers during the training phase is used to generate a testing set. The testing set is Klatt-synthesized vowels with no noise added (clean) and Klatt-synthesized vowels with babble noise added at SNRs of 20 dB, 10 dB, 5 dB, 0 dB, and −5 dB. For vowels with added babble, a Klatt-synthesized vowel (sig_{kl}) and a random 300-ms section of the babble noise (sig_{noi}) recording from the Aurora2 dataset (Pearce and Hirsch, 2000) are chosen at each iteration. The two signals are added together after determining the proper coefficient to multiply by the noise to get the desired SNR. The final input signal is $sig_{fin} = sig_{kl} + sig_{noi}\sqrt{\frac{p_{sig}}{p_{noi}}}10^{\frac{-SNR_{dB}}{20}}$ where the power of the 300-ms Klatt-synthesized vowel is p_{sig} and the power of the 300-ms segment of babble noise is p_{noi}. This iteration is done 100 times for each of the ten vowels for a total of 1000 inputs at each SNR. 1000 vowel inputs or 100 for each vowel are also generated in the clean condition. The Klatt-synthesized vowels in the testing set are then fed through the model and compared to the template databases built during training for each training set. Furthermore, recorded vowels produced from two native male English speakers for four of the ten vowels are also compared against the completely random template database. The recorded vowels are /ae, ɜ, i, u/.

RESULTS

The model requires 400,000 iterations to build the training database. The total number of templates compared to iteration number in each training set is shown in Fig. 2a. The number of templates stored for four hundred thousand iterations at semitone intervals of 3, 1, 0.25, 0.1, and random is 189, 505, 1009, 1294, and 1324, respectively. The recognition accuracy by the end of the four hundred thousand iterations for the training sets at 3, 1, 1/4, 1/10, and random semitone intervals is 96.8%, 98.8%, 99.5%, 99.9%, and 99.9%, respectively. The addition of finer frequency resolution adds fine pitch information that is not needed for high rates of recognition for the vowels. The recognition rate drops only 3% from selecting the fundamental frequency using 3 semitone intervals (5 frequencies total) to completely random selection of the fundamental. Furthermore, the number of templates stored drops substantially from 1324 with the completely random training set to only 189 for the training set at 3 semitone intervals. This fine pitch resolution, although not being necessary for recognition of American English vowels, is necessary for tonal languages and emotional prosody and could be stored in the templates if required. With fine pitch resolution stored in the templates, the model is also capable of detecting pitch at the accuracy of highly trained musicians (around 0.1 of a semitone) (Moore, 2003).

The percentage of the total templates (1324) stored for each vowel is shown in Fig. 2b. The most added templates are for the vowels /ʊ/ and /ɛ/, which both overlap other vowels in the first and second formant space (Peterson and Barney, 1952). These two vowels are among the worst performers in being classified in Peterson and Barney (1952) as well. The vowel with the least added templates is /i/, which also causes the least perceptual confusions with other vowels for human listeners in both quiet and noise (Peterson and Barney, 1952; Mi *et al.*, 2013). Also, /i/ is the least overlapping vowel in the formant space as seen in Fig. 1. This vowel was expected to require less templates and cause less confusions than any other vowel. Fig. 2c shows the percentage that each vowel contributes to the total number of misclassifications. This shows that the vowels that overlap the most not only require the most templates for correct recognition but also cause the most false classifications.



(a) **(b)**

(c)

Fig. 2: Vowel training performance. (a) Total number of templates stored compared to the training iteration number. This is training with fundamental frequency semitone interval resolution of 3, 1, $\frac{1}{4}$, and $\frac{1}{10}$ semitones. Also training with random fundamental frequencies. Each new template corresponds to a misclassification error in the training set. (b) The percentage of the total stored templates for each vowel. The training is using random fundamental frequencies shown in Fig. 2a. (c) The percentage of the total misclassifications for each vowel. A misclassification means that a vowel is classified instead of the intended vowel that should be classified for that input. The training is using random fundamental frequencies shown in Fig. 2a.

The Klatt-synthesized template database from the randomly selected fundamental frequencies is then tested with Klatt-synthesized vowels with babble speech added at SNRs of 20 dB, 10 dB, 5 dB, 0 dB, and -5 dB. The results using 100 test inputs for each vowel (1000 total inputs) at SNRs of 20 dB, 10 dB, 5 dB, 0 dB, and -5 dB are 99.7%, 99.6%, 97.0%, 77.6%, and 54.0%, respectively. The Klatt-synthesized template database is also tested against two male native English speakers' recordings of four vowels (/ae, ɜ, i, u/) for a total of 8 vowel recordings. The results with no noise show that all 8 of the recorded vowels are classified correctly when using the Klatt-synthesized template database. Although this is a very small sample size of recorded speech, this is a promising result.

CONCLUSION

The model trains to near 100% recognition performance on the 10 Klatt-synthesized vowels with a total of 1324 templates stored. The template database can be reduced by 86% with only a 3% loss in recognition accuracy by using sparse fundamental-frequency resolution. This reduced storage requires only 12 Mb of memory, which is well within the memory storage requirements to compute on parallel architectures such as FPGAs and GPUs. Furthermore, a 10-ms input passed through the model with 300 filter channels compared to a spectral template requires roughly 2 µs. The total time required for the recognition decision would then be dependent on the particular hardware chosen but is roughly the comparison for one spectral template (2 µs) times the number of templates divided by the number of processors. Therefore, the recognition decision does not become a very time-limiting step in the computation on a parallel architecture with sufficient cores, and the model can perform with near real-time performance. The model also performs exceptionally well when tested with Klatt-synthesized vowels with babble noise added at SNRs of 20 dB, 10 dB, 5 dB, 0 dB, and -5 dB. The vowel recognition rates at each SNR are 99.7%, 99.6%, 97.0%, 77.6%, and 54.0%, respectively. Furthermore the Klatt-synthesized vowel template database correctly recognizes recorded speech from two male speakers for the four vowels (/ae, ɜ, i, u/). The further exploration of the computational mechanisms in the model may elucidate how the brain adapts to learn language.

REFERENCES

De Wachter, M., Matton, M., Demuynck, K., Wambacq, P., Cools, R., and Van Compernolle, D. (**2007**). "Template-based continuous speech recognition", IEEE T. Audio Speech, **15**, 1377-1390.

Deng, L. and Strik, H. (**2007**). "Structure-based and template-based automatic speech recognition – Comparing parametric and non-parametric approaches", in *Interspeech 2007*, pp. 2608-2611.

Fairbanks, G., and Grubb, P. (**1961**). "A psychophysical investigation of vowel formants", J. Speech Hear. Res., **4**, 203-219.

Hawks, J.W., and Miller, J.D. (**1995**). "A formant bandwidth estimation procedure for vowel synthesis", J. Acoust. Soc. Am., **97**, 1343-1344.

Hillenbrand, J., and Gayvert, R.T. (**1993**). "Identification of steady-state vowels synthesized from the Peterson and Barney measurements", J. Acoust. Soc. Am., **94**, 668-674.

Kalinli, O., Seltzer, M.L., Droppo, J., and Acero, A. (**2010**). "Noise adaptive training for robust automatic speech recognition", IEEE T. Audio Speech, **18**, 1889-1901.

Klatt, D.H. (**1980**). "Software for a cascade/parallel formant synthesizer", J. Acoust. Soc. Am., **67**, 971-995.

Klatt, D.H., and Klatt, L.C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers", J. Acoust. Soc. Am., **87**, 820-857.

Krishnamoorthy, K., and Mathew, T. (**2009**). "The multivariate normal distribution", in *Statistical Tolerance Regions: Theory, Applications, and Computation* (John Wiley & Sons, Inc.), pp. 225-247.

McLachlan, N. (**2009**). "A vomputational model of human pitch strength and height judgments.", Hear. Res., **249**, 23-35.

McLachlan, N., and Wilson, S. (**2010**). "The central role of recognition in auditory perception: a neurobiological model", Psychol. Rev., **117**, 175-196.

McLachlan, N. (**2011**). "A neurocognitive model of recognition and pitch segregation", J. Acoust. Soc. Am., **130**, 2845-2854.

Mi, L., Tao, S., Wang, W., Dong, Q., Jin, S.-H., and Liu, C. (**2013**). "English vowel identification in long-term speech-shaped noise and multi-talker babble for English and Chinese listeners", J. Acoust. Soc. Am., **133**, EL391-EL397.

Moore, B.C.J. (**2003**). *An Introduction to the Psychology of Hearing*, 3rd Ed. (Academic Press).

Neel, A.T. (**2008**). "Vowel space characteristics and vowel identification accuracy", J. Speech Lang. Hear. Res., **51**, 574-585.

Oliphant, T.E. (**2007**). "Python for scientific computing", Comput. Sci. Eng., **9**, 10-20.

Pearce, D., and Hirsch, H.-G. (**2000**). "The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions", in *ISCA ITRW ASR-2000* (Paris, France), pp. 181-188.

Peterson, G.E., and Barney, H.L. (**1952**). "Control methods used in a study of the vowels", J. Acoust. Soc. Am., **24**, 175-184.

Rabiner, L.R. (**1989**). "A tutorial on hidden Markov models and selected applications in speech recognition", P. IEEE, **77**, 257-286.

Schouten, B., Gerrits, E., and Van Hessen, A. (**2003**). "The end of categorical perception as we know it", Speech Commun., **41**, 71-80.

Slaney, M. (**1993**). "An efficient implementation of the Patterson-Holdsworth auditory filter bank", Apple Computer Technical Report #35.

Viemeister, N.F., and Wakefield, G.H. (**1991**). "Temporal integration and multiple looks", J. Acoust. Soc. Am., **90**, 858-865.

Is hearing-aid signal processing ready for machine learning?

BERT DE VRIES[1,2,*] AND ANDREW DITTBERNER[3]

[1] *GN ReSound, Het Eeuwsel 6, 5612 AS, Eindhoven, Netherlands*

[2] *Eindhoven University of Technology, Den Dolech 2, Eindhoven, Netherlands*

[3] *GN ReSound, 2601 Patriot Blvd., Glenview, IL 60026, USA*

In the hearing-aids community, machine-learning technology enjoys a reputation as a potential performance booster for signal-processing issues such as environmental steering, personalization, algorithm optimization, and speech detection. In particular in the area of in situ hearing aid personalization, the promise is steep but clear success stories are still hard to come by. In this contribution, we analyze the 'personalizability' of typical hearing-aid signal-processing circuits. We discuss a few salient properties of a very successful adaptable and personalized signal-processing system, namely the brain, and we discover that among some other issues, the lack of a probabilistic framework for hearing-aid algorithms hinders interaction with machine-learning techniques. Finally, the discussion leads to a set of challenges for the hearing-aid research community in the quest towards in situ personalizable hearing aids.

INTRODUCTION

In this paper, we distinguish three groups of hearing-aid (HA) algorithm designers. By a designer we mean any entity that is capable to affect the input-output behavior of an HA algorithm. The first designer group entails the *professionals*: engineers, scientists, and dispensing audiologists. The professionals deal with *ex situ* design. Roughly speaking, engineers and scientists define the algorithm *structure* (i.e., the equations), whereas audiologists set the HA algorithm *parameters* during a fitting session. After a patient has been fitted and he walks away with an operational hearing aid, there still remain two entities that are capable of changing the HA algorithm under *in situ* conditions. The second designer group is the *patient* himself who can update an HA algorithm through (machine-learning-based processing of) preference feedback. For instance, patient feedback, collected through a volume-control wheel, could be used to change some gain parameters of the hearing aid. Finally, the *acoustic environment* could in principle be recruited to change parameters or structure of the HA algorithm. With a sample rate of 16 kHz and a 16-bit code per sample, about one million bits of acoustic data get recorded every four seconds by the hearing aid. One could imagine that machine-learning methods take advantage of these in situ acquired acoustic data streams, e.g., to train an environmental classifier.

In general, the field of *machine learning* refers to methods that aim to improve the

*Corresponding author: bdevries@gnresound.com

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

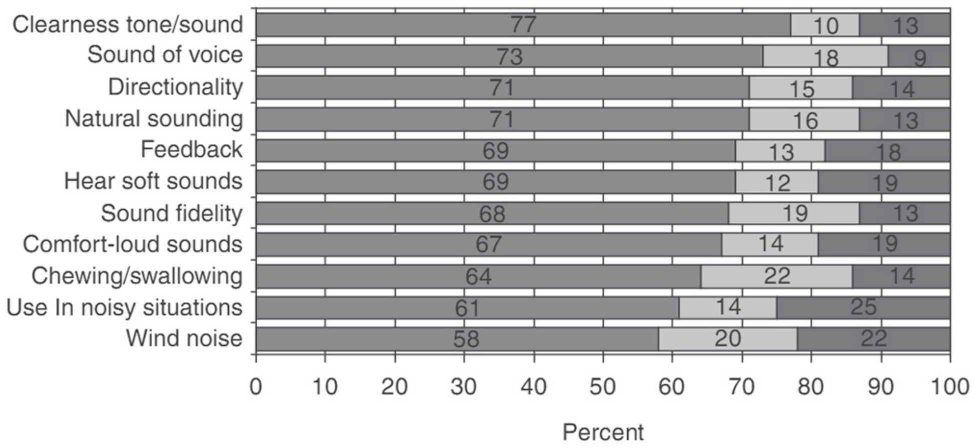


Fig. 1: Percentages of patient satisfaction with sound processing in hearing aids. Figure from Kochkin *et al.* (2010).

performance of a device through (learning from) experience with that device. In the hearing-aid context, in situ updates of the algorithm by patient preference feedback or by the acoustic environment could be considered machine-learning-based design. Today, HA design is almost exclusively the domain of the professionals. Why has in situ machine-learning-based design not yet claimed a substantial role in the HA design process? In this paper we will discuss some fundamental signal-processing issues that hinder application of machine-learning-based design to HA algorithms.

WHY IN SITU MACHINE LEARNING?

How satisfied are users of hearing aids? The graph in Fig. 1 is from a large study in 2010 on the hearing-aids market by Kochkin *et al.* (2010). The horizontal bars represent patient satisfaction rates with various aspects of sound processing in hearing aids. The bars on the left reflects the percentage of people that are happy, dark grey (right) indicates dissatisfaction, and light grey (middle) relates to a neutral opinion. Let's keep this simple: about 20% of hearing-aid patients are not happy with the sound processing performance of their devices.

This is a remarkable number because over the past decade, we, the engineers and scientists in the hearing-aids industry and in academic environments, have collectively spent a few thousand man-years on improving the sound processing in hearing aids. Apparently, despite a very extensive collective engineering effort, one out of five patients remains not satisfied. The performance of sound processing in hearing aids seems to have plateaued.

There is a plausible explanation for this observation. When an engineer designs a hearing aid, he does not know yet who the patient will be, he doesn't know the hearing-loss portrait of that patient, nor does he know in which acoustic environments the patient will spend his time. To complicate matters, this type of knowledge changes

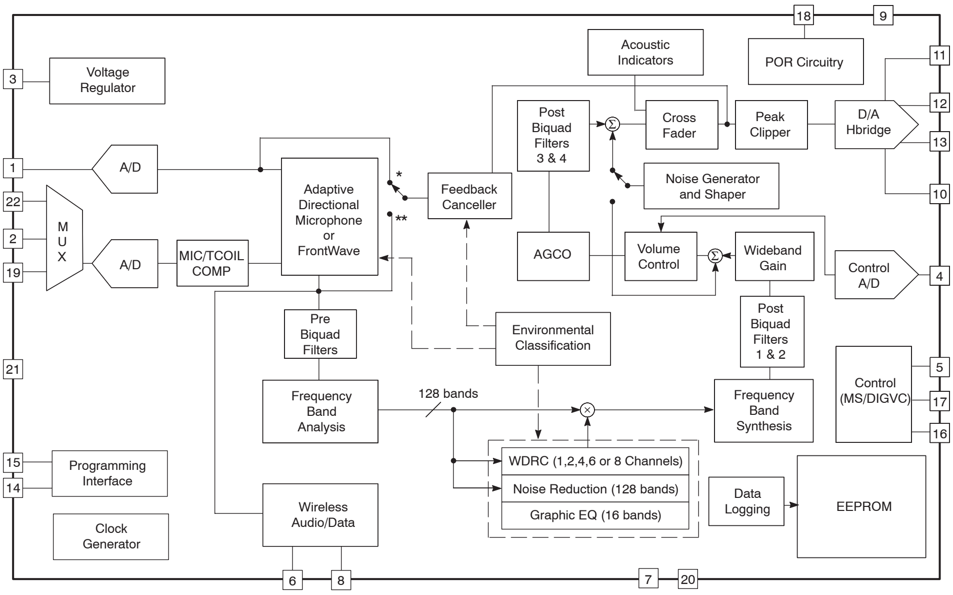


Fig. 2: Block diagram of the AYRE-SA3291 hearing-aid algorithm, ON-Semiconductor (2013).

over time. Every time a patient puts in his hearing aid, the physical placement of the device will be a bit different from last time, leading to an altered acoustical situation in and around the ear. In other words, when the engineer designs the sound-processing properties of a hearing aid, he has to deal with many unknowns about the actual circumstances where the hearing aid will be used. It won't help to ask the engineer to work harder or do his extra very best this time, since these future in situ conditions are simply unknown. Instead, we must provide the patient will tools to solve problems right there when they occur on the spot.

In order to get an idea of what we are up against, have a look at Fig. 2, which is a block diagram of a commercial hearing-aid algorithm by ON-Semiconductor (2013). We use this particular block diagram because it is publicly available but the discussion applies generally to the signal-processing algorithms of commercially available hearing aids. Most blocks in this graph hide sub-algorithms that are at least as complex as this top-level diagram. Now suppose that you are at a cocktail party and you can't understand your conversation partner. You would like to make a small change to this circuit and test a few variants, but how? If you pull a wire, this circuit will likely crash and no output get generated. Which wire should you pull anyway? Or should you add a wire somewhere? In order words, how do you bring about variation as a means for experimentation in this system? Even if you succeed in improving the quality for your current situation, will that change still be an improvement later, after the party is over? In practice, the way to update systems like this one is to give it to an expert signal-processing engineer and let him tinker with it; then take it back after a few months and hope that it works better. But that's not what we are interested in here. If this system doesn't work to your full satisfaction in the field, you might be willing to invest maximally one minute to make it sound better. And it should sound better because if it doesn't, you will be less willing to spend that minute the next time. This signal processing circuit is not fit for that purpose.

Based on the foregoing discussion, let us state an important challenge for the HA industry. How can we build tools that facilitate *fast and easy (machine-learning-based) re-design of hearing-aid algorithms driven by end users and the environment*. There are three very challenging aspects about this goal. Whereas a normal design update by a signal-processing expert in his laboratory environment may take a few months, in this challenge the aim is to execute an incremental design update (1) by a (non-expert) user, (2) under normal operational conditions, and (3) within a minute. In search for answers, we are inspired by research from others on how the brain processes information. In this paper we will discuss some aspects of computation in brains that in our opinion should influence the future engineering practice of HA algorithm design. However, before we turn to the brain, let us discuss an important engineering lesson for the design of systems with large uncertainties.

DESIGN FOR REDESIGN – THE FLIGHT OF THE GOSSAMER CONDOR

In 1959, the British industrialist Henry Kremer announced a prize of £50,000 (in today's money worth about 1 million euros) for the first successful human-powered flight around a figure-eight course with the two turning points placed half a mile apart. A second prize of £100,000 was created for the first human-powered flight across the English channel.

Many years and 50 failed attempts passed. In 1977, the British aviation engineer Paul MacCready took on the challenge and noticed a common pattern when he studied the records of past attempts. Previous engineering teams had often invested more than a

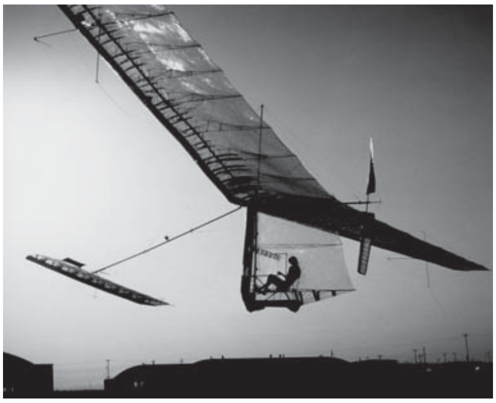


Fig. 3: The Gossamer Albatross, the first human-powered airplane that crossed the English channel. Figure from Raskin (2011).

year to carefully design a prototype plane based on elaborate theories and conjecture. Then, a few seconds after take-off of the maiden flight, a year's work would crash on the ground and obliterate the massive effort.

MacCready came to a crucial insight. The past efforts were focused on solving the wrong problem. The essential problem was not how to design a human-powered airplane. Instead, the essential problem was that *they did not understand the problem* (Raskin, 2011). Rather than attempting to design an optimal aircraft, MacCready reformulated the problem as the quest to design an airplane that could be re-built in hours, not months. His team started building planes from cheap and light aluminum tubing, mylar, wires, and scotch tape. In MacCready's approach, design was to be interpreted as an experiment to learn more about the problem. The first flight failed right away. But the team learned from the crash and delivered a second prototype just a few hours later. This process of fast iterative redesign continued for about half a year until 23 August 1977, when Bryan Allen of MacCready's team pedaled the Gossamer Condor for the 223rd time and cleared the finish line 7 minutes and 27 seconds after take-off. Two years later, Allen flew a further evolution of the Gossamer (the 'Albatross') across the English channel to claim the second Kremer prize, cf. Fig. 3.

Where other teams had failed for more than 17 years, MacCready's fast-iterations approach turned out to be the key to solving poorly-understood engineering problems. While this story has on the surface little to do with hearing-aid design, the underlying challenge to cope with a poorly-understood problem is the same for both tasks. This story illuminates the *engineering need* to focus on fast redesign of hearing-aid sound-processing algorithms, instead of a research focus on the optimal algorithm per se.

INFORMATION PROCESSING AND THE BRAIN

Engineers study the brain for its usability to design artificial systems. Since the brain is our most crucial instrument in our drive to survive, it must work today and yet be fully prepared to adapt to unforeseen new circumstances. In the next sections we will discuss a few salient properties that enable the brain to execute fast redesign iterations so as to cope with a world where the problems keep changing in unpredictable ways.

Probability theory

If the brain is a system that processes information then there must be some computing rules that the brain adheres to. There is strong scientific support for the claim that brains compute with the rules of probability theory (e.g., Friston, 2009). This is the same probability theory that we all got to love and hate in high school.

We can use probability theory to predict the future, based on observations from the past. For instance, if we observe 100 coin tosses and 96 out of 100 throws came up tails, then we predict that the 101st observation will come up tails with higher probability than for heads. Intuitively this happens by extrapolating past observations.

Technically, in order to predict the future we need to build a *model* to summarize regularities that were present in past observations and use that model to predict the future. We humans need to have some capacity to predict the future, because we want to avoid to be surprised by the physical world around us. For instance, we must be able to make predictions on what's edible or hostile to us. More generally, any large surprise in the physical world could possibly kill us. So, a key task of the brain is to build a model for the world in which we live and use that model to make predictions about that world.

Probability theory can be used to make optimal predictions about future (data) observations by

$$\underbrace{\mathrm{Pr}(\,\text{future}\,|\,\text{data}\,)}_{\text{data-based prediction of future}} = \sum_{\text{all models}} \underbrace{\mathrm{Pr}(\,\text{future}\,|\,\text{model}\,)}_{\text{model-based prediction of future}} \times \underbrace{\mathrm{Pr}(\,\text{model}\,|\,\text{data}\,)}_{\text{model based on past observations}} \qquad \text{(Eq. 1)}$$

The expression Pr(.) here is mathematical notation for a probability mass function, but we will not bother with explaining the details of the formula, other than to point out that something as complex as predicting the future can be captured by a single-line equation. The left-hand side states that we want to predict the future from past data. The data refer to observations from the outside world that enter the brain through sensory organs like the eyes or ears. The right-hand side states how predictions of data relate to a model and past observations. The model can be implemented by a brain or by a computer program. The right-most factor, $\mathrm{Pr}(\,\text{model}\,|\,\text{data}\,)$, captures what the model has learned from past data. By another rather simple manipulation with probability theory we can express how models learn from data:

$$\underbrace{\mathrm{Pr}(\,\text{model}\,|\,\text{data}\,)}_{\text{model after learning}} = \frac{\overbrace{\mathrm{Pr}(\,\text{data}\,|\,\text{model}\,)}^{\text{model based predictions}} \times \overbrace{\mathrm{Pr}(\,\text{model}\,)}^{\text{model before learning}}}{\underbrace{\mathrm{Pr}(\,\text{data}\,)}_{\text{evidence}}} \qquad \text{(Eq. 2)}$$

In probability theory this equation is known as *Bayes rule*. Bayes rule describes how we learn about the world. It doesn't matter if the observations relate to music, video, or even financial stock rates: Bayes rule applies and tells us how to optimally update our knowledge about a phenomenon based on new observations about that phenomenon. Bayes rule is basically a prediction-correction method. The model gets updated on the basis of differences between actual and (synthesized) predicted observations.

If a human brain were capable of executing Bayes rule, then our concept of what a tree looks like would get updated every time when we see a tree. The more trees we see, the better we understand what a tree looks like. It seems that it would be very useful for a brain to be able to process sensory information by Bayes rule, because it would enable us to learn a model about the world just by looking at the world. Apparently, using the same rules from probability theory we can then use that model to make predictions about the world, which are so crucial for us to stay alive.

It can be shown that, under some very agreeable assumptions, Bayes rule prescribes the *optimal* method for learning from observations (Jaynes, 2003). So there is no need to look for a specialized learning algorithm that works particularly well for any specific problem. The simplicity of Bayes rule is a strength. Whether we have to learn a language or learn about how to repair a bicycle, Bayes rule is how we *should* learn. If the brain computes with probability theory then there is no need to invent new prediction or detection methods when the outside world changes. It doesn't matter if the observed signals are of acoustic or visual nature, the difficulty lies mostly in how to *implement* Bayes rule, both in brains and computers.

The probabilities that we discussed relate data to models and back. Models and data are very much the core issues for engineered signal-processing systems. Next we take a look at how the brain deals with models from the perspective of adaptability.

Models and structures

Signal-processing algorithms can be intuitively visualized by block diagrams like in Fig. 2. A block diagram consists of a set of blocks (nodes) and links (edges) that connect the blocks. With each link we associate a variable in the system. In a block, mathematical relations between the connected variables are described. Often, we may find another block diagram in a block, so blocks can be used to hide details of the algorithm. The algorithm *structure* refers to the mathematical relations between the variables that are described by a block diagram. We also like to distinguish between variables whose values change as time moves on (the *state* variables) and those (the *parameters*) whose values are expected to stay fixed or change much slower than the rate of change of the states. In neural terms, the structure relates to the neuronal network of the brain, the parameters are represented by the strength of synaptic connections between neurons, and the state relates to the electric fields in the brain. In particular, our perception of the world is represented by the state variables. The model structure and parameter values provide constraints on how the states (read: our perception) will change over time. If our perceptions and prediction of future perceptions are accurate enough, we can stay alive.

Unfortunately, unexpected things will happen and we will need to change the algorithm structure and parameter values so as to keep our model of the world sufficiently accurate.

It is clear that if we change a structure at one location, we do not want that change to have serious consequences on variables in another location of the network. If the network were now to be adapted at the latter location, this could have effects elsewhere again and thus lead to a *snowball* effect of unpredictable changes, likely followed by a crash of the algorithm. Therefore, *modularity* is an essential characteristic of complex yet adaptable networks. A modular network is composed of sub-networks called modules with more dependencies within the modules than between the modules. The relative independence of modules prevents the snowball effect of changes to escalate.

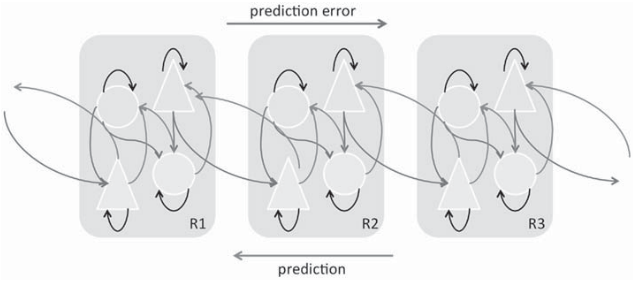


Fig. 4: An example flow graph of hierarchical modularity across three cortical regions. Figure from Friston (2009).

On the other hand, some communication between modules is necessary to generate behavior that transcends the functional complexity of individual modules. In order to avoid the snowball effect, modules should preferably depend on other modules that are *more stable* than themselves. Let's assume the opposite, namely that module A depends on module B and the natural rate of change for B is faster than for A. In that case, A will have to adapt each time that B changes, which is more often than A's natural rate of change. The idea that the snowball-of-changes effect can be avoided by constraining intermodule communication to flow from more to less stable structures leads to *hierarchical* networks.

Technically, probability theory supports hierarchical modularity almost effortlessly. Bayes rule decomposes into a hierarchy of four modules by

$$\begin{aligned}
\Pr(\text{ model}\,|\,\text{data}) &\propto \Pr(\text{data}\,|\text{ model}) \times \Pr(\text{ model}) = && \text{(Eq. 3)}\\
&\Pr(\text{data}\,|\,\text{states}, \text{parameters}, \text{structure}) && \text{(now)}\\
&\times \Pr(\text{states}\,|\,\text{parameters}, \text{structure}) && \text{(short-term memory)}\\
&\times \Pr(\text{parameters}\,|\,\text{structure}) && \text{(mid-term)}\\
&\times \Pr(\text{structure}) && \text{(long-term)}
\end{aligned}$$

In the final result of the computation, the left-hand side $\Pr(\text{ model}\,|\,\text{data})$, the model depends directly on fast fluctuations in the observed data. Straight implementation leads to an undesired network structure. However, after the hierarchical decomposition, at each level, variables only depend on other variables that are more stable than themselves. We now have an answer to our question on *how* to implement Bayes rule. Through hierarchical modularity the snowball effect of changes is avoided. This property is crucial when in situ structural algorithm changes are demanded.

We can think of many reasons why modularity is the most prominent feature of adaptable systems. But how would the brain know that? Is there an evolutionary drive for brains to develop modular structures? If we accept that the brain is mostly an engine for probabilistic reasoning, then it would help if probability theory would prefer modular over densely coupled structures (all else being equal). This is indeed the case. The factor Pr(data) in Bayes rule, known as the *evidence*, can be used to evaluate how well a model summarizes a set of observations. It can be mathematically shown that the (logarithm of the) evidence decomposes into a sum of two terms, namely *accuracy* plus *model simplicity*:

$$\begin{aligned}\log(\,\text{evidence}\,) &= \text{accuracy} + \text{simplicity} \qquad \text{(Eq. 4)}\\ &\approx \text{‘works today’} + \text{‘works tomorrow’}\end{aligned}$$

The first term, accuracy, measures how well the model predicts past observations. If we need to predict future observations, it makes sense that we prefer models that performed best on past data, so we want models with high accuracy. A system that scores high at accuracy works well today. However, the second term, the simplicity term, favors models that are simple and adaptable. Indeed, it can be shown that modular structures score higher simplicity values than densely coupled systems. As we discussed, modular systems are more adaptable than coupled systems. Therefore, probability theory prefers structures that balance excellent performance today (high accuracy) against adaptability for tomorrow (high simplicity). We also conclude that if brains would follow probability theory, then it's no surprise that brains are both excellent performers today and yet remain very adaptable. After all, both properties are highly prioritized by straight probability theory. The brain has no choice but to optimize both for today *and* an unknown future.

Driven by data

We discussed how probabilities and models relate to information processing in the brain. The third term in the equations for learning and prediction is called the data or observations.

Data are observed through sight, hearing, taste, smell, and touch, collectively known as our senses. Observations inform us about the current state of the world around us. We use probability theory to summarize observations in models and use these models to predict how the world evolves.

One of the most interesting aspects of our brain is how much we seem to learn from just a few teaching events. After a mother has showed her two-year old daughter a few times what a tree looks like, the girl is able to identify new trees that she has not seen before and also to discriminate trees from other plants in general. Considering the various shapes, sizes, and colors that apply to trees, it would be impossible for a child to learn to reliably recognize trees from just a few remarks by her mother. Instead, a child learns what trees look like through building models straight from incoming visual data. There is no teacher involved here. The interaction with her mother just added a label (‘tree’) to the concept of a tree that had already been acquired through modeling the world in an unconscious fashion. In the machine-learning field, learning without a teacher is called ‘unsupervised learning’.

The human cortex holds about 10^{14} configurable synapses, which can be considered

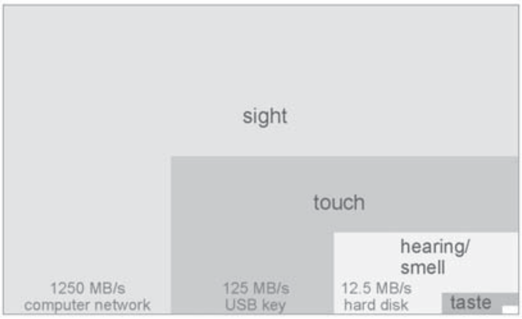


Fig. 5: Data rates for the senses relative to bandwidth of computer networks (McCandles, 2010).

parameters of the brain. We live about 10^9 seconds, so on average there is room to train about $100,000$ synapses every second. Indeed, brains receive a massive amount of data through the senses, for instance the retina sends more than 10 million bits of data to the brain every second. The role of teachers, parents, books, and other sources of abstract information is mostly to help us sort out which parts of these incoming data streams are important or should be ignored. In other words, teachers help us to select and label data streams that are used to train a model of the world. Crucially, in order to cope with a world where the settings and problems keep changing, a massive amount of unsupervised learning must always be going on.

In Fig. 5, data visualization artist Dave McCandles, based on work by the Danish science writer Tor Nørretranders, graphically displayed the amount of information that the different senses pass on to the brain in comparison to the bandwidth of computer networks (McCandles, 2010). Clearly, vision is the dominant sense. The white box in the lower-right corner represents the amount of data (0.7%) that is processed consciously in relation to the colored planes that refer to unconscious processing. Apparently, almost all incoming data is processed unconsciously. Building models of the world, including creating a model for what a tree looks like, is mostly an unconscious process.

Summary of information processing in the brain

In order to adapt to unforeseen changing conditions, brains need to iterate quickly through new model proposals for explaining the world. In the past three sections on information processing in the brain, we found three crucial ingredients for iterating quickly though signal processing system proposals. The first ingredient concerns probability theory as a foundational calculus. Probability theory prescribes how to learn and predict in a world where noise obscures the signals, where observations are scarce and where people's preferences change. The second principle relates to hierarchical modularity. In order to discover better algorithms, we need to test alternatives to existing algorithms and at the same time remain operational. We can only introduce a change to an existing algorithm if the effect of the change does not cause other parts of the algorithm to crash. We must survive the change and modularity is a crucial structural element so as to limit the impact of changes throughout the algorithm. Finally, when talking about the data we noted that the structure of real world data is so rich and volatile that we cannot rely on teachers, parents, scientists, and engineers to design and update the algorithm. Surviving in the real world implies a massive amount of unsupervised learning, which is always going on in the background. In engineering terms, continuous calibration is essential.

HEARING-AID SYSTEMS THAT WORK TODAY AND TOMORROW

Most of this paper has been dedicated to a review of data processing in the brain. Let us now get back to the engineering practice. We left this topic about half an hour ago when we were stuck with a block diagram of a hearing-aid algorithm. You were at a party and did not understand your conversation partner. You then wanted to test some variants of the hearing- aid algorithm right there when the problem occurred, but the system looked so complicated that any ideas on how to change the circuit were hard to come by. Our feeling was that if you would change anything, the algorithm would probably crash.

But let us assume that you have managed the dependencies between modules in such a way that you have enough confidence that a small change will not kill the algorithm. Then you can introduce some small changes to the hearing-aid algorithm and with a bit of luck you can improve your listening experience at the party. The next question is now whether the hearing aid should stick to this new configuration after you have left the party. You gave the hearing aid some new information, namely you showed the hearing aid how to behave when you are at a cocktail party. How relevant was that information for other acoustic environments? When the party is over, and you are in your car driving home, the hearing aid has two possible algorithms to choose from: the one that you came to the party with and the other algorithm that you preferred while the party was alive. Since you don't want to keep fiddling with your hearing aid every time when something changes in the acoustic environment, we want the hearing aid to decide for you.

In order to answer this question, the hearing aid would have to consider what features of the cocktail-party environment were so favorable for the second rather than the first algorithm and it would have to consider if or how much of these features remain active in the current car environment. In other words, the hearing aid should have access to a model of the acoustic world and it should be capable to answer what-if questions based on information that is preserved by the model. The hearing aid should have built such a world model by unsupervised training on past acoustic observations. In principle, this seems possible since a hearing-aid microphone records one million bits of acoustic data every four seconds. This continuous data stream should be summarized by a hierarchically organized structure, which is a necessary ingredient for the model to stay changeable, so it *can* adapt as new data get recorded. We have also discussed that the model should practice Bayesian reasoning in order to assess *how much* to adapt.

Unfortunately, these necessary ingredients for in situ learning are not part of today's HA signal-processing algorithms. As a result, rational adaptation of HA algorithms based on in situ acquired evidence is limited today. In our opinion, the key hearing-aid signal-processing challenge for the next decade will be to absorb the discussed additional features into our algorithms.

DISCUSSION

In this paper, we have taken a high-level perspective on the design and in situ re-design of hearing-aid algorithms. We have tried to make an argument for why fast in situ re-design of HA algorithms is crucial if we want to break through the 20% barrier of unsatisfied end users. Wireless links to remote control devices and fancy user interfaces lead to impressive products, but in the end all patient interactions should result in rational algorithm updates based on the evidence. As it turns out, while today's hearing-aid algorithms keep roughly 80% of end users satisfied, they are not suited for fast in situ experimentation and adaptation in case the patient is not happy. We then identified three salient properties of a very successful adaptable and personalized signal-processing system, namely the brain, that are absent in today's HA signal-processing structures. Specifically we discussed (1) learning through strict application of probability theory, (2) a hierarchically modular algorithm structure, and (3) continuous calibration. The absence of these properties hinder machine-learning-based re-design of today's HA algorithms. On the other hand, an emerging trend of cross-fertilization of ideas between the computational neuroscience, machine-learning, and signal-processing communities should make us mildly optimistic that significant progress towards in situ HA design can be achieved over the next decade.

REFERENCES

Friston, K. (**2009**). "The free-energy principle: a rough guide to the brain?" Trends Cogn. Sci., **13**, 293-301.

Jaynes, E.T. (**2003**). *Probability Theory: The Logic of Science*. Cambridge University Press.

Kochkin, S. (**2010**). "MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing," Hearing Journal, **63**, 19-20,22,24,26,28,30-32.

McCandles, D. (**2010**). "The beauty of data visualization", Talk at TED Global Oxford (http://goo.gl/7MzQ). Based on work by Tor Nørretranders.

ON Semiconductor (**2013**). "Datasheet for AYRE SA3291 Preconfigured DSP System for Hearing Aids," http://onsemi.com.

Raskin, A. (**2011**). *Wanna Solve Impossible Problems? Find Ways to Fail Quicker*. http://goo.gl/9zX3L.

Modeling auditory evoked brainstem responses to speech syllables. Can variations in cochlear tuning explain argued brainstem plasticity?

FILIP M. RØNNE[1,2,*], J AMES HARTE[1,3], AND TORSTEN DAU[1]

[1] *Centre for Applied Hearing Research, Technical University of Denmark, DK-2800 Lyngby, Denmark*

[2] *Eriksholm Research Centre, Rørtangvej 20, 3070 Snekkersten, Denmark*

[3] *Institute of Digital Healthcare, WMG, University of Warwick, Coventry, UCVA 7AL, UK*

Hornickel *et al.* (2009) and Skoe *et al.* (2011) measured and analyzed brainstem responses (ABRs) in response to the synthetic syllables /ba/, /da/ and /ga/, in normal and learning-impaired children. They reported a co-variation between the differences in average phase lag between the three syllable-evoked responses (called average phase-shifts), and speech-intelligibility performance (used as a predictor for learning-impairment). It was argued that, due to the reported normal peripheral hearing of both groups, the co-variation was evidence for neural differences in the brainstem, likely related to brainstem plasticity. They suggested brainstem functionality can be influenced by cortical structures to increase the difference between syllable responses. This study developed an ABR model capable of simulating ABRs to a variety of stimuli. The model was used to investigate whether the state of the peripheral hearing could be another possible explanation for the decreased average phase shifts observed for the learning-impaired children. Specifically, by changing the cochlear tuning of the model and evaluating the simulations based on models with broad versus sharp tuning (yet keeping all tuning estimates within normal audiometrical and wave-V latency range), it was observed that broader tuning systematically lead to smaller phase-shifts between the syllable-evoked ABRs.

INTRODUCTION

Auditory evoked potentials (AEP) have been used to assess the neural encoding of sound both for clinical and research purposes. Most studies have focused on the auditory brainstem response (ABR) as they are less affected by attention and sleep than potentials with origin at higher neural stages. The ABR has also been observed to be unaffected by training. However, a number studies have recently investigated and found evidence of plasticity[1] of the complex ABR (cABR), both

*Corresponding author: fmr@eriksholm.com

[1] physiological changes of the nervous system due to, e.g., learning

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

considering short term training effects and long-term experience effects. Hornickel *et al.* (2009) and Skoe *et al.* (2011) measured brainstem responses to the synthetically created syllable-stimuli /ba/, /da/, and /ga/, in normal and learning-impaired children. Both groups of children were reported to have normal audiometric thresholds and ABR wave-V latencies. Skoe *et al.* (2011) developed a 'cross-phaseogram' from the time-varying cross-power-spectral-density between two ABR recordings. When analyzed in time-frames, the outcome was a spectrogram-like representation of the phase-lag as a function of time and frequency. From the cross-phaseogram an averaged phase-shift between two syllable-evoked responses was obtained. The average phase-shift was shown to correlate with reading abilities and speech-in-noise perception, such that large phase-shifts correlated with good performance in the speech-in-noise test. Hornickel *et al.* (2009) and Skoe *et al.* (2011) argued that this result was evidence for plasticity in the brainstem, as the group with the good behavioral performance had undergone long-term learning. Thus, better performance was an indication of learning that had affected both the behavioral performance and the electrophysiological brainstem recordings. This paper challenges the reasoning behind this interpretation. By modeling it attempts to show that individual variations in cochlear tuning, all within normal-hearing boundaries, significantly affect the average phase-shifts, thus showing that the measures of the peripheral hearing chosen by Hornickel *et al.* (2009) are not sufficient to conclude that the individual spread in the peripheral hearing does not affect the average-phase shift group differences between normal and learning-impaired children.

METHOD

ABR model

The ABR model used in this study was similar to the model of Rønne *et al.* (2012). However, the auditory-nerve (AN) model used to compute the summed activity pattern was updated such that the Zilany *et al.* (2009) AN model was used instead of the Zilany and Bruce (2007) model. This update was made as the Zilany *et al.* (2009) has an improved IHC-AN stage producing more realistic adaptation properties. As the syllable-stimuli are of longer duration, a precise adaptation is beneficial. The change of the AN model required a recalculation of the unitary response (UR). The UR (based on standard cochlear filter tuning) was calculated, following Rønne *et al.* (2012), as the deconvolution of a 95.2 dB peSPL grand average click-evoked ABR recording (Elberling *et al.*, 2010; Rønne *et al.*, 2012) and the summed activity pattern obtained by simulating the response to an identical click-stimulus.

The simulated cABRs were at the output filtered with a 2nd order band-pass filter with cutoff frequencies at 70 Hz and 2 kHz. These filter settings were identical to the output filters of Hornickel *et al.* (2009) and Skoe *et al.* (2011).

Stimuli

Synthetic /ba/, /da/, and /ga/ syllables (Hornickel *et al.*, 2009; Skoe *et al.*, 2011) were used, that only differ in the frequency content of the second formant, f_2, of the first 60 ms, corresponding to the consonant part of the stimuli. The second formants decrease in the [ga] stimulus from 2480 Hz, in the [da] from 1700 Hz, and increased in the [ba] stimulus from 900 Hz, reaching a steady-state frequency (corresponding to the /a/ part of the syllable) of 1240 Hz in all 3 stimuli. The /a/ vowel-part of the syllables was the same for the three syllables, consisting of the formant frequencies f_0 = 100 Hz, f_1 = 720 Hz, f_2 = 1240 Hz, f_3 = 2500Hz, f_4 = 3300 Hz, f_5 = 3750 Hz and f_6 = 4900 Hz. All three stimuli were calibrated to have a root-mean-square (RMS) level of 1, and were presented to the model at a level corresponding to 80 dB SPL, which was also used in the study by Skoe *et al.* (2011).

Cross-phaseogram

Skoe *et al.* (2011) proposed a cross-phaseogram to illustrate the phase-differences and thus the time delays between two cABR recordings. Each recording was divided into 20-ms time frames with 19-ms overlap. A Hanning window was applied, resulting in a 3-dB main lobe width of 141 Hz. The cross power spectrum density, i.e., the power spectrum density of the cross correlation, was computed between each pair of frames from the two recordings. An artificial frequency resolution of 4 Hz was obtained by zero padding, effectively acting as a smoothing operation. Finally, the unwrapped phase (in radians) was extracted and plotted as a function of time (midpoint of the 20-ms frames) and frequency. Skoe *et al.* (2011) also proposed the average phase-shift to simplify the cross-phaseogram into a single number that could be compared to other measures, such as psychoacoustic speech-in-noise performance. The average phase-shift (in π radians) was calculated on the formant transition period (15 to 60 ms) of the syllable-evoked cABR in the frequency range of 70 to 1100 Hz.

Weighted cross-phaseogram

The cross-phaseogram weights time-frequency bins with little activity as high as bins with much activity. This limits the use of the cross-phaseogram, as it is impossible to distinguish between time-frequency bins of presumable little importance due to low activity from bins of major importance due to large activity. A weighted cross-phaseogram is therefore suggested here. It was created by deriving the energy from each of the two syllable-evoked cABRs in similar time-frequency bins as those chosen in the Skoe *et al.* (2011) cross-phaseogram. The two resulting matrices were summed and normalized with the average bin activity. This matrix was then multiplied bin-per-bin with the original cross-phaseogram.

Variability of cochlear filter tuning

Cochlear filter tuning and basilar-membrane (BM) delay are inherently related (Eggermont, 1979; Bentsen *et al.*, 2011; Verhulst *et al.*, 2013), such that broader filters lead to shorter delays. Elberling and Don (2008) measured derived-band latencies from a total of 81 normal-hearing subjects (hearing thresholds $<$ 15 dB HL), at four different band center frequencies (bCF; 710, 1400, 2800, and 5700). ABR wave-V latency and an inter-subject standard deviation (SD) were derived. The BM delay was achieved by subtracting the wave I-V delay (4.1 ms) and the synaptic delay (1 ms). A representation of the variation of cochlear filter tuning in normal-hearing subjects can be obtained from the mean latencies ± 1 standard deviation. The stimulus of Elberling and Don (2008) was a click presented at approximately 90 dB peSPL.

Eggermont (1979) derived a theoretical relation between the cochlear filter tuning, Q_{10}, and the average number of cycles in the impulse response up to the latency (minus 1 ms of synaptic delay) of the derived band CAP, N_{av};

$$N_{av} = \frac{0.5}{\pi^2}\left(\frac{5(1+\gamma)(2+\gamma)}{12\gamma}Q_{10} - 1\right)\left(2 + ln\frac{5(1+\gamma)(2+\gamma)}{12\gamma} + lnQ_{10}\right) \qquad \text{(Eq. 1)}$$

where N_{av} can be calculated as $(CF/1000) * \tau_{CF}$, where τ is the BM latency of at the CF. $\gamma = 2$ is representative of a normal cochlea (Eggermont, 1979), and Q_{10} values can thus be derived. To convert the Q_{10} values into Q_{ERB} values, the conversion from Ibrahim and Bruce (2010) was applied:

$$Q_{ERB} = \frac{Q_{10} - 0.2085}{0.505} \qquad \text{(Eq. 2)}$$

Fig. 1 shows the Q_{ERB} values derived from Elberling and Don (2008)'s measured delays ±1 SDs and ±2 SDs. New tuning-curve estimates were obtained from the ±1 SD and ±2 SD based Q-estimates, by multiplying the Shera *et al.* (2002) estimates by a constant offset (broader tuning-estimates multiplied by 0.80 and 0.60, sharper tuning-estimates by 1.15 and 1.28). The four suggested tuning curves were implemented in the ABR model. For each simulated condition, a new UR was calculated.

RESULTS

Table 1 shows the average phase-shifts obtained in Skoe *et al.* (2011) and the corresponding values obtained from the simulations.[2] Both experimental results and simulations show the largest phase-shift between /ga/ and /ba/, which also differs most in their frequency spectrum. Also, both the data and the simulations show that the phase-shift between /ga/ and /da/ is smaller than the phase-shift between /da/ and /ba/. In Fig. 2, weighted average phase-shifts for all syllable comparisons and all five different tuning-curve implementations are shown. Although the growth of the phase-shift with increasing tuning amount is non-monotonic, a trend is clearly observed, where sharp tuning leads to larger phase-shifts. This confirms that the state of the

[2] Coloured cross-phaseograms describing the results in details are shown on the poster (available from `http://www.eriksholm.com/~asset/cache.ashx?id=26052&type=14&format=web`).

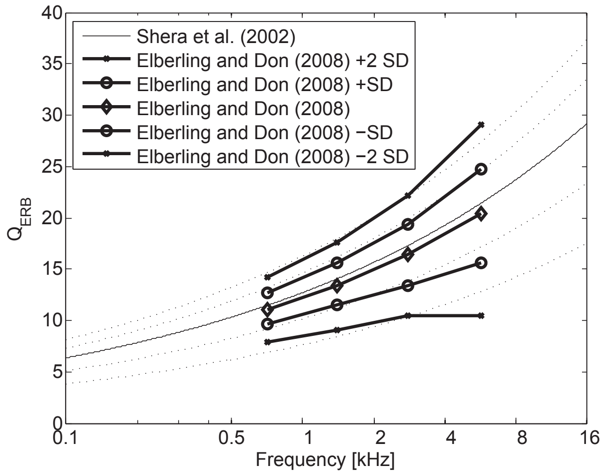


Fig. 1: Q_{ERB}'s calculated based on Elberling and Don (2008)'s measured derived band latencies (diamonds). In circles and triangles, Q_{ERB} estimates based on Elberling and Don (2008)'s measured latencies ±1 SD an ±2 SD. Also shown is the Shera *et al.* (2002) tuning (solid line) which is implemented in the standard ABR model. The alternative tuning curves (dotted lines) are fitted to the Elberling and Don (2008) based tuning (±1 SD and ±2 SD) and implemented in the model.

auditory periphery affects the cross-phaseogram and weighted average phase-shifts. The implications for the Hornickel *et al.* (2009) and Skoe *et al.* (2011) studies are discussed further below.

	Skoe et al. (2011)	Simulations	Simulations (weighted)
/ga/-/ba/	0.317 ± 0.040	0.353	3.040
/da/-/ba/	0.288 ± 0.031	0.243	2.163
/ga/-/da/	0.208 ± 0.028	0.141	1.660

Table 1: Average phase-shifts of Skoe *et al.* (2011) recordings (left column), simulated average phase-shifts (center column), and weighted average phase-shifts (right column). The average is taken across the region from 15 to 60 ms, and from 70 to 1100 Hz.

DISCUSSION

Unweighted versus weighted cross-phaseogram

The cross-phaseogram and the average phase-shifts were developed by Skoe *et al.* (2011) and have proven to be valuable tools for investigating phase-shifts between different frequency components of the recorded (or simulated) cABR. However, the

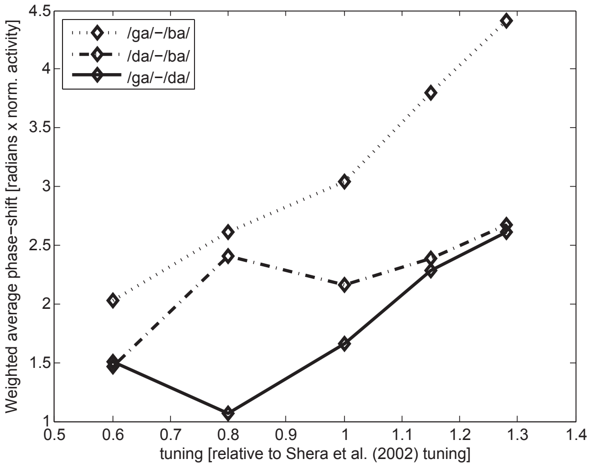


Fig. 2: Weighted average phase-shifts for each of the syllable combinations, for both broad (0.60 and 0.80), standard (1.00) and sharp (1.15 and 1.28) tuning.

equal weighting of all time-frequency bins limits the value of the average phase-shift (Skoe *et al.*, 2011), since a bin with little activity will hardly influence the cABR generation. In fact, a time-frequency bin with little energy is likely to be dominated by measurement noise, and the average measure might thus emphasize noise. In the simulations presented in this study, noise is not included. This makes a comparison between simulations and data in the terms of the average phase-shift difficult, as a systematic phase-shift at bins with little activity will be included in the simulated average phase-shift, whereas such a phase-shift is likely to be influenced or masked by measurement noise in the data-derived average phase-shift. This could be solved by adding noise to simulations. However, this would imply that the model would no longer be deterministic, which has not been considered in the present study.

Implications of changing cochlear tuning on Skoe et al. (2011) conclusions

Hornickel *et al.* (2009) and Skoe *et al.* (2011) found correlations between learning-impairments of children, and recorded cross-phaseogram phase-shifts (peak latencies in Hornickel *et al.*, 2009) between syllable-evoked cABRs, such that a small average phase-shift was an indication of learning-impairment. A basic assumption of Hornickel *et al.* (2009) was that the two groups of normal and learning-impaired children have equally good peripheral hearing (equal audiograms and ABR wave-V latencies). Hornickel *et al.* (2009) argued that this was the case as all subjects had audiometric thresholds below 20 dB HL and had normal ABR wave-V latencies. However, given the possible variation of ‘normal’ BM tuning, an alternative explanation for the Hornickel *et al.* (2009) results can be hypothesized. A broad cochlear tuning leads to shorter peak-latencies for all three stimuli. Further, the traveling-wave delay decreases logarithmically with increasing stimulus frequency (e.g., Neely *et al.*, 1988; Elberling *et al.*, 2010). A broad tuning would thus lead to a decreased difference between the cABR peaks, and thus a smaller phase-shift. Phase-shift differences similar to the one Skoe *et al.* (2011) finds between the groups of normal and learning-impaired children, could thus be hypothesized to also be found when measuring cABRs to two normal-hearing groups but with different cochlear tuning.

The results from this modeling study showed that there is indeed a relation between filter tuning and weighted averaged cross-phaseogram values, where sharper tuning leads to larger phase-shifts. Although this relation was not strictly monotonic, it does indicate that the phaseograms are sensitive to changes in the auditory periphery. Whether this finding offers an alternative explanation for the results of Hornickel *et al.* (2009) and Skoe *et al.* (2011) is, however, questionable. That would require the assumption that the group of learning-impaired children had significantly overall broader cochlear tuning than the normal children. Although this hypothesis is not unlikely, this study cannot verify such a claim. That would require a major study, where the cochlear tuning of learning-impaired and normal subjects were measured carefully and correlated with weighted average phase-shifts. Thus, the conclusion of this study is that the huge spread of normal-hearing cochlear-tuning likely leads to a huge spread in weighted average phase-shifts.

Skoe *et al.* (2011) concluded that the correlation between learning-impairment and average phase-shifts showed plasticity of brainstem. This conclusion was based on the assumption that the state of the auditory periphery was equal (i.e., normal hearing) in both groups. However, this study has indicated that the cochlear tuning of the normal-hearing subjects does have a significant effect on the average phase-shift, and does thus challenge the underlying assumption of the conclusions from Hornickel *et al.* (2009) and Skoe *et al.* (2011). Further, this study has shown that the use of audiograms and click-evoked ABR wave-V latencies are unlikely to be precise enough to claim that the cochlear tuning is similar between two groups.

SUMMARY AND CONCLUSION

This study evaluated the performance of an ABR model to simulate cABR responses to three synthetic syllables. The ABR model was shown to predict phase-shifts between the responses to the three syllable stimuli. It was shown that altering the cochlear tuning influenced the simulated phase-shifts, illustrating that the state of the auditory periphery is crucial when analyzing responses based on the cross-phaseogram. The results suggests that the assumption of Hornickel *et al.* (2009) and Skoe *et al.* (2011) that the peripheral hearing was similar between their two groups of test subjects might be flawed, and that the following conclusion that the larger phase-shifts for the non-learning-impaired children was the consequence of plasticity might thus be wrong.

Bentsen, T., Harte, J.M., and Dau, T. (**2011**), “Human cochlear tuning estimates from stimulus-frequency otoacoustic emissions,” J. Acoust. Soc. Am., **129**, 3797-3807.

Eggermont, J. (**1979**), "Narrow-band AP latencies in normal and recruiting human ears," J. Acoust. Soc. Am., **65**, 463-470.

Elberling, C., and Don, M. (**2008**). "Auditory brainstem responses to a chirp stimulus designed from derived-band latencies in normal-hearing subjects," J. Acoust. Soc. Am., **124**, 3022-3037.

Elberling, C., Callø, J., and Don, M. (**2010**). "Evaluating auditory brainstem responses to different chirp stimuli at three levels of stimulation," J. Acoust. Soc. Am., **128**, 215-223.

Hornickel, J., Skoe, E., Nicol, T., Zecker, S., and Kraus, N. (**2009**). "Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception," Proc. Natl. Acad. Sci. USA, **106**, 13022-13027.

Ibrahim, R.A., and Bruce, I.C. (**2010**), "Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues," in *Neurophysiological bases of auditory perception*. Edited by E.A. Lopez-Poveda and A.R. Palmer, Med Elect, Hear Life, 15th International Symposium on Hearing, Salamanca, Spain, June 2009, pp. 429-438.

Neely, S., Norton, S., Gorga, M., and Jesteadt, W. (**1988**). "Latency of auditory brainstem responses and otoacoustic emissions using tone-burst stimuli," J. Acoust. Soc. Am., **83**, 652-656.

Rønne, F., Harte, J., Elberling, C., and Dau, T. (**2012**). "Modelling auditory evoked brainstem responses to transient stimuli," J. Acoust. Soc. Am., **131**, 3903-3913.

Rønne, F., Harte, J., and Dau, T. (**2013**). "Modelling human auditory evoked brainstem responses to speech syllables," Proc. Meet. Acoust., **19**, International Congress on Acoustics, Montréal, Canada, pp. 050120.

Shera, C., Guinan, J., and Oxenham, A.J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. USA, **99**, 3318-3323.

Skoe, E., Nicol, T., and Kraus, N. (**2011**). "Cross-phaseogram: Objective neural index of speech sound differentiation," J. Neurosci. Meth. **196**, 308-317.

Verhulst, V., Bharadwaj, H., Mehraei, G., and Shinn-Cunningham, B. (**2013**). "Understanding hearing impairment through model predictions of brainstem responses," Proc. Meet. Acoust., **19**, International Congress on Acoustics, Montréal, Canada, pp. 050182.

Zilany, M.S.A., and Bruce, I.C. (**2007**). "Representation of the vowel (epsilon) in normal and impaired auditory nerve fibers: Model predictions of responses in cats," J. Acoust. Soc. Am., **122**, 402-417.

Zilany, M.S.A., Bruce, I.C., Nelson, P.C., and Carney, L.H. (**2009**). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics," J. Acoust. Soc. Am., **126**, 2390-2412.

Tinnitus: maladaptive plasticity?

Jos J. Eggermont*

Department of Physiology and Pharmacology / Department of Psychology,
University of Calgary, Calgary, Alberta, Canada

Tinnitus is a symptom, not a disease. Tinnitus is often accompanied by hyperacusis as well as hearing loss. Tinnitus is foremost not an auditory disorder but a particular consequence of hearing loss, and then only in about 1/3 of the cases. Tinnitus can also result from insults such as whiplash, via somatic-auditory interaction in the dorsal cochlear nucleus. These are examples of bottom-up mechanisms that may underlie tinnitus. Much is known about necessary neural substrates of tinnitus, but much less about the sufficient ones. I will review proposals from animal research for these neural correlates, i.e., increased spontaneous firing rates, increased neural synchrony and reorganized cortical tonotopic maps. These can occur following noise trauma, but also following long-term exposure to non-traumatic (< 70 dBA) sounds. Homeostatic plasticity may play a role. I will compare these findings with what is known from human imaging and electrophysiology in tinnitus patients, and suggest that animal studies and human findings related to tinnitus are so far not fully compatible.

INTRODUCTION

Tinnitus, defined as the percept of sound in the absence of external sounds, is common. Its average prevalence ranges from about 7% in adolescents to about 17% in the elderly. The most common cause is hearing loss, in particular noise-induced hearing loss. However, head and neck injuries also constitute a large percentage, presumably through the interaction of somatosensory and auditory inputs in the dorsal cochlear nucleus (DCN). Ototoxic drugs that do not cause permanent hearing loss such as salicylates present only a small fraction of the etiology. Furthermore, stopping their use typically ends the tinnitus. One of the conundrums is that only 30% of people with hearing loss develop tinnitus, whereas in those that develop it, at most half find the tinnitus bothersome. This suggests that top-down influences, such as attention, effects of stress, and potentially central gating mechanisms play a role in the tinnitus percept (Roberts *et al.*, 2010; 2013).

Tinnitus is a conscious percept, namely, people who have tinnitus are aware of it and can express to others how it sounds. Consciousness most likely has a solid neural correlate (De Ridder *et al.*, 2011). One of the burning questions facing animal research into tinnitus must thus be: Are animals conscious of their tinnitus? According to Ward (2011) conscious percepts are thalamocortical based, thereby putting mammals firmly in possession of the putative neural substrate. But can they express the presence of their tinnitus? Behavioral tests in animals generally do not

*Corresponding author: eggermon@ucalgary.ca

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

rely heavily on thalamocortical activity; however, they may reflect subthalamic changes in spontaneous activity or in synaptic gain, or both. For instance, cortical ablation generally allows relearning of conditioned response and hardly affects prepulse (or gap) startle reflexes (Eggermont, 2013). Understandably, tests that can unambiguously indicate whether an animal perceives tinnitus are essential to advance tinnitus research.

ANIMAL MODELS OF TINNITUS

Laboratory studies have shown that tinnitus may develop in humans almost immediately after exposure to loud traumatic sounds. Animal studies can be used to discover the neural substrates related to such early-onset, and often transient, tinnitus. After traumatic noise, prolonged exposure to occupational or recreational noise, or following slowly acquired losses during aging, tinnitus may over time develop from an intermittent presence to a chronic status, and likely acquire a dominant central contribution.

So far, animal models of tinnitus have concentrated on acute or chronic application of salicylate and on acute and chronic exposure to traumatic noise. Neural correlates of these applications form presumed substrates for tinnitus. Currently, most animal research is combined with behavioral tests. As I have outlined elsewhere (Eggermont, 2013), the results of these tests are not straightforward for the determination of the presence of tinnitus.

Spontaneous activity

Let us focus on the neurobiological correlates of noise-induced hearing loss, in particular those that relate to spontaneous activity, as this most likely relates to tinnitus. A potential neural correlate of tinnitus is increased spontaneous firing rate (SFR). Typically, SFR does not change in animals with aging, neither in dorsal cochlear nucleus (Caspary *et al.*, 2005) nor in auditory cortex (Turner *et al.*, 2005). Thus, aging in itself is unlikely to be a tinnitus-inducing factor, albeit that it may enhance pre-existing tinnitus given the increased incidence of tinnitus with age.

After noise trauma, the SFR in cat auditory nerve fibers was significantly reduced (Liberman and Kiang, 1978). *In vivo* experiments in hamster dorsal cochlear nucleus indicated massive increases in SFR 5-180 days after noise exposure (Kaltenbach *et al.*, 2000). Complete or nearly complete section after 4 weeks of ascending (Zacharek *et al.*, 2002) or descending inputs (Zhang *et al.*, 2006) did not significantly affect the magnitude of SFR in the dorsal cochlear nucleus, suggesting that increased SFR is either a self-contained neural network phenomenon or reflects intrinsic cell changes. The increase in SFR in hamster dorsal cochlear nucleus correlated with the strength of the behavioral index of tinnitus (Kaltenbach *et al.*, 2004). Vogler *et al.* (2011) investigated SFRs in the ventral cochlear nucleus (VCN) of guinea pigs exposed for 2 h to a 10-kHz tone presented at 124 dB SPL. After a 2-week recovery period, the mean SFR in noise-exposed ears was significantly elevated (by a factor of about two) compared to sham controls. This was more evident in primary-like and onset categories of neurons.

The independence of SFR from cochlear input demonstrated in the DCN (see above) could not be replicated for recordings in the central nucleus of the inferior colliculus (ICC) in noise-exposed (10-kHz tone at 124 dB SPL for 1 h) guinea pigs. The increase in SFR ceased after cochlear ablation, cochlear cooling, or perfusion with a pre-synaptic transmitter release inhibitor, or after destroying the post-synaptic receptors with kainic acid (Mulders and Robertson, 2009).

The time of onset of increased SFR in ICC was present by 12 h post acoustic trauma, whereas data obtained within approximately 4 h of the cessation of acoustic trauma showed no evidence of hyperactivity. These data suggest that hyperactivity in the inferior colliculus (IC) is a relatively rapid plastic event beginning within hours rather than days post cochlear trauma. Hyperactivity did not show any further systematic increase between 12 h and up to 2 weeks post acoustic trauma. At recovery times of 12 and 24 h, hyperactivity was widespread across most regions of the IC, but at longer recovery times it became progressively more restricted to ventral regions corresponding to the regions of the cochlea where there was persistent damage (Mulders and Robertson, 2013).

Recovery after acoustic trauma resulted in more neurons with high SFR compared to control animals, resulting in an increase in the average SFR. At recovery times up to 4 weeks after the exposure, the increased SFR disappeared when cochlear input to the ICC was destroyed. Thus, the hyperactivity in the ICC after acoustic trauma is dependent on activity in the contralateral cochlea. How this could happen, with the persisting hyperactivity in the DCN after cochlear ablation at about the same post-recovery time, is unclear. However, the VCN may provide the dominant input to the ICC and determine the SFR. This is likely, as we have seen that after chronic trauma SFRs are increased in the VCN (Vogler *et al.*, 2011). When the recovery time after acoustic trauma is extended to 8 and 12 weeks, cochlear ablation does not significantly decrease the increased spontaneous activity measured in the IC. This demonstrates that central hyperactivity that develops after acoustic trauma evolves from an early stage, when it is dependent on continued peripheral afferent input, to a later stage in which the hyperactivity is intrinsically generated within the central nervous system (Mulders and Robertson, 2011).

In cat primary auditory cortex (AI), a significant increase in SFR occurred at least 2 hours after the trauma, but not immediately (< 15 min) following it (Noreña and Eggermont, 2003). At least 3 weeks after the trauma, the SFR was significantly higher than in controls at all characteristic frequencies (CFs) tested, so increased SFR in AI is not restricted to the region of the hearing loss, although that region showed a more pronounced increase (Noreña and Eggermont, 2006).

The degree to which spike firing from two different, simultaneously recorded, neurons is time-locked or synchronized can be quantified by the cross-correlogram (Eggermont, 1992). Effects of acute noise trauma on neural synchrony were studied by Noreña and Eggermont (2003) in AI. A significant increase in peak cross-correlation coefficients was apparent within 15 minutes of the trauma, and increased by a further 50% at 2 h after the trauma (Fig. 1). This suggests an important role for neural synchrony in the generation of tinnitus, potentially eclipsing that of increased SFR. Several weeks to months after the trauma, all neuron pairs in the reorganized region of auditory cortex showed significant neural correlations (Noreña and Eggermont, 2005). Weisz *et al.* (2007) proposed that gamma band activity, which is increased in tinnitus patients, may reflect the synchronous firing of neurons within the auditory cortex and constitute the neural code of tinnitus.

Long-term exposure to different types of non-traumatic acoustic environments also results in changes in SFR activity in the cat AI (Munguia *et al.*, 2013). Four different groups of adult cats were exposed to moderate-level (~70 dB SPL) behaviorally irrelevant sounds for several weeks to months, and their SFRs were compared with those in control cats. The sounds consisted of random multi-frequency tone pip ensembles with various bandwidths (2-4 kHz, 4-20 kHz, and a pair of third-octave bands centered at 4 and 16 kHz), as well as a "factory noise". Auditory brainstem response (ABR) thresholds, ABR wave-3 amplitudes at ~55 and 75 dB SPL, and distortion product otoacoustic emission (DPOAE) amplitudes were unaffected by the exposure. However, we found that the SFR decreased within the exposure frequency range and increased outside the exposure range. This increased SFR for units with characteristic frequencies outside the exposure frequency range, which was slow to reverse after the exposure offset, suggests a mechanism for tinnitus in the absence of hearing loss.

Stimulus evoked activity

Stimulus-induced neural responses are also altered following noise-induced hearing loss (NIHL). Significant effects reflecting central gain changes have been found. Despite a reduction in the compound action potential amplitude of the auditory nerve and in the local field potential of the cochlear nucleus following noise trauma in the rat, the local field potential amplitude in the IC was typically enhanced at higher intensity levels (Wang *et al.*, 2002), and so was the local field potential in auditory cortex (Yang *et al.*, 2007).

Tonotopic maps are representations of the distribution of CF as a function of spatial coordinates in an auditory nucleus or cortex. Local mechanical damage to the cochlea, ototoxic-drug damage to the cochlea, and NIHL all cause tonotopic map changes in AI (Eggermont and Roberts, 2004). The map changes are not causally related to the hearing loss (Noreña and Eggermont, 2005), but are always accompanied by increased SFR and increased neural synchrony, pointing to their correlative rather than causal nature. We suggested that this prolonged synchronization would induce the perception of tinnitus (Noreña and Eggermont, 2003; Seki and Eggermont, 2003).

Several stages of cortical reorganization can be differentiated. The first relates to the unmasking of normally inhibited connections (Calford, 2002). This unmasked excitatory activation could be the result of loss of GABA-mediated inhibition (Wang

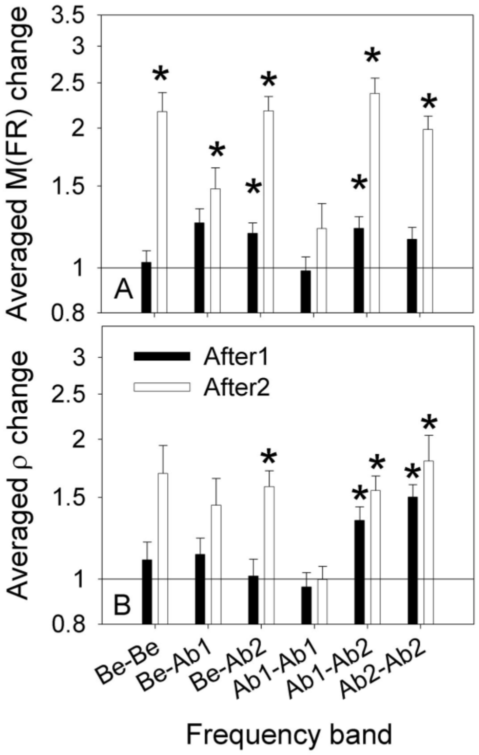


Fig. 1: Effect of the acoustic trauma on cross-correlation coefficient (ρ). (A) Change in M(FR) averaged (geometric mean) into six frequency bands. (B) Change in ρ averaged (geometric mean) into six frequency bands, immediately (After1) and a few hours (After2) after the acoustic trauma (± S.E.M., * p < 0.0083). Immediately after the acoustic trauma (black bars), one notes that ρ is significantly increased in the Ab2-Ab2 group whereas M(FR) is not. Be: below the trauma-tone frequency (TTF). Ab1: within 1 octave of the TTF. Ab2: 1-2 octaves above TTF. From Noreña and Eggermont (2003).

et al., 2011). A second stage involves structural changes such as axonal sprouting, as well as alterations in synaptic strength. Finally, use-dependent plasticity might lead to additional changes based on Hebbian learning and long-term potentiation. Tonotopic map changes do not occur if, immediately after noise trauma, a compensatory complex sound that mimics the frequency range of the hearing loss in bandwidth and level is presented for several weeks (Noreña and Eggermont, 2005). It is assumed that during the presentation of this compensatory sound the down regulation of inhibition that usually follows NIHL (Milbrandt *et al.*, 2000) does not occur, and that the unmasking of new excitatory inputs (Noreña and Eggermont, 2003) does not happen or is reversed. When this ‘unmasking’ trigger for tonotopic map reorganization is absent, map changes do not occur, despite a remaining hearing loss. Furthermore, no increases in SFR and neural synchrony were seen (Noreña and Eggermont, 2006).

WHERE IN THE BRAIN IS TINNITUS?

Auditory system

A recent study by Gu *et al.* (2010) allowed an identification of the auditory brain areas involved in generating tinnitus. They reported physiological correlates of two perceptual abnormalities in the auditory domain that very frequently co-occur: tinnitus and hyperacusis. Despite receiving identical sound stimulation levels, subjects with hyperacusis showed elevated evoked activity in the auditory midbrain, thalamus, and primary auditory cortex compared with subjects with normal sound tolerance. This reflects the increased gain for processing external auditory stimuli. Primary auditory cortex, but not subcortical centers, showed elevated activation specifically related to tinnitus, i.e., in the absence of hyperacusis. The results directly link both hyperacusis and tinnitus to hyperactivity within the central auditory system.

Langers *et al.* (2012) investigated tonotopic maps in primary auditory cortex of 20 healthy controls and 20 chronic subjective tinnitus patients. The goal was to test the hypothesis, proposed on basis of animal and previous human studies (Eggermont and Roberts, 2004) that tinnitus results, among others, from an abnormal tonotopic organization of the auditory cortex. All participants had normal or near-normal hearing up to 8 kHz. The study found no evidence for a reorganization of cortical tonotopic maps in these tinnitus patients. This is perhaps not surprising since there was no appreciable hearing loss. It had been previously shown (Fig. 2) that in animals there is no reorganization of the cortical tonotopic map for hearing losses ≤ 25 dB (Rajan, 1998; Seki and Eggermont, 2002). However, Langers *et al.* (2012) clearly did demonstrate that reorganized tonotopic maps in auditory cortex are not a requirement for tinnitus to occur.

Although tinnitus is a percept of sound in the absence of external stimulation, whereas hyperacusis is an increased response to external stimulation, they are often co-occurring. The prevalence of hyperacusis in tinnitus patients can be as high as 79% (Dauman and Bouscau-Faure, 2005). Jastreboff and Hazell (1993) described hyperacusis as a ‘manifestation of increased central gain’, which may cause enhanced perception of peripheral signals. Threshold measures are not sensitive indicators, as Kujawa and Liberman (2009) demonstrated that cochlear (inner hair cell ribbon synapses) and nervous damages (high-threshold auditory nerve fibers) can occur in the presence of normal audiometric thresholds.

Non-auditory brain regions

Amplifying on a prescient model of Jastreboff (1990), Rauschecker *et al.* (2010) proposed the first consistent model that incorporates the interaction between the limbic and auditory system: “(1) In most, if not all, cases, the process leading to tinnitus is triggered by a lesion to the auditory periphery, e.g., a loss of hair cells in the inner ear resulting from acoustic trauma or aging. (2) Loss of input in the lesioned frequency range leads to an overrepresentation of lesion-edge frequencies, which causes hyperactivity and possible burst-firing in central auditory pathways, constituting the initial tinnitus signal. (3) Under normal circumstances, the tinnitus signal is cancelled out at the level of the thalamus by an inhibitory feedback loop originating in paralimbic structures: activity from these structures reaches the thalamic reticular nucleus, which in turn inhibits the medial geniculate nucleus. If, however, paralimbic regions are compromised, inhibition of the tinnitus signal at the thalamic gate is lost, and the signal is relayed all the way to the auditory cortex, where it leads to permanent reorganization and chronic tinnitus." In essence, Rauschecker *et al.* (2010) proposed that normally, the unwanted SFR (noise signal) is identified by the limbic system and eliminated from perception by feeding it back to the (inhibitory) thalamic reticular nucleus, which subtracts it from the afferent auditory signal. This mechanism would then fail in about 30% of people with NIHL, but why it would do so is unknown.

TINNITUS AS MALADAPTIVE PLASTICITY IN THE CENTRAL NERVOUS SYSTEM

A common hypothesis is that tinnitus results from an imbalance between excitation and inhibition as a result of a maladaptive down-regulation of inhibitory amino-acid neurotransmission in the central auditory pathway. This loss of inhibition may be a compensatory response to loss of afferent input such as that caused by acoustic insult and/or age-related hearing loss, the most common causes of tinnitus in people. Compensatory plastic changes may result in pathologic neural activity that underpins tinnitus (Wang *et al.*, 2011). Homeostatic mechanisms stabilize the mean firing activity of a neuron over a time period of a few days, and typically do so by scaling the efficacy of the neuron's synapses (Turrigiano, 1999). An important aspect of synaptic scaling is that the direction of change in the synaptic strength depends on both the nature of the synapse and the nature of the postsynaptic neuron. Cortical pyramidal neurons are embedded in networks with extensive recurrent excitatory and inhibitory feedback. Pyramidal-neuron firing rates reflect not only their excitatory drive, but also the balance between excitatory inputs from other pyramidal neurons and inhibitory inputs from GABAergic interneurons.

In the healthy auditory system, homeostatic plasticity could help to ensure that the working point of auditory neurons is within the right range of firing rates independent of the prevailing acoustic environment. Homeostatic plasticity in auditory neurons might also prevent us from perceiving normal spontaneous neuronal activity as sound. Schaette and Kempter (2006; 2009) modeled the effects of homeostatic plasticity by a change in a gain factor proportional to the deviation of the mean activity from a certain target rate. In their model, homeostatic plasticity restores the mean firing rate of neurons in the DCN after hearing loss. Thus, both stimulus-driven and spontaneous mean firing rates are scaled upward to the pre-noise exposure target level. This applies to all affected neurons along the auditory pathway. Restoring the mean rate therefore likely increases the spontaneous rate throughout the auditory system. Knipper et al. (2012) suggested that "two divergent kinds of hyperactivity at the level of the DCN may differently influence higher brain areas after auditory trauma. Hyperactivity in sound-driven pathways may be regarded in the context of a rather typical compensatory response of a healthy system that, after sensory deprivation, adapts the synaptic strength toward original levels through homeostatic scaling".

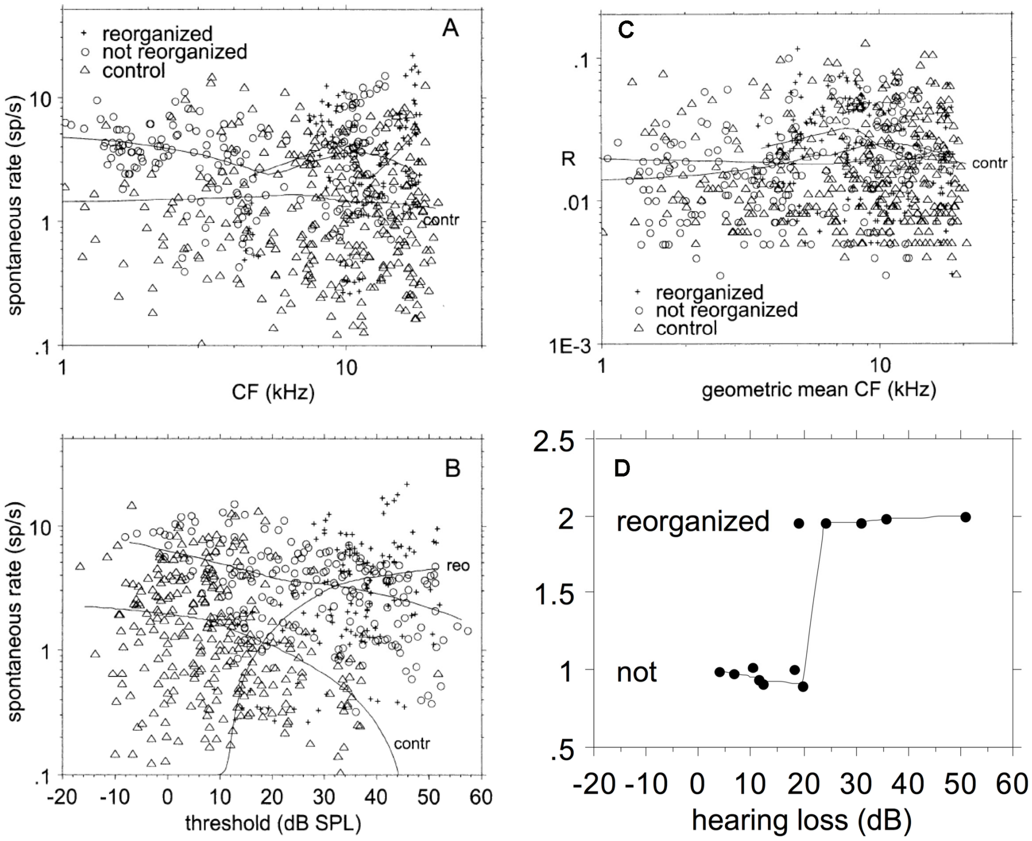


Fig. 2: Dependence of SFR (A, B) and synchrony (C) on CF and threshold is not dependent on the presence of tonotopic map reorganization. (D) Presence of map reorganization on the average hearing loss measured by ABR above 6 kHz. From Seki and Eggermont (2002; 2003).

Analogies with phantom pain

The cause of phantom pain experience has also commonly been attributed to maladaptive plasticity: following loss of sensory input, e.g., the deprived hand area of the primary sensorimotor cortex becomes responsive to inputs from cortical neighbors (for example the face), thereby triggering pain representations relating to the hand. However, Makin *et al.* (2013) showed that, while loss of sensory input is generally characterized by structural and functional degeneration in the deprived sensorimotor cortex, the experience of persistent pain is associated with preserved structure and functional organization in the former hand area. Furthermore, phantom pain is associated with reduced inter-regional functional connectivity in the primary sensorimotor cortex. Makin *et al.* (2013) therefore proposed that, contrary to the maladaptive model, cortical plasticity associated with phantom pain is driven by powerful and long-lasting subjective sensory experience, such as triggered by nociceptive or top-down inputs. They suggested that phantom pain be best understood in terms of experience-dependent plasticity, with chronic phantom pain providing the experience.

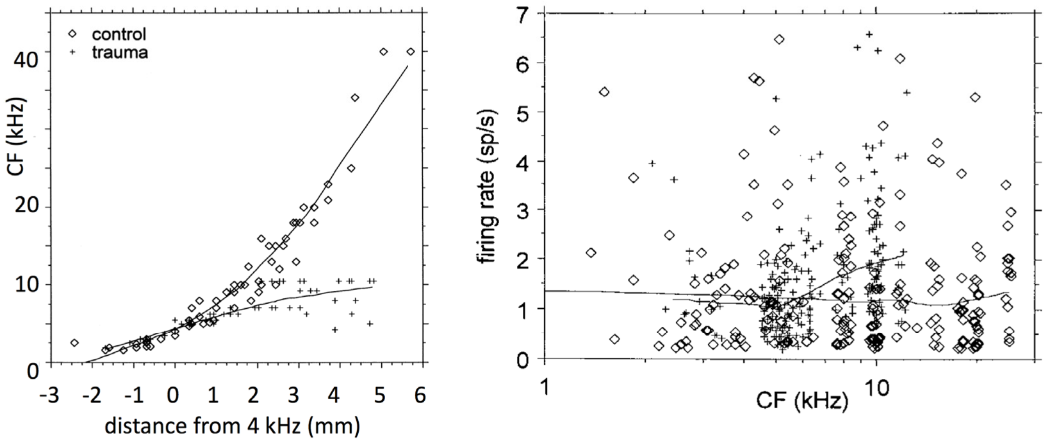


Fig. 3: Tonotopic map changes > 2 months after noise trauma (left panel). Note that in the reorganized cortex no units with CFs > 10 kHz occur, albeit that these neurons in the region with pre-trauma CFs > 10 kHz showed enhanced spontaneous activity (right panel). After Eggermont and Komiya (2000).

Makin *et al.*'s suggestion, translated to tinnitus, implies that the chronic tinnitus experience, which may be triggered either by bottom-up increased SFRs and neural synchrony or by top-down inputs from auditory-related brain areas, including limbic areas, drives plasticity because it maintains local cortical representations and disrupts inter-regional connectivity. We have seen that tinnitus does occur in the absence of tonotopic map reorganization. Local cortical representation implies a somatic memory for pitch. The missing frequencies still generate the remembered pitch as reflected in the tinnitus spectrum. This would mean that it is the continuing input to the cortex from subcortical structures that activates the auditory frequency-representation memories prior to the hearing loss, and so explains the pitch or tinnitus spectrum of tinnitus (Noreña *et al.*, 2002; Roberts *et al.*, 2008; Mulders and Robertson, 2011; Langers *et al.*, 2012). This does not violate the presence of a reorganized tonotopic map (Fig. 3), defined as the representation of CFs on the cortex, which is basically a reflection of how these neurons respond to sound just above threshold, not how their spontaneous activity is perceived. The increase in spontaneous activity in the reorganized area is referred to the reorganized CFs in

Fig. 3. The interpretation of disrupted inter-regional connectivity could then be that the connectivity between tonotopic areas and non-tonotopic areas becomes different for spontaneous activity compared to that for stimulus-induced activity.

Summarizing, homeostatic plasticity does not need to be maladaptive because the chronic tinnitus percept may either be caused by a malfunctioning gate downstream from auditory cortex, or is the result of experience-dependent plasticity with a percept engrained in memory as a result of continuous attention to it.

ACKNOWLEDGEMENTS

This work was supported by the Alberta Innovates-Health Solutions, by the Natural Sciences and Engineering Research Council of Canada, and by the Campbell McLaurin Chair for Hearing Deficiencies.

REFERENCES

Calford, M.B. (**2002**). “Dynamic representational plasticity in sensory cortex,” Neuroscience, **111**, 709-738.

Caspary, D.M., Schatteman, T.A., and Hughes, L.F. (**2005**). “Age-related changes in the inhibitory response properties of dorsal cochlear nucleus output neurons: role of inhibitory inputs,” J. Neurosci., **25**, 10952-10959.

Dauman, R., and Bouscau-Faure, F. (**2005**). “Assessment and amelioration of hyperacusis in tinnitus patients,” Acta Oto-Laryngol., **125**, 503-509.

De Ridder, D., Elgoyhen, A.B., Romo, R., and Langguth, B. (**2011**). “Phantom percepts: Tinnitus and pain as persisting aversive memory networks,” Proc. Natl. Acad. Sci. USA, **108**, 8075-8080.

Eggermont, J.J. (**1992**). “Neural interaction in cat primary auditory cortex. Dependence on recording depth, electrode separation and age,” J. Neurophysiol., **68**, 1216-1228.

Eggermont, J.J., and Komiya, H. (**2000**). “Moderate noise trauma in juvenile cats results in profound cortical topographic map changes in adulthood,” Hear. Res., **142**, 89-101.

Eggermont, J.J., and Roberts, L.E. (**2004**). “The neuroscience of tinnitus,” Trends Neurosci., **27**, 676-682.

Eggermont, J.J. (**2013**). “Hearing loss, hyperacusis, and tinnitus: what is modeled in animal research?” Hear. Res., **295**, 140-149.

Gu, J.W., Halpin, C.F., Nam, E.C., Levine, R.A., and Melcher, J.R. (**2010**). “Tinnitus, diminished sound-level tolerance, and elevated auditory activity in humans with clinically normal hearing sensitivity,” J. Neurophysiol., **104**, 3361-3370.

Jastreboff, P.J. (**1990**). “Phantom auditory perception (tinnitus): mechanisms of generation and perception,” Neurosci. Res., **8**, 228-251.

Jastreboff, P.J., and Hazell, J.W.P. (**1993**). “A neurophysiological approach to tinnitus: clinical implications,” Br. J. Audiol., **27**, 7-17.

Kaltenbach, J.A., Zhang, J., and Afman, C.E. (**2000**). “Plasticity of spontaneous neural activity in the dorsal cochlear nucleus after intense sound exposure,” Hear. Res., **147**, 282-292.

Kaltenbach, J.A., Zacharek, M.A., Zhang, J., and Frederick, S. (**2004**). "Activity in the dorsal cochlear nucleus of hamsters previously tested for tinnitus following intense tone exposure," Neurosci. Lett., **355**, 121–125.

Knipper, M., Müller, M., and Zimmermann, U., (**2012**). "Molecular mechanisms of tinnitus," in *Tinnitus, Springer Handbook of Auditory Research 47*. Edited by J.J. Eggermont, F.-G. Zeng, A.N. Popper, and R.R. Fay (Springer Science+Business Media, New York), pp. 59-82.

Kujawa, S.G., and Liberman, M.C. (**2009**). "Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss," J. Neurosci., **29**, 14077-14085.

Langers, D.M., de Kleine, E., and van Dijk, P. (**2012**). "Tinnitus does not require macroscopic tonotopic map reorganization," Front. Syst. Neurosci., **6**, 2.

Liberman, M.C., and Kiang, N.Y. (**1978**). "Acoustic trauma in cats. Cochlear pathology and auditory-nerve activity," Acta Oto-Laryngol. Suppl., **358**, 1-63.

Makin, T.R., Scholz, J., Filippini, N., Slater, D.H., Tracey, I., and Johansen-Berg, H. (**2013**). "Phantom pain is associated with preserved structure and function in the former hand area," Nat. Comm. **4**, 1570.

Milbrandt, J.C., Holder, T.M., Wilson, M.C., Salvi, R.J., and Caspary, D.M. (**2000**). "GAD levels and muscimol binding in rat inferior colliculus following acoustic trauma," Hear. Res., **147**, 251-260.

Mulders, W.H. and Robertson, D. (**2009**). "Hyperactivity in the auditory midbrain after acoustic trauma: dependence on cochlear activity." Neurosci., 164, 733-746.

Mulders, W.H., and Robertson, D. (**2011**). "Progressive centralization of midbrain hyperactivity after acoustic trauma," Neurosci., **192**, 753-760

Mulders, W.H.A.M., and Robertson, D. (**2013**). "Development of hyperactivity after acoustic trauma in the guinea pig inferior colliculus," Hear. Res., **298**, 104-108.

Munguia R, Pienkowski, M., and Eggermont, J.J. (**2013**). "Spontaneous firing rate changes in cat primary auditory cortex following long-term exposure to non traumatic noise. Tinnitus without hearing loss?" Neurosci. Lett., **546**, 46-50.

Noreña, A., Micheyl, C., Chery-Croze, S., and Collet, L. (**2002**). "Psychoacoustic characterization of the tinnitus spectrum: implications for the underlying mechanisms of tinnitus," Audiol. Neuro-Otol., **7**, 358-369.

Noreña, A.J., and Eggermont, J.J. (**2003**). "Changes in spontaneous neural activity immediately after an acoustic trauma: implications for neural correlates of tinnitus," Hear. Res., **183**, 137-153.

Noreña, A.J., and Eggermont, J.J. (**2005**). "Enriched acoustic environment after noise trauma reduces hearing loss and prevents cortical map reorganization," J. Neurosci., **25**, 699-705.

Noreña, A.J., and Eggermont, J.J. (**2006**). "Enriched acoustic environment after noise trauma abolishes neural signs of tinnitus," Neuroreport, **17**, 559-563.

Rajan, R. (**1998**). "Receptor organ damage causes loss of cortical surround inhibition without topographic map plasticity," Nat. Neurosci., **1**, 138-143.

Rauschecker, J.P., Leaver, A.M., and Mühlau, M. (**2010**). "Tuning out the noise: limbic auditory interactions in tinnitus," Neuron, **66**, 819-826.

Roberts, L.E., Moffat, G., Baumann, M., Ward, L.M., and Bosnyak, D.J. (**2008**). “Residual inhibition functions overlap tinnitus spectra and the region of auditory threshold shift,” J. Assoc. Res. Oto., **9**, 417-435.

Roberts, L.E., Eggermont, J.J., Caspary, D.M., Shore, S.E., Melcher, J.R., and Kaltenbach, J.A. (**2010**). “Ringing ears: the neuroscience of tinnitus,” J. Neurosci., **30**, 14972–14979.

Roberts, L.E., Husain, F., and Eggermont, J.J. (**2013**) “Role of attention in the generation and modulation of tinnitus,” Neurosci. Biobehav. R., **37**, 1754-1773.

Schaette, R., and Kempter, R. (**2006**). “Development of tinnitus-related neuronal hyperactivity through homeostatic plasticity after hearing loss: a computational model,” Eur. J. Neurosci., **23**, 3124-3138.

Schaette, R., and Kempter, R. (**2009**). “Predicting tinnitus pitch from patients audiograms with a computational model for the development of neuronal hyperactivity,” J. Neurophysiol. **101**, 3042-3052.

Seki, S., and Eggermont, J.J. (**2002**). “Changes in cat primary auditory cortex after minor-to-moderate pure-tone induced hearing loss,” Hear. Res., **173**, 172-186.

Seki, S. and Eggermont, J.J. (**2003**). “Changes in spontaneous firing rate and neural synchrony in cat primary auditory cortex after localized tone-induced hearing loss,” Hear. Res., **180**, 28-38.

Turner, J.G., Hughes, L.F., and Caspary, D.M. (**2005**). “Divergent response properties of layer-V neurons in rat primary auditory cortex,” Hear. Res., **202**, 129-140.

Turrigiano, G. (**1999**). “Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same,” Trends Neurosci., **22**, 221-227.

Vogler, D.P., Robertson, D., and Mulders, W.H.A.M. (**2011**). “Hyperactivity in the ventral cochlear nucleus after cochlear trauma,” J. Neurosci., **31**, 6639–6645.

Wang, H., Brozoski, T.J., Ling, L., Hughes, L.F., and Caspary, D.M. (**2011**). “Impact of sound exposure and aging on brain-derived neurotrophic factor and tyrosine kinase B receptors levels in dorsal cochlear nucleus 80 days following sound exposure,” Neurosci., **172**, 453-459.

Wang, J., Ding, D., and Salvi, R.J. (**2002**). “Functional reorganization in chinchilla inferior colliculus associated with chronic and acute cochlear damage,” Hear. Res., **168**, 238-249.

Ward, L.M. (**2011**). “The thalamic dynamic core theory of conscious experience,” Conscious. Cogn., **20**, 464-486.

Weisz, N., Müller, S., Schlee, W., Dohrmann, K., Hartmann, T., and Elbert, T. (**2007**). “The neural code of auditory phantom perception,” J. Neurosci., **27**, 1479-1484.

Yang, G., Lobarinas, E., Zhang, L., Turner, J., Stolzberg, D., Salvi, R., and Sun, W. (**2007**). “Salicylate induced tinnitus: behavioral measures and neural activity in auditory cortex of awake rats,” Hear. Res., **226**, 244-253.

Zacharek, M.A., Kaltenbach, J.A., Mathog, T.A., and Zhang, J. (**2002**) “Effects of cochlear ablation on noise induced hyperactivity in the hamster dorsal cochlear nucleus: implications for the origin of noise induced tinnitus,” Hear. Res., **172**, 137-143.

Zhang, J.S., Kaltenbach, J.A., Godfrey, D.A., and Wang, J. (**2006**). “Origin of hyperactivity in the hamster dorsal cochlear nucleus following intense sound exposure,” J. Neurosci. Res., **84**, 819-831.

Use of tinnitus masking functions to support or refute the presence or absence of auditory plasticity

JOSE LUIS BLANCO AND MICHAEL J. NILSSON*

Oticon A/S, Clinical Communication and Evidence, Smørum, Denmark

Tinnitus, the perception of sounds that do not have a peripheral correlate, is often hypothesized to be associated with cortical reorganization that over-emphasizes baseline cortical activity and is perceived as these phantom signals. But there are several issues that suggest this explanation may not be universal (if the system is plastic, why can't tinnitus be eliminated by another plastic change?). A potential technique to distinguish tinnitus that may be correlated with auditory plasticity versus tinnitus associated directly with peripheral damage will be evaluated. Narrow bands of noise will be used to determine masking thresholds across frequencies. Thresholds will be plotted relative to the tinnitus pitch to determine whether the frequency of optimal masking is aligned with the frequency of tinnitus, which does not support plasticity, or with adjacent frequencies, supporting the existence of auditory plasticity. Subjects with tinnitus frequency less than 6 kHz will be recruited, and a test battery will be collected, including DPOAE, tinnitus frequency, TEN test to detect possible dead zones, as well as masking thresholds with narrow bands of noise around the tinnitus frequency. Case studies will be presented to demonstrate the threshold functions found in a small sampling of tinnitus patients. Implications for treatment will be discussed.

INTRODUCTION

Auditory plasticity has been defined as a change to the tonotopic arrangement of the auditory cortex caused by lack of stimulation (Engineer *et al.*, 2011). Evidence for plasticity has been caused by ablation of regions of the basilar membrane in animal studies. The evidence shows re-allocation of cortical responses from ablated frequencies to adjacent frequencies that are still audible, so the cortical mapping becomes distorted with larger cortical area allocated to audible frequencies, and less or none allocated to ablated frequencies. This demonstrates how the cortex will map to the nearest frequency with some useful input, since the cells are responding to stimulation from the auditory pathways.

Tinnitus is often hypothesized to be associated with cortical reorganization, where baseline cortical activity is over-emphasized or becomes synchronized when additional cortical areas become associated with frequencies already covered in the standard tonotopic mapping (Engineer *et al.*, 2013). This hypothesis raises several questions such as:

*Corresponding author: mss@oticon.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

- Does cortical reorganization require ablation or dead zones?
- If the system is plastic, why can't tinnitus be eliminated by another plastic change?

For this hypothesis to be true at the simplest level, tinnitus should occur at adjacent frequencies compared to dead zones where the tonotopic mapping has been distorted and too many cells respond together at a baseline level. A potential technique to try and distinguish tinnitus that may be correlated with auditory plasticity using clinically available materials has begun to be evaluated to see if some answers are possible.

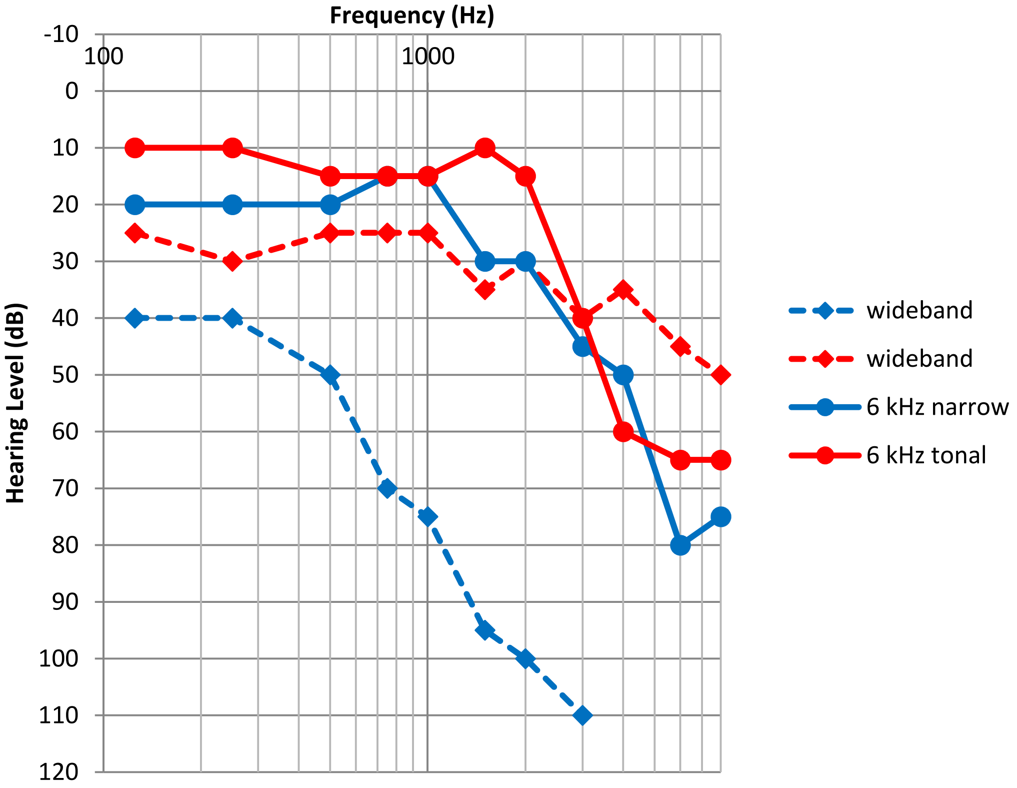


Fig. 1: Audiograms for the four subjects, with the type of tinnitus and characteristic frequency identified in the legend. The colors were chosen to identify the alternate wideband or tonal subject, and are not to identify the ear involved.

METHOD

Subjects were recruited randomly at clinics in Spain with complaints of tinnitus. Standard audiograms were measured from 125 Hz to 8000 Hz, including half octaves, and are reported in Fig. 1 for the ear with the loudest tinnitus. Tinnitus pitch and level in the ear with the loudest tinnitus were determined, as well as masking thresholds using narrowband noise and broadband noise from an audiometer. The masking threshold would help to see what frequencies have the greatest impact on the audibility of the tinnitus, and how these are related to the tinnitus itself. The test battery included distortion product otoacoustic emissions (DPOAEs) and threshold-equalizing-noise (TEN) test results to detect possible dead zones.

Subjects

Four subjects (2 male, 2 female) were recruited with clinical complaints of tinnitus accompanying sensorineural hearing loss (no conductive component). They had a mean age 53 years, with a range from 42 to 64 years of age. The subjects identified the quality of their tinnitus and, where applicable, used pitch matching to identify the nearest audiometric frequency to their tinnitus. The four subjects included two with broadband tinnitus, one with narrowband tinnitus centered at 6 kHz, and one with pure-tone tinnitus centered at 6 kHz.

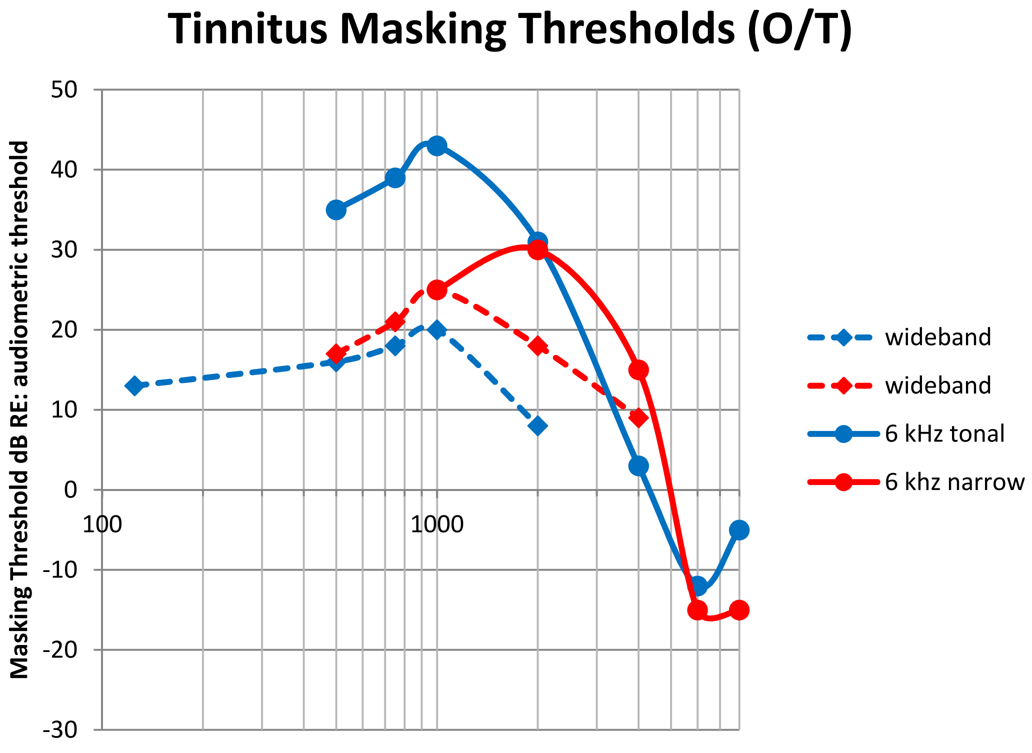


Fig. 2: Plot of Masking Thresholds relative to audiometric thresholds

All showed sloping high-frequency hearing loss of varying degrees. As a note, the tonal tinnitus occurred in the subjects with the best low-frequency hearing, when analyzing audiograms to look for any defining features. For the two subjects with any tonal quality, the tinnitus frequency appears to coincide with the frequency of maximum hearing loss.

ANALYSIS

Thresholds were plotted relative to the tinnitus pitch (Fig. 2) to determine whether the frequency of optimal masking is aligned with the frequency of tinnitus, which does not support plasticity, or with adjacent frequencies, supporting the existence of auditory plasticity.

For the wideband tinnitus, thresholds drop towards higher frequencies, but cannot be measured at the highest frequencies because of equipment limitations along with increasing hearing loss. These subjects do not meet the expectation to support or refute any auditory plasticity hypothesis, but are interesting none the less in the similarity in the function of masking thresholds relative to audiometric frequencies. It appears that masking thresholds, if audible, will fall towards 6 kHz, which would be an interesting finding.

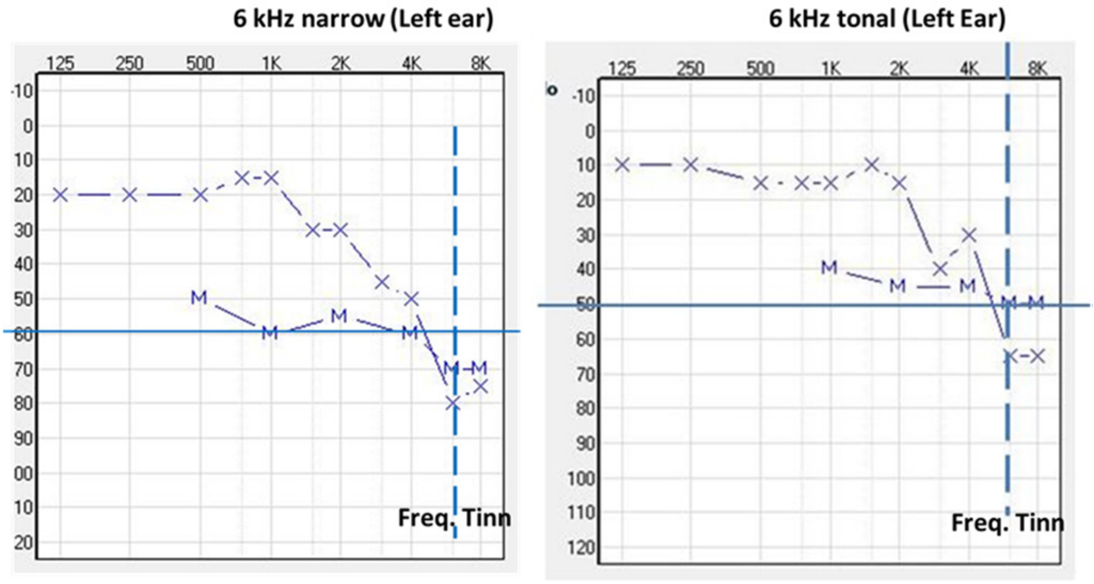
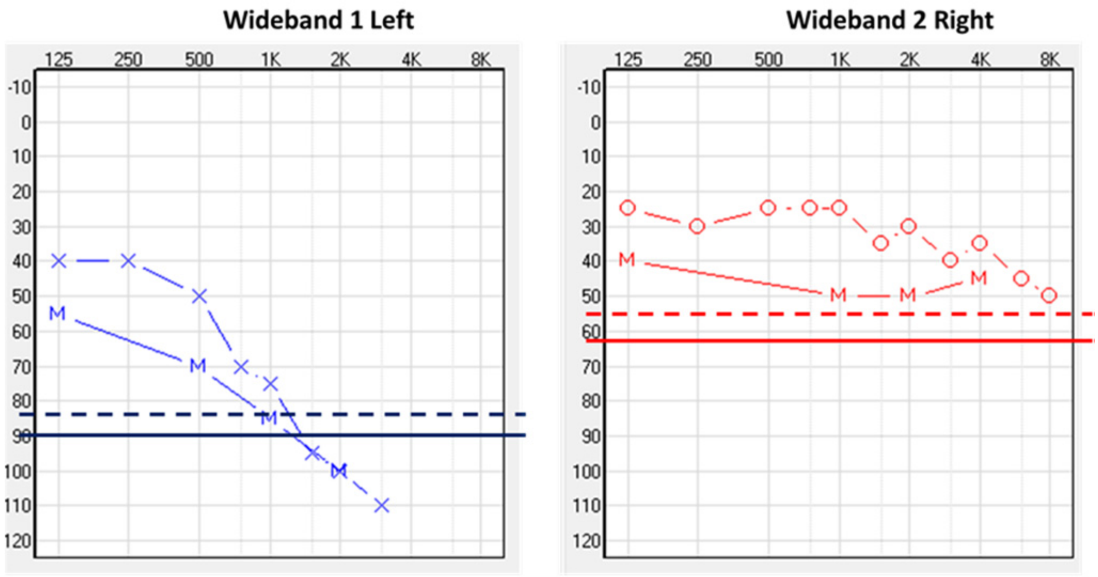


Fig. 3: Audiometric thresholds (X), masking thresholds (M), and the frequency (dashed line) and level (solid line) of the tinnitus in the two tonal subjects tested.

For these two tinnitus subjects with tonal tinnitus (Fig. 3), what is surprising is the fact that thresholds are lower than audiometric thresholds at 6 and 8 kHz. Both show masking by narrow band noise at 6 kHz and 8 kHz at lower levels than hearing thresholds at these frequencies. But this means that the masking signal should be inaudible even though they still have masking of their tinnitus. Stimulation occurring at audible adjacent frequencies from the bandwidth of the noise is not possible based on the adjacent audiometric thresholds and the skirts of the narrow band noise used, so some other explanation is needed.

Both of these subjects have positive results of distortion products in their left ear at 6 kHz, but only at high stimulus levels (70 dB and 75 dB, respectively). These masking thresholds show optimal masking at the frequency of the tinnitus, suggesting that dead zones and reallocation to adjacent frequencies is not likely. The OAEs also support some outer-hair-cell function, so the traditional plasticity model is not possible.

The two patients with broadband tinnitus were not the target for this study, but testing was completed on any subject that could be recruited. In these subjects, masking occurs only at audible levels (Fig. 4), above the audiometric thresholds, but often at a level lower than the tinnitus level. When tested with a broadband masker, broadband masking occurred at a level above the level of the tinnitus.


Wideband 1 Left

Wideband 2 Right

Fig. 4: Audiometric thresholds (X and O), masking thresholds (M), and the level of the tinnitus (dashed line) and level (solid line) of the broadband masking threshold.

Narrowband masking occurs at levels below the level of the tinnitus until the hearing loss exceeds that level and pushes the masking thresholds to very high levels. So this begs the question, which is the best masking noise for these patients? Should it be white noise, because it is similar to the tinnitus noise? Or should it be 1-kHz or 4-kHz narrowband noise, respectively, because those are the bands with the minimum difference between masking and audiometric thresholds? Or should the patient be allowed to alternate between these two (or other) noises? Tinnitus treatment methods have only proven that preferred and optimal sounds have not been shown to be predictable based on clinical measures, but a systematic approach like this highlights some interesting effects. And the results do not explain how the narrowband noise was often lower in level than the broadband noise, as well as the tinnitus, yet still masked the tinnitus. How can this be?

CONCLUSIONS

We raised many questions and did not find conclusive answers. No dead zones were identified (up to the limits of the TEN test at 4 kHz), so should we have expected any cortical reorganization to have occurred? If cortical reorganization does not require cochlear dead zones, then the auditory mapping must constantly be changing, suggesting that tinnitus should disappear as easily as it begins. But this has not been reported in the cases severe enough to seek treatment. In the current data, the tinnitus was aligned with hearing loss, but is this a resolution problem with the audiometric stimuli? It is possible that dead regions may be smaller than audiometric resolution, and therefore an adjacent frequency being associated with tinnitus is just too close to the test frequencies to measure with clinical stimuli.

There are clearly differences in the impact of various simple maskers on the perception of tinnitus, and these differences varied between subjects. The only useful conclusion is that simple clinical measurements may be useful to guide selection of sounds for tinnitus treatment (if masking thresholds are useful to improve treatment), but supporting or refuting cortical plasticity is still not possible. To that end, subject testing will continue.

REFERENCES

Engineer, N.D., Riley, J.R., Seale, J.D., Vrana, W.A., Shetake, J.A., Sudanagunta, S.P., Borland, M.S., and Kilgard, M.P. (**2011**). “Reversing pathological neural activity using targeted plasticity,” Nature, **470**, 101-104.

Engineer, N.D., Møller. Å.R., and Kilgard, M.P. (**2013**). “Directing neural plasticity to understand and treat tinnitus,” Hear. Res., **295**, 58-66.

Experience-related changes in the adult auditory system

KEVIN J. MUNRO[1,2,*] , PIERS DAWES[1], AND MICHAEL MASLIN[1]

[1] *School of Psychological Sciences, University of Manchester, Manchester M13 9PL, UK*

[2] *Central Manchester University Hospitals NHS Foundation Trust Academic Health Science Centre, Manchester M13 9WL, UK*

Changes in the auditory environment, as a result of deprivation or stimulation, modify our sensory experience and may result in experience-related or learning-induced reorganisation within the central nervous system. Electrophysiological and imaging techniques have revealed reorganisation of the adult human auditory map, for example, after sudden unilateral hearing loss. In parallel to these studies, there is behavioural evidence that auditory function can be modified by changing the acoustic environment; for example, experience with amplification may have consequences for long-term performance. Future studies could usefully unite these behavioural and advanced objective techniques in order to provide a direct link between changes in perception and reorganisation of the auditory system. In this paper, we summarise our work investigating changes in perceptual and physiological measures, in adult humans, after the sensory environment has been modified by: (i) amplification, (ii) short-term sound treatment, and (iii) unilateral deafness. The findings are consistent with the growing body of literature that shows that the mature central auditory system is malleable and is modified by experience.

INTRODUCTION

Changes in the sensory environment, as a result of deprivation or stimulation, modify our sensory experience and may result in experience-related or learning-induced reorganisation within the central nervous system. Probably the most spectacular (and most commonly cited) example of injury-induced reorganisation is ‘phantom limbs’, a term coined by Silas Weir Mitchell to describe the sensation that an amputated limb is still attached to the body and moving appropriately. In 1871, Weir Mitchell described an amputee with a phantom limb as follows: *“A person in this condition is haunted… by a phantom of himself… an unseen ghost of the lost part.”* Ramachandran *et al.* (1992) suggested that phantom limb sensations could be due to reorganization in the somatosensory cortex. Yang *et al.* (1994) were the first to demonstrate direct evidence of cortical reorganisation and its perceptual correlate: stimulation of the face and hand resulted in cortical activity in the area vacated by the amputated hand and this was perceived by the amputee as stimulation of the phantom limb.

*Corresponding author: kevin.munro@manchester.ac.uk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Electrophysiological and imaging techniques have also revealed plasticity in sensory systems including the adult human auditory system. In parallel to these studies, there is behavioural evidence that auditory function can be modified by changing the acoustic environment; for example, experience with amplification may have consequences for long-term performance. In this paper, we summarise our work investigating changes in perceptual and physiological measures, in adult humans, after the sensory environment has been modified by: (i) experience with amplification, (ii) short-term use of sound treatments (earplug or hearing aid), and (iii) sudden and severe unilateral deafness. The findings are consistent with the growing body of literature showing that the mature central auditory system is malleable and is modified by experience. An understanding of the underlying mechanisms associated with experience-related changes in the normal and impaired auditory system is a pre-requisite to the development of more effective treatments for hearing and hearing-related disorders.

CHANGES INDUCED BY AMPLIFICATION

Deprivation and acclimatization refer to the concept that the ability to use auditory information may be affected by listening experience: Deprivation implies that the absence of experience leads to a decline in ability whereas acclimatization implies that auditory experience leads to an improvement in auditory ability (Arlinger *et al.,* 1996). Hearing aids change the sensory environment by stimulating a deprived auditory system, and therefore may induce changes within the auditory system. The earliest studies that investigated improvements in performance following hearing-aid use were motivated by the clinical need to know when best to measure hearing-aid benefit (i.e., at the time of fitting or after a period of hearing-aid use). More recent studies have been motivated by a desire to understand the dynamic nature of the mature auditory system.

Most acclimatization research has focused on loudness perception and speech recognition in noise (see reviews by Munro, 2008; Palmer *et al.,* 1998; Turner *et al.,* 1996). Gatehouse (1989) reported that speech recognition in the fitted ear of unilateral hearing-aid users was better than in the non-fitted ear for high presentation levels, while recognition was worse in the fitted ear than the non-fitted ear for low presentation levels. Gatehouse concluded that acclimatization involved adjustment to a dynamic range consistent with the gain provided by hearing aids.

Gatehouse (1992) then tested speech recognition in four new hearing-aid users over the first few months of hearing-aid use. Various listening conditions were simulated over headphones. Gatehouse reported improvements in speech recognition in listening conditions that matched the pattern of amplification provided by the hearing aid, but not in conditions with an unfamiliar pattern of amplification or in unaided listening conditions.

Munro and Lutman (2003) also tested speech recognition in the first months following hearing-aid fitting. They reported greater improvements in aided speech recognition for higher intensity speech stimuli over lower ones. In line with

Gatehouse's (1989) suggestion that changes relate to particular listening conditions, Munro and Lutman (2003) concluded that acclimatization effects occur specifically for those aspects of the stimulus that have been altered by the hearing aid and which have not been usually experienced in daily life prior to amplification. In parallel with behavioural studies, other studies have shown evidence of asymmetric ABR and cortical ERPs in experienced unilateral hearing-aid users (Gatehouse and Robinson, 1996; Munro *et al.*, 2007; Bertoli *et al.*, 2011).

Evidence for acclimatization to hearing aids is inconsistent, however. One review concluded that acclimatization effects – if they do exist – were likely to be insignificantly small and not of clinical relevance (Turner and Bentler, 1998), while others contend that acclimatization effects do have clinical relevance (Palmer *et al.*, 1998). Various aspects of experimental design may explain the inconsistency in research findings, such as inclusion of participants with previous hearing-aid experience (e.g., Bentler *et al.*, 1993) or participants with a heterogeneous mix of signal processing or fitting schemes (e.g., Saunders and Ceinkowski, 1997) or relatively mild levels of hearing loss (Palmer *et al.*, 1998), use of a control condition rather than a separate control group (e.g., Gatehouse, 1992; Munro and Lutman, 2003) or lack of control group (e.g., Reber and Kompis, 2005), and delays in the initial testing following fitting (e.g., Taylor, 1993).

We recently completed a longitudinal study of new hearing-aid users followed over the first months of hearing-aid use. New hearing-aid users had hearing loss of at least 1-year duration and symmetrical losses of at least 40 dB HL at 2 kHz and above and no previous experience with hearing-aid use. All new users were fitted with hearing aids with identical signal processing and fit to the same fitting formula. Accuracy and stability of fit over the study period was confirmed with real ear measures of gain. Initial testing occurred within 7 days of first fitting with retesting after 12 weeks hearing-aid use. A control group of experienced hearing-aid users was tested over the same timescale as new hearing-aid users. Tests included speech recognition in noise, spatial release from masking, auditory brainstem response (ABR), cortical auditory evoked potential (CAEP), and questionnaire measures of real life benefit. On average, new hearing-aid users showed no statistically significant changes in aided speech recognition (Dawes *et al.*, in press) or spatial release from masking (Dawes *et al.*, 2013a) over the first 12 weeks of hearing-aid use (compared to the control group). There were also no changes in ABR (Dawes *et al.*, 2013b) or cortical responses (Dawes *et al.*, submitted). New hearing-aid users did however report significant improvements in aided listening on a questionnaire measure, while no such improvements were reported by the control group (Fig. 1). This may relate to an aspect of adjustment to hearing aids not measured in this study, such as greater confidence or familiarity with hearing aids.

One possible explanation for the lack of effects in our recent acclimatization studies compared to earlier ones such as Gatehouse's is that earlier studies utilized linear-gain hearing aids while our recent studies used non-linear amplification. Non-linear hearing aids provide less amplification for higher intensity inputs than linear hearing aids. Acclimatization effects may be less robust for non-linear amplification. Our

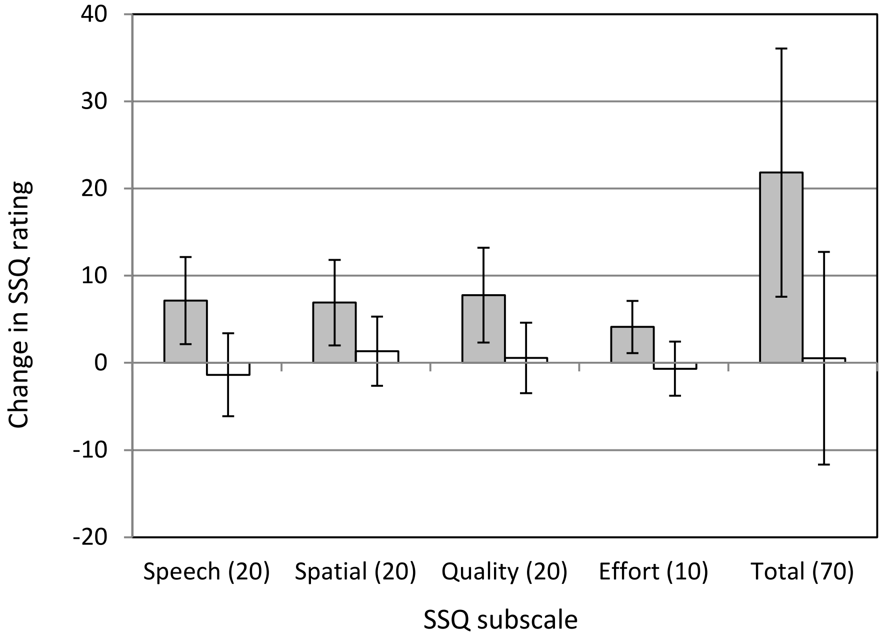


Fig. 1: Changes in self-rated hearing-aid performance over 12 weeks from first fitting, based on the Spatial, Speech and Qualities of Hearing Questionnaire (Gatehouse and Noble, 2004). Experienced hearing-aid users (control group), open columns; New hearing-aid users, filled columns. Positive values represent improvement and values in brackets display the maximum score for each subscale. Error bars show ±1 standard deviation.

study was statistically powered to detect changes of the size reported by previous acclimatization studies. However, as with previous studies, there was wide variability in outcome between participants, and this may obscure small average acclimatization effects. Our conclusion was that if they do exist, acclimatization effects with non-linear hearing aids are probably too small to be of clinical relevance (at least for older adult first-time hearing-aid users and for the outcome measures used in our studies). Despite our null findings with non-linear hearing aids and despite the controversy in the literature concerning the rate, extent and clinical significance of the acclimatization effect, there remains some evidence that a deprived auditory system may be modified by experience with hearing-aid use.

CHANGES INDUCED BY SHORT-TERM SOUND TREATMENTS

The previous section described studies involving adults with age-related hearing loss whereas this section involves studies using normal-hearing participants. The participants were provided with a short-term monaural sound treatment: either an earplug or a low-gain hearing aid. The measures used in the studies include the middle-ear muscle reflex and categorical loudness ratings. Measurements were made at baseline and within 1-2 weeks of commencing the treatment.

The middle-ear reflex is a brainstem reflex that involves bilateral contraction of the middle-ear muscles in response to a high sound level presented to either ear (Borg, 1973). In order to measure the acoustic reflex threshold (ART), short sound stimuli were initially presented below threshold and increased in intensity until there was a repeatable decrease in compliance ≥ 0.02 cm^3.

Loudness judgements were obtained using the Contour Test of Loudness Perception (Cox *et al.*, 1997). Listeners used a response pad to assign one of seven loudness categories to a train of tones. The exact details varied between studies but generally involved initially presenting tones close to hearing threshold. After the listener allocated a loudness category to the stimulus, the presentation level was increased in 5-dB steps and the process repeated until a response was recorded at the highest category, i.e., uncomfortably loud.

In our first study (Munro and Blount, 2009) 11 normal-hearing listeners were asked to use a monaural earplug continuously for 7 days. When hearing levels were measured with the earplug inserted, thresholds showed a mean increase of 22 dB at 0.25 kHz and 46 dB at 8 kHz. After seven days of earplug use, the level of a 2-kHz and 4-kHz tone required to elicit the acoustic reflex in the ear with the earplug had decreased by 5-7 dB, relative to pre-earplug levels. Measurements made 7 days after removing the earplug showed that the ART had returned to baseline values.

In our next study (Maslin *et al.*, 2013a), a new group of 11 normal-hearing listeners wore a monaural earplug continuously for seven days. The mean attenuation of the earplug, measured using real-ear insertion gain (REIG), i.e., the difference in response between the plugged and unplugged conditions, was < 10 dB at 0.25 kHz to > 30 dB at 3 and 4 kHz. Whereas Munro and Blount tested acoustic reflexes with two high-frequency stimuli an octave apart, in this study reflexes were tested with a high (4 kHz) and a low (0.5 kHz) frequency pure tone to elicit the reflex. The hypothesis was that a greater decrease in the ART should be observed for higher frequency stimuli because ear plugging provided greater attenuation of input for high frequencies. We found that the level required to elicit an acoustic reflex in the treatment ear decreased by 3 dB at 0.5 kHz and by 7 dB at 4 kHz but the difference between frequencies was not statistically significant.

Munro and Blount (2009) and Maslin *et al.* (2013a) only measured acoustic reflexes so it is unknown if there is a relationship between any changes in perceived loudness and changes in ART. Our most recent study (Munro et al., submitted) addressed this issue. We provided 18 normal-hearing participants with a monaural earplug for 7 days. ARTs were measured with a high (2 kHz) and low (0.5 kHz) frequency tone and with broadband noise. Categorical loudness ratings were obtained at 0.5 kHz and 2 kHz. All measurements were made at baseline and after 7 days use. Further measurements were taken 1 and 7 days after removal of the earplug in order to characterise the time course of recovery. After 7 days of unilateral auditory deprivation, acoustic reflexes were obtained at a lower sound pressure level in the ear that had been fitted with an earplug and at a higher sound pressure level in the not-fitted control ear. In contrast, stimuli were reported as louder after earplug experience in both ears. The relationship between changes to the ART and changes in loudness was not statistically significant. For both ARTs and loudness, changes had essentially disappeared within 24 hours of earplug removal and this is consistent with homeostatic plasticity (see later).

In our final study, Munro and Merrett (2013) provided 21 normal-hearing listeners with a monaural hearing aid that provided a REIG of 20 dB at 2-4 kHz. ARTs were measured with a 2-kHz and 0.5-kHz pure tone and with broadband noise. After five days of hearing aid use, ARTs were elicited with a higher sound pressure level of 3-4 dB at both 0.5 and 2 kHz, relative to the pre-treatment baseline. The changes occurred in the opposite direction to those reported after sensory deprivation, and this is consistent with experience-driven auditory plasticity. On the categorical loudness task, stimuli were reported as less loud after hearing-aid use but the relationship with changes to the ART was not statistically significant.

Our studies investigating short-term sound treatments provide evidence of plasticity in the adult human auditory system. This plasticity may be explained by a gain control mechanism mediated by a process operating at the level of the brainstem, although this could be controlled from higher levels. A potential function of this gain control mechanism could be to counteract changes in input in order to stabilize the overall level of neuronal activity in the central auditory system. This would require an increase in gain after deprivation and a decrease in gain after additional stimulation, as observed by the changes in sound level required to elicit an acoustic reflex. The lack of relationship between changes in ARTs and loudness, and the different pattern of findings with each measure, suggests multiple gain mechanisms.

The mechanism underlying the changes in ARTs is unknown, but a reasonable candidate is homeostatic plasticity which is thought to stabilize the mean activity of the neuron (Turrigiano, 1999). In response to sensory deprivation, the strength of excitatory synapses is scaled up and the strength of inhibitory synapses is scaled down, resulting in increased neural response gain, which could lead to lower ART thresholds. Conversely, in response to sensory stimulation, the strength of the excitatory synapses is scaled down and the strength of the inhibitory synapses is scaled up, resulting in decreased neural gain, possibly increasing ARTs.

The findings of these studies may have implications for some patients with tinnitus and/or sound tolerance problems. Computational models have illustrated how auditory deprivation may result in an increase in neural gain as homeostatic plasticity attempts to restore average neuronal activity (Schaette *et al.*, 2012). Tinnitus, the perception of a sound in the absence of a corresponding sound source, may be a side-effect of 'over-amplification' of spontaneous neural activity due to increased neuronal gain. Likewise, increased gain could cause an 'over-amplification' of stimulus-evoked neural activity, leading to sound tolerance problems. This would support the use of sound treatments to 'reset' gain.

CHANGES INDUCED BY PROFOUND UNILATERAL DEAFNESS

The studies in the previous sections describe the effect of environmental modification of auditory input via hearing aids or earplug manipulations. The current section refers to the effect of profound unilateral deafness on auditory processing in adult humans. The studies all have the same basic design: The pattern of auditory activity to stimulation of the intact ear is compared (i) before and after the onset of unilateral deafness, or (ii) with that of control participants receiving monaural stimulation. The outcome measures were CAEPs and ABRs.

Normally, stimulation of one ear produces a bilateral but asymmetrical activation within the central auditory system. This is because the ascending contralateral pathway contains more nerve fibres and fewer synapses. However, after unilateral deafness the hemispheric asymmetry disappears as the nerve fibres previously innervated by the deafened ear adapt to become more sensitive to the remaining, intact ear. What is puzzling is that, while this change has consistently been shown in animal models (for review see Moore and King, 2004), the evidence in humans has been less consistent. Studies using fMRI have demonstrated reduced hemispheric asymmetries in humans (e.g., Scheffler *et al.*, 1998; Langers *et al.*, 2005), although those using CAEPs (or CAEFs) have not (e.g., Vasama *et al.*, 2001; Hine *et al.*, 2008). This led us to suspect that there was some aspect of CAEP/Fs methodology, e.g., calculation of hemispheric asymmetries, that could be causing the inconsistency. Our first study compared CAEPs from 18 individuals with unilateral deafness to 18 controls (Maslin *et al.*, 2013b). We focused on the N1 response and measured the asymmetry using Dipole Source Analysis. We controlled for more variables than previous studies. The results revealed an overall increase in the amplitude of N1, and a reduction in the normally observed hemispheric asymmetry.

Individuals due to undergo translabyrinthine surgery (for removal of a unilateral acoustic neuroma) provide an opportunity to study the time course of injury-induced plasticity. Baseline readings from the intact ear can be obtained in advance of the surgery-induced profound unilateral deafness. We have monitored the time course of changes in N1 (and P1 and P2) in five adults from baseline pre-surgery through to 36 months post-surgery (Maslin *et al.*, 2013c; Maslin *et al.,* in prep.). The results showed that even at baseline some changes had already taken place in comparison to the control group, presumably because of some hearing loss in the tumour ear. However, a series of further changes in all three cortical components could be observed post-surgery. The P1 was significantly different to baseline at 1 month post-surgery, whereas changes in N1 and P2 did not reach statistical significance until 6 months post-surgery. Recent data at 36 months post-surgery do not appear to show any further significant changes. The time-course of changes after surgery suggests a range of physiological mechanisms: Some are relatively fast acting (at least within 1 month), and others are more gradual (6 months). Candidate mechanisms include functional disinhibition (i.e., removal of inhibitory input normally acting on the intact ear) and up-regulation of existing synapses and proliferation of new synapses favouring the input from the intact ear.

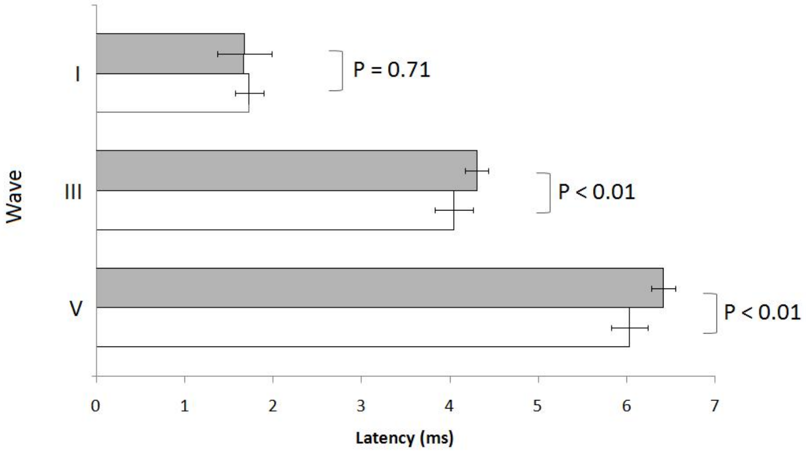


Fig. 2: Mean latencies of waves I, III, and V of the ABR from 7 individuals pre- (filled) and post-labyrinthectomy (open). Error bars show ±1 standard deviation.

Our most recent study has focussed on identifying when the surgery-induced changes can be first measured (Maslin *et al.,* in prep.). Very rapid changes may occur if the physiological mechanism is disinhibition or rapid intra-cellular signalling. We have been conducting ABR measures during surgery. ABRs are unaffected by anaesthesia, and have the added bonus of providing sub-cortical information. The results showed a rapid (within minutes) reduction in ABR latencies for waves III and V post-labyrinthectomy. So far, we have measured responses from seven individuals (see Fig. 2) and are in process of completing testing on control subjects undergoing non-auditory neurosurgery.

Further work is needed to elucidate the perceptual consequences of the physiological changes such as improvements in localisation (Slattery and Middlebrooks, 1994). It is also possible that the physiological changes result in maladaptive changes including tinnitus and hyperacusis. Hence it may be clinically relevant to understand, and potentially manipulate, injury-induced plasticity for therapeutic gain.

CONCLUSIONS

Despite the controversy in the literature concerning the rate, extent, and clinical significance of the acclimatization effect, there is evidence that the deprived auditory system of some listeners can be modified with hearing-aid experience. The findings from our studies involving short-term monaural sound treatments provide evidence of plasticity in the adult human auditory system and are consistent with a neural gain control mechanism. These studies, along with the more extreme example of profound unilateral deafness, may shed light on the underlying mechanisms causing aberrant auditory perceptions such as tinnitus and hyperacusis, as well as the capacity of the adult auditory system to recover function. Our current studies aim to identify the potential benefits of plasticity.

ACKNOWLEDGEMENTS

Many of the hearing-aid acclimatisation studies were funded by the Starkey Hearing Research Centre. The work was facilitated by the Manchester Biomedical Research Centre, and a Studentship from the Medical Research Council.

REFERENCES

Arlinger, S., Gatehouse, S., Bentler, R.A., Byrne, D., Cox, R.M., Dirks, D., Humes, L.E., Neuman, A., Ponton, C., Robinson, K., Silman, S., Summerfield, A.Q., Turner, C.W., Tyler, R.S., and Willott, J F. (**1996**). “Report of the Eriksholm workshop on auditory deprivation and acclimatization,” Ear Hearing, **17**, 87S-90S.

Bentler, R.A., Neibuhr, D.P., and Getta, J.P. (**1993**). “Longitudinal study of hearing aid effectiveness. I Objective measures,” J. Speech Lang. Hear. Res., **36**, 808-819.

Bertoli, S., Probst, R., and Bodmer, D. (**2011**). “Late auditory evoked potentials in elderly long-term hearing-aid users with unilateral or bilateral fittings,” Hear. Res., **280**, 58-69.

Borg, E. (**1973).** “On the neural organisation of the acoustic middle ear reflex,” Brain Res., **49**, 101-123.

Cox, R.M., Alexander, G.C., Taylor, I.M., and Gray, C.A. (**1997**). “The contour test of loudness perception,” Ear. Hearing, **18**, 338-400.

Dawes, P., Munro, K.J., Kalluri, S., and Edwards, B. (**2013a**). “Unilateral and bilateral hearing aids, spatial release from masking and auditory acclimatization,” J. Acoust. Soc. Am., **134**, 596-606.

Dawes, P., Munro, K.J., Kalluri, S., and Edwards, B. (**2013b**). “Brainstem processing following unilateral and bilateral hearing-aid amplification,” NeuroReport, **24**, 271-275.

Dawes, P., Munro, K.J., Kalluri, S., and Edwards, B. (**submitted**). “Hearing aid use-related auditory acclimatization: Late auditory evoked potentials and speech recognition following unilateral and bilateral hearing-aid amplification”, J. Assoc. Res. Oto.

Dawes, P., Munro, K.J., Kalluri, S., and Edwards, B. (**in press**). “Acclimatization to hearing aids,” Ear Hearing.

Gatehouse, S. (**1989**). “Apparent auditory deprivation effects of late onset: The role of presentation level,” J. Acoust. Soc. Am., **86**, 2103-2106.

Gatehouse, S. (**1992**). “The time course and magnitude of perceptual acclimatization to frequency responses: Evidence from monaural fitting of hearing aids,” J. Acoust. Soc. Am., **92**, 1258-1268.

Gatehouse, S., and Noble, W. (**2004**). “The Speech, Spatial and Qualities of Hearing Scale (SSQ),” Int. J. Audiol., **43**, 85-99.

Gatehouse, S., and Robinson, K. (**1996**). “Acclimatization to monaural hearing aid fitting – effects on loudness functions and preliminary evidence for parallel electrophysiological and behavioural effects,” in *Psychoacoustics, Speech and Hearing Aids*. Edited by B. Kollmeier (World Scientific, Singapore), pp. 319-330.

Hine, J., Thornton, R., Davis, A., and Debener, S. (**2008**). “Does long-term unilateral deafness change auditory evoked potential asymmetries?,” Clin. Neurophysiol., **119**, 576-586.

Langers, D.R., van Dijk, P., and Backes, W.H. (**2005**). “Lateralization, connectivity and plasticity in the human central auditory system,” Neuroimage, **28**, 490-499.

Maslin, M.R.D., Munro, K.J., Lim, V.K., Purdy, S.C., and Hall, A.D. **(2013a)**. “Investigation of cortical and sub-cortical plasticity following short-term unilateral auditory deprivation in normal hearing adults,” NeuroReport, **24**, 287-291.

Maslin, M.R., Munro, K.J., and El-Deredy, W. (**2013b**). “Source analysis reveals plasticity in the auditory cortex: evidence for reduced hemispheric asymmetries following unilateral deafness,” Clin. Neurophysiol., **124**, 391-399.

Maslin, M.R., Munro, K.J., and El-Deredy, W. (**2013c**). “Evidence for multiple mechanisms of cortical plasticity: A study of humans with late-onset profound unilateral deafness,” Clin. Neurophysiol., **124**, 1414-1421.

Moore, D.R., and King, A.J. (**2004**). “Plasticity of binaural systems,” in *Springer Handbook of Auditory Research*. Edited by T.N. Parks, E.W. Rubel, R.R. Fay, and A.N. Popper (Springer-Verlag, New York), pp. 96-172.

Munro, K.J., and Lutman, M.E. (**2003**). “The effect of speech presentation level on measurement of auditory acclimatization to amplified speech,” J. Acoust. Soc. Am., **114**, 484-495.

Munro, K.J., Pisareva, N.Y., Parker, D.J., and Purdy, S.C. (**2007**). “Asymmetry in the auditory brainstem response following experience of monaural amplification,” NeuroReport, **18**, 1871-1874.

Munro, K.J. (**2008**). “Reorganization of the adult auditory system: Perceptual and physiological evidence from monaural fitting of hearing aids,” Trends Ampl., **12**, 254-271.

Munro, K.J., and Blount, J. **(2009)**. “Adaptive plasticity in brainstem of adult listeners following earplug-induced deprivation (L),” J. Acoust. Soc. Am., **126**, 568-571.

Munro, K.J., and Merrett, J.F. **(2013)**. “Brainstem plasticity and modified loudness following short-term use of hearing aids,” J. Acoust. Soc. Am., **133**, 343-349.

Munro, K.J., Turtle, C., and Schaette, R. (**submitted**). “Sub-cortical plasticity and modified loudness following short-term unilateral deprivation: evidence of multiple neural gain mechanisms within the auditory system,” J. Acoust. Soc. Am.

Palmer, C.V., Nelson, C.T., and Lindley, G.A. (**1998**). “The functionally and physiologically plastic adult auditory system,” J. Acoust. Soc. Am. **103**, 1705-1721.

Ramachandran, V.S., Stewart, M., and Rogers-Ramachandran, D.S. (**1992**). “Perceptual correlates of massive cortical reorganization,” Neuroreport, **3**, 583-586.

Reber, M.B., and Komopis, M. (**2005**). “Acclimatization in first-time hearing aid users using three different fitting protocols,” Auris Nasis Larynx, **32**, 345-351.

Saunders, G.H., and Cienkowski, K.M. (**1997**). “Acclimatization to hearing aids,” Ear Hearing, **18**, 129-139.

Schaette, R., Turtle, C., and Munro K.J. (**2012)**. “Reversible induction of phantom auditory sensations through simulated unilateral hearing loss,” PLos ONE, **7**, e35238.

Scheffler, K., Bilecen, D., Schmid, N., Tschopp, K., and Seelig, J. (**1998**). “Auditory cortical responses in hearing subjects and unilateral deaf patients as detected by functional magnetic resonance imaging,” Cereb. Cortex, **8**, 156-163.

Slattery, W.H., 3rd, and Middlebrooks, J.C. (**1994**). “Monaural sound localization: acute versus chronic unilateral impairment," Hear. Res., **75**, 38-46.

Taylor, K.S. (**1993**). “Self-perceived and audiometric evaluations of hearing aid benefit in the elderly,” Ear Hearing, **14**, 390-394.

Turner, C.W., Humes, L.E., Bentler, R.A., and Cox, R.M. (**1996**). “A review of past research on changes in hearing aid benefit over time,” Ear Hearing, **17**, 14S-25S.

Turner, C.W., and Bentler, R.A. (**1998**). “Does hearing aid benefit increase over time?” J. Acoust. Soc. Am., **104**, 3673-3674.

Turrigano, G.G. (**1999**). “Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same,” Trends Neurosci., **22**, 221-227.

Vasama, J.P., Marttila, T., Lahin, T., and Makela, J.P. (**2001**). “Auditory pathway function after vestibular schwannoma surgery,” Acta Oto-Laryngologica, **121**, 378-383.

Weir Mitchell, S., (**1871**). “Phantom Limbs,” Lippincott’s Mag., **8**, 563-569.

Yang, T.T., Gallen, C., Schwartz, B., Bloom, F.E., Ramachandran, V.S., and Cobb, S. (**1994**), “Sensory maps in the human brain,” Nature, **368**, 592-593.

Unilateral conductive hearing loss causes impaired auditory information processing in neurons in the central auditory system

JENNIFER L. THORNTON, KELSEY L. ANBUHL, AND DANIEL J. TOLLIN*

University of Colorado School of Medicine, Department of Physiology, Aurora, CO, USA

Temporary conductive hearing loss (CHL) during development and in adults can lead to hearing impairments that persist beyond the CHL. Despite decades of studies, there is little consensus on the mechanisms responsible. Here we introduced 6 weeks of unilateral CHL to adult chinchillas via a foam earplug. Single-unit recordings from inferior colliculus (IC) neurons indicated that the CHL caused a decrease in the efficacy of inhibitory input to the IC contralateral to the earplug and an increase of inhibitory input ipsilateral to the earplug. The changes were seen after removal of the CHL. Sensitivity to interaural-level-difference (ILD) cues to location in IC neurons was shifted by ~10 dB relative to controls. In both ICs, the direction of the shift was consistent with a compensation of the altered ILDs due to the CHL. IC neurons responses carried ~33% less information (mutual information) about ILDs after CHL than normals. Experiments examining cochlear anatomy and peripheral evoked responses confirmed that the results did not arise from damage to the periphery. The CHL-induced shifts of ILD sensitivity suggest a compensatory form of plasticity occurring by at least the level of the IC. The neurons were also impaired in their abilities to encode information about the spatial attributes of sound. How these physiological changes may lead to impaired hearing will be discussed.

INTRODUCTION

Conductive hearing loss (CHL) during development can change auditory system structure and function (see reviews by Moore and King, 2004; Tollin, 2010; Whitton and Polley, 2011). Early life exposure to CHL, particularly unilateral, can lead to impairments in binaural hearing even after resolution of the CHL and hearing sensitivity in both ears returns to normal. The persistently-impaired binaural hearing often recovers, but this can take months or years. During recovery, a child may present as audiologically normal, yet speech perception in noisy, reverberant environments may continue to be compromised. As language is often learned in such environments, these impairments may contribute to deficits in language acquisition. Decades of studies of the neural, anatomical and behavioral consequences of experimentally-induced CHL in animal models have revealed effects related to the timing of onset, the duration, and the severity of the deprivation. Regarding neural

*Corresponding author: daniel.tollin@ucdenver.edu

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

processing, these studies have generally demonstrated how only the most basic of neural response properties are altered by early CHL. Yet the persistent binaural behavioral deficits in humans have generally defied explanation based simply on these basic response properties. Currently it is not known whether or how CHL alters the *neural information carrying* capabilities of the auditory system. Towards this goal, we use the novel framework of information theory (Dayan and Abbott, 2001) to investigate how CHL alters information processing in the central nucleus of the inferior colliculus (ICC). Similar persistent impairments in binaural hearing have been reported in human adults that had experienced chronic CHL. Thus, to begin this new line of inquiry, we examine how neural information processing is altered when a CHL is induced in *adult* animals.

METHODS

Eleven young (~P70) adult chinchillas were used for the deprivation experiments while 19 normal-hearing animals were used for control data. Following the method of Lupo *et al.* (2011) a small foam earplug (AO Safety, Indianapolis, IN, USA) was cut to fit snugly into the external ear canal of the animal and was then inserted into the left ear canal for 6 weeks.

Cochlear microphonic (CM) recordings

Animals were anesthetized and prepared for electrophysiology as described by Jones *et al.* (2011). Briefly, a hole (2-3 mm diameter) was made in each bulla through which electrodes were placed on the round windows and fixed in place with dental acrylic, resealing the bullae. The CM was differentially amplified, filtered, and verified by oscilloscope. To quantify the magnitude of the CHL due to the earplug, free-field CM (and compound action potential, CAP) measurements were taken for both the left (plugged) and right (normal) ears for two different conditions: with the earplug in place (left ear) and after the earplug was removed. Stimuli consisted of 10-ms sinusoids (2.5-ms rise/fall, 5-ms plateau) with octave steps from 0.25-20 kHz. Each stimulus was presented at least 25 times with a 40-ms interstimulus period.

Electrophysiological methods

Single unit, extracellular responses were recorded from neurons in the ICC. All recordings in the group of animals with CHL were performed the same day as the earplug removal. Frequency-intensity response areas were measured with tone pips to estimate the characteristic frequency (CF) and threshold. Neuronal ILD sensitivity was examined using 50 repetitions of 50-ms duration CF tones by holding the signal level to the contralateral ear (~20 dB re: threshold) constant and varying the level in 5-dB steps at the ipsilateral ear from at least 25 dB below to 25 dB above ipsilateral threshold. The rate vs. ILD for each neuron was fitted with a 4-parameter sigmoid, $\text{rate(ILD)} = y_0 + \alpha/(1\text{-exp}(-(\text{ILD-ILD}_0)/\beta))$; before fitting, the data were normalized to the maximum rate. The fits described the data for all neurons ($R > 0.9$). The fit parameters were used for analysis; half-max ILD is the ILD at 50% of the maximal rate, rate-ILD slope (spikes/s/dB, not normalized) was computed at half-max ILD, and ILD dynamic range was defined between 90-10% of max rate.

Neural information analysis – Mutual information computation

The mutual information (MI) is a measure of the strength of the association between two random variables, such as a spike count, *r*, and a given stimulus, *S* (Dayan and Abbott, 2001). MI is given by

$$MI(r,S)=\sum_i\sum_j p(S_j)p\langle r_i\,|\,S_j\rangle\log_2\left[\frac{p\langle r_i\,|\,S_j\rangle}{p(r_i)}\right] \qquad \text{(Eq. 1)}$$

where $p(S_j)$ is the probability that the stimulus (S) had a particular value [S values (i.e., ILDs) were presented with equal probability], $p(r_i)$ is the probability that the count was r_i at any value of S, and $p(r_i|S_j)$ is the probability that the count was r_i when the stimulus was S_j. Intuitively, MI will be high when the count variability is larger when computed across different stimuli than the variability computed within single presentations of a particular stimulus. The MI represents the upper bound on the information that even the best 'decoder' could represent. Thus, if CHL changes the information carrying capacity then the MI will capture and quantify it.

RESULTS

Conductive hearing loss due to earplug does not alter periphery

To quantify the CHL caused by the earplug, as well as assay the function of the peripheral auditory system, sound-evoked CM responses were measured while the earplug was still in place and also immediately after earplug removal (see Lupo *et al.*, 2011 and Thornton *et al.*, 2012; 2013 for detailed methods). CM data from the right (unplugged) ear was used as a control; unilateral CHL does not cause residual deficits in the normal-hearing ear (Larsen *et al.*, 2010) and the present data is consistent with this finding. The CHL was ~10-15 dB for frequencies < 4 kHz increasing to ~30 dB > 4 kHz consistent with Lupo *et al.* (2011); the CHL with earplugs was qualitatively similar to CHL due to experimental middle-ear effusion in chinchilla and CHL in children due to effusion (Thornton *et al.*, 2012; 2013). With earplugs, the mean CM thresholds across frequencies and animals were 46.2 ± 7.1 dB. After removal of the earplug, CM thresholds were reduced to 30.7 ± 8.3 dB. Thus, the plug produced an across-frequency attenuation of 15.5 dB. A two-way repeated-measures ANOVA revealed that there was a significant decrease in attenuation after the earplug was removed ($F_{1,10} = 103.3$, $p < 0.0001$). There was no significant difference between the CM thresholds in the control ear and thresholds in the experimental ear after the earplug was removed ($F_{1,15} = 3.71$, $p = 0.073$). The return of CM thresholds to normal levels after earplug removal indicates that the hearing loss induced by the plug was reversible, a finding reinforced by normal amplitudes and thresholds of the CAPs (not shown). Cochlear surface preparations revealed that the integrity of the cochlea was normal and that earplugging did not cause hair-cell death or other cochlear abnormalities. These data indicate that the earplug-induced CHL produced a reversible hearing loss without damaging the periphery.

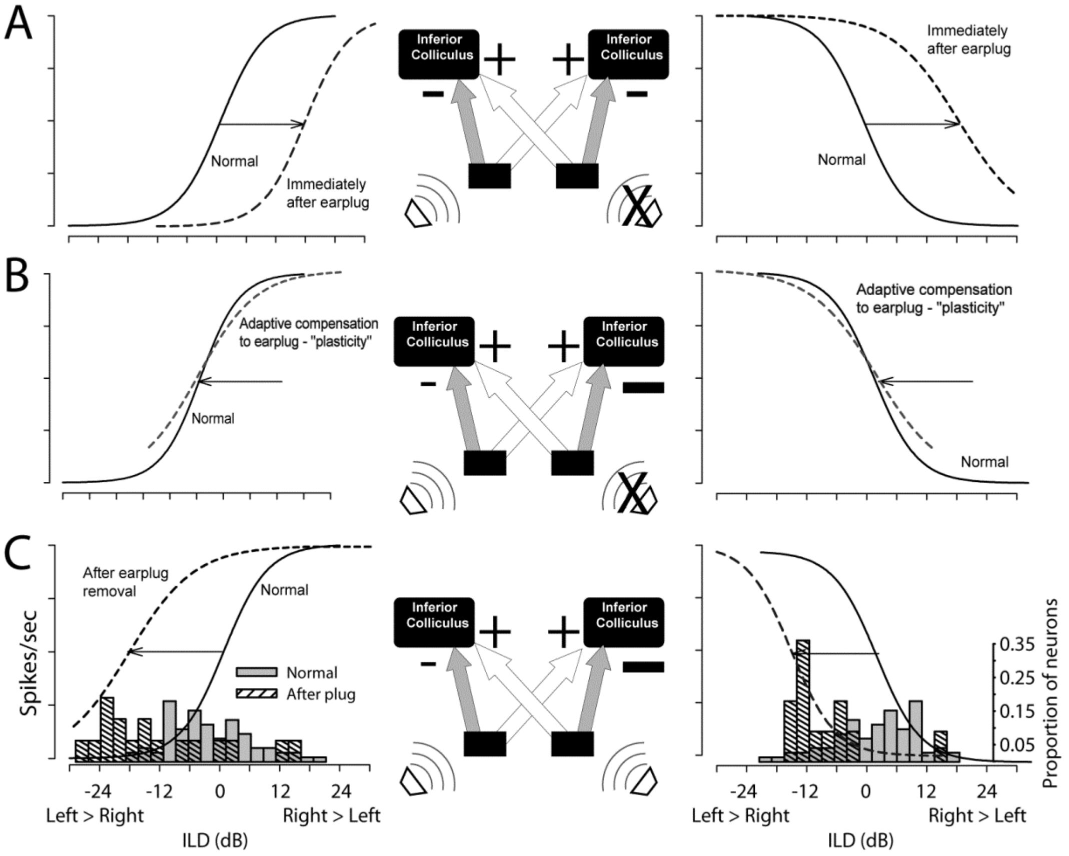


Fig. 1: Hypothesized changes due to CHL (right ear, 'X') in circuit function and the sensitivity to ILDs in the left (contra) and right (ipsi) ICC. The simplified circuit shows ipsi inhibitory ('−') and contra excitatory ('+') inputs to the ICC (strengths indicated by sizes of the symbols). **A**: Normal hearing (solid lines). Neural ILD coding shifts to the right due to CHL (dashed). **B**: Shifts if circuit plasticity has compensated for CHL. **C**: Upon CHL removal, ILD coding shifts left due to the CHL-induced altered circuitry. Histograms: empirical half-max ILDs form IC neurons in normals (n = 31) and after plug removal (n = 31).

Neural coding of interaural-level-difference cues is altered by unilateral CHL

Half-max ILD values were compared between normal animals and animals that received a unilateral earplug as adults. A shift in half-max ILD indicates a shift in the entire rate-ILD curve, signifying that that specific neuron encodes a different range of ILDs. For normal animals, the mean half-max ILD was 1.9 ± 8.3 dB (n = 31 units), with a median value of 0.94 dB and a range of from −20.6 to 17.1 dB (Fig. 1C, left and right panels, grey bars). For neurons in the ICC contralateral to the ear with CHL the mean half-max ILD value was shifted to −10.26 ± 12.2 dB (n = 19 neurons) with a median value of −14.1 dB. The overall range of ILDs encoded by these neurons was shifted toward negative ILDs, ranging from −28.4 to 15.0 dB (Fig. 1C, left panel, gray hatched bars). Relative to controls, the CHL produced an effective shift in ILD coding of 12.2 dB for neurons contralateral to the CHL. An unpaired *t*-test indicated a significant difference in half-max ILDs between normal and earplugged neurons [$t_{(44)} = 3.85$, $p = 0.0004$]. Similarly, for neurons ipsi-lateral to the CHL, the half-max ILDs were shifted to −6.52 ± 8.5 dB (median: −9.4 dB), which was significantly different than controls [$t_{(37)} = 2.92$, $p = 0.006$]. The CHL produced an effective shift in ILD coding of 8.4 dB. The ILD dynamic range of the rate-ILD curves was also impacted by CHL. The mean dynamic range in normals was 26.1 ± 10.1 dB (median: 25.7 dB). For neurons in the ICC contralateral to the CHL, the mean dynamic range was 19.4 ± 9.8 dB (median: 17.6 dB), significantly lower than controls [$t_{(44)} = 2.24$, $p = 0.031$]. Similarly, for neurons ipsilateral to the CHL, the mean dynamic range was reduced to 17.1 dB, significantly different from controls [$t_{(37)} = 2.56$, $p = 0.016$]. Finally, CHL altered the modulation of discharge rate due to ILD. In normals, over the range of ILDs tested the rate was modulated re:

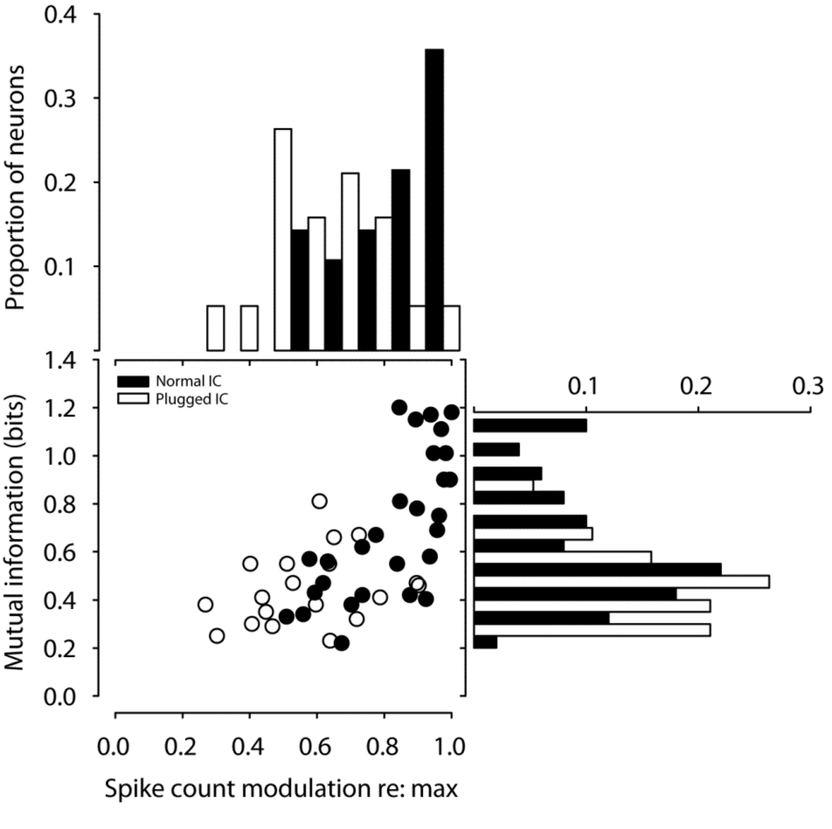


Fig. 2: Mutual information between spike count and ILD in ICC neurons in normal-hearing (black) adults and after 6 weeks of unilateral CHL (grey) is plotted as a function of the spike count modulation by ILD (re: max count).

max rate by 82 ± 16%. Neurons contralateral to the CHL were modulated by 58 ± 18%, significantly less than controls [$t_{(44)}$ = 4.9, p < 0.0001], while neurons ipsilateral to the CHL were modulated by 76 ± 13%, which was not significantly different than controls [$t_{(37)}$ = 1.13, p = 0.26]. Neurons ipsilateral to the CHL were significantly more modulated by ILD than neurons contralateral to the CHL [$t_{(29)}$ = 3.03, p = 0.005]. The reduction in overall rate-modulation by ILD for neurons contralateral to the CHL was due to an increase in the minimum rate at ILDs that should cause inhibition, which is consistent with an overall reduction in inhibition.

IC responses carry less information regarding ILD cues following CHL

Figure 2 shows mutual information computed for 31 neurons from normal animals (black bars, symbols) and for 19 neurons measured in the ICC contralateral to the CHL (white bars and symbols). The mean MI was significantly reduced [$\mathrm{t}_{(48)}$ = 3.38, p = 0.001] after CHL from 0.7 ± 0.29 (median: 0.64) bits for normals to 0.44 ± 0.15 (median: 0.41) bits. The responses of ICC neurons thus carried ~37% less mutual information regarding ILD cues following a unilateral CHL as compared to normal controls. Several factors were examined to account for the reduction in information. For neurons contralateral to the CHL, reduction in the information-carrying capacity was consistent with the significant reduction in the amount by which ILD modulated the spike count relative to the max count revealed in the earlier section. This is consistent with an effective reduction in inhibition due to the CHL.

DISCUSSION

Altered inputs to the auditory system can result in anatomical, physiological, and behavioural changes that persist beyond the hearing impairment (reviewed by Moore and King, 2004; Tollin, 2010; Whitton and Polley, 2011). The majority of evidence for CHL-induced plasticity in the auditory system comes from developmental studies in humans and animals. However, studies in adult humans and animals have also suggested that CHL can drive plasticity and that subjects can adapt to altered auditory inputs particularly via behavioural training paradigms. The data presented here suggest a compensatory mechanism for plasticity by at least the level of the inferior colliculus as well as altered information processing. Figure 1 illustrates our general hypothesis regarding compensatory changes in the ascending circuitry to the IC in response to a unilateral CHL. In normal-hearing circuitry (Fig. 1A, solid lines), spike rate is modulated by ILD sigmoidally with maximum responses for ILDs favouring the excitatory contralateral ear and reduced responses for ILDs favouring the inhibitory ipsilateral ear. Immediately after introduction of a CHL (in this case, earplug insertion), the rate-ILD curves would shift toward the right simply because the acoustical input from the contralateral ear has been attenuated (i.e., less effective excitatory input). This is represented by the dashed lines in Fig. 1A.

Figure 1B illustrates the hypothesized circuit changes that would occur if mechanisms were to compensate for the altered sound localization cues due to CHL (see Lupo *et al.*, 2011 and Thornton *et al.*, 2012; 2013). Compensatory mechanisms would work to shift the rate-ILD curves back towards normal (dashed line overlapping normal curves). To achieve this kind of adaptive compensation in the ICC contralateral to the CHL, the strength (or gain) of inhibitory input from the ipsilateral normal-hearing ear (left side in example) is hypothesized to be reduced and/or the strength (or gain) of the excitation from the contralateral CHL-ear increased; the size of the '+' and '−' symbols have been adjusted in Fig. 1B to illustrate this change. After removing the CHL, the effective changes to the ILD-coding pathways to the ICC can be revealed. If the circuit had been altered as in Fig. 1C, then after CHL removal the rate-ILD curves are hypothesized to shift toward the left (Fig. 1C, black dashed line, left column), demonstrating a reduced ipsilateral inhibitory (and/or increased contralateral excitatory) response when compared to normal. Our data is consistent with this hypothesis.

A similar compensatory response is expected in ICC neurons that are ipsilateral to the CHL. Immediately after introduction of a CHL, there will be an effective reduction in the strength of inhibition to the ICC ipsilateral to the CHL simply due to the attenuation of sound (Fig. 1A, dashed line, right column). If adaptive compensation occurs, the strength of excitatory contralateral inputs will be reduced in order to match the reduced inhibitory inputs and/or an increase in the strength of the ipsilateral inhibitory input to match the normal contralateral excitation. These changes would effectively shift the rate-ILD curves back to normal with the CHL in place (Fig. 1B, dashed line, right column). Immediately after CHL removal, an overall large inhibitory response would remain, causing the rate-ILD curves to shift to the left of normal (Fig. 1C, dashed line, right column). The results are also in agreement with this compensatory model of plasticity.

The present results disagree with previous physiological results in the ICC of animals with unilateral CHL (summarized by Moore and King, 2004; Tollin, 2010; Whitton and Polley, 2011). Prior studies in rats demonstrated that CHL persistently reduced the effectiveness of inputs to the two ICCs from the ear with the CHL, a finding that produces data consistent with illustrations in Fig. 1A as opposed to Fig. 1C. One possible reason for this may be that the experiments in the current study were performed in the chinchilla which is a precocious species (Jones *et al.*, 2011) that also has good low-frequency hearing. Additionally, the results of the prior studies could potentially be due to an altered periphery due to the CHL, such as a residual CHL even after its removal, which would also yield results as in Fig. 1A (dashed lines) even without central auditory-system plasticity. More studies are needed to reveal the sources of the differences in the results.

While compensatory plasticity may or may not occur as illustrated in Fig. 1, there is no doubt that CHL exerts a persistent effect on the neural coding of spatial information in ICC neurons as demonstrated by the 37% reduction in the capacity of neurons to carry information about ILDs (Fig. 2). Reduced MI may suggest alterations in the responsiveness (spike rates), reliability (spike rate variability), as well as the general sensitivity of ICC neurons and/or their inputs to the cues to location, including ILD. The results suggest that at least for ICC neurons contralateral to the CHL a reduction in the capacity of ILD to modulate spiking was correlated with a reduction in information-carrying capacity of these neurons. The impaired neural information processing demonstrated here may provide a basis for the persistent behavioural deficits in binaural and spatial hearing tasks that have been observed clinically after chronic CHL both during development and in adulthood. Since we have found persistent reductions in the ability of critical neural circuits in the ascending auditory pathway to encode spatial attributes of sound, it may logically follow that there will be a similar reduction in the perceptual capabilities as well. Towards this end, on-going studies are examining the behavioural consequences of reduced information processing due to CHL during development and in adults.

ACKNOWLEDGMENTS

Supported by NIH NIDCD grants R01-DC011555 (DJT) and F31-DC011198 (JLT).

REFERENCES

Dayan, P., and Abbott, L.F. (**2001**). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, Cambridge MA).

Jones, H.G., Koka, K., and Tollin, D.J. (**2011**). "Postnatal development of cochlear microphonic and compound action potentials in a precocious species, *chinchilla lanigera*," J. Acoust. Soc. Am., **130**, 38-43.

Larsen, E, and Liberman, M.C. (**2010**). "Contralateral cochlear effects of ipsilateral damage: No evidence for interaural coupling," Hear. Res., **260**, 70-80.

Lupo, J.E., Koka, K., Thornton, J.L., and Tollin, D.J. (**2011**). "The effects of experimentally induced conductive hearing loss on the spectral and temporal aspects of sound transmission through the ear," Hear. Res., **272**, 30-41.

Moore, D.R., and King, A.J. (**2004**). "Plasticity of binaural systems," in *Development of the Auditory System, Springer Handbook of Auditory Research.* Edited by T.N. Parks, E.W. Rubel, R.R. Fay, and A.N. Popper (Springer-Verlag, New York), pp. 96-172.

Thornton J.L., Chevallier K.M., Koka K., Lupo, J.E., and Tollin, D.J. (**2012**). "The conductive hearing loss due to an experimentally-induced middle ear effusion alters the interaural level and time difference cues to sound location," J. Assoc. Res. Otolaryngol., **13**, 641-654.

Thornton, J.L., Chevallier, K.M., Koka, K., and Tollin, D.J. (**2013**). "Conductive hearing loss induced by experimental middle ear effusion in a chinchilla model reveals impaired TM-coupled ossicular chain movement," J. Assoc. Res. Otolaryngol., **14**, 451-464.

Tollin, D.J. (2010). "The development of sound localization mechanisms," in *Oxford Handbook of Developmental Behavioral Neuroscience*. Edited by M.S. Blumberg, J.H. Freeman, and S.R. Robinson (Oxford University Press, New York), pp. 262-282.

Whitton, J.P., and Polley, D.B. (2011). "Evaluating the perceptual and pathophysiological consequences of auditory deprivation in early postnatal life: a comparison of basic and clinical studies," J. Assoc. Res. Otolaryngol. **12**, 535-546.

Cortical plasticity and reorganization in hearing loss

ANU SHARMA, HANNAH GLICK, AND LAUREN DURKEE*

Brain and Behavior Laboratory, Department of Speech, Language, and Hearing Science, University of Colorado, Boulder, Colorado, USA

Hearing-impaired adults and children who receive intervention with hearing aids and cochlear implants provide a platform to examine the trajectories and characteristics of deprivation-induced and experience-dependent plasticity in the central auditory system. We review the evidence for sensitive periods for development of the central auditory pathways. A sensitive period in early childhood appears to coincide with the period maximal synaptogenesis in the auditory cortex. Implantation within this sensitive period provides the auditory experience needed for refinement of essential synaptic pathways. Cross-modal recruitment is another aspect of plasticity that is apparent in deaf children. In long-term congenital deafness, somatosensory and visual stimuli activate higher-order auditory areas. Overall, it appears that the functional activation of cognitive circuitry resulting from cortical reorganization in deafness is predictive of outcomes after intervention. A better understanding of cortical development and reorganization in auditory deprivation has important implications for optimal intervention and habilitation of these patients.

DEVELOPMENT AND CORTICAL AUDITORY EVOKED POTENTIALS

Normal trajectory of central auditory system development

Cortical auditory evoked potentials (CAEPs) are averaged electroencephalography recordings of cortical brain activity in response to sound. With age, CAEP waveforms undergo major morphological changes. In infants, the response is dominated by a large, broad positivity referred to as the P1 component. As a child ages, an invagination known as the N1 and a second positive peak called the P2 appears (Sharma *et al.*, 2007). These new components can be observed in children as young as 3 to 5 years using slow stimulation rates and are consistent by preadolescence at standard stimulation rates (Gilley *et al.*, 2006).

Latency of the P1 response represents the summation of the synaptic delays throughout the central and peripheral auditory pathways (Eggermont *et al.*, 1997). In normal-hearing children, it decreases systematically and chronically with age and thus it has been used as a biomarker for auditory brain maturation (Sharma *et al.*, 2002a). Sharma and colleagues (2002b) established norms for typical P1 latency as a function of age. The P1 component occurs around 300 milliseconds in newborns then rapidly decreases over the first years of life to a latency of around 125 milliseconds in 3-year-olds. Afterwards, latency levels off at about 60 milliseconds

*Corresponding author: lauren.durkee@gmail.com

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

in adults. Auditory thalamic and cortical sources have been identified as generators and it has been suggested that P1 represents the first recurrent auditory cortex activity (Liegeois-Chauvel *et al.*, 1994; Kral and Eggermont, 2007).

Effects of deprivation

Congenitally-deaf cats are commonly used to study the effects of auditory deprivation on the brain. Kral *et al.* (2000) demonstrated layer-specific deficits in synaptic activity in electrically-stimulated deaf cats compared with hearing cats and proposed that similar deficits were likely in deaf children. As predicted, the research in cats shows significant parallels with results in humans (Kral and Sharma, 2012). A significant delay was found in the P1 latencies of prelingually deafened cochlear-implant users compared to age-matched normal-hearing subjects (Ponton *et al.*, 2000a,b; Eggermont and Ponton, 2002; 2003). Interestingly, Ponton and colleagues also found that after cochlear implantation, there is clear evidence of cortical maturation, suggesting that for the first few years of life the potential for normal auditory development is maintained in deaf children.

A SENSITIVE PERIOD FOR AUDITORY DEVELOPMENT

In a study of 104 (later 235) congenitally-deaf children, those who were fitted with cochlear implants before approximately 3.5 years had age-appropriate P1 response latencies within 6 months while those with periods of deprivation of more than 7 years had abnormal CAEP responses. Children with an intermediate deprivation duration — between 3.7 and 7 years — showed a more variable performance (Sharma *et al.*, 2002a, 2009). These results suggest that the auditory system has a sensitive period of optimal plasticity up until 3.5 years of deprivation. Plasticity decreases after that age, but does remain in some children up to age 7. These results point to the importance of early implantation within the 3.5 year period. Indeed, the approved clinical guideline has moved from age 4 in 1990 to 12 months presently, taking maximal advantage of a highly plastic central auditory system in early childhood. Interestingly, the established sensitive period cut-offs correspond to the end of the period of synaptic overshoot at approximately age 3.5 to 4 years (Conel, 1939-1967; Huttenlocher and Dabfholkar, 1997; Kral and Eggermont, 2007) and the development of adult-like myelin by age 7 to 8 (Su *et al.*, 2009; Eggermont and Moore, 2012). Implantation within this brief sensitive period provides the auditory experience needed for the establishment and refinement of essential synaptic pathways necessary for auditory-based learning to occur.

CORTICAL REORGANIZATION FOLLOWING SENSORY DEPRIVATION

Cross-modal reorganization in hearing loss

Research indicates that auditory deprivation persisting beyond the end of the sensitive period may facilitate a functional decoupling of primary auditory cortex from higher-order auditory cortex. In deaf cats implanted at the end of the sensitive period (approximately 4 months), a delay of activation of supragranular layers of the cortex and reduced activation at infragranular layers (V and VI) has been demonstrated when compared to normal-hearing cats (Kral *et al.*, 2000, 2002, 2005, 2006). These changes suggest deficient or partial development of inhibitory synapses between layer IV and supragranular layers (Kral *et al.*, 2000, 2002, 2005, 2006). Such a partial or complete decoupling between primary auditory cortex and secondary auditory cortex is also supported by FDG-PET imaging studies demonstrating decreased functional connectivity of primary auditory cortex to adjacent regions in older compared to younger pre-lingually deaf children (Kang *et al.*, 2003). The fact that a majority of children implanted after the sensitive period never develop a normal N1 CAEP response while children implanted before the age of 3.5 demonstrate an N1 response with normal morphology and latency further substantiates the decoupling hypothesis (Sharma and Dorman, 2006). Since the N1 component is presumed to arise from secondary auditory cortex, a missing N1 response would indicate improper cortico-cortical activation between primary and secondary auditory cortices (Kral and Eggermont, 2007; Kral and Sharma, 2012).

While primary auditory cortex may still retain basic facilities to process auditory information, higher-order representations linked to incoming auditory stimuli may not be effectively established if top-down modulatory processing is altered (Kral *et al.*, 2001, 2005). Because these top-down cortico-cortical pathways provide modulatory feedback, such a decoupling between primary and higher-order auditory areas may significantly affect perception as well as learning.

Given that higher-order cortex is multi-modal in nature, a decoupling between primary and secondary auditory cortex may also lead to extensive cross-modal reorganization. In the case of auditory deprivation persisting beyond the sensitive period, there is evidence that higher-order auditory areas may be re-purposed by other sensory modalities such as vision (Nishimura *et al.*, 1999; Bavelier and Neville, 2002; Lee *et al.*, 2003) and somatosensation (Sharma *et al.*, 2007). This is corroborated by evidence of atypically-distributed networks in multi-modal auditory areas in late-implanted children (Gilley *et al.*, 2006). It is well documented that early-implanted children demonstrate better speech and language outcomes relative to children implanted after age 6 to 7 years, and it has been suggested that changes in neural resource allocation (i.e., cross-modal recruitment by other sensory modalities) may indeed explain poorer behavioural outcomes with implants associated with late-implanted children (Svirsky *et al.*, 2004; Doucet *et al.*, 2006; Geers, 2006).

More recently, signs of cross-modal plasticity have been indicated within the context of adult hearing impairment. Animal studies suggest that inputs from other sensory modalities may significantly influence neurons in auditory areas, which may account for some of the functional deficits observed in adult implant and hearing-aid users. For instance, an increase in multisensory neurons in the auditory cortex and anterior auditory field in adult ferrets with moderate hearing loss compared to normal-hearing adult ferrets suggests that cross-modal reorganization may facilitate compensatory plasticity, negatively affecting important processes necessary to speech understanding such as multisensory integration (Meredith *et al.*, 2012). In this sense, cross-modal reorganization may at least partially explain poorer outcomes associated with this population of late-deafened adults who receive cochlear implants.

CLINICAL APPLICATIONS OF THE P1

There is an abundance of research supporting the clinical utility of the P1 biomarker of central auditory maturation in children (Rance *et al.*, 2002; Golding *et al.*, 2007; Pearce *et al.*, 2007; Cardon *et al.*, 2012; Cardon and Sharma, 2013). Because normal P1 latency varies as a function of age, normative data provide a standard from which P1 responses in congenitally deaf children and congenitally-deaf children fit with cochlear implants at various ages can be evaluated (Sharma *et al.*, 2002b).

The P1 biomarker can serve as an objective candidacy and/or outcome measure for children who receive hearing aids or cochlear implants. For example, Sharma *et al.* (2005) used P1 latency to determine the benefit of hearing aids in hearing-impaired children. If P1 latency was within normal limits for the child's age, then it was assumed that the hearing aid was providing sufficient stimulation for normal development of auditory pathways. However, if P1 latency did not decrease after regular hearing-aid use, then other options such as alternative hearing-aid settings or cochlear implants were considered. Thus, the P1 biomarker may aid in the clinical decision-making process, particularly in determination of cochlear-implant candidacy. Similarly, the P1 can be used as an outcome measure in children fit with hearing aids and cochlear implants. Tracked over time, the P1 can be used to evaluate the developmental progress of the cortical maturation in these children after receiving intervention (Sharma *et al.*, 2002a; 2009).

In special cases like auditory neuropathy spectrum disorder (ANSD), cortical auditory development can be assessed by examination of the P1 CAEP. Recent findings from our laboratory suggest a shorter sensitive period (approximately 2 years) for central auditory maturation after cochlear implantation in children with ANSD as compared to the sensitive period for congenitally-deaf children (i.e., 3.5 years) reviewed earlier (Cardon and Sharma, 2013). Therefore, in children with ANSD the P1 response may be especially important in the evaluation of efficacy of intervention. Moreover, it is very possible that ANSD and other disorders of the nervous system that co-exist with hearing loss (i.e., Fragile X Syndrome and Rett's Syndrome) may alter sensitive periods given developmental differences in underlying neuronal maturation. A clearer understanding of the existence and time courses of P1 development in this population may lead to improved intervention and treatment options for these children (Sharma *et al.*, 2013). While the existence and difference in sensitive periods for these individual disorders are not well understood, the P1 biomarker nevertheless provides normative data against which the developmental trajectories of children with these disorders receiving various forms of intervention can be assessed.

It is well documented that children with multiple disabilities account for a substantial percentage of children with hearing loss (Fortnum *et al.*, 2002). Many of these children are also difficult to condition to traditional behavioural threshold techniques (i.e., visual reinforcement audiometry). Often, life-threatening co-morbid health conditions make obtaining thresholds via auditory brainstem response (ABR) difficult to perform in this population since sedation under anaesthesia is not a viable option (Edwards, 2007). Additionally, as a significant proportion of children with multiple handicaps concomitant with hearing loss who receive cochlear implants never achieve closed- or open-set speech discrimination abilities, the ability to document outcomes post-intervention is additionally limited (Trimble *et al.*, 2008). While the resolution to implant a child with multiple disabilities is a multi-sided decision in which the complex medical, social-emotional, and developmental needs of the child need to be considered, the P1 CAEP response is non-invasive, easy to record, requires no anaesthesia, and proves a useful tool in assessing developmental status, hearing-aid benefit, and cochlear-implant outcomes in these cases (Sharma *et al.*, 2013).

CASE STUDY

In the next section of this paper, a case study demonstrating the clinical capability of the P1 biomarker in objectively assessing cochlear-implant outcomes will be presented.

Procedures

The stimulus used to elicit the CAEP response was a speech syllable /ba/ presented at a comfortable level through a speaker located at 45 degrees azimuth at a suprathreshold level. All testing took place in an electromagnetically-shielded sound booth. The subject was seated comfortably in a reclining chair during the recording and was allowed to watch a video or cartoon of her choice with the audio muted. For all testing, the subject's cochlear implants were set to their usual settings.

CAEPs were recorded using a standard electrode montage, recording parameters, and test procedures used routinely in our laboratory and outlined in previous studies (Sharma *et al.*, 1997; Sharma *et al.*, 2002a; 2002b). Cochlear-implant electrical artifact was removed via a common mode rejection technique detailed in a study from our group (Gilley *et al.*, 2006). The latency of the P1 component of the CAEP response was identified using the grand average waveforms for each subject.

Results

The subject was a female child identified with a bilateral hearing loss at age 27 months following a case of spinal meningitis at age 10 to 12 months. The child's hearing loss was progressive in nature. She received bilateral cochlear implants sequentially, the first implant in her left ear at age 32 months and her second implant around age 5.

In this case of a child who received cochlear implants sequentially, P1 responses were used to evaluate whether the implant was allowing for normal cortical auditory maturation. CAEP responses using the P1 biomarker were recorded at age 11 years.

As seen in Fig. 1A, a robust P1 response was present. The latency of the P1 response in the left ear fell within the 95% confidence intervals for normal development of the P1 response, indicating age-appropriate development of the central auditory pathway (Fig. 2). This robust P1 response with normal latency and morphology for the left ear clearly demonstrates that the left implant is providing adequate stimulation for cortical auditory development. These findings are consistent with the fact that the child received her left cochlear implant within the sensitive period as well as behavioural results from the child's audiologist indicating that the subject performs better on speech perception measures in the right implanted ear.

As shown in Fig. 1B, a P1 response recorded via stimulation of the right ear was present, but with a morphology that is not appropriate given the child's age. The latency of the P1 response in the right ear fell outside of normal limits, indicating abnormal or delayed development of the central auditory pathway. While the presence of a P1 response indicates that the right cochlear implant is providing adequate stimulation, the morphology of this response was not age-appropriate, likely due to the fact that the child was implanted at the end of the sensitive period around age 5. These results are consistent with findings from Sharma *et al.* (2005) which showed that cochlear implantation in one ear may not necessarily facilitate an extended sensitive period in the later implanted ear.

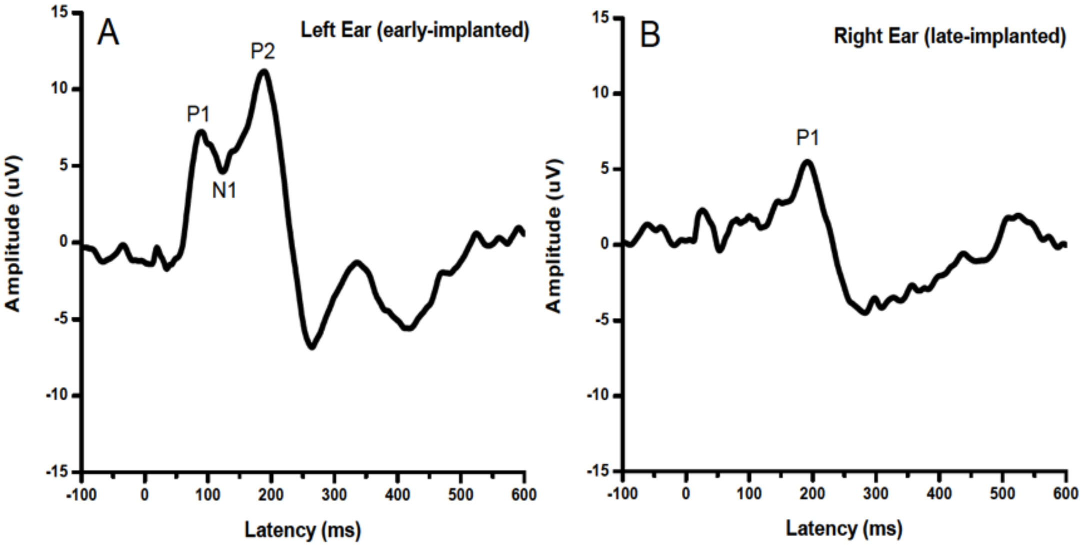


Fig. 1: Grand average CAEP response in the early-implanted left ear (top) and the late-implanted right ear (bottom) for 11-year-old sequentially-implanted subject.

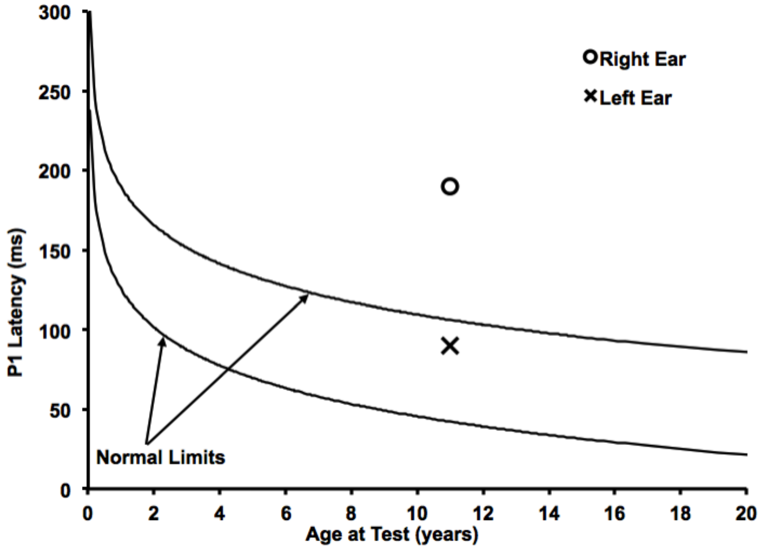


Fig. 2: Average P1 latency as a function of child's age plotted against 95% confidence limits for normal-hearing children for 11-year-old sequentially-implanted subject.

FUTURE DIRECTIONS

The P1 biomarker has proven to have real clinical value in assessing central auditory development following intervention via hearing aids and cochlear implants in congenitally-deaf children, children with ANSD, and children with multiple disabilities concomitant with hearing loss. The P1 response has great clinical capability of providing a measure of cortical auditory development, with potential applications in cochlear-implant candidacy and objective outcomes following intervention via hearing aids or cochlear implants. The maladaptive consequences of cross-modal reorganization in hearing loss are still not well understood. Future research should focus on grasping such cross-modal auditory-visual and auditory-somatosensory changes that take place in a deprived auditory system and the extent to which these changes are reversible following treatment and/or intensive rehabilitation, as findings from these studies may contribute to better outcomes for some children with hearing loss who receive intervention. Though differences in the time window of central auditory plasticity has been documented in specific cases such as ANSD, a clearer understanding of differences in sensitive periods in cases of specific disabilities is critically lacking. This knowledge may help us better understand variable outcomes in implanted children and may lead to more timely intervention for this population.

REFERENCES

Bavelier, D., and Neville, H. (**2002**). “Cross-modal plasticity: where and how?” Nat. Rev. Neurosci., **3**, 443-452.

Cardon, G., Campbell, J., and Sharma, A. (**2012**). “Plasticity in the developing auditory cortex: evidence from children with sensorineural hearing loss and auditory neuropathy spectrum disorder,” J. Am. Acad. Audiol., **23**, 396-411.

Cardon G, and Sharma, A. (**2013**). “Central auditory maturation and behavioral outcome in children with auditory neuropathy spectrum disorder who use cochlear implants,” Int. J. Audiol., **52**, 577-586.

Conel, J. (**1939-1967**). *The Post-Natal Development of Human Cerebral Cortex*, Vols. I–VIII (Cambridge, MA: Harvard University Press).

Doucet, M., Bergeron, F., Lassonde, M., Ferron, P., and Lepore, F. (**2006**). “Cross-modal reorganization and speech perception in cochlear implant users.” Brain, **129**, 3376-3383.

Edwards, L.C. (**2007**). “Children with cochlear implants and complex needs: a review of outcome research and psychological practice,” J. Deaf Stud. Deaf Edu., **12**, 258-268.

Eggermont, J., Ponton, C., Don, M., Waring, M., and Kwong, B. (**1997**). “Maturational delays in cortical evoked potentials in cochlear implant users,” Acta Oto-Laryngol., **117**, 161-163.

Eggermont, J., and Ponton, C. (**2002**). “The neurophysiology of auditory perception: From single units to evoked potentials,” Audiol. Neuro-Otol., **7**, 71-99.

Eggermont, J., and Ponton, C. (**2003**). “Auditory-evoked potential studies of cortical maturation in normal hearing and implanted children: Correlations with changes in structure and speech perception,” Acta Oto-Laryngol., **123**, 249-252.

Eggermont, J., and Moore, J. (**2012**). “Morphological and functional development of the auditory nervous system,” in *Human Auditory Development*. Edited by L. Werner, R. Fay, and A.N. Popper (New York: Springer), pp. 61-105.

Fortnum, H.M., and Marshall, D.H., and Summerfield, A.Q. (**2002**). “Epidemiology of the UK population of hearing-impaired children, including characteristics of those with and without cochlear implants – audiology, aetiology, cormobiditiy and affluence,” Int. J. Audiol., **41**, 170-179.

Geers, A.E. (**2006**). “Factors influencing spoken language outcomes in children following early cochlear implantation,” Adv. Oto-Rhino-Laryng., **64**, 50-65.

Gilley, P., Sharma, A., Dorman, M., and Martin, K. (**2006**). “Abnormalities in central auditory maturation in children with language-based learning problems,” Clin. Neurophysiol., **117**, 1949-1956.

Golding, M., Pearce, W., Seymour, J., Cooper, A., Ching, T., and Dillon, H. (**2007**). “The relationship between obligatory cortical auditory evoked potentials (CAEPs) and functional measures in young infants,” J. Am. Acad. Audiol., **18**, 117-125.

Huttenlocher, P., and Dabholkar, A. (**1997**). “Regional differences in synaptogenesis in human cerebral cortex,” J. Comp. Neurol., **387**, 167-178.

Kang, E., Lee, D., Lee, J., Kang, H., Hwang, C., Oh, S., and Zilles, K. (**2003**). "Developmental hemispheric asymmetry of interregional metabolic correlation of the auditory cortex in deaf subjects," NeuroImage, **19**, 777-783.

Kral, A., Hartmann, R., Tillein, J., Heid, S., and Klinke, R. (**2000**). "Congenital auditory deprivation reduces synaptic activity within the auditory cortex in a layer-specific manner," Cereb. Cortex, **10**, 714-726.

Kral, A., Hartmann, R., Tillein, J., Heid, S., and Klinke, R. (**2001**). "Delayed maturation and sensitive periods in the auditory cortex," Audiol. Neuro-Otol., **6**, 346-362.

Kral, A., Hartmann, R., Tillein, J., Heid, S., and Klinke, R. (**2002**). "Hearing after congenital deafness: central auditory plasticity and sensory deprivation," Cereb. Cortex, **12**, 797-807.

Kral, A., Tillein, J., Heid, S., Hartmann, R., and Klinke, R. (**2005**). "Postnatal cortical development in congenital auditory deprivation," Cereb. Cortex, **15**, 552-562.

Kral A., Tillein J., Heid S., and Klinke R. (**2006**). "Cochlear implants: cortical plasticity in congenital deprivation," Prog. Brain Res., **157**, 283-313.

Kral, A., and Eggermont, J. (**2007**). "What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity," Brain Res. Rev., **56**, 259-269.

Kral, A., and Sharma, A. (2012). "Developmental neuroplasticity after cochlear implantation," Trends Neurosci., **35**, 111-122.

Lee, J.S., Lee, D.H., Oh, S.H., Kim, C.S., Kim, J.-W., Hwang, C.H., Koo, J., Kang, E., Chung, J.-K., and Lee, M.C. (**2003**). "PET evidence of neuroplasticity in adult auditory cortex of postlingual deafness," J. Nucl. Med., **44**, 1435-1439.

Liegeois-Chauvel, C., Musolino, A., Badier, J., Marquis, P., and Chauvel, P. (**1994**). "Evoked potentials recorded from the auditory cortex in man: Evaluation and topography of the middle latency components," Electroen. Clin. Neuro., **92**, 204-214.

Meredith, M., Keniston, L., and Allman, B. (**2012**). "Multisensory dysfunction accompanies crossmodal plasticity following adult hearing impairment," Neuroscience, **214**, 136-148.

Nishimura, H., Hashikawa, K., Doi, K., Iwaki, T., Watanabe, Y., Kusuoka, H., and Kubo, T. (**1999**). "Sign language 'heard' in the auditory cortex," Nature, **397**, 116.

Pearce, W., Golding, M., and Dillon, H. (**2007**). "Cortical auditory evoked potentials in the assessment of auditory neuropathy: two cases," J. Am. Acad. Audiol., **18**, 380-390.

Ponton, C., Eggermont, J.J., Kwong, B., and Don, M. (**2000a**). "Maturation of human central auditory system activity: Evidence from multi-channel evoked potentials," Clin. Neurophysiol., **111**, 220-236.

Ponton, C., Eggermont, J., Don, M., Waring, M., Kwong, B., Cunningham, J., and Trautwein, P. (**2000b**). "Maturation of the mismatch negativity: Effects of profound deafness and cochlear implant use," Audiol. Neuro-Otol., **5**, 167-185.

Rance, G., Cone-Wesson, B., Wunerlich, J., and Dowell, R. (**2002**). “Speech perception and cortical event related potentials in children with auditory neuropathy,” Ear Hearing, **23**, 239-253.

Sharma, A., Kraus, N., McGee, T., and Nicol, T., **1997**. “Developmental changes in P1 & N1 auditory responses elicited by consonant-vowel syllables.” Clin. Neurophysiol. **104**, 540–545.

Sharma, A., Dorman, M., and Spahr, A. (**2002a**). “A sensitive period for the development of the central auditory system in children with cochlear implants: Implications for age of implantation,” Ear Hearing, **23**, 532-539.

Sharma A., Dorman M, Spahr A, and Todd N. (**2002b**). “Early cochlear implantation in children allows normal development of central auditory pathways.” Ann. Oto. Rhinol. Laryn. Suppl., **189**, 38-41.

Sharma, A., Dorman, M., and Kral, A. (**2005**). “The influence of a sensitive period on central auditory development in children with unilateral and bilateral cochlear implants,” Hear. Res., **203**, 134-143.

Sharma, A., and Dorman, M. (**2006**). “Central auditory development in children with cochlear implants: clinical implications,” Adv. Oto-Rhino-Laryng., **64**, 66-88.

Sharma, A., Gilley, P.M., Dorman, M.F., and Baldwin, R. (**2007**). “Deprivation-induced cortical reorganization in children with cochlear implants,” Int. J. Audiol., **46**, 494-499.

Sharma, A., Nash, A., and Dorman, M. (**2009**). “Cortical development, plasticity, and re-organization in children with cochlear implants,” J. Commun. Disord., **42**, 272-279.

Sharma, A., Glick, H., Campbell, J., and Biever, A. (**2013**). “Central auditory development in children with hearing loss: clinical relevance of the P1 CAEP biomarker in hearing-impaired children with multiple disabilities,” Hear. Balance Commun., **11**, 110-120.

Su, P., Kuan, C., Kaga, K., Sano, M., and Mima, K. (**2008**). “Myelination progression in language-correlated regions in brain of normal children determined by quantitative MRI assessment,” Int. J. Pediatr. Otorhi., **72**, 1751-1763.

Svirsky, M.A., Teoh, S.-W., and Neuburger, H. (**2004**). “Development of language and speech perception in congenitally, profoundly deaf children as a function of age at cochlear implantation,” Otol. Neurotol., **9**, 224-233.

Trimble, K., Rosella, L.C., Propst, E., Gordon, K.A., Papaioannou, V., and Papsin, B.C. (**2008**). “Speech perception outcome in multiply disabled children following cochlear implantation: Investigating a predictive score,” J. Am. Acad. Audiol., **19**, 602-611.

Challenges associated with participation and compliance in auditory training

ROBERT SWEETOW*

University of California, San Francisco, California, USA

When individuals have hearing loss, physiological changes in their brain interact with relearning of sound patterns. Some individuals utilize compensatory strategies that may result in successful hearing aid use. Others, however, are not so fortunate. Aural rehabilitation has long been advocated to enhance communication but has not been considered time or cost-effective. Home-based, interactive adaptive computer therapy programs are available which are designed to engage the adult hearing impaired listener in the hearing aid fitting process, provide listening strategies, build confidence, and address cognitive changes. Despite the availability of these programs, many patients and professionals are reluctant to engage in and complete therapy. In this presentation reasons for the lack of compliance with therapeutic options will be identified and possible solutions to maximizing participation and adherence will be offered.

INTRODUCTION

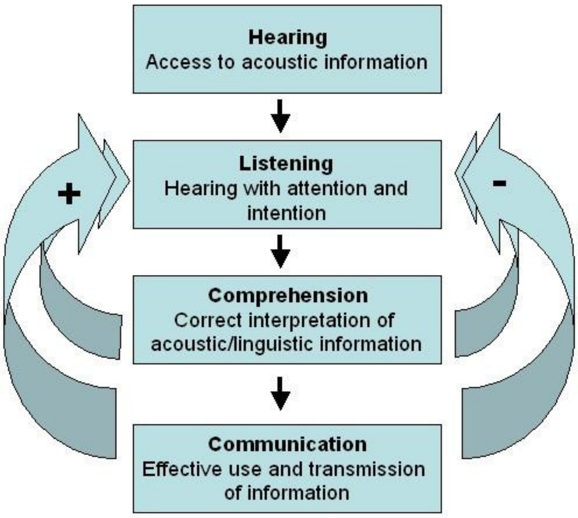
The long held myth that the brain is a fixed, immutable system has been clearly dispelled and replaced by the notion that it is indeed plastic. It is now obvious that neural connections can be altered and that these modifications, whether considered refinements or weaknesses, can manifest themselves as behavioral changes. Research has demonstrated that peripheral dysfunction and attenuation, including hearing loss, leads to subsequent neuroplastic changes. Secondary plasticity may also occur following remedial efforts, such as provided by amplification, but problems persist due to limitations in hearing aids and cognitive deficits. Other attempts at remediation, including auditory training (AT), also results in plasticity, but there has been a reluctance by both patients and professionals to adopt this as a regular part of aural rehabilitation (AR). Few audiologists would argue with the notion that additional training beyond the use of wearable amplification could potentially benefit patients. Unfortunately, despite the logic and growing body of evidence supporting this position, most audiologists do not offer or prescribe additional therapies, and most patients do not ask for, or even wish to participate in additional rehabilitation. There are many possible reasons for this bilateral reluctance. In this paper, reasons for resistance, opportunities for change, and suggestions for greater compliance will be explored.

*Corresponding author: robert.sweetow@ucsf.edu

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

WHY DO PATIENTS SEEK OUR HELP?

It would be too simplistic to assume that patients request advice from audiologists simply because they are having difficulty hearing. Indeed, few patients seek assistance because they are unable to detect birds chirping or other environmental sounds. Rather, patients seek intervention (although they don't state it as such) because of breakdowns in auditory communication. A number of elements comprise the hierarchy ranging from hearing to communication. The most basic step is hearing, which for the purpose of this discussion, can be defined as access to acoustic information. Ability to hear should (but does not always) lead to the ability to listen. This is because listening requires attention and intention. Listening is an active process requiring effort. Listening enables (but does not guarantee) comprehension, which presumes the accurate establishment of meaning. This results, in many cases, in communication, which entails the bidirectional transfer of information, meaning, and intent (Kiessling *et al.*, 2003; Sweetow and Sabes, 2004). Potential impediments to achieving mastery of these elements include peripheral hearing loss, progressive neurodegeneration (Kim *et al.*, 1997, Morest *et al.*, 1998), global cognitive decline, maladaptive compensatory behaviors, and loss of confidence (Sweetow and Sabes, 2010a). These elements are displayed in Fig. 1.


Hearing
Access to acoustic information

Listening
Hearing with attention and intention

Comprehension
Correct interpretation of acoustic/linguistic information

Communication
Effective use and transmission of information

Fig. 1: Elements of communication. Adapted from Kiessling *et al.*, 2003; Sweetow and Sabes, 2004.

If adequate communication is not achieved, remedial efforts, including the purchase and use of hearing aids is impacted, both when owners refuse to wear their hearing aids, and when hearing aids are returned for credit. Returns and exchanges average in the double digits for hearing aids. Reasons include inaudibility, poor benefit/cost ratio, unrealistic expectations and inadequate counselling, neural plasticity, cognitive changes, and poor listening habits. Some of these factors can be eliminated or minimized. For example, the use of verification via probe microphone measures can mitigate inaudibility, and the use of realistic, time-based expectations can lower unrealistic patient expectations. The reality, however, is that there are numerous limitations to what modern hearing aids are capable of correcting. For example, hearing aids themselves cannot resolve impaired frequency resolution, rectify impaired temporal processing, undo maladaptive listening strategies, produce accurate proper localization cues which can be vital for navigating auditory space, 'properly' reverse neural plastic effects, or correct for changes in cognitive function that coincide with aging. This latter cause is particularly relevant because about two-thirds of people age 70 and older have hearing loss and older adults with hearing loss have a 24% higher risk of cognitive impairment. Lin *et al.* (2011) have speculated that this could be related to common cause hypothesis (shared neural pathways) leading to extra resource expenditure and isolation.

Imaging studies of word identification in unfavorable signal-to-noise ratios have revealed greater activation of memory and attention brain regions in older adults compared with younger adults (Wong *et al.*, 2009). To compensate for reduced audibility or deficits in temporal processing (Anderson *et al.*, 2013), older adults draw more on cognitive resources than younger adults (Wong *et al.*, 2010). Despite this, older adults often have a diminished cognitive reserve when trying to communicate in a complex listening environment. Pichora-Fuller and Singh (2006) evaluated the role of the auditory-cognitive system in speech-in-noise perception in older adults. They evaluated the strength of contributions from cognitive function (memory and attention), peripheral hearing status (audiometric thresholds and distortion product otoacoustic emissions), and neural processing (subcortical measures of pitch and response fidelity) to speech-in-noise perception. They also included a life experiences factor comprised of musical training because of its known long-term effects on speech-in-noise perception and memory (Parbery-Clark *et al.*, 2009). They found that cognitive function and neural processing were the biggest contributors to variance in speech-in-noise perception, but life experiences also had an effect. Interestingly, the contribution of hearing thresholds was not significant. This finding is consistent with previous work demonstrating that the audiogram is not a good predictor of speech-in-noise perception.

EFFECTS OF TRAINING

As stated earlier, plasticity occurs when there are peripheral deficits (Willott, 1993). But secondary plasticity can occur as a result of auditory training (Kraus *et al.*, 1995; Tremblay *et al.*, 1997; Menning *et al.*, 2000). Physiologic changes post training have been demonstrated in a number of studies and a variety of ways. For example, cortical thickening in older adults (Engvig, 2010); changes in mismatched negativity response (Recanzone *et al.*, 1993; Kraus *et al.*, 1995); changes in auditory evoked magnetic fields - (Vasama and Mäkelä, 1995); enhanced NI-P2 on novel speech sounds and demonstrated training effects (Tremblay *et al.*, 2001). Tremblay *et al.* (2009) attributed training related physiological changes to a greater number of neurons responding in the sensory field, and improved neural synchrony. They hypothesized that training decorrelates activity between neurons, making each neuron as different as possible in its functional specificity.

Training effects, in order to be truly beneficial, however, must extend beyond physiological changes and must be reflected in behavioral changes. Here too, there is ample evidence promoting the use of auditory training, both in individualized and group formats (Beynon *et al.*, 1997; Chisolm *et al.*, 2004; Hawkins, 2005). Sweetow and Palmer (2005), and more recently Henshaw and Ferguson (2013), conducted evidence-based reviews of the literature on individualized auditory rehabilitation and training in adults. Both reviews reached similar conclusions. They included: 1) less than 5% of studies published on auditory training meet rigorous evidenced based criteria; 2) auditory training resulted in improved performance for trained tasks in nearly all the articles that met evidenced-based criteria; 3) although significant generalization of learning was shown to untrained measures of speech intelligibility, cognition, and/or self-reported hearing abilities, the improvements were variable, relatively small and not robust, though retention of learning was shown at post-training. This individual variability in results is likely a product of protocol, but in addition, subtle reorganization could produce diverse presentations by scattering the deficit in neural space, and individuals' brain anatomy differ (i.e., variations in fissural patterns and propensities for adaptation and recovery).

Synthetic (top-down) training refers to training based on recognition of the overall meaning of discourse. Data indicate that it is capable of teaching hearing-impaired individuals to better use active listening strategies that can translate into improved psychosocial function. Some studies further support the finding that speech recognition skills, particularly in noise, can be improved by synthetic training. Uncertainty remains regarding the contribution of analytic training (bottom-up exercises using small segments of the speech signal such as phonemes or syllables). However, a number of issues may account for the lack of definitive results. Among these issues are the sensitivity of the outcome measures used in formulating conclusions and doubts regarding whether the optimal analytic training parameters have yet to be identified.

But while the improvements in speech recognition reflect a relatively modest statistic, the practical benefits may be larger than suspected. Consider, for example, that normal hearing people generally require a +2 dB signal to noise ratio for 50% recognition of words in sentences (while people with hearing loss require a +8 dB signal to noise ratio) (Killion, 2002) and that even a 1-dB reduction in SNR has been equated with a 6-8% improvement in sentence recognition (Crandell, 1991). Yet improvements, as reflected in off-task measures such as the QuickSIN for certain training protocols (as described below) may show group averages in excess of 3 dB and double-digit improvements in individual improvements. In addition, it has been shown that higher benefit from training is significantly correlated with reduced listening effort (Olson *et al.*, 2013); new hearing-aid-user groups experience the largest improvement (Olson *et al.*, 2013); patients with more severe handicap show greater benefit (Henderson-Sabes and Sweetow, 2007; Hickson *et al.*, 2007) and patients with more severe handicap are more likely to comply with therapeutic recommendations (Henderson-Sabes and Sweetow, 2007).

AUDITORY TRAINING PROGRAMS

As recently as ten years ago, the state of the art dictated therapy had to be performed in a face to face condition, thus rendering it less than cost effective. But now in the digital age, we have the means to provide therapy via computer-aided auditory rehabilitation so that it can be performed in a private, non-threatening environment, proceed at the individual's optimal pace, and progress assessment can be done automatically. A number of computerized auditory training programs are available. A partial list is shown in Table 1.

- CAST (Computer Assisted Speech Training)
- LACE (Listening and Communication Enhancement)
- Read My Quips
- Seeing and Hearing Speech
- Sound Auditory Training (Chermak, Musiek, and Weihing)
- Sound and Beyond
- SPATS (Speech Assessment and Training System)

Table 1: Partial list of available auditory training programs (in alphabetical order).

Of these programs, one in particular has been designed to engage the adult hearing-impaired listener in the hearing-aid fitting process, provide listening strategies, build confidence, and address cognitive changes characteristic of the aging process. LACE (Listening and Communication Enhancement) provides exercises in the types of situations most difficult for hearing-impaired listeners (Sweetow and Sabes, 2006). It utilizes an adaptive training algorithm so that the training difficulty level occurs near the individual's skill threshold and proceeds at the patient's optimal pace. The training combines listening training (analytic) with repair strategies (synthetic), and gives the patient feedback regarding performance. LACE provides a variety of tasks that are divided into three main categories (degraded speech, cognitive skills, and communication strategies). In a multi-site study of the effectiveness of a pilot version of LACE on 65 subjects, significant improvements were reported, not only on the training tasks, but also on a variety of 'off-task' standardized outcome measures including the QuickSIN (Etymotic Research, 2001; Killion *et al.*, 2004), Hearing Handicap Scale for the Elderly (HHIE) (Ventry and Weinstein, 1982), and Communication Scale for Older Adults (CSOA) (Kaplan *et al.*, 1997). Sixty percent of the subjects improved in all of the training tasks. Eighty-three percent of subjects improved in all but one of the training tasks. Subjects improved on the off-task outcome measures as well. Trained subjects improved an average of 2.2 dB SNR loss and 1.5 dB SNR loss on the QuickSIN test, presented at 45 dB and 70 dB, respectively. Eighty-five percent and seventy-four percent of the subjects showed improvement, with 46% and 42% of subjects showing clinically significant improvements on the QuickSIN (> 1.6 dB SNR loss improvement) for the 45 dB and 70 dB presentations, respectively.

Moreover, Song *et al.* (2012) evaluated the effects of LACE on 60 normal-hearing adults using both the QuickSIN) and HINT (Nilsson *et al.*, 1994), and concluded that "LACE training generalizes to standardized clinically-utilized measures of speech-in-noise perception – a critical factor if (auditory) training is to have an impact on real-world listening". They further stated that 'naturalistic training' that combines sensory and cognitive elements can enhance the central nervous system's ability to encode acoustic pitch-related fundamental frequency (FF) and second formant (F2) cues.

PROBLEMS

All of this is good news. Here's the bad news. Less than 20% of new users (and less than 10% of experienced users) receive any form of audiologic rehabilitation (beyond hearing aids) and only 2-5% are provided with formal retraining opportunities (Kochkin, 2009). Considering the fact that the profession of audiology was first formally conceived in 1946 for the purpose of providing rehabilitation for hearing-impaired veterans returning from World War II (Carhart, 1960), it is quite a disappointment that the use of formal rehabilitative services beyond hearing aids has reduced to such a level. What happened to aural rehabilitation? Ross (1997) has speculated that it declined beginning in the 1960s because outcome measures concentrated on analytic auditory training (difficult to achieve considering the limited bandwidth produced by hearing aids in those days) and speech-reading, and did not consider emotional and psychological by-products. In addition, many professionals consider it to be rather boring to administer, believe it is too time consuming, are reluctant to ask patients to spend more time or money, and are not convinced by the data supporting its efficacy. Each of these theories are quite tenuous. There is, however, validity to the belief that there is an undeniable, and unfortunate, lack of reimbursement.

Let us consider each of the arguments against providing auditory training.

Boring to administer: Many audiologists, including this author, initially attracted to the profession by the glamour and promise of technology, are underwhelmed by the tedium of plotting lesson plans and spending hours of individualized therapy. Yet indeed, auditory training in the 1950s though 2000 was comprised of exactly that. Now, however, the bulk of AT is conducted via computerized training that not only includes adaptive training to optimize individual learning rates, but automatic scoring.

Too time consuming: Since the bulk of training is done via computer, there is no need for the professional to spend significant time in the training phase (with the exception, of course, of initial instructions, occasional monitoring, and follow-up counseling). Establishing the protocol and collecting materials for both AT and group AR is also no longer an onerous task because there are numerous materials available via the web; e.g. Active Communication Education (Hickson *et al.*, 2007); Learning to Hear Again (Wayner and Abrahamson, 1996); Mayo Clinic Group AR (Hawkins, 2005).

Reluctance to ask patients to spend more time or money: Given the substantial cost of hearing aids, incorporating the relatively small additional monetary expenditure may seem insignificant to patients, or it can be included in the bundled pricing structure. Asking the patient to spend more time in the rehabilitation process is a somewhat trickier issue. If the audiologist is not convinced AT and AR will help, there may be a reluctance to ask the patient to participate in what could become a frustrating task. However, requesting patient participation in even more difficult, and sometimes uncomfortable, therapy such as physical therapy post-surgery is commonplace and a well-accepted component of such rehabilitation.

Not convinced by the data supporting its efficacy: As stated earlier, very few studies meeting evidence-based criteria have been published on AT efficacy. Those that have been published often have poor control or inadequate sample size. In addition, even some studies that support AR and AT can be misinterpreted. For example, Chisolm *et al.* (2004) indicated that hearing-aid users participating in an AR program performed better on a communication profile than those with no group AR experience at the conclusion of the program. However, there were no significant differences between the groups after one year. This finding may be interpreted as suggesting AR did not help. However, given the importance of hearing-aid uptake and usage to the overall AR process, the first month (the trial period) during which patients decide whether or not to keep and continue wearing hearing aids will be highly influenced by success that might be attributed to the AR classes. If patients do not recognize some early success, they may indeed cease amplification usage, thus increasing the likelihood of an unsuccessful rehabilitation. Uncertainties regarding the optimal training parameters required to drive secondary plasticity in the proper direction also account for the lack of belief in the value of therapy. Thus, in order for professionals to embrace the concept of the need for AR and AT, research data must be gathered and presented in a compelling scientific manner, and disseminated by established and respected investigators and clinicians.

But even if audiologists recognized the importance of providing AT, there is still the task of convincing patients to participate. Clinical data from over 3,000 individuals reported that adherence (defined as completion of at least half of the recommended number of AT sessions) was less than 30% (Sweetow and Sabes, 2010a). Similarly, in a study of home-based computerized AT for cochlear-implant users, Stacey and Summerfield (2005) reported that about 1/3 of their users completed less than 1/3 of the recommended training. It should be mentioned that the profession of audiology is not unique when it comes to non-compliance with recommendations. Non-compliance with prescribed medication regimens for hypotensive treatment ranges from 5% to 80% among glaucoma patients (Olthoff *et al.*, 2005). Vincent (1971) reported that 43% of glaucoma patients refused to take the physician-ordered measures necessary to prevent blindness, even when that refusal had already led to impairment in one eye.

It is also difficult to determine which patients will comply with recommendations. Intelligence, age, gender, and economic background are not correlated with compliance (Cameron, 1996). There are, however, some social and psychological factors believed to influence compliance. They include: knowledge and understanding including communication, quality of the patient-provider interaction, social and family support, and factors associated with the illness and the treatment including the duration and the complexity of the regimen (Cameron, 1996).

Six predictors of positive compliance cited by Laplante-Lévesque *et al.* (2012) are: higher socioeconomic status, greater initial self-reported hearing disability, lower pre-contemplation stage (denial), greater action stage of change, lower chance locus of control, and greater hearing disability perceived by others and self. In addition, motivation to improve, lifestyle, available free time, desire to please family members, and readiness for change are vital factors.

The following suggestions may improve compliance: 1) provide clear and understandable information about the condition and progress in a sincere and responsive way; 2) simplify instructions and therapy regimens as much as possible; 3) have systems in place to generate patient treatment or appointment reminders; and 4) for home based AT programs, conduct the first session with the patient in the clinic. This can be done by an assistant to maximize the professional's time. Data collected on compliance with the LACE program indicate that the number of participants completing the prescribed regimen increased by 20% when the first session was done in the clinic.

CHALLENGES AND CONCLUSIONS

The popularity of 'brain-training' programs continues to increase. Programs such as Lumosity, Fit Brains, and Brain HQ from Posit Science enjoy widespread usage. The challenge is to attain similar acceptance and popularity for AT. To do so, a number of improvements in current programs should be considered. Among them are: conduct large-scale, multi-site studies with adequate control groups and large sample sizes; develop mobile AT apps; incorporate videos, animations, and graphical interfaces into AT programs; and create more exciting and enjoyable training protocols. To this end, a number of studies suggest non-speech training materials can be of use. Music training can lead to better processing of speech in the auditory brainstem and cortex and to better understanding of speech in noise across age groups (Parbery-Clark *et al.*, 2009).

In fact, older musicians do not have the same brainstem timing delays in their speech-evoked responses as older nonmusicians do (Kraus and Anderson, 2013). The concept of using music as a stimulus for AT is supported by the Patel's acronym OPERA (Patel, 2012), which stands for:

- **O**verlap: in the anatomy and physiology for speech and music
- **P**recision: more precision is required for music processing than speech
- **E**motions: strong emotions evoked by music may induce plasticity via the brain's reward centers
- **R**epetition: extensive practice tunes the auditory system
- **A**ttention: focused attention to details of sound is required when playing an instrument

There is a great need to have better diagnostic and prognostic assessments. Computerized training may not be feasible for every patient. It would be useful to predict which subjects are more likely to commit to participating in, and then ultimately completing a training program. It is currently not possible to predict outcomes based on initial data. Therefore, clinical expertise and experience, as well as information obtained from counseling, is important when deciding who should participate in aural rehabilitation.

Many unresolved questions remain. What are the best training parameters and modes? What sequences of specific inputs will change the brain in desired ways? In training, should one use analytic microtraining (bottom-up) or synthetic macro-training (top-down) approaches, or a combination? Will training generalize to "real-life" experiences? Will training improve the acceptance of hearing aids? Will results be magnified when training is introduced in conjunction with introduction or changes to amplification? When should AT be offered (before hearing-aid fitting, during trial, after trial)? Will training last over extended time periods? Will audiologists be resolute in recommending training? And perhaps most important to convincing audiologists and patients about the efficacy of AT, what are appropriate outcome measures and how should success be measured? Certainly group mean data do not reflect individual variations in improvements from AT. Should success be defined by on-task improvement, generalized speech recognition performance, subjective communication confidence (Sweetow and Sabes, 2010b), or quality of life? The answers to these questions will solidify the place for computerized auditory training and aural rehabilitation in the clinical audiology practice. Research must lead the way to acceptance that hearing aids are one, but not the only, component of AR.

DISCLOSURE

The author has a financial interest in Neurotone, Inc., the company that produces LACE.

REFERENCES

Anderson, S., Parbery-Clark, A., White-Schwoch, T., and Kraus, N. (**2013**). "Reversal of age-related neural timing delays with training," Proc. Natl. Acad. Sci. USA, **110**, 4357-4362.

Beynon G.J., Thornton, F.L., and Poole, C. (**1997**). “A randomized controlled trial of the efficacy of a communication course for first time hearing aid users,” Br. J. Audiol., **31**, 345-351.

Cameron, C.J. (**1996**). “Patient compliance: recognition of factors involved and suggestions for promoting compliance with therapeutic regimens,” Adv. Nurs., **24**, 244-250.

Carhart, R. (**1960**). “Auditory Training,” in *Hearing and Deafness*. Edited by H. Davis and R. Silverman, 2nd ed. (Holt Rinehart and Winston, New York, NY), pp. 346-359.

Chisolm, T., Abrams, H., and McArdle, R. (**2004**). “Short- and long-term outcomes of adult audiological rehabilitation,” Ear. Hearing, **25**, 464-477.

Crandell, C. (**1991**). “Individual differences in speech recognition ability: implications for hearing aid selection,” Ear. Hearing, **12**, 100S-108S.

Engvig, A. (**2010**). “Effects of memory training on cortical thickness in the elderly,” NeuroImage, **52**, 1667-1676.

Etymotic Research. (**2001**). *QuickSIN Speech in Noise Test, v. 1.3*. Elk Grove Village, IL.

Hawkins, D. (**2005**). “Effectiveness of counseling-based adult group aural rehabilitation programs: a systematic review of the evidence,” J. Am. Acad. Audiol., **16**, 485-493.

Henderson-Sabes, J., and Sweetow R. (**2007**). “Variables affecting outcome on Listening and Communication Enhancement (LACE^{TM}) training,” Int. J. Audiol., **46**, 374-383.

Henshaw, H., and Ferguson, M. (**2013**). “Efficacy of individual computer-based auditory training for people with hearing loss: a systematic review of the evidence,” PLoS One, **8**, e62836.

Hickson, L., Worrall, L., and Scarinci, N. (**2007**). “A randomized controlled trial evaluating the Active Communication Education program for older people with hearing impairment,” Ear. Hearing, **28**, 212-230.

Kaplan, H., Bally, S., Brandt, F., Busacco, D., and Pray, J. (**1997**). “Communication Scale for Older Adults (CSOA),” J. Am. Acad. Audiol., **8**, 203-217.

Kiessling, J., Pichora-Fuller, M.K., Gatehouse, S., Stephens, D., Arlinger, S., Chisholm, T., Davis, AC., Erber, N.P., Hickson, L., Holmes, A., Rosenhall, U., and von Wedel, H. (**2003**). “Candidature for and delivery of audiological services: Special needs of older people,” Int. J. Audiol., **42**, S92-S101.

Killion, M. (**2002**). “New thinking on hearing in noise: A generalized articulation index,” Sem. Hear., **23**, 57-75.

Killion, M., Niquette, P., Gudmundson, G., Revit, L., and Banerjee, S. (**2004**). “Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners,” J. Acoust. Soc. Am., **116**, 2395-2405.

Kim, J., Gross, J., Morest, D., and Potashner, S. (**2004**). “Quantitative study of degeneration and new growth of axons and synaptic endings in the chinchilla cochlear nucleus after acoustic overstimulation,” J. Neurosci. Res., **15**, 829-842.

Kochkin, S. (**2009**). "MarkeTrak VIII: 25 year trends in the hearing health market," Hearing Review, **16**, 12-31.

Kraus, N., McGee, T., Carrell, T.D., King, C., Tremblay, K., and Nicol, T. (**1995**). "Central auditory system plasticity associated with speech discrimination training," J. Cogn. Neurosci., **7**, 25-32.

Kraus, N., and Anderson, S. (**2013**). "Hearing matters: Music training: an antidote for aging?" Hearing Journal, **66**, 52.

Laplante-Lévesque, A., Hickson, L., and Worrall, L. (**2012**). "What makes adults with hearing impairment take up hearing aids or communication programs and achieve successful outcomes?" Ear Hearing **33**, 79-93.

Lin, F., Metter, J., O'Brien, R., Resnick, S., Zonderman, A., and Ferrucci, L. (**2011**). "Hearing loss and incident dementia," Arch Neurol., **68**, 214-220.

Menning, H., Roberts, L.E., and Pantev, C. (**2000**). **"**Plastic changes in the auditory cortex induced by intensive frequency discrimination training," Neuroreport, **11**, 817-822.

Morest, D., Kim, J., Potashner, S., and Bohne, B. (**1998**). "Long-term degeneration in the cochlear nerve and cochlear nucleus of the adult chinchilla following acoustic overstimulation," Micro. Res. Tech., **41**, 205-216.

Nilsson, M., Soli, S., and Sullivan, J. (**1994**). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am., **95**, 1085-1099.

Olson, A.D., Preminger, J.E., and Shinn, J.B. (**2013**). "The effect of LACE DVD training in new and experienced hearing aid users," J. Am. Acad. Audiol., **24**, 214-230.

Olthoff, C.M., Schouten, J.S., van de Borne, B.W., and Webers, C.A. (**2005**). "Non-compliance with ocular hypotensive treatment in patients with glaucoma or ocular hypertension: an evidence-based review," Opthalmology, **112**, 953-961.

Patel, A.D. (**2012**). "The OPERA hypothesis: assumptions and clarifications," Ann. N. Y. Acad. Sci., **1252**, 124-128.

Parbery-Clark, A., Skoe, E., and Kraus, N. (**2009**). "Musical experience limits the degradative effects of background noise on the neural processing of sound," J. Neurosci., **29**, 14100-14107.

Pichora-Fuller, M.K., and Singh, G. (**2006).** "Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation," Trends Amplif., **10**, 29-59.

Recanzone, G.H., Schreiner, C.E., and Merzenich, M.M. (**1993**). "Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys," J. Neurosci., **13**, 87-103.

Ross, M. (**1997**). "A retrospective look at the future of aural rehabilitation," J. Acad. Rehab. Audiol., **30**, 11-28.

Song, J., Skoe, E., Banai, K., and Kraus, N. (**2012**). "Training to improve speech in noise: Biological mechanisms;" Cereb. Cortex, **22**, 1180–1190.

Stacey, P., and Summerfield, A. (**2005**). "Auditory-perceptual training using a simulation of a cochlear implant system: A controlled study," ISCA Workshop on Plasticity in Speech Perception (PSP2005), London, UK, pp. 143-145.

Sweetow, R.W., and Sabes, J.H. (**2004**). "The case for LACE, individualized listening and auditory communication enhancement training," Hearing Journal, **57**, 32-40.

Sweetow, R.W., and Palmer, C.V. (**2005**). "Efficacy of individual auditory training in adults: a systematic review of the evidence;" J. Am. Acad. Audiol., **16**, 494-504.

Sweetow, R.W., and Sabes, J.H. (**2006**). "The need for and development of an adaptive Listening and Communication Enhancement (LACE) program," J. Am. Acad. Audiol., **17**, 538-558.

Sweetow, R.W., and Sabes, J.H. (**2010a**). "Auditory training and challenges associated with participation and compliance," J. Am. Acad. Audiol., **21**, 586-593.

Sweetow, R.W., and Sabes, J.H. (**2010b**). "The Communication Confidence Profile: A vital, but often overlooked domain," Hearing Journal, **63**, 17-24.

Tremblay, K., Kraus, N., Carrell, T.D., and McGee, T. (**1997**). "Central auditory system plasticity: generalization to novel stimuli following listening training," J. Acoust. Soc. Am., **102**, 3762-3773.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., and Otis, B. (**2001**). "Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training," Ear Hearing, **22**, 79-90.

Tremblay, K., Shahin, A., Picton, T., and Ross, B. (**2009**). "Auditory training alters the physiological detection of stimulus-specific cues in humans," **120**, 128-135.

Vasama, J.P., and Mäkelä, J.P. (**1995**). "Auditory pathway plasticity in adult humans after unilateral idiopathic sudden sensorineural hearing loss," Hear. Res., **87**, 132-140.

Ventry, I., and Weinstein, B. (**1982**). "The hearing handicap inventory for the elderly: a new tool," Ear Hearing, **3**, 128-134.

Vincent, P. (**1971**). "Factors influencing patient noncompliance: a theoretical approach," Nurs. Res., **20**, 509–516.

Wayner, D.S., and Abrahamson, J.E. (**1996**). *Learning to hear again: An audiologic rehabilitation curriculum guide.* Hear Again, Austin, TX.

Willott, J.F., Aitkin, L.M., and McFadden, S.L. (**1993**). "Plasticity of auditory cortex associated with sensorineural hearing loss in adult C57BL/6J mice," J. Comp. Neurol., **329**, 402-411.

Wong, P.C., Jin, J.X., Gunasekera, G.M., Abel, R., Lee, E.R., and Dhar, S. (**2009**). "Aging and cortical mechanisms of speech perception in noise," Neuropsychologia, **47**, 693-703.

Wong, P.C., Ettlinger, M., Sheppard, J., Gunasekera, G., and Dhar, S. (**2010**). "Neuroanatomical characteristics and speech perception in noise in older adults," Ear Hearing, **31**, 471-479.

Cognitive aspects of auditory plasticity across the lifespan

MARY RUDNER[1,*] AND THOMAS LUNNER[1,2]

[1] *Linneaus Centre HEAD, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden*

[2] *Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark*

This paper considers evidence of plasticity resulting from congenital and acquired hearing impairment as well as technical and language interventions. Speech communication is hindered by hearing loss. Individuals with normal hearing in childhood may experience hearing loss as they grow older and use technical and cognitive resources to maintain speech communication. The short- and medium-term effects of hearing-aid interventions seem to be mediated by individual cognitive abilities and may be specific to listening conditions including speech content, type of background noise, and type of hearing-aid signal processing. Furthermore, some aspects of cognitive function may decline with age and there is evidence that age-related hearing impairment is associated with poorer long-term memory. It is not yet clear whether improving audition through hearing-aid intervention can prevent cognitive decline. Profound deafness from an early age implicates a set of critical choices relating to possible restoration of the auditory signal through the use of prostheses including cochlear implants and hearing aids as well as to mode of communication, sign or speech. These choices have an influence on the organization of the developing brain. In particular, while the cortex may display sensory reorganization in response the linguistic modality of choice, cognitive organization seems to prevail.

INTRODUCTION

For the majority of the population, speech is the main mode of communication. Because the auditory signal provides the main channel of speech reception, any impairment of the auditory system makes speech communication more difficult. This has consequences that differ according to the time of life at which hearing impairment occurs and the compensatory choices made by individuals with hearing impairment and their significant others. Hearing aids represent a technical form of compensation that acts directly on the auditory channel, while use of sign language is a sociocultural form of compensation that is independent of the need for auditory processing. Both technical and sociocultural compensation may cause plasticity of the neurocognitive mechanisms that support communication. Further, the nature and degree of any such plasticity may depend both on the timing and efficiency of compensation as well as the onset, nature, progression, and severity of hearing loss.

*Corresponding author: mary.rudner@liu.se

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

SPEECH COMMUNICATION UNDER ADVERSE CONDITIONS

Hearing loss is just one of a wide range of suboptimal or adverse conditions for speech communication that include unfamiliar language, unfamiliar speaker characteristics and signal degradation as a result of external noise and reverberation, and internal adverse conditions as a consequence of hearing impairment such as masking, filtering, and distortion as well as the individual cognitive limitations of the listener, including fatigue and cognitive load (Mattys *et al.*, 2012). It goes without saying that each and any of these additional adverse conditions may make speech communication even more problematic for the listener with hearing loss, for whom the target signal is already attenuated and distorted as a consequence of physiological degeneration. Loss of sensitivity can often be compensated by amplification and parts of the distortion (the abnormal loudness growth imposed by sensorineural hearing impairment) may be compensated by non-linear signal processing. However, even with hearing aids, the segregation abilities of persons with hearing impairment are not as good as those of persons with normal hearing. Furthermore, the technologies they use to optimise hearing may generate additional distortion of the speech signal. Thus, although hearing aids may ameliorate some adverse conditions, others they cannot influence; indeed hearing aids may even generate adverse conditions of their own. When speech communication takes place under adverse conditions, high level cognitive resources such as working memory (WM) are brought into play.

WORKING MEMORY FOR COMMUNICATION

WM is the ability to keep relevant information in mind briefly while at the same time processing it. This ability is fundamental to many mental activities including language processing. For example, to achieve comprehension, individual words may have to be kept in mind until a particular statement is complete. WM capacity is limited and may differ substantially between individuals. Even under ideal conditions, most people cannot retain more than about seven unrelated words or other items of information (Miller, 1956), while some exceptional individuals may retain as few as five and others as many as nine. Short-term retention of words, which is part and parcel of language comprehension, is often conceived of in terms of the phonological loop of WM (Baddeley, 1986). Loop capacity can be measured using simple span tests such as digit span in which spoken digits in series of increasing length are presented for immediate serial recall until performance breaks down. However, simple span tests which merely tap individual storage capacity tend not to be predictive of the ability to perform challenging language tasks (Unsworth and Engle, 2007). On the other hand, complex span tests such as reading span (Daneman and Carpenter, 1980), which require simultaneous storage and processing capacity, are reliable predictors of language processing under challenging conditions, probably because they demand the ability to strategically deploy cognitive resources online. Current versions of the task (Rönnberg *et al.*, 1989) typically require first a semantic judgment of each sentence in a set of sentences followed by cued recall of the words occurring at a particular position in each sentence in the set. As set size increases, more storage is required while ongoing semantic processing competes for limited resources. In a review article, Akeroyd (2008) identified the reading span task as a good cognitive predictor of the ability to understand speech in noise, especially in older individuals with hearing impairment.

Recent models of working memory (Baddeley, 2012; Rönnberg *et al.*, 2013) are characterized by an episodic buffer component whose function is the integration and processing of multimodal representations based on input from multiple sources including the senses and long term memory (Rudner and Rönnberg, 2008). Whereas Baddeley's (2012) model is a general working memory model, the WM model for Ease of Language Understanding (ELU, Rönnberg *et al.*, 2013) specifically addresses cognition for communication. The ELU model proposes that the episodic buffer deals with Rapid, Automatic, Multimodal integration of PHOnology and is thus referred to as RAMBPHO. RAMBPHO function is smooth when communication conditions are optimal, and as a result, speech understanding proceeds rapidly and automatically. However, when adverse conditions prevail, explicit, or consciously recruited, cognitive processing resources are brought into play. Thus, speech communication relies not only on an efficient RAMBPHO but also on the ability to strategically deploy explicit processing resources. It is this dual ability that is tapped by the reading span task.

HEARING AIDS AND SPEECH PERCEPTION IN NOISE

Hearing aids are designed to help persons with hearing loss hear better. One of the technologies used to achieve this is Wide Dynamic Range Compression (WDRC) that compensates the abnormal growth of loudness resulting from sensorineural hearing loss. However, speech intelligibility may not be improved if the parameters of the WDRC scheme do not suit the characteristics of the individual (Lunner *et al.*, 2009). One critical individual characteristic seems to be WM capacity. Ten years ago it was established that the benefit obtained from WDRC was contingent on an interaction between cognitive ability and the time-constants of the compression system (Gatehouse *et al.*, 2003; Lunner, 2003). Since then, it has been shown that this relationship is influenced by type of background noise (Foo *et al.*, 2007; Lunner and Sundewall-Thorén, 2007; Rudner *et al.*, 2008) and the type of target speech material (Foo *et al.*, 2007; Rudner *et al.*, 2009; 2011). The combination of modulated noise and fast-acting compression seems to provide a particular challenge to cognitive resources (Lunner and Sundewall-Thorén, 2007; Rudner *et al.*, 2008; 2009; 2011; 2012) especially when the predictability of the target speech is low (Rudner *et al.*, 2011). Furthermore, these complex relations change over time (Cox and Xu, 2010; Rudner *et al.*, 2009; 2011) suggesting plasticity. In particular, it seems that the disadvantage of WRDC initially experienced by persons with low WM capacity may become less apparent after a period of familiarization (Rudner *et al.*, 2011). This suggests that persons with lower WM may experience more plastic change than persons with high working-memory capacity.

The work reviewed here, relating to the role of cognition in WDRC benefit, was conducted by investigating the relation between independently-measured cognitive capacity and speech reception thresholds measured in the traditional manner at relatively poor signal-to-noise ratios (SNR). The disadvantage of this approach is that although poor SNRs may occur in exceptional circumstances they are not representative of everyday communication (e.g., Smeds *et al.*, 2012) and thus may be misleading in terms of day-to-day functioning. There is much more to communication than just perceiving target speech. Above all, the message has to be understood and retained for further processing. Thus, in order to determine the efficacy of hearing-aid signal processing in terms of everyday communication it may be more useful to assess the ability to retain and process audible information presented in a relatively low level of background noise. This ability may be termed Cognitive Spare Capacity (CSC, Mishra *et al.*, 2010).

COGNITIVE SPARE CAPACITY

Sarampalis *et al.* (2009) showed that hearing-aid signal processing in the form of noise reduction can improve retention of heard speech in adults with normal hearing thresholds. This finding was recently extended to persons with hearing impairment (Ng *et al.*, 2013a). Experienced hearing-aid users listened to sets of sentences with high intelligibility and repeated the final word of each sentence. At the end of each set, they were prompted to recall all those words. Despite high intelligibility, background noise disrupted recall ability. However, the noise reduction processing (Wang *et al.*, 2009) reduced the negative effect of noise on recall. This effect was particularly marked for participants with good WM capacity and for sentence final words that occurred towards the end of each sentence set. A follow-up study replicated the positive effects of noise reduction on memory for sentence final words (Ng *et al.*, 2013b) and showed that this effect was similar in magnitude to that obtained by replacing native-language competing talkers by foreign-language (Chinese) talkers. The follow-up study also showed that when the memory load was reduced by decreasing sentence set size, beneficial effects of noise reduction generalized to individuals with lower WM capacity. These findings show that hearing-aid signal processing can improve retention of heard information, even when intelligibility is good, and demonstrate the need for new tools to study CSC (Rudner and Lunner, 2013).

The Cognitive Spare Capacity Test (CSCT, Mishra *et al.*, 2013a; 2013b) was developed to meet this need. In particular, it provides a tool to measure the ability to maintain and process intelligible information. In the CSCT, sets of spoken two-digit numbers are presented and the participant is required to report back at least two of those numbers depending on specific instructions designed to elicit executive processing of those numbers. Two executive processes are targeted: updating and inhibition. These two particular executive processes are likely to be engaged during speech understanding in adverse conditions. Updating ability is likely to be required to strategically replace the contents of WM with relevant material while inhibition ability is likely to be brought into play to keep irrelevant information out of WM. In the CSCT, WM load is manipulated by requiring participants to hold an additional dummy number in mind during high-load conditions. In everyday interaction, visual information can enhance speech perception by several dB and to determine the influence of visual cues on CSC, the CSCT manipulates whether the talker's face is visible or not. Finally, the CSCT can be administered in quiet or in noise. Results of studies employing this paradigm are beginning to delineate the nature of CSC (Mishra *et al.*, 2013a; 2013b; Rudner *et al.*, 2013b). For adults with normal hearing, provision of visual cues actually reduces performance in quiet conditions, probably because visual cues provide superfluous information that causes distraction when target information is highly intelligible (Mishra *et al.*, 2013a; 2013b). However, in noisy conditions, visual cues do not reduce performance, probably because they help segregate the target signal, resulting in richer cognitive representations (Mishra *et al.*, 2013b). At high intelligibility levels, steady-state noise reduces CSCT performance when visual cues are not provided, but modulated noise does not reduce performance for adults with normal hearing (Mishra *et al.*, 2013b). Older adults with mild hearing loss demonstrate lower CSC than young adults, even with individualised amplification, and this effect is most notable in noise and when memory load is high (Rudner *et al.*, 2013b). Visual cues do not reduce performance for this group. Interestingly, although CSC and WM do not seem to be strongly related, there is evidence that age-related differences in WM and executive function do influence CSC (Rudner *et al.*, 2013b).

PHONOLOGICAL DISTINCTIVENESS

We have seen that the RAMBPHO component of the ELU model deals with phonological integration (Rönnberg *et al.*, 2013). Phonology refers to the sublexical structure of language and is manifest in the sound patterns of speech and corresponding cognitive representations in the mental lexicon. Equivalent representations based on the gestural patterning of sign language suggest phonology can be understood at an abstract level (MacSweeney *et al.*, 2013). Access to the mental lexicon is faster when the phonological representation is more distinct because of fewer phonological neighbours (Luce and Pisoni, 1998). Severe hearing impairment may lead to more diffuse representation of speech phonology in the long term reflected in poorer visual rhyme judgement ability (Andersson, 2002; Classon *et al.*, 2013c) and verbal fluency (Classon *et al.*, 2013a). Individuals with poor phonological representations due to severe long-term hearing impairment can compensate for this deficit by good WM capacity measured by reading span performance (Classon *et al.*, 2013c). An early ERP signature of hearing loss was recently found in just such a task, likely reflecting use of a compensatory strategy, involving increased reliance on explicit mechanisms (Classon *et al.*, 2013b). However, this compensation comes at the cost of poorer long-term storage (Classon *et al.*, 2013c).

SEMANTIC CONTEXT

Language understanding is about grasping the gist of the message. Use of available semantic context can facilitate speech understanding under adverse conditions and has been shown to recruit language processing networks in left posterior inferior temporal cortex and inferior frontal gyri bilaterally (Rodd *et al.*, 2005). Rudner *et al.*, (2011) found that although the role of WM in speech understanding with WDRC was clearly apparent with matrix-type sentences (Hagerman and Kinnefors, 1995) this was not the case with Swedish Hearing In Noise Test (HINT) sentences (Hällgren *et al.*, 2006). Although the Hagerman sentences are semantically coherent they have low ecological validity; the five-word syntactic structure is always identical and each individual word comes from a closed set of ten items, but no particular item can be predicted from sentence context. The HINT sentences, on the other hand, range in length and syntactic structure as well as semantic coherence. It was suggested that the low redundancy of the Hagerman sentences increases reliance on the details in the speech signal.

THE AGING BRAIN

Cognitive function declines with advancing age and the mechanisms behind this have been traced to both genetic and lifestyle factors (Nyberg *et al.*, 2012). Sensory functions also decline with age and there are several different theories explaining the relation between sensory and cognitive decline. The common cause hypothesis (Baltes and Lindenberger, 1997) suggests that a general reduction in the efficiency of physiological function drives both phenomena, while the information degradation hypothesis (Schneider *et al.*, 2002) suggests that cognitive processes function less efficiently when sensory input is less well defined due to declining sensory function. The Compensation-Related Utilization of Neural Circuits Hypothesis (Reuter-Lorenz and Cappell, 2008) suggests that older adults compensate for less effective use of neural resources, such as the prefrontal cortex, by engaging them at lower task loads than younger adults. However, potential activation levels in these regions are lower for older compared to younger individuals. This suggests that any factor that can reduce cognitive load during speech understanding under adverse conditions, including good hearing-aid fitting, phonological distinctness, and semantic context, is likely to become even more important with advancing age. This is supported by emerging results relating to CSC (Rudner *et al.*, 2013b).

It is important to gain an objective understanding of the link between sensory and cognitive function from a rehabilitation perspective. If hearing impairment drives cognitive decline then auditory rehabilitation becomes doubly important: satisfactory treatment of hearing loss may not only improve speech communication but also be able to prevent cognitive decline. There is accumulating evidence of a specific association between hearing impairment and cognition. Epidemiological studies show that individuals with hearing loss are at increased risk of cognitive impairment and that rate of cognitive decline as well as risk of cognitive impairment are associated with severity of hearing loss (Lin *et al.*, 2013). Analysis of data from hearing-aid users participating in the Betula study of cognitive aging (Nilsson *et al.*, 1997) showed that individuals with more hearing loss had poorer long term memory and that this cognitive deficit was not restricted to the visual domain (Rönnberg *et al.*, 2011). Importantly there was no significant association between loss of vision and cognitive function or between hearing loss and WM. These findings show that there is a link between sensory loss specifically in the auditory domain and cognitive decline that is limited to long-term memory without affecting WM.

MODALITY SPECIFICITY

We have seen that both acquired hearing impairment and hearing-aid use may result in changes in neurocognitive representation that may be susceptible to neurocognitive compensation. We have also seen that WM capacity modulates the neurocognitive processes involved in phonological processing and the integration of contextual information during speech understanding under adverse conditions. Further, there is a link between sensory and cognitive status with advancing age that seems to be specific in the sense that it is related to the auditory channel at a sensory level and to a modality-general long-term memory system. However, it is not clear how auditory deprivation and cognitive experience mediate this relation at the end of the lifespan. On the other hand, there is evidence that both auditory deprivation and cognitive experience drive neural plasticity during early development.

Parents of deaf children are faced by a set of critical choices. These include technical interventions influencing sensory input in the auditory domain and mode of communication. Hearing aids and cochlear implants can provide auditory input to facilitate development of speech communication but sign language provides a mode of communication that can develop independently of the auditory channel given adequate communicative input. Auditory deprivation as a result of congenital deafness results in recruitment of auditory cortex for visual processing (Fine *et al.*, 2005; Lomber *et al.*, 2010). However, it was not clear until very recently how auditory deprivation, on the one hand, and language choice, on the other, contribute to cortical plasticity. Cardin *et al.* (2013) dissociated these factors in a study that included two groups of congenitally profoundly-deaf adults: one group consisted of native sign-language users, that is, persons born into deaf families where signing was used as regular means of communication, and the other group who used speech communication and had no knowledge of sign language. Speakers with normal hearing constituted a reference group. All participants were scanned using fMRI while watching a model signing. The experimental design allowed separation of the effects of auditory deprivation and language experience. It was found that while sign-language experience drove recruitment of superior temporal cortex in both cerebral hemispheres, auditory deprivation drove recruitment of this region in the right hemisphere only. This shows that, although auditory deprivation from birth leads to a change in the sensory function of superior temporal cortex, the left lateralized cognitive function of language processing is preserved. Other evidence shows largely similar neural organization of cognitive function for sign and speech with some language-modality-specific differences that may be attributable to sensorimotor differences but also to modality-specific differences in the relationship between phonological and semantic processing (Rudner *et al.*, 2013a).

CONCLUSION

A wide range of factors conspire to make communication more or less successful across the lifespan. Hearing loss may hinder speech communication and this may be compounded by other adverse conditions. However, properly fitted hearing aids, phonological distinctness, and semantic context may all support speech understanding under adverse conditions. Limited cognitive resources are used in the very act of listening, and thus any factors supporting speech understanding may reduce cognitive load, effectively increasing cognitive spare capacity. Supporting speech understanding and thus reducing cognitive load is probably particularly important in older adults. Evidence suggests that given adequate experience, the neurocognitive organisation of congenitally deaf adults who are native signers is similar to that of adults with normal hearing who use speech communication. Neurocognitive organisation in deaf native signers, who do not experience age-related auditory decline, may provide a useful benchmark for understanding the complex interactions between age-related sensory and cognitive decline as well as audiological, cognitive, and social interventions aimed at supporting speech communication.

REFERENCES

Akeroyd, M.A. (**2008**). “Are individual differences in speech perception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing impaired adults,” Int. J. Audiol. Suppl., **47**, S125-S143.

Andersson, U. (**2002**). “Deterioration of the phonological processing skills in adults with an acquired severe hearing loss,” Eur. J. Cogn. Psychol., **14**, 335-352.

Baddeley, A. (**1986**). *Working memory* (Oxford: Clarendon Press).

Baddeley, A. (**2012**). “Working memory: Theories, models, and controversies,” Ann. Rev. Psychol., **63**, 1-29.

Baltes, P., and Lindenberger, U. (**1997**). “Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging?” Psychol. Aging, **12**, 12-21.

Cardin, V., Orfanidou, E., Rönnberg, J., Capek, C.M., Rudner, M., and Woll, B. (**2013**). “Dissociating cognitive and sensory neural plasticity in human superior temporal cortex,” Nature Communications, **4**, 1473.

Classon, E., Löfkvist, U., Rudner, M., and Rönnberg, J. (**2013a**). “Verbal fluency in adults with postlingually acquired hearing impairment,” Speech, Language and Hearing, Available online, DOI:10.1179/2050572813Y.0000000019.

Classon, E., Rudner, M., Johansson, M., and Rönnberg, J. (**2013b**). “Early ERP signature of hearing impairment in visual rhyme judgment,” Front. Auditory Cogn. Neurosci., **4**, 241.

Classon, E., Rudner, M., and Rönnberg, J. (**2013c**). “Working memory compensates for hearing related phonological processing deficit,” J. Comm. Dis., **46**, 17-29.

Cox, R.M., and Xu, J. (**2010**). “Short and long compression release times: speech understanding, real world preferences, and association with cognitive ability,” J. Am. Acad. Audiol., **21**, 121-138.

Daneman, M., and Carpenter, P.A. (**1980**). “Individual differences in working memory and reading,” J Verb. Learn. Verb. Be., **19**, 450-466.

Fine, I., Finney, E.M., Boynton, G.M., and Dobkins, K.R. (**2005**). “Comparing the effects of auditory deprivation and sign language within the auditory and visual cortex,” J. Cog. Neurosci., **17**, 1621-1637.

Foo, C., Rudner, M., Rönnberg, J., and Lunner, T. (**2007**). ”Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity,” J. Am. Acad. Audiol., **18**, 553-566.

Gatehouse, S., Naylor, G., and Elberling, C. (**2003**). “Benefits from hearing aids in relation to the interaction between the user and the environment,” Int. J. Audiol. **42**, S77–S85.

Hagerman B., and Kinnefors C. (**1995**). “Efficient adaptive methods for measuring speech reception threshold in quiet and in noise,” Scand. Audiol., **24**, 71-77.

Hällgren, M., Larsby, B., and Arlinger, S.A. **(2006)**. “Swedish version of the hearing in noise test (HINT) for measurement of speech recognition,” Int. J. Audiol., **45**, 227-237.

Lin, F.R., Yaffe, K., Xia, J., Xue, Q.L., Harris, T.B., Purchase-Helzner, E., Satterfield, S., Ayonayon, H.N., Ferrucci, L., and Simonsick, E.M.. (**2013**). “Hearing loss and cognitive decline in older adults,” JAMA Intern. Med., **173**, 293-299.

Lomber, S.G., Meredith, M.A., and Kral, A. (**2010**). “Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf,” Nat. Neurosci., **13**, 1421-1427.

Luce, P.A., and Pisoni, D.B. (**1998**). “Recognizing spoken words: the neighborhood activation model,” Ear Hearing, **19**, 1-36.

Lunner T. (**2003**). “Cognitive function in relation to hearing aid use,” Int. J. Audiol., **42**, S49-S58.

Lunner, T., and Sundewall-Thorén, E. (**2007**). “Interactions between cognition, compression, and listening conditions: effects on speech-in-noise performance in a two-channel hearing aid,” J. Am. Acad. Audiol., **18**, 604-617.

Lunner, T., Rudner, M., and Rönnberg, J. (**2009**). “Cognition and hearing aids,” Scand. J. Psychol., **50**, 395-403.

MacSweeney, M., Goswami, U. and Neville, H. (**2013**). “The neurobiology of rhyme judgment by deaf and hearing adults: An ERP study,” J. Cogn. Neurosci., **25**, 1037-1048.

Mattys, S.L., Davis, M.H., Bradlow, A.R., and Scott, S.K. (**2012**). “Speech recognition in adverse conditions: A review,” Lang. Cogn. Proc., **27**, 953-978.

Miller, G.A. (**1956**). “The magic number seven, plus or minus two: Some limits on our capacity for processing information,” Psychol. Rev., **63**, 81-93.

Mishra, S., Rudner, M., Lunner, T., and Rönnberg, J. (**2010**). “Speech understanding and cognitive spare capacity,” in *Binaural processing and spatial hearing.*

Edited by J.M. Buchholz, T. Dau, J.C. Dalsgaard and T. Poulsen (ISAAR: Elsinore, Denmark), pp. 305-313.

Mishra, S., Lunner, T., Stenfelt, S., Rönnberg, J., and Rudner, M. (**2013a**). "Visual information can hinder working memory processing of speech," J. Speech Lang. Hear. Res., **56**, 1120-1132.

Mishra, S., Lunner, T., Stenfelt, S., Rönnberg, J., and Rudner, M. (**2013b**). "Executive processing at high speech intelligibility levels in adults with hearing loss: A measure of cognitive spare capacity," under review.

Ng, E.H.N., Rudner, M., Lunner, T., Syskind Pedersen, M., and Rönnberg, J. (**2013a**). "Improved cognitive processing of speech for hearing aid users with noise reduction," Int. J. Audiol., **52**, 433-441.

Ng, E.H.N., Rudner, M., Lunner, T., and Rönnberg, J. (**2013b**). "Noise reduction improves memory for target language speech in competing native but not foreign language speech," under review.

Nilsson, L.-G., Bäckman, L., Erngrund, K., Nyberg, L., Adolfsson, R., Bucht, G., Karlsson, S., Widing, M., and Winblad, B. (**1997**). "The Betula prospective cohort study: Memory, health, and aging," Aging Neuropsychol. Cogn., **4**, 1-32.

Nyberg, L., Lövdén, M., Riklund, K. Lindenberger, U, and Bäckman, L. (**2012**). "Memory aging and brain maintenance;" Trends Cogn. Sci., **16**, 292-305.

Reuter-Lorenz, P.A., and Cappell, K.A. "Neurocognitive aging and the compensation hypothesis," Curr. Direct. Psychol. Sci., **17**, 177-182.

Rodd, J.M., Davis, M.H., and Johnsrude, I.S. (**2005**). "The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity," Cer. Cor., **15**, 1261-1269.

Rönnberg, J., Arlinger, S., Lyxell, B., and Kinnefors, C. (**1989**). "Visual evoked potentials: relation to adult speechreading and cognitive function," J. Speech Lang. Hear. Res., **32**, 725-735.

Rönnberg, J., Danielsson, H., Rudner, M., Arlinger, S., Sternäng, O., Wahlin, Å., and Nilsson, L-G. (**2011**). "Hearing loss is negatively related to episodic and semantic long-term memory but not to short-term memory," J. Speech Lang. Hear. Res., **54**, 705-726.

Rönnberg, J., Lunner, T., Zekveld, A.A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M.K., and Rudner, M. (**2013**). "The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances," Front. Systems Neurosci., **7**, 31.

Rudner, M., Foo, C., Sundewall Thorén, E., Lunner, T., and Rönnberg, J. (**2008**). "Phonological mismatch and explicit cognitive processing in a sample of 102 hearing aid users," Int. J. Audiol., **47**, S163-S170.

Rudner, M., and Rönnberg, J. (**2008**). "The role of the episodic buffer in working memory for language processing," Cogn. Proc., **9**, 19-28.

Rudner, M., Foo, C., Rönnberg, J., and Lunner, T. (**2009**). "Cognition and aided speech recognition in noise: specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids," Scand. J. Psychol., **50**, 405-418.

Rudner, M., Rönnberg. J., and Lunner, T. (**2011**). ”Working memory supports listening in noise for persons with hearing impairment,” J. Am. Acad. Audiol., **22**, 156-167.

Rudner, M., Lunner, T., Behrens, T., Sundewall Thorén, E., and Rönnberg, J. (**2012**). “Working memory capacity may influence perceived effort during aided speech recognition in noise;” J. Am. Acad. Audiol., **23**, 577-589.

Rudner, M., Karlsson, T., Gunnarsson, J., and Rönnberg, J. (**2013a**). “Levels of processing and language modality specificity in working memory,” Neuropsychologia, **51**, 656-666.

Rudner, M., and Lunner, T. (**2013**). “Cognitive spare capacity as a window on hearing aid benefit,” Seminars in Hearing, **34**, 298-307.

Rudner, M., Mishra, S. Stenfelt, S., Lunner, T., and Rönnberg, J. (**2013b**). “Age-related individual differences in working memory capacity and executive ability influence cognitive spare capacity,” Aging and Speech Communication, Indiana University, Bloomington, Indiana, October 6-9 2013.

Sarampalis, A., Kalluri, S, Edwards, B., and Hafter, E. (**2009**). “Objective measures of listening effort: Effects of background noise and noise reduction,” J. Speech Lang. Hear. Res., **52**, 1230-1240.

Schneider, B.A., Daneman, M., and Pichora-Fuller, M.K. **(2002)**. “Listening in aging adults: from discourse comprehension to psychoacoustics,” Can. J. Exp. Psychol., **56**, 139-152.

Smeds, K., Wolters, F., and Rung, M. (**2012**). “Estimation of realistic signal-to-noise ratios,” International Hearing Aid Research Conference 2012 (IHCON), Lake Tahoe, California, August 8-12, 2012.

Unsworth, N., and Engle, R.W. (**2007**). “On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities,” Psychol. Bull., **133**, 1038-1066.

Wang, D., Kjems, U., Pedersen, M.S., Boldt, J.B., and Lunner, T. (**2009**). “Speech intelligibility in background noise with ideal binary time-frequency masking,” J. Acoust. Soc. Am., **125**, 2336-2347.

Auditory training strategies for adult users of cochlear implants

PAULA STACEY[1,*] AND QUENTIN SUMMERFIELD[2]

[1] *Division of Psychology, Nottingham Trent University, Nottingham, England*

[2] *Department of Psychology, University of York, York. England*

There has been growing interest recently in whether computer-based training can improve speech perception among users of cochlear implants (Fu *et al*., 2005; Oba *et al*., 2011; Ingvalson *et al*., 2013). This paper reports a series of experiments which first evaluated the effectiveness of different training strategies with normal-hearing participants who listened to noise-vocoded speech, before conducting a small-scale study with users of cochlear implants. Our vocoder studies revealed (1) that 'High-Variability' training led to greater generalisation to new talkers than training with a single talker, and (2) that word- and sentence-based training materials led to greater improvements than an approach based on phonemes in nonsense syllables. Informed by these findings, we evaluated the effectiveness of a computer-based training package that included word- and sentence-based tasks, with materials recorded by 20 talkers. We found good compliance with the training protocol, with 8 out of the 11 participants completing 15 hours of training as instructed. Following training, there was a significant improvement on a consonant test, but in general the improvements were small, highly variable, and not statistically significant. A large-scale randomised controlled trial is needed before we can be confident that computer-based auditory training is worthwhile for users of cochlear implants.

INTRODUCTION

Cochlear implantation is highly effective at improving speech perception among adults with severe to profound hearing impairment. Developments in cochlear implant technology are constantly being made, with improved speech processing strategies and electrode design enhancing outcomes for adults who receive cochlear implants. In addition to these developments, auditory training is another way in which outcomes for cochlear implant users can possibly be maximised. By the mid-1990s, most hospital-based cochlear-implant programmes in the UK had ceased to provide an extensive amount of face-to-face aural rehabilitation to adult patients because there was little evidence in support of its effectiveness (e.g., Gagne *et al*., 1991). However, interest in auditory training was revived by studies in US by Gfeller *et al*. (2002) and Fu *et al*. (2005) which suggested that intensive computer-based auditory training can improve the timbre-recognition and speech-perception

[*] Corresponding author: ....

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

skills of cochlear-implant users. Further research has added to these findings in recent years, and suggests that computer-based training may be valuable for adults who use cochlear implants (Ingvalson *et al*., 2013; Oba *et al*., 2011; Zhang *et al*., 2012) and hearing aids (Sweetow and Sabes, 2006; Olson *et al*., 2013).

An important consideration when evaluating the effectiveness of training regimes is to separate training-related learning from 'incidental' learning. Incidental learning refers to improvements that occur independent of the auditory training task, through procedural learning of task demands (Robinson and Summerfield, 1996), or perceptual learning resulting from repeated exposure to test materials. A further type of incidental learning has also been documented. Amitay *et al*. (2006) reported larger improvements in frequency discrimination for control participants who played a purely visual computer game (Tetris) between successive tests than for control participants who did not engage in an intervening task. These results suggest that maintaining attention and arousal, without explicit training, may be sufficient to lead to improvements on perceptual tasks. In order to evaluate the extent to which a training task has contributed to improvements in performance, it is therefore important to factor out improvements related to 'incidental learning'.

Although many of the studies into the effectiveness of computer-based training for users of cochlear implants have produced positive results, there is no consensus about how training materials should be structured to achieve maximal effectiveness. In this paper, we will (1) summarise results from a series of simulation studies which evaluated the effectiveness of different training strategies, before (2) assessing whether there is any evidence that our training materials are effective for users of cochlear implants.

SIMULATION STUDIES: HOW SHOULD TRAINING MATERIALS BE DESIGNED?

In this section, we will discuss the results from some of our studies with normal-hearing participants who listened to speech processed by a noise-excited vocoder (Shannon *et al*., 1995). A noise-excited vocoder can be used to mimic the speech processing that occurs in a cochlear implant system, and it allows the spectral and temporal information that is transmitted to the listener to be manipulated. These studies allowed us to carry out controlled experiments that compared the effectiveness of different training strategies, in a way which would have been difficult with users of cochlear implants themselves. Specific issues that we addressed were (1) whether variability is needed in the speech materials, and (2) which training strategies appear to be most effective.

Variability in training materials

One issue in designing training materials for use by cochlear-implant users is whether it is important to incorporate variability in the training materials. Previous research from Lively and colleagues (Lively *et al*., 1993) conducted with Japanese Americans learning to distinguish between /r/ and /l/ has suggested that training with several talkers is more effective than training with a single talker. For example, Lively *et al*. (1993) found that when training materials were recorded by five talkers there was significant generalisation to new talkers, but this was not the case when training materials were recorded by a single talker. It is important to know whether variability is an important consideration for cochlear-implant users, who listen to speech with reduced spectral and temporal cues which are important in differentiating between talkers (Fu *et al*., 2004).

Our study

We (Stacey and Summerfield, 2007) undertook a study to investigate whether variability is important. Experiment 1 included sixteen participants who listened to speech processed by an 8-channel noise-excited vocoder (frequency range 350 to 5500 Hz). To make the speech more difficult to understand, and following Rosen *et al*. (1999), signals were spectrally-shifted upwards to simulate a tonotopic misalignment of 6 mm according to the Greenwood (1990) function. This meant that the centre frequencies of the 8 bands were shifted upwards by approximately one and a half octaves.

Auditory training was provided by a 2-alternative forced-choice task. At the start of each trial, two words were presented orthographically on the left and right of the touch screen. The target word was then presented acoustically. Participants responded by touching the word corresponding to the target. Visual feedback on accuracy was given. In the 'High-Variability' (H-V) conditions, materials were recorded by 10 talkers, and in the 'Single-Talker' (S-T) conditions, materials were recorded by a single male talker. A visual control task was also implemented in order to separate training-related learning from 'incidental' learning. The visual control task was based on the same 2-alternative forced-choice task, but in this case participants responded to visually presented targets which were degraded by visual noise.

The experiment took place over the course of three days. During Test sessions 1, 2, and 3, participants completed tests of speech perception. There were 4 groups of participants. Groups 1 and 2 received H-V training, while groups 3 and 4 received S-T training. A cross-over design was used in which groups 1 and 3 received auditory training between Test Sessions 1 and 2, whilst groups 2 and 4 received auditory training between Test Sessions 2 and 3 (Table 1).

Figure 1 shows the results from this experiment. We can see that larger improvements in sentence perception were found following the auditory training task than following the visual control task. This difference was statistically significant (auditory training mean improvement: 7.98%, SD: 3.95%, visual control task mean improvement: 2.02%, SD: 4.46%, $t_{15} = 3.13$, $p < 0.01$).

Group	Day 1	Day 2		Day 3	
1	Test Session 1	H-V train	Test Session 2	Control	Test Session 3
2	Test Session 1	Control	Test Session 2	H-V train	Test Session 3
3	Test Session 1	S-T train	Test Session 2	Control	Test Session 3
4	Test Session 1	Control	Test Session 2	S-T train	Test Session 3

Table 1: Study design of Experiments 1 and 2 in Stacey and Summerfield (2007). H-V train = High-Variability auditory training, S-T train = Single-Talker auditory training.

However, Experiment 1 found no advantage for High-Variability training over Single-Talker training (High-Variability mean improvement: 7.16%, SD: 4.71%, Single-Talker mean improvement: 8.8%, SD: 3.09%, t_{14} = 0.83). We reasoned that this may have been because a spectral shift of 6 mm was too extreme to allow participants to differentiate between talkers. Therefore, Experiment 2 replicated Experiment 1, but simulated a tonotopic misalignment of 3 mm (signals were shifted upwards by approximately 3/4 octave). Thirty-two participants were recruited.

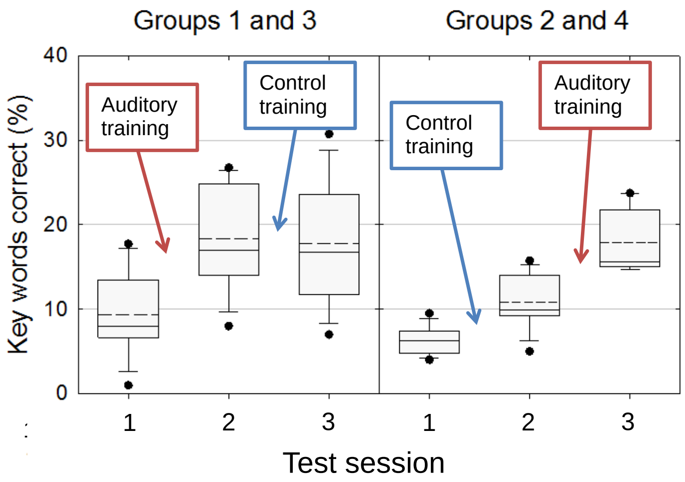


Fig. 1: Results from Stacey and Summerfield (2007), Experiment 1. Percentage of key words correctly identified in IEEE sentences according to test session and training group. The mean value is represented by the dashed line in the box, the median by the solid line. The box spans the interquartile range. Outliers are plotted as dots beyond the 10th-90th percentile whiskers.

Experiment 2 again found evidence that auditory training led to greater improvements in sentence perception than the control task, but it additionally found an advantage for High-Variability training over Single-Talker training (t_{30} = 2.38, $p < 0.05$, Fig. 2). The IEEE sentence test we used was recorded by 10 talkers, 5 of whom also recorded the training materials ('old' talkers), and 5 of whom did not ('new' talkers). The advantage of High-Variability over Single-Talker training was greater for the 'new' talkers (High-Variability mean improvement: 13.1%, Single-Talker mean improvement: 8.6%, t_{30} = 2.03, p = 0.051) than for the 'old' talkers (High-Variability mean improvement: 9.9%, Single-Talker mean improvement: 5.9%; t_{30} = 1.31, ns), thereby suggesting that High-Variability training leads to greater generalisation to new talkers than Single-Talker training.

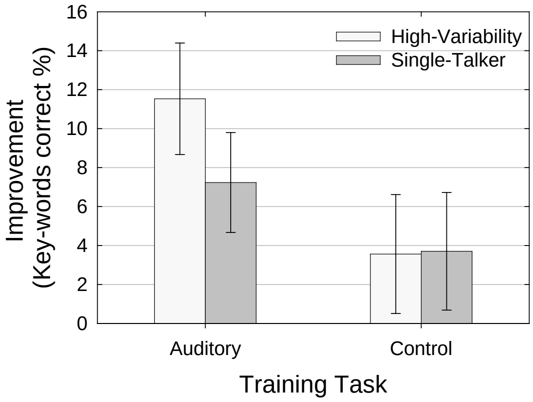


Fig. 2: Results from Stacey and Summerfield (2007), Experiment 2. Advantage of High-Variability training over Single-Talker training. Error bars indicate 95% confidence intervals.

This study therefore shows that there appear to be advantages for including materials recorded by several talkers, even when speech is noise-vocoded and spectrally shifted.

Different training strategies

Although we have found evidence that variability may be important when creating training materials for cochlear implant users, there remains uncertainty about what type of materials should be used for training. There is a long standing argument about whether training materials should be 'analytic' or 'synthetic'. Those who argue that training works best with analytic (bottom-up) approaches based on isolated phonemes claim that if people are trained with the basic building blocks of language, then greater generalisation to new materials can be obtained (e.g., Moore *et al.*, 2005). However, other studies which have found that lexical information plays an important role in the perceptual learning of speech (Norris *et al.*, 2003; Davis *et* *al.*, 2005) suggest that synthetic (top-down) approaches using word and sentence materials will work better.

These differing views on the best approach to training are reflected in the diversity of training materials used in recent studies into the effectiveness of computer-based training for users of hearing aids and cochlear implants. For example, the study by Fu *et al.* (2005) created materials based on phonemes in words, Oba *et al.* (2011) used digits in noise, and Stecker *et al.* (2006) based their training materials on nonsense syllables.

Our study

We (Stacey and Summerfield, 2008) compared the effectiveness of three different approaches to training, based on words, sentences, and phonemes (see below).

i. Word-based training

Two-alternative forced-choice task (as described previously). Materials were recorded by a single male talker.

ii. Sentence-based training

The sentence-based training task was designed as a computer-based Connected Discourse Tracking procedure (CDT, De Fillipo and Scott, 1978; Rosen *et al.*, 1999). The training task used IEEE sentences recorded by a single male talker. In this task, participants heard the target sentence, and their task was to decide which 3 out of 6 words displayed orthographically on the computer screen were in the sentence they had just heard.

iii. Phoneme-based training

For phoneme training we used Phonomena (Mindweavers, 2003; Moore *et al.*, 2005). We used 10 sets of sounds, each of which consisted of a continuum which ranged from one nonsense syllable to another (e.g., ‘i’ to ‘e’, ‘va’ to ‘wa’, ‘sa’ to ‘sha’). At the extremes of each continuum was a synthesized example of that sound, and these examples were acoustically warped into one another so that each continuum consisted of 96 stimuli. The training task consisted of an XAB two-alternative forced-choice procedure, in which participants heard a target sound (X) and were asked to decide which of two following sounds (A or B) was the same as the target. The training package increased or decreased the separation between stimuli adaptively in order to track 71% correct performance (see Stacey and Summerfield (2008) for more detail).

Eighteen participants took part in the study, and speech was processed by a vocoder which simulated a 6-mm shift. Training was provided during nine 20-minute sessions. Tests of speech perception were administered at the beginning of the study and following every hour (3 sessions) of training thereafter.

The main results are shown in Fig. 3. This figure shows the average overall improvement following the word-, sentence-, and phoneme-based training approaches on each of the tests of speech perception. Overall, we can see that there were larger improvements following word- and sentence-based training than following phoneme-based training. The word- and sentence-based approaches led to significant improvements on the BKB and IEEE sentence tests and the consonant test, while the only significant improvement following phonetic training was found on the consonant test.

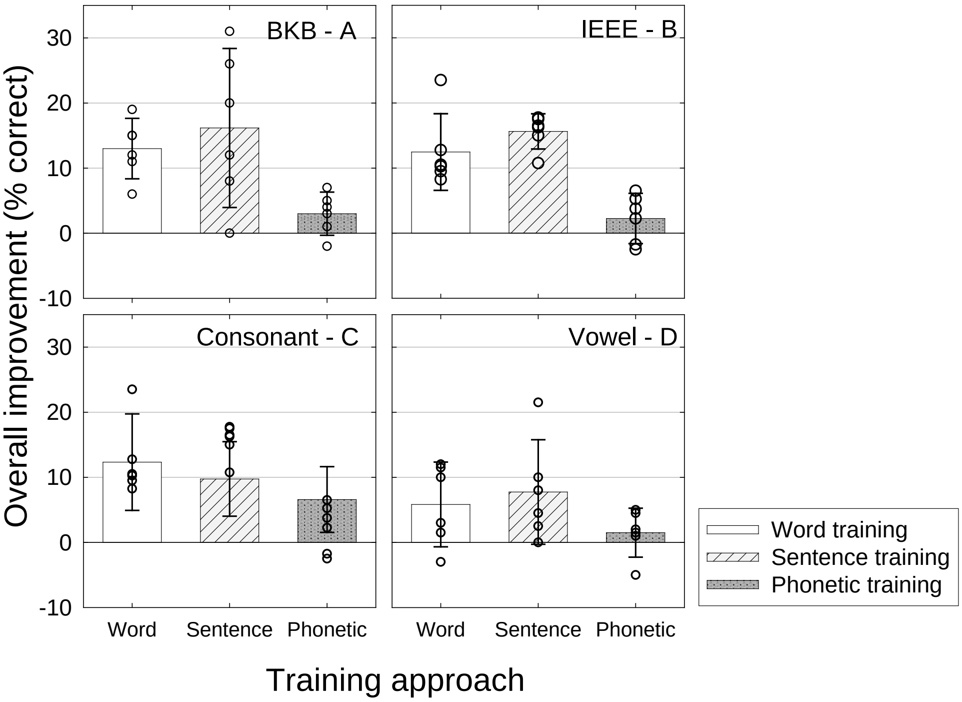


Fig. 3: Results from Stacey and Summerfield (2008). Overall improvements (following 3 hours of training) on the BKB sentence test (Panel A), the IEEE sentence test (Panel B), the consonant test (Panel C), and the vowel test (Panel D). Error bars indicate 95% confidence intervals, and improvements for individual participants are overlaid on the plots.

The approach to controlling for incidental learning we took in this experiment was to repeatedly administer these tests of speech perception at baseline, until an asymptote in performance was reached (see Fu *et al*., 2005). Participants' baseline level of performance was taken to be their highest level of performance during any of the baseline tests. Figure 4 shows the consequences of exercising this control, rather than taking baseline performance to be the first time participants completed the speech tests. We can see that if no control was exercised over the effects of repeated testing we would (1) find that improvements were much larger in magnitude, and (2) that many of the improvements would reach statistical significance.

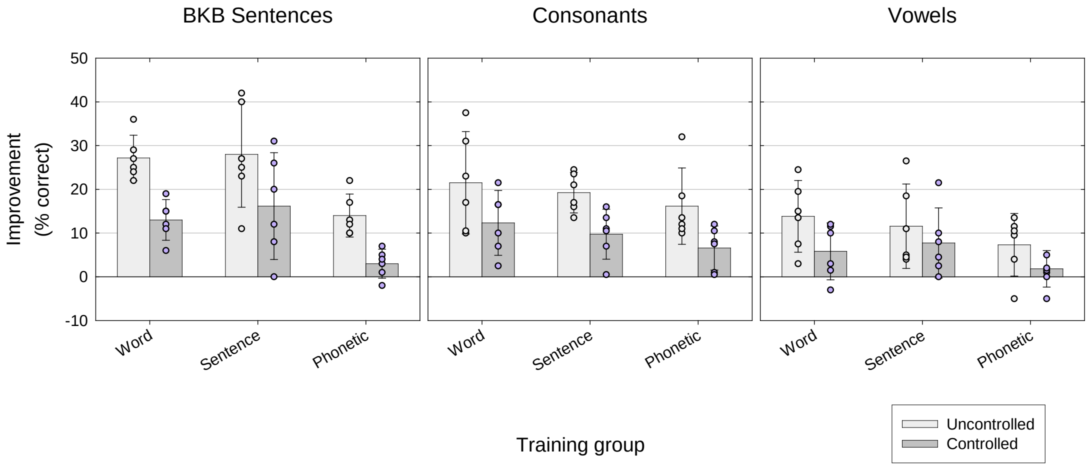


Fig. 4: Results from Stacey and Summerfield (2008). Improvements in performance on the BKB sentence test, the consonant test, and the vowel test. The white bars (uncontrolled) show the overall level of improvement between the first time tests were completed in the baseline session and the final testing session. The light grey bars (controlled) show the level of improvement between the 'highest baseline' and the final testing session. Error bars denote 95% confidence intervals.

The results from this study support the view that lexical information is important in the perceptual learning of speech (Norris *et al*., 2003; Davis *et al*., 2005). Our results are also consistent with the results from Faulkner *et al*. (2005) who found that training generalised best to similar tests, and that analytic training approaches did not lead to improvements in sentence perception.

EFFECTIVENESS OF TRAINING FOR COCHLEAR-IMPLANT USERS

Our study

Informed by our earlier results, we created a training package for use by adult users of cochlear implants (Stacey *et al*., 2010). The training package was designed to be user-friendly for people with limited experience of computers, and minimal computer skills were required. The training package contained the word- and sentence-based training materials, and materials were recorded by 20 talkers.

Eleven people who had used cochlear implants for over 3 years took part in the study (average duration of use: 5.73 years, SD: 2.69). They were aged between 23 and 71 years (average age: 58.82 years, SD: 18.89), and their latest score on the BKB sentence test recorded by the implant teams ranged from 34-81% correct (average BKB score: 59.82%, SD: 18.82).

Participants were asked to complete one hour of training a day, 5 days per week, over a period of three weeks (totalling 15 hours). Training was self-administered in participants' own homes, and they were visited by the first author at the beginning of the study who administered tests of speech perception repeatedly until an asymptote in performance was reached. Participants were then visited again following every week of training, when they completed further tests of speech perception and a questionnaire (the Glasgow Benefit Inventory, Robinson *et al.*, 1996) which measured whether training had benefitted participants in their everyday lives.

We found good compliance with the training protocol, with eight of the eleven participants completing 14-16 hours of training as instructed. The following analyses were limited to these 8 participants who completed the required amount of training.

The average overall improvements following three hours of training are shown in Fig. 5. We can see that there is a significant improvement consonant perception following training (mean improvement: 8.06%, SD: 6.90, $t_7 = 3.31$, $p < 0.05$), but the improvements did not reach significance for the other speech tests and there was a great deal of variability between participants.

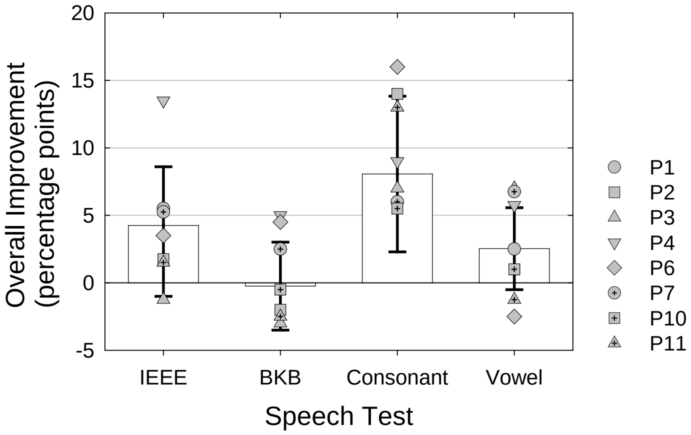


Fig. 5: Results from Stacey *et al.* (2010). Overall improvements following three hours of training on each test of speech perception. Error bars indicate 95% confidence intervals, and improvements by individual participants are overlaid on the plots.

Baseline levels of performance are shown in Table 2. There was large variability in participants' performance levels prior to training. For example, performance on the IEEE test ranged from 10% correct to 49%, and performance ranged from 20 to 61% correct on the consonant test. Given the diversity in the baseline level of performance of our participants, we tested whether there was an association between performance levels and levels of improvements (Table 2). These analyses revealed no significant associations, although it must be recognised that these analyses were limited by the small sample size.

Finally, there was no evidence that auditory training led to improvements in everyday life (as measured by the GBI).

		IEEE	*BKB*	*Consonant*	*Vowel*
Baseline performance	Average	22.20	48.69	43.84	47.47
	St. dev.	15.43	15.11	13.19	6.78
	Range	10 to 49	25 to 71	20 to 61	36 to 51
Correlations		$r_7 = 0.32$, $p = 0.48$	$r_8 = 0.63$, $p = 0.09$	$r_8 = 0.11$, $p = 0.80$	$r_8 = 0.50$, $p = 0.21$

Table 2: Average baseline levels of performance from Stacey and Summerfield (2010). The numbers indicate percentage correct, and baseline performance levels consisted of participants' average performance during the last two tests they completed in the first session. This table also shows the results of the correlation analyses.

Comparison with previous research

Overall, the results from this study do not give strong evidence that computer-based training is a worthwhile intervention for adult users of cochlear implants. This is in contrast to the generally positive outcomes reported by other studies (Fu *et al.*, 2005; Oba *et al.*, 2011; Zhang *et al.*, 2012; Ingvalson *et al.*, 2013). This disparity may be in part due to the highly variable levels of pre-training performance of the participants in our study, and more positive results may have been found if our participants had performed more poorly overall.

Evidence from other studies that auditory training can produce positive effects is encouraging, and these studies give useful insight into the type of training materials that might be effective for use by cochlear implant users. However, before we can be confident that auditory training is responsible for the improvements that have been found, a larger-scale randomised controlled study is needed (see Henshaw and Ferguson, 2013, for a review). The most common type of design used in this area to date has been a repeated measures design (Fu *et al.*, 2005; Oba *et al.*, 2011; Zhang *et al.*, 2012; Ingvalson *et al.*, 2013). Some studies have sought to control for the effects of incidental learning by employing the methodology we used here of repeatedly administering tests at baseline, so that participants act as their own controls (e.g., Fu *et al.*, 2005; Ingvalson *et al.*, 2013). However, this methodology makes the assumption that incidental learning is a short-term phenomenon, and is complete by the end of baseline testing. This methodology fails to take into account the fact that incidental learning might continue over the course of the study, and performance may improve because participants are engaged in *any* task, rather than one we would expect to improve speech perception. Ideally therefore, a control group who completes an alternative to auditory training should also be included.

CONCLUSION

Our simulation studies suggested that word- and sentence-based training materials recorded by several talkers will be most effective for use by cochlear-implant users. However, our small-scale study with cochlear-implant users found only minimal evidence of improvements following training. Although previous studies have found positive effects of training for users of cochlear implants, more robust evidence in the form of a large-scale randomised controlled trial is needed before we can be confident that computer-based auditory training is worthwhile for users of cochlear implants.

REFERENCES

Amitay, S., Irwin, A., and Moore, D.R. (**2006**). "Discrimination learning induced by training with identical stimuli," Nat. Neurosci., **9**, 1446-1448.

Davis, M.H., Hervais-Adelman, A., Taylor, K., McGettigan, C., and Jonsrude, I.S. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen., **134**, 222-241.

De Filippo, C.L., and Scott, B.L. (**1978**). "A method for training and evaluating the reception of ongoing speech," J. Acoust. Soc. Am., **63**, 1186-1192.

Faulkner, A., Rosen, S., and Jackson, A. (**2005**). "Relative effectiveness of training methods for shifted speech," Poster presented at the British Society of Audiology short-papers meeting, Cardiff, UK.

Fu, Q.-J., Chinchilla, S., and Galvin, J.J. (**2004**). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," J. Assoc. Res. Oto., **5**, 253–260.

Fu, Q.-J., Galvin, J.J., Wang, X., and Nogaki, G. (**2005**). "Moderate auditory training can improve speech performance of adult cochlear implant patients," Acoust. Res. Lett. Online, **6**, 106-111.

Gagne, J.P., Parnes, L.S., LaRocque, M., Hassan, R., and Vidas, S. (**1991**). "Effectiveness of an intensive speech perception training program for adult cochlear implant recipients," Ann. Otol. Rhinol. Laryngol., **100**, 700-707.

Gfeller, K., Witt, S., Ademek, M., Mehr, M., Rogers, J., Stordahl, J., and Ringgenberg, S. (**2002**). "Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients," J. Am. Acad. Audiol., **13**, 132-145.

Greenwood, D.D. (**1990**). "A cochlear frequency-position function for several species – 29 years later," J. Acoust. Soc. Am., **87**, 2592-2605.

Henshaw, H., and Ferguson, M.A. (**2013**). "Efficacy of individual computer-based auditory training for people with hearing loss: A systematic review of the evidence," PLoS One, **8**, e62836.

Ingvalson, E.M., Leea, B., Fiebigb, P., and Wonga, P.C.M. (**2013**). “The effects of short-term computerized speech-in-noise training on postlingually deafened adult cochlear implant recipients,” J. Speech Lang. Hear. Res., **56**, 81-88.

Lively, S.E., Logan, J.S., and Pisoni, D.B. (**1993**). “Training Japanese listeners to identify English /r/ and /l/. II. The role of phonetic environment and talker variability in learning new perceptual categories,” J. Acoust. Soc. Am., **94**, 1242-1255.

Moore, D.R., Rosenberg, J.F., and Coleman, J.S. (**2005**). “Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children,” Brain Lang., **94**, 72-85.

Norris, D., McQueen, J.M., and Cutler, A. (**2003**). “Perceptual learning in speech,” Cogn. Psychol., **47**, 204-238.

Oba S.I., Fu, Q.-J., and Galvin, J.J. (**2011**). “Digit training in noise can improve cochlear implant users’ speech understanding in noise,” Ear Hearing, **32**, 573-581.

Olson, A.D., Preminger, J.E., and Shinn, J.B. (**2013**). “The effect of LACE DVD training in new and experienced hearing aid users,” J. Am. Acad. Audiol., **24**, 214-230.

Robinson, K., Gatehouse, S. and Browning, G.G. (**1996**). “Measuring patient benefit from otorhinolaryngological surgery and therapy,” Ann. Otol. Rhinol. Laryngol., **105**, 415-422.

Robinson, K., and Summerfield, A.Q. (**1996**). “Adult auditory learning and training.” Ear Hearing, **17**, 51S-65S.

Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). “Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants,” J. Acoust. Soc. Am., **106**, 3629-3636.

Shannon, R.V., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). “Speech recognition with primary temporal cues,” Science, **270**, 303-304.

Stacey, P.C., and Summerfield, A.Q. (**2007**). “Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech,” J. Acoust. Soc. Am., **121**, 2923-2935.

Stacey, P.C., and Summerfield, A.Q. (**2008**). “Comparison of word-, sentence-, and phoneme-based training strategies in improving the perception of spectrally-distorted speech,” J. Speech Lang. Hear. Res., **51**, 526-538.

Stacey, P.C., Raine, C.H, O’Donoguhe, G.M., Tapper, L., Twomey, T., and Summerfield, A.Q. (**2010**). “Effectiveness of computer-based auditory training for adult users of cochlear implants,” Int. J. Audiol., **49**, 347-356.

Stecker, G.C., Bowman, G.A., Yund, E.W., Herron, T.J., Roup, C.M., and Woods D.L. (**2006**). “Perceptual training improves syllable identification in new and experienced hearing aid users,” J. Rehabil. Res. Dev., **43**, 537-552.

Sweetow, R.W., and Sabes, J.H. (**2006**). “The need for and development of an adaptive Listening and Communication Enhancement (LACE) Program,” J. Am. Acad. Audiol., **17**, 538-558.

Zhang, T., Dorman, M.F., Fu, Q.-J., and Spahr, A.J. (**2012**). “Auditory training in patients with unilateral cochlear implant and contralateral acoustic stimulation,” Ear Hearing, **33**, e70–e79.

Perception of music and speech in adolescents with cochlear implants – A pilot study on effects of intensive musical ear training

BJØRN PETERSEN[1,2,*], STINE DERDAU SØRENSEN[3],
ELLEN RABEN PEDERSEN[4], AND PETER VUUST[1,2]

[1] *Center of Functionally Integrative Neuroscience, Aarhus University Hospital, DK-8000 Aarhus, Denmark*

[2] *Royal Academy of Music, Aarhus/Aalborg, DK-8000 Aarhus, Denmark*

[3] *Department of Aesthetics and Communication, Aarhus University, DK-8000 Aarhus, Denmark*

[4] *The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, DK-5230 Odense M, Denmark*

The purpose of this study was to examine 1) perception of music and speech of pre-lingually deaf adolescent cochlear-implant (CI) users, 2) the potential effects of an intensive musical ear training program, and 3) these adolescents' music engagement. Eleven adolescent CI users participated in a short intensive training program involving music-making activities and computer-based listening exercises. Before and after the program they completed music and speech tests. In addition, the participants filled out a questionnaire which examined music listening habits and enjoyment. A normally-hearing (NH) group provided reference data at the same points of time, but received no training. CI users significantly improved their overall music perception and discrimination of melodic contour and rhythm in particular. The NH reference group produced marginally-lower music discrimination scores at the second test. No effect of the music training was found on discrimination of emotional prosody or speech. The CI users described levels of music engagement and enjoyment that were comparable to the NH reference. The CI participants showed great commitment, but found computer-based training less relevant than music-making activities. The findings are an indication of not only the potential of training but also of the plastic potential in the young brain.

BACKGROUND

Cochlear implants (CIs) have revolutionized the lives of persons with severe or profound hearing loss (HL), but auditory processing in general and music perception in particular are hampered in CI users (Gfeller *et al.*, 2005; Gfeller *et al.*, 2007; Cooper *et al.*, 2008; Petersen *et al.*, 2012). Nevertheless, there are examples of CI-users who seem to enjoy music after repeated listening (Gfeller *et al.*, 2000), and

[*] Corresponding author: bpe@musikkons.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

some studies show significantly improved music discrimination after training (Petersen *et al.*, 2012). While previous studies primarily examined implant outcome in adult CI-recipients with an acquired HL, perception of music and speech in the growing population of adolescent CI users with a congenital HL has not been thoroughly investigated. Recent studies, however, indicate that to keep pace with their normal-hearing peers, supplementary measures of rehabilitation are in demand throughout adolescence. Music training may provide a strong, motivational, and beneficial method of strengthening not only music perception, but also linguistic skills, particularly the prosodic properties of speech. With this pilot study we aimed to investigate the potential impact of intensive musical training on adolescent CI-users' discrimination of music and speech and compare these measures with a normally-hearing (NH) reference. Furthermore, we aimed to examine music listening habits and music enjoyment among adolescent CI users. Finally, we intended to develop and evaluate new musical methods and materials aimed at adolescent CI users.

METHODS

Eleven adolescent CI users (six girls, five boys, M_{age} = 17.0 y, age range: 15.6-18.8 y) participated in a group-based music training program, consisting of active music making supplemented with daily computer-based listening exercises. The active training part was scheduled over six days, distributed over two weeks, adding to a total of 20 hours. The program was formed by three elements: rhythm-training, singing and ear training. The computer-based training (presented as musical quiz games) trained discrimination of melodic contour, timbre (musical instruments), and rhythm. Nine of the CI users had bilateral implants and two had unilateral implants. The mean implant experience was 9.5 years (range: 1.8-15.2 y). Ten NH peers (two girls, eight boys, M_{age} = 16.2 y, age range: 15.3-17.0 y) formed a reference group, who followed their standard school schedule and received no music training. Before and after the intervention period, both groups completed a set of tests for perception of music, speech, and emotional prosody. The participants were all recruited from Frijsenborg Efterskole in Denmark, an independent residential school, at which adolescents from the ages of 14 to 18 years can spend one or two years to finish their primary education.

Music tests

The battery of music tests used in the study was adopted from Petersen *et al.* (2012).

The musical-instrument identification test is an eight-alternative-forced-choice test which measures identification of musical timbre represented by eight musical instruments belonging to different families (brass, woodwind, string, and pitched percussion). After a brief familiarization with the instruments, the participant is required to identify the instrument playing parts of a famous Danish children's song.

Melodic-contour identification requires the participant to identify the melodic contour of a five-note sequence. The contour is either rising, falling, flat, rising-falling, or falling-rising in (i) scale steps (subtest 1) or (ii) semitones (subtest 2).

Rhythmic discrimination measures the ability to discriminate rhythmic pairs in a same/different paradigm. The rhythm patterns represented different levels of musical complexity.

Pitch ranking requires the participant to judge which of two tones played in sequence is the higher. The tones are presented in three different registers and with three different notes distances (one, three, and five semitones).

Language tests

Emotional prosody recognition tests the participant's ability to recognize three different emotions: *happy*, *sad*, and *angry*. The emotions are expressed in short everyday sentences and words like 'yes' and 'no', spoken by four different speakers (two females and two males). The test has 30 trials, ten of each emotion presented in random order.

The Danish speech material Dantale II (Wagener *et al.*, 2003) measures the participant's ability to understand spoken five-word sentences in background noise. For each of the five words presented, the participant is required to select from ten alternative words or click an 'I do not know'-button. The test is organized in lists of ten sentences each. Here, the participants completed three lists, one training list and two trial lists. The test adapts to the respondents' performance by increasing or decreasing the volume of the speech. The result of the test is given as the speech reception threshold (SRT) corresponding to a level of 50% correct responses – the lower the SRT, the better the ability to hear speech in noise.

Music listening questionnaire

Prior to the intervention period, both groups completed an online questionnaire including questions concerning music listening habits and music enjoyment. Furthermore, the CI participants were required to rate the overall quality of music through the implant using five 100-point scales (0-100), each anchored with bipolar adjectives. The adjective pairs were: unpleasant-pleasant, complex-simple, fuzzy-clear, hard to follow-easy to follow, and dislike-like.

Evaluation

Shortly after the conclusion of the training program, the CI participants were asked to evaluate the music training program by filling out an online questionnaire. The questionnaire included questions about the content of the program, their perceived outcome, and use of computer applications. Scoring was done by use of five-point Likert rating scales with five as the most positive and one as the most negative.

RESULTS

Musical skills

The users in the CI group obtained higher mean test scores post-training in all five music tests compared to pre-training. The most marked progress was found in the melodic-contour identification subtest 1 ($z = -2.094$; $p = 0.036$) and in the rhythm discrimination test ($z = -2.310$; $p = 0.021$). Furthermore, the overall music progress (mean music post scores − mean music pre scores) was statistically significant ($z = -1.956$; $p = 0.050$).

Except for pitch ranking, the NH participants achieved lower mean music discrimination scores in the post-tests than in the pre-tests. Furthermore, the NH reference group obtained mean scores that were significantly higher than those of the CI participants in all music tests except the melodic-contour identification post-test (subtest 1).

Speech performance

In the emotional prosody recognition test, the CI users obtained lower mean test scores both pre- and post-training compared to the NH participants. Furthermore, the CI users showed a non-significant decline in their performance in the post-test, whereas the NH group obtained non-significantly higher mean test scores at the end of the intervention period.

In the Dantale II test, the CI users had significantly higher mean SRTs than the NH participants, which in this case indicates a poorer performance. Both groups improved their SRT values at the post-test, for the NH group with a statistical significance ($z = -2.293$; $p = 0.022$) (Fig. 1). CI performance showed a high variability with SRTs ranging from −3.9 to 10.9 dB SNR.

Music listening

The questionnaire responses showed that despite poor music-discrimination skills, the majority of the adolescent CI users enjoy listening to music and do so often (three out of 11) or every day (six out of 11). This pattern was comparable to the responses of the NH reference group. Furthermore, a majority of the CI users stated that they appreciated to listen to music (mean Likert-scale score: 3.7).

Music quality

The adolescent CI users in general gave positive ratings of the quality of music through their implant. Figure 2 shows the mean values for the five adjective pairs. The average value across the five pairs was 80.9 points.

PROJECT EVALUATION

A majority of the participants found that the content of the program was relevant. Four of the participants stated that the duration of the program was too short. The participants in general stated that they felt their participation in the music training positively influenced their music listening outcome (mean Likert-scale score: 3.7).

On average, the participants used the computer applications for home training approximately one hour during the entire training period, which was much less than the requested 15 minutes per day. The responses concerning the potential usefulness of such applications for training of music-discrimination abilities were moderately positive. Two stated that the programs could not be useful at all.

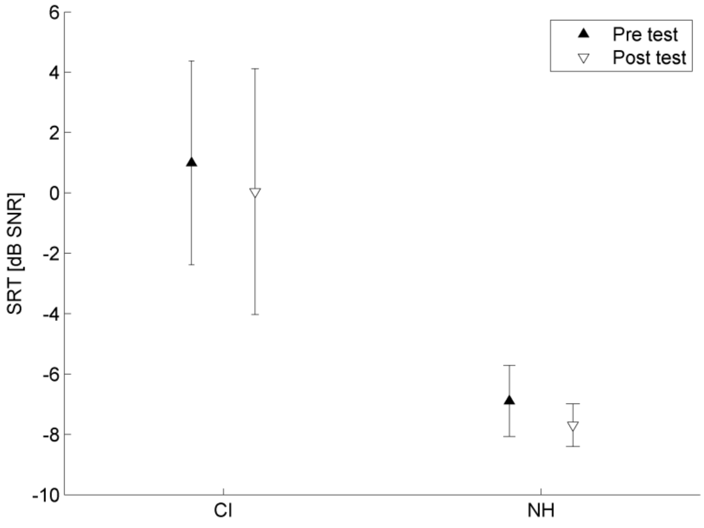


Fig. 1: Mean scores for 11 CI users and ten normal-hearing peers obtained with the Dantale II test. The error bars indicate one standard deviation.

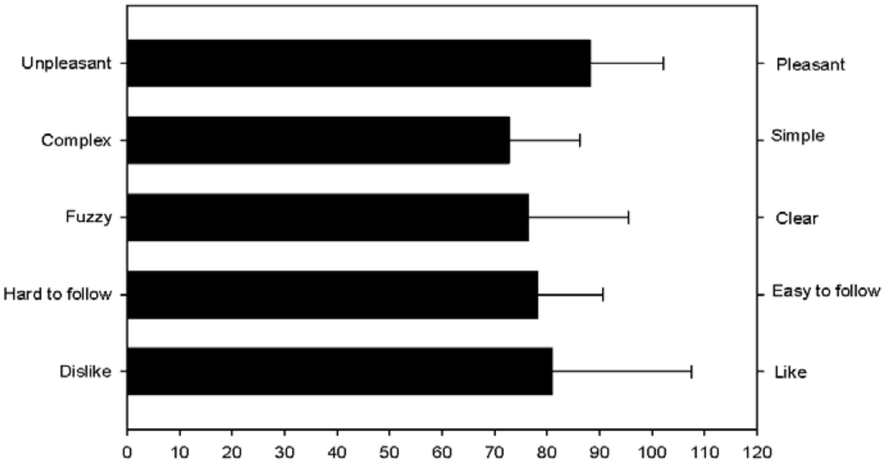


Fig. 2: Mean values for adjective pairs.

The participants agreed to a high degree that being exclusively CI users in the training group was a positive aspect. Their individual comments reflected that the feeling of having equal prerequisites for the different musical tasks was prevalent. Also, absence of embarrassment of not being able to sing in tune was regarded a positive factor.

The participants rated their general satisfaction with the program with a mean score of 86 points on a 100-point visual analog scale and their willingness to participate again with a mean score of 75 points.

DISCUSSION

On average, the adolescent CI users in this study improved their discrimination skills within all musical domains after training, which was reflected in a significant overall progress. Furthermore, they particularly improved their discrimination of rhythm and melodic contour significantly. These findings are consistent with a previous study with recently implanted adult CI users (Petersen *et al.*, 2012), and suggests that changes in patterns of rhythm and in the direction of a sequence of notes are properties of music that are well preserved in the implant's sound transmission and thus most efficiently influenced by training. The NH reference group produced poorer scores at the end of the intervention period, which suggests that the CI progress is an effect of the music training and not a test learning effect. Given the brevity of the program and the duration and profound nature of these adolescents' deafness, these findings are an encouraging indication of not only the potential of training but also of the plastic potential in the young brain.

Speech perception

As expected, the NH participants all scored close to ceiling at both occasions of the emotional prosody recognition test, while the CI users produced significantly poorer results. Contrary to our hypothesis, we saw no progress in the CI users' performance, which may have several explanations. First, the progress in music discrimination was modest, which naturally lessens the chance of a transfer effect. Second, because a majority of the CI users also mastered manual and visual modes of communication, aural expressions of emotion as presented in the test may have been unfamiliar. Finally, though intensive, the training period was very short. It is possible that much more training is necessary and maybe of a more targeted nature, aiming at musical features more specifically relevant to emotional prosody.

The fact that the NH participants performed significantly better and with much less variability in the speech-in-noise test comes as no surprise. The performance gain observed in both groups, however, was unexpected and indicates an effect of learning rather than of the music training. For familiarization, all participants were given one training list in advance of the actual test at both occasions, which may have been insufficient for an optimal performance (Hernvig and Olsen, 2005; Pedersen and Juhl, 2013). However, since the test is comprehensive and loads heavily on attention and working memory, we presumed that a total of three lists was a maximum. Although the Dantale II test is a strong tool for evaluating perception of speech in challenging conditions, we speculate that supplementary measurements of speech perception with more direct phonological focus might provide further documentation of the ability to detect aspects of music in speech and the possible benefit from music training.

Music listening and quality

The CI users performed significantly poorer than their normally-hearing peers in all music tests except the melodic-contour identification subtest 1 post-test. This confirms that perception of music through a cochlear implant is very far from the normal-hearing experience. Nevertheless, we saw no difference between the music listening habits of the CI users and their NH peers. This is in line with Gfeller *et al.* (2012), who in a survey including 31 adolescent CI users, found that a large majority of the respondents engaged in music listening and rated music as an important or very important factor in their life. Since perception of pitch, timbre, and harmony is poor with a CI, we speculate that musical features linked to timing, such as pulse, meter, form, and groove, maybe in combination with lyrics, are the main sources to these adolescents' music enjoyment and engagement. Moreover, streamed music videos on the Internet, accompanied by strong visual elements, may provide an extra dimension to the music experience, thus adding further to their musical interest.

The adolescent CI users rated the quality of music through the implant quite positively and on average significantly higher than the average rating reported in a recent survey among 163 adult CI users (Petersen *et al.*, 2013). There may be manifold causes to this difference, one of them being the point of perspective. The post-lingually deaf adult CI users may tend to compare the quality of musical sounds with their recollection of what it used to sound like, while the young pre-lingually deaf CI users have no reference and therefore are less restricted in their judgment.

Evaluation

For most of the young CI users, this project was their first experience with structured and targeted music making and indeed challenging. Nevertheless, they generally responded with great enthusiasm and engagement to the different exercises and tasks. This positive attitude was further documented in the written feedback, expressed in their ratings of the program's relevance and their general satisfaction. Many of the participants even stated that the training had positively affected their perception of music and speech. Though such positive feedback may have causes of a more psychological nature and should be interpreted with caution, we conclude that active music making involving singing, rhythm and ear training is an absolutely relevant and instructive activity for adolescents with CIs.

According to our feedback, the CI participants only used the computer applications sparsely and less than requested. Obviously, being part of a community such as a boarding school, offering a variety of social activities, may leave little time for homework. The lack of commitment to this part of the program might also be due to fatigue after long and, especially for CI users, tiring school days. However, the most plausible explanation is probably that the applications provided too little excitement. Despite instant feedback and progressive design, the quizzes offered nothing with regard to animation, graphics, and competition in comparison with current computer games. We firmly believe in the potential in digital learning, also in the domain of expanding hearing capabilities, but acknowledge that to succeed in the new digital generation such applications must be fast, adaptive, competitive, offer a social dimension, and preferably be instantly accessible on a smartphone or a tablet computer.

ACKNOWLEDGEMENTS

The authors wish to acknowledge all the participants and the staff at Frijsenborg Efterskole for their commitment to this study. The study was supported by a grant from the Danish Ministry of Culture's Research Committee.

REFERENCES

Cooper, W.B., Tobey, E., and Loizou, P.C. (**2008**). "Music perception by cochlear implant and normal hearing listeners as measured by the Montreal Battery for Evaluation of Amusia," Ear. Hearing, **29**, 618-626.

Gfeller, K, Christ, A., Knutson, J.F., Witt, S., Murray, K.T., and Tyler, R.S. (**2000**). "Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients," J. Am. Acad. Audiol., **11**, 390-406.

Gfeller, K., Olszewski, C., Rychener, M., Sena, K., Knutson, J.F., Witt, S., and Macpherson, B. (**2005**). "Recognition of 'real-world' musical excerpts by cochlear implant recipients and normal-hearing adults," Ear. Hearing, **26**, 237-250.

Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., and Olszewski, C. (**2007**). "Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise," Ear. Hearing, **28**, 412-423.

Gfeller, K., Driscoll, V., Smith, R.S., and Scheperle, C. (**2012**). "The music experiences and attitudes of a first cohort of prelingually-deaf adolescents and young adults CI recipients," Semin. Hear., **33**, 346-360.

Hernvig, L.H., and Olsen, S.O. (**2005**). "Learning effect when using the Danish Hagerman sentences (Dantale II) to determine speech reception threshold," Int. J. Audiol., **44**, 509-512.

Pedersen, E.R., and Juhl, P.M. (**2013**). "Examination of the learning effect with the Dantale II Speech Material," Poster presented at ISAAR 2013.

Petersen, B., Mortensen, M.V., Hansen, M., and Vuust, P. (**2012**). "Singing in the key of life: A study on effects of musical ear training after cochlear implantation," Psychomusicology, **22**, 134-151.

Petersen, B., Hansen, M., Sørensen, S.D., Ovesen, T., and Vuust, P. (**2013**). "Aspects of music with cochlear implants – Music listening habits and appreciation in Danish cochlear implant users," Poster presented at ISAAR 2013.

Wagener, K., Josvassen, J.L., and Ardenkjaer, R. (**2003**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol., **42**, 10-17.

Auditory training

SUNE THORNING KRISTENSEN[1] AND CARSTEN DAUGAARD[2,*]

[1] *University of Southern Denmark, SDU Campusvej, 5020 Odense SV, Denmark*

[2] *DELTA, Teknisk-Audiologisk Laboratorium, Edisonvej 24, 5000 Odense C, Denmark*

The mapping of a sound pattern to a linguistic context is the base of acoustical communication. This process is taking place whenever language skills are acquired. However, sound cues might be changed or lost in amplification, thereby changing the sound pattern. Adaptation is required to reconnect sound with context. Focused training on this connection will speed up and improve the process. The necessity of this training is evident where hearing is restored from deafness, but a training effect is also expected in rehabilitation of gradually emerging hearing loss. Programs training speech recognition and cognitive skills exist for English speakers. They are used with some success, however the criteria for who will benefit from training are unclear. From sensory perception evaluation, training the attention to sound details and developing a language about sound attributes is well known, but the use of non-speech stimuli in auditory training has not yet been given much attention. Looking at the hearing-aid fitting process, an improved fitting could be expected if sound description ability is improved within the framework of specialized training. Music as a part of an auditory training program may increase sound property awareness to the benefit of cognitive skills also related to speech perception. Adding music improves the fun and thus the motivation of the training sessions.

BACKGROUND

Auditory training links naturally to hearing rehabilitation. The attention to the field grew in the USA around World War II, where better diagnostic capabilities and means of rehabilitation of hearing casualties from military service were severely needed. Skills such as lip-reading and 'listening practice' would accompany the prescription of hearing aids to minimize the perceived handicap of the hearing loss. As hearing aids were improved during the eighties, the auditory training as a unique part of the rehabilitation disappeared. In the late nineties, however, auditory training in the USA had a revival based on computer-controlled learning programs and new scientific results.

Auditory training has traditionally been focusing on enhancing speech understanding. However, with the complexity of modern hearing aids another training opportunity is the vocabulary of words describing sound. The link between the impression of sound and a word expressing it could be important in the process

*Corresponding author: cd@delta.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

of adjusting the hearing aid to the user. Sound impressions could be created by music samples supplementing the speech stimuli, adding diversity and amusement to the speech training sessions.

The basic concept which makes the training of hearing possible is the auditory plasticity – reorganizing neural connections in the brain on the basis of input – and behavioral changes (Musiek, 2002). The argument is that a ski-sloping hearing loss, for example, deprives the stimulation of sound at high frequencies thus causing the neurons to reorganize based on a bass-dominated input. Restoring the treble by means of a hearing aid will not find the right path in the brain until the connections regarding treble input are restored. Training might improve the speed of these changes.

COMMUNICATION MODEL

Sweetow and Sabes (2006) have introduced a hierarchical communication model illustrating that the basis for acoustical communication is hearing, i.e., that there is access to the information on the acoustical level (Fig. 1). Next step is to pay attention to the acoustic signal, to listen with the purpose of understanding the signal. If - when listening concentrated - the acoustical signal bears any meaning, comprehension is the third step. Comprehension can be aided by increased listening and by request of repeating part of the acoustic signal with the purpose of more critical listening. Finally when comprehension reaches a state where meaningful information can be derived from the acoustic signal, answers or rephrasing of information can be formulated and communication has been established. Taking this hierarchy into account, it is fair to propose that auditory training with speech signals aims at promoting the understanding at the higher levels in the hierarchy, while music training probably will enhance the ability to listen when introduced in the communication model.

AUDITORY TRAINING METHODS

In the theory four different learning concepts to approach the understanding of spoken language are described (Blamey and Alcántara, 1994). These are:

Bottom-up: Identification of all cues individually, come together as a meaningful sentence.

Top-down: Identification of a still increasing number of cues in an informational stream are compared with similar meaningful structures until one correct meaning is chosen.

Pragmatic: Introduction to different hearing strategies, also including distance to speaker, reformulate questions, etc.

Eclectic: A combination of the three concepts mentioned above, fitted to target group and to the purpose of the training. The American auditory training program LACE is an example of this concept.

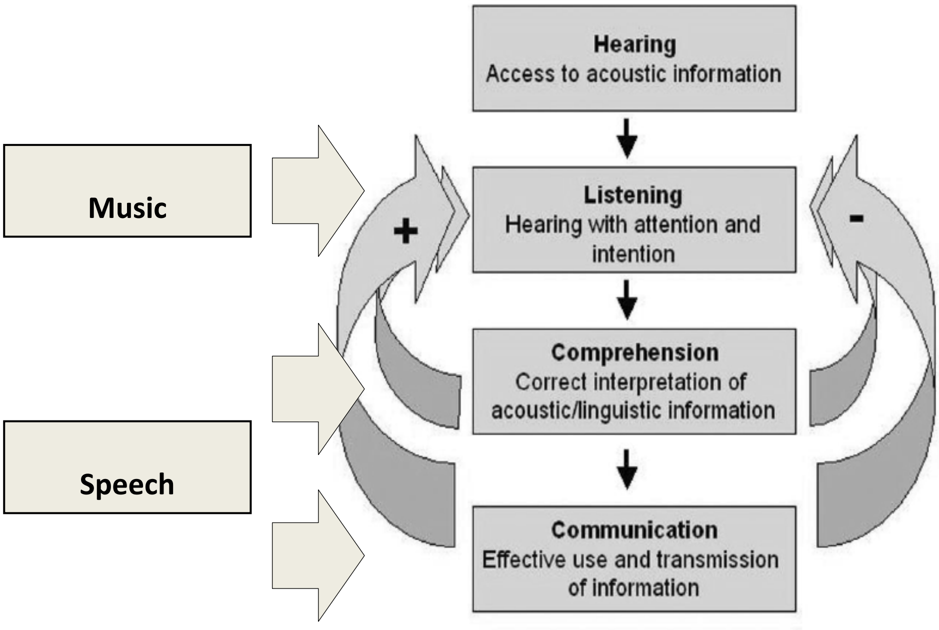

Hearing
Access to acoustic information
Listening
Hearing with attention and intention
Comprehension
Correct interpretation of acoustic/linguistic information
Communication
Effective use and transmission of information
Music
Speech
+
-

Fig. 1: Communication model modified from Sweetow and Sabes (2004).

SPEECH PERCEPTION TRAINING

Based upon the principle of the brain's plasticity, the purpose of speech perception training is to exercise the neural pathways related to decode the acoustical structure of speech into comprehensive information as it is modeled in Fig. 1. This can be obtained by listening to speech material with different degrees of degradation, enabling a correlation of speech cues with the information of the acoustic presentation. This probably also exercises the cognitive abilities, and thus the fundamental task of figuring out the meaning of an acoustical message. The three most commonly used degradations are masking the speech, speeding up the speech, and distorting the speech. Adding a masker corresponds, depending on the nature of the masker, to the real-life situation of noisy places or competing speech, e.g., at a cocktail party. Time compression is another type of degradation, which represents fast-talking speech signals in real life. Degradation could also be harmonic distortion representing real-life situations where the speech signal is unclear due to transmission through electrical circuits (phone, etc.). Also other types of degradation could be imagined: frequency shaping, amplitude variations, etc. Most speech training programs also employ the knowledge of different communication strategies to help reach the goal of a better speech understanding (eclectic training).

MUSIC AND THE BRAIN

There is no tradition of using music as a part of computer-based auditory training for hearing rehabilitation. In the music industry, however, auditory training with the purpose of producing and reproducing sound has been practiced for many decades

(Letowski, 1985; Brixen, 1993). In the later years some projects direct attention towards the perception of music and development of language and communication (Holst, 2009; Petersen *et al.*, 2012). Petersen *et al.* apply traditional musical training to cochlear implant (CI) users, in order to investigate the effect especially on speech understanding ability. They find that musical training speeds up the process of learning speech understanding for the CI users, but more striking is the fact that there is a large interest in participation in the music-oriented tasks. This enthusiasm of music might be explained by investigations of music and the brain, acknowledging music as a rewarding stimulus triggering dopamine in the brain (Salimpoor and Zatorre, 2011). This study also shows that music causes brain activity in the frontal lobe, indicating that music also creates intellectual stimuli (such as pattern recognition). This finding might support an idea of music tasks also supporting higher order processes used in speech understanding. Williamson *et al.* (2010) investigate the short term memory of verbal and musical sounds and find some correspondence in the way these two are processed in memory. This again supports the combination of music and speech stimuli in auditory training. In the last few decades sensory evaluation of sound has been investigated. In this field a vocabulary of sound-describing words is established enabling communication about sensory experiences. Although this does not directly enhance the ability to understand speech, the opportunity to express the acoustical experiences through a hearing aid might lead to a better fitting and thus better sound quality (Daugaard *et al.*, 2011).

WORDS FOR SOUND

Describing the subjective impression of a sound might be difficult, especially if the description should be consistent and useful in changing the sound. In sensory evaluation, development of the right attributes describing the sensation and covering all aspects of the sound impressions is an important issue. Several approaches are possible. ‘Word elicitation’, leaving the creation, grouping, and description of attributes to a focus group of assessors; this is not a useful approach in the context of a fitting situation. Using pre-selected attributes based on earlier investigations is the obvious short-cut. Several investigations have suggested sets of attributes; comparing those leaves us with a handful of common sound characteristics relevant to different kinds of listening situations (mobile phones, stereo reproduction, hearing aids, etc.) (Pedersen and Zacharov, 2008). From sensory evaluation of food, comprehensive groupings of taste attributes are used for the evaluation of products. Best known is perhaps the wine wheel, describing the nuanced taste of red wine. DELTA is in the process of developing corresponding ‘sound wheels’ for different listening situations.

In the dictionary ‘semantic space of sound’ (Pedersen, 2008), 17 profiles or primary descriptors of classes of words describing sound are defined (Table 1). These attributes are emerged from the words in the dictionary and are scalable values. These could also be used as a base for training of word description, or as inspiration, as it is done in this project.

Sound profiles (Pedersen, 2008):		
Loudness	Tempo	Pitch strength
Amplitude variation	Regularity	Pitch
Impulse prominence	Roughness	Tone prominence
Duration	Sharpness	Polyphony
Decay	Presence	Harmony
Frequency variation	Localised in space	

Table 1: The list of sound profiles described in 'semantic space of sound'.

IMPLEMENTATION

A Power-Point based training program was developed, focusing on the inclusion of the right combination of tasks rather than the computer implementation. The program included adaptive training, post-trial re-hearing, variation, and immediate feedback on tasks. The implementation is easy to distribute and install, easy to use, fairly interactive and cost effective (development time and money). The program is based upon a top-down learning strategy, as this approach seems to be the most straight-forward to implement and use, as it does not require special knowledge or interest in linguistics from the user.

Speech exercises

The speech exercises in the Danish material are played with a competing noise consisting of one or two other talkers. The purpose of the exercises is to improve the users' ability to extract the speech information in the noise. The sentences are from the Danish DAT material developed at CAHR, DTU: "Dagmar tænkte på et skind og en lynlås i går" ("Dagmar thought of a hide and a zipper yesterday") (Nielsen *et al.*, 2011). This material is the best available sentence-based Danish word material. The speech exercises consist of seven simple tasks containing two simultaneously played speech tracks, the target sentence (i.e., Dagmar) and the masker (i.e., Asta), and four difficult tasks containing three simultaneously played speech tracks, one the target sentence (i.e., Dagmar) and the other two maskers (i.e., Asta and Tine).

Music exercises

The spectrum and amplitude of several music recordings was modified to represent different degrees of the three attributes roughness, tone, and vibration. At total of 12 exercises with one or more manipulated music recordings was produced. The purpose of the music exercises is training the recognition of the three attributes and associating them with their describing name. Furthermore the goal is to exercise the ability to rank each of the attributes according to the applied signal processing (degree of degradation). The three musical attributes included in the distributed program are roughness, tone, and vibration. The words are from the category 1 (direct descriptions) or 2 (words borrowed from other sensory domains) in the 'semantic space of sounds' (Pedersen, 2008). These attributes were selected as realistic effects experienced by hearing-aid users, as well as characteristics related to a psychoacoustic description. Signal processing was applied to music pieces corresponding to the impression of each attribute. The appropriate signal processing was determined by judgment of one of the authors.

Roughness: Distortion of the sound. Represented by the psychoacoustic attributes *roughness* and *sharpness*. The attribute is made by choosing the VST effect 'BJ overdrive' in the sound editor 'Audacity' software for PC.

Tone: A raise in the midrange of the soundstage (bandpass filtering around 1 kHz with bandwidths from 0.4 to 4.5 kHz) represented by the psychoacoustic attributes *tone prominence* and (possibly) *sharpness*. The effect is achieved by manual adjustment of an equalizer in the sound-editor 'Audacity' software for PC.

Vibration: Amplitude variation, changes in frequency and modulation depth, represented by the psychoacoustic attribute *amplitude variation*. The attribute is made by choosing the effect 'tremolo' in the sound-editor 'Audacity' software for PC.

The music exercises are based upon instrumental classical pieces (with one rhythmic music exception). This is thought of as a neutral choice of music focusing the attention on sound quality and the altered characteristics.

PILOT STUDY

To evaluate the effect of auditory training, a small group of three hearing-aid users and one CI user was selected to go through the 23 exercises in four weeks. They were presented with a test battery prior to the exercises, and again after a month, when the exercises were expected to be completed. The test battery consisted of a speech test, a test in musical vocabulary, and the Abbreviated Profile of Hearing Aid Performance (APHAP; Purdy and Jerram, 1998) questionnaire.

The speech test was the DAT test in Danish developed at DTU, presented (as the exercises) with one or two competing speakers as masking noise. The result for one-speaker masking noise shows improvement for three of four participants, while the two-speaker masking noise was either too difficult to perform or showed no improvement. Due to the limited data material no statistics were calculated. And thus the significance of the results is not determined.

The attribute recognition was tested by ranking tests and description of sound examples. Results show improved skills in both tasks after the exercises were completed.

The APHAP profile did not show improvement after the training, which indicates that the test subjects do not experience any improvement in their communication situation due to the training, or at least the APHAP questionnaire is not able to register that change.

	Tasks	Set-up
1-4	Find the reference, attribute name not mentioned	Double blind triple stimulus with hidden reference
5-7	Rank the changes, attribute name is presented	Ranking method
8-10	Rank the changes, compare/differentiate attributes	Ranking method
11-12	Rank + match sound and attribute	Ranking method + naming the right attribute

Table 1: Overview of the 12 music exercises.

As a part of the pilot test the participants were interviewed on the experience of the tests and their test activity. In general they were positive, and the majority had repeated each of the tests from one up to eight times in order to experience an improvement in their listening skills.

The distribution of the tests via computer, and thus the possibility to access the exercises at any given time, was approved, as well as the music tests, which all users reported added to the motivation.

DISCUSSION

The purpose of this project was first and foremost to establish a Danish auditory training program, also including music exercises. An evaluation of the program was preferable, but time limitations, as well as challenges finding the right parameters to evaluate, made it difficult to make a thorough evaluation. Under the heading ‘Pilot study’ the outcomes of the limited evaluation are summarized. Results are not that encouraging, but still indicate that some benefits could be obtained. It is possible that some groups of hearing-aid users could benefit more from auditory training than others. An obvious question then becomes: How to select the right candidates? One criterion for selecting candidates is their motivation, of course. Another one is prior knowledge of music and linguistics. Also it could be speculated that, for first time users of hearing aids, the training would help in the fitting process as well as in the general acclimatization to hearing-aid use.

PERSPECTIVES

The current project introduces the concept of computer-based auditory training in Danish language. It is modeled on the principles of English auditory training programs and their speech-focused training. The current project also introduces training of a vocabulary for auditory events, aiming at increasing the user’s ability to express the experience of the sound through his hearing aids and thus making the fitting process easier. In the project a program for self-training of auditory awareness is implemented, focusing on learning concepts introduced in the English training programs. To establish the effect of the auditory training further investigation is needed. It is possible that some people (first time users) gain more effect of the training and it would be a subject of further investigations to show whether that is the case.

REFERENCES

Blamey, P.J., and Alcántara, J.I. (**1994**). "Research in auditory training," J. Acad. Reh. Suppl., **27**, 161-191.

Brixen, E.B. (**1993**). "Spectral ear training," Audio Engineering Society's 94th Convention, Berlin, 1-18.

Daugaard, C., Jørgensen, S.L., and Elmelund, L. (**2011**). "Benefits of common vocabulary in hearing aid fitting," in *Speech Perception and Auditory Disorders*. Edited by T. Dau, M.L. Jepsen, T. Poulsen, and J.C. Dalsgaard (Danavox Jubilee Fndn., Ballerup), pp. 432-440.

Holst, F. (**2009**). *Musik, Sprog og Integration – Evalueringsrapport 08-09.* From http://www.musiksprogogintegration.dk – (http://www.finnholst.dk/pjts/MSI_1aarsrapport_smnf.pdf).

Letowski, T. (**1985**). "Development of technical listening skills: Timbre solfeggio," J. Audio Eng. Soc., **33**, 240-244.

Musiek, F.E. (**2002**). "Auditory plasticity: What is it, and why do clinicians need to know?" Hearing Journal, **55**, 70-71.

Nielsen, J.B., Neher, T., and Dau, T. (**2011**). "Towards a Danish speech material for speech-on-speech masking investigation," in *Speech Perception and Auditory Disorders*. Edited by T. Dau, M.L. Jepsen, T. Poulsen, and J.C. Dalsgaard (Danavox Jubilee Fndn., Ballerup), pp. 175-181.

Pedersen, T.H. (**2008**). *The Semantic Space of Sounds.* DELTA.

Pedersen, T.H., and Zacharov, N. (**2008**). "How many psycho-acoustic attributes are needed?" *Paper præsenteret i forbindelse med "Acoustics '08, Paris"*. DELTA Acoustics & Senselab, Hørsholm, Denmark.

Petersen, B., Mortensen, M.V., Hansen, M., and Vuust, P. (**2012**). "Singing in the key of life: A study on effects of musical ear training after cochlear implantation," Psychomusicology: Music, Mind and Brain, **22**, 134-151.

Purdy, S.C., and Jerram, J.C. **(1998)**. "Investigation of the profile of hearing aid performance in experienced hearing aid users," Ear Hearing, **19**, 473-480.

Salimpoor V.N, and Zatorre R.J. (**2013**). "Neural interactions that give rise to musical pleasure," Psychology of Aesthetics, Creativity, and the Arts, **7**, 62-75.

Sweetow, R.W., and Sabes, J.H. **(2004)**. "The case for LACE: Listening and Auditory Communication Enhancement training," Hearing Journal, **57**, 32-40.

Sweetow, R.W., and Sabes, J.H. (**2006**). "The need for and development of an adaptive listening and communication enhancement (LACE^{TM}) program," J. Am. Acad. Audiol., **17**, 538-558.

Williamson, V.J., Baddeley, A.D., and Hitch, G.J. (**2010**). "Musicians and nonmusicians short term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity," Mem. Cognition, **38**, 163-175.

Verbal fluency naming in children with CIs: What can we learn from children with CIs on sensitive periods for language?

DEENA WECHSLER-KASHI[1,2,*], RICHARD G. SCHWARTZ[3,4], AND MIRANDA CLEARY

[1] *Department of Communication Sciences and Disorders, Ono Academic College, Kiryat Ono, Israel*

[2] *Gonda Brain Research Center, Bar Ilan University, Ramat Gan, Israel*

[3] *Program in Speech-Language-Hearing Sciences, CUNY Graduate Center, New York, NY, USA*

[4] *New York Eye and Ear Infirmary, New York, NY, USA*

This study examined lexical retrieval processes as a possible underlying language mechanism responsible for language deficits in some children with cochlear implants (CIs). Lexical retrieval processing was examined using phonological and semantic verbal fluency (VF) naming tasks. In the VF tasks, children were given one minute to generate as many words as they can that begin with a given sound (/t/, /l/, /f/) or that belong to a certain semantic category (animals, food). Twenty children with CIs and twenty age- and IQ-matched normal-hearing (NH) children aged 7-10 participated in this study. Children with CIs generated fewer words on the VF tasks. In addition, qualitative differences were found in the performance of the two groups on these tasks. Children with CI seem to process words at a slower rate compared to NH children. Children with CIs showed significance differences compared to NH children in the phonological VF task on measures of the number of switches and the number of words produced in the first 15 seconds of the task. Age at implantation was significantly correlated with performance on the semantic part of the VF task. Younger implanted children performed better (named more words) on the semantic VF task. These correlations might suggest that early implantation is advantageous for certain aspects of lexical performance. Taken together the data support recent work suggesting that the development of certain aspects of language may have an earlier sensitive period than other linguistic skills.

INTRODUCTION

Research findings show a great enhancement rate of language development in young hearing-impaired children who have been implanted with a CI (Svirsky *et al.*, 2000; Blamey *et al.*, 2001; Le Normand *et al.*, 2003). However, there is a need to examine more specific aspects of language in order to learn more about the language processing abilities of children with CIs.

*Corresponding author: deenawk@gmail.com

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

The present study used phonological and semantic verbal fluency (VF) naming tasks. These tasks have been used extensively with typically developing children and also with children with language and reading impairments (Frith *et al.*, 1995; Nation *et al.*, 2001; Weckerly *et al.*, 2001; Koren *et al.*, 2005). However, these tasks have never been applied to hearing-impaired children who use CIs. In addition, this study is designed to look at more specific parameters related to optimal performance on VF tasks. Analyzing responses on phonological and semantic VF naming can aid in identifying differences in word retrieval processes that elucidate the organization of words in the mental lexicons of children with CIs, and point to specific areas in language processing where children with CIs may differ from NH children.

METHODS

Participants

Twenty children with CIs and twenty age- and IQ- matched NH children aged 7;10 to 10;2 participated in this study. All NH children passed an audiological screening test. In the CI group, inclusion criteria was a hearing impairment diagnosed before the age of 3;0 and a minimum of eight months experience with the CI device. All participants had TONI nonverbal IQ scores above 80. See Wechsler-Kashi (2011) and Wechsler-Kashi *et al.* (2013) for a complete description.

Stimuli and scoring procedure

In the VF task, children were given one minute to generate as many words as possible beginning with a particular speech sound (phonological VF) or from a specific semantic category (semantic VF).

Additional detailed clustering and switching analyses of the subjects' responses in the VF tasks were conducted. The rules for defining and scoring clusters were based on Troyer (2000), Troyer *et al.*, (1997), and Koren *et al.*, (2005). The analysis included both semantic and phonological clusters. Semantic clusters consist of words with related meanings that belong to the same subcategory (e.g., sea animals '…seal, dolphin, whale, fish…' or jungle animals '…lion, giraffe, monkey...') according to lists of common subcategories of *animals* and *food* listed in Troyer (2000), Troyer *et al.*, (1997), and Koren *et al.*, (2005). Phonological clusters consist of words that share similar phonemes (e.g., words that begin with /fr/ '…fright, fraud, free, fry…' or phonological neighbors; words with the same initial and final phonemes '…fat, feet, foot, fit…').

The analyses also included the number of switches within each subject's response. Switches were defined as transitions from one word, or a group of words (cluster) to the next word (or cluster). Additional analyses included measurements of reaction times to first-retrieved-words in each subtask. Reaction time was measured using Sound Forge 4.5 (1998) from the starting point of the task (press of the stopwatch) to the initiation of the verbal response. The score for the number of words produced during the first 15 seconds of the task was also obtained. This was measured by counting the number of words generated in the initial 15 seconds time frame of the response (setting this point using Sound Forge 4.5, 1998). The score for the proportion of words produced during the first 15 seconds of the child's response was attained by calculating the percentage of words produced during the first 15 seconds with respect to the total number of words in this subtask. The mean cluster size (MCS) measure was calculated by averaging the cluster size scores across each task. For each of the measures, a separate score was calculated for the phonological task and a separate score was calculated for the semantic task. See Wechsler-Kashi (2011) and Wechsler-Kashi *et al.* (2013) for a complete description.

RESULTS

As reported in Wechser-Kashi (2011) and Wechsler-Kashi *et al.* (2013), children with CIs named significantly less words on both phonological and semantic VF tasks. These findings are illustrated below in Fig. 1.

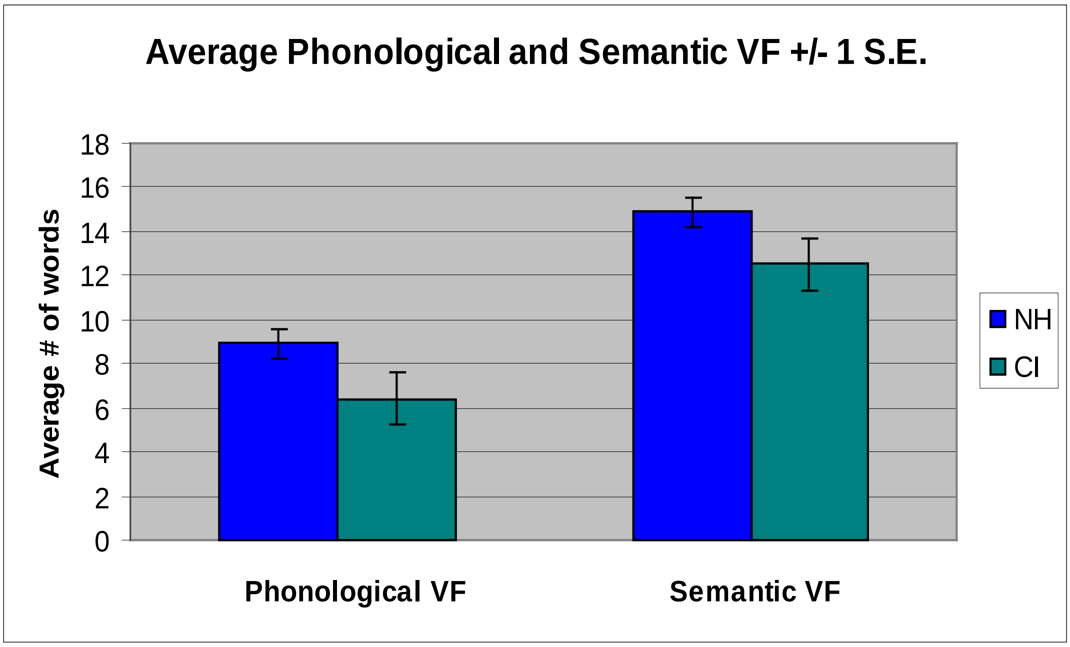


Fig. 1: Average number of words and standard errors (S.E.) on phonological and semantic VF naming tasks.

Pearson product-moment correlation coefficients were computed between results in the VF experiment and variables related to background factors in the CI group, age at implantation, and years of CI use. These correlations are summarized in Table 1. As can be seen in Table 1, age at implantation and years of CI use were significantly correlated with performance on the semantic part of the VF task. Younger implanted children performed better on the semantic VF task (named more words on the semantic task). Similarly, more years of CI use was positively correlated with performance on the semantic VF task. Children who had used their implants for a longer duration of time performed better on the semantic VF task.

	Phonological VF task	Significance	Semantic VF task	Significance
Age at implantation	$r = 0.335$	$p > 0.05$	$r = -0.463$	$p < 0.05$
Years of CI use	$r = -0.109$	$p > 0.05$	$r = 0.514$	$p < 0.05$

Table 1: Pearson correlation coefficients between age at implantation and years of CI use and performance on VF experiment.

Results of the detailed analyses of the verbal fluency responses are summarized below in Table 2.

	Phonological			Semantic		
	CI	NH	Significance	CI	NH	Significance
Number of clusters	3.75 (0.50)	4.95 (0.54)	$p = 0.06$	8.55 (0.86)	8.4 (0.61)	$p > 0.05$
Number of switches	12.7 (1.49)	18.1 (1.39)	$p < 0.05$	12.6 (1.25)	14 (0.80)	$p > 0.05$
Number of words in first 15 s of task	3.06 (0.24)	4.58 (0.26)	$p < 0.05$	5.3 (0.44)	6.25 (0.37)	$p > 0.05$
Latency (RT) in ms to first word produced	1643 (246)	1037 (204)	$p > 0.05$	2009 (749)	1200 (147)	$p > 0.05$
Proportion of words in 15 s	40% (3.00)	42% (1.85)	$p > 0.05$	51% (2.74)	50% (1.37)	$p > 0.05$
Mean cluster size	2.04 (0.13)	2.21 (0.14)	$p > 0.05$	2.58 (0.09)	2.73 (0.07)	$p > 0.05$

Table 2: Cross group comparisons for the analyses of the VF responses. Standard errors are provided in parentheses. ANOVA significance levels are also presented.

DISCUSSION

Research findings show that children with CIs seem to access words less efficiently than NH peers. Moreover, the differences found between performance on phonological and semantic VF tasks in children with CIs implies that their phonological memory is more susceptible to auditory limitations. Age at implantation was significantly correlated with performance on the semantic part of the VF task. Younger implanted children performed better (named more words) on the semantic VF task. The results support recent work suggesting that the development of certain aspects of language may have an earlier sensitive period than other linguistic skills. Further studies, examining performance of children with CIs on VF naming tasks at different ages can aid in better defining these time frames.

REFERENCES

Blamey, P.J., Sarant, J.Z., Paatsch, L.E., Barry, J.G., Bow, C.P., Wales, R.J., Wright, M., Psarros, C., Rattigan, K., and Tooher, R. (**2001**). "Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing," J. Speech Lang. Hear. Res., **44**, 264-285.

Frith, U., Karin, L., and Frith, C. (**1995**). "Dyslexia and verbal fluency: More evidence for a phonological deficit," Dyslexia, **1**, 2-11.

Koren, R., Kofman, O., and Berger, A. (**2005**). "Analysis of word clustering in verbal fluency of school-age children," Arch. Clin. Neuropsych., **20**, 1087-1104.

Le Normand, M.-T., Ouellet, C., and Cohen, H. (**2003**). "Productivity of lexical categories in French-speaking children with cochlear implants," Brain Cognition, **53**, 257-262.

Nation, K., Marshall, C.M., and Snowling, M.J. (**2001**). "Phonological and semantic contributions to children's picture naming skill: Evidence from children with developmental reading disorders," Lang. Cognitive Proc., **16**, 241-259.

Svirsky, M.A., Robbins, A.M., Iler-Kirk, K., Pisoni, D.B., and Miyamoto, R.T. (**2000**). "Language development in profoundly deaf children with cochlear implants," Psychol. Sci., **11**, 153-158.

Troyer, A.K. (**2000**). "Normative data for clustering and switching on verbal fluency tasks," J. Clin. Exp. Neuropsych., **22**, 370-378.

Troyer, A.K., Muscovitch, M., and Winocur, G. (**1997**). "Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults," Neuropsychology, **11**, 138-146.

Wechsler-Kashi, D. (**2011**). "Lexical processing during naming in children with cochlear implants," Dissertation Abstracts International: Section B: The Sciences and Engineering, p. 6734.

Wechsler-Kashi, D., Schwartz, R. G., and Cleary, M. (**2013**). "Lexical processing during naming in children with cochlear implants," Ear Hearing (under revision).

Weckerly, J., Wulfeck, B., and Reilly, J. (**2001**). "Verbal fluency deficits in children with specific language impairment: Slow rapid naming or slow to name?" Child Neuropsychol., **7**, 142-152.

Factors behind the 'cleaning' of the auditory pathway in late implantation of prelingual oral deaf adults

J. TILAK RATNANATHER[1,*] AND CHARLES J. LIMB[2]

[1] *Center for Imaging Science and Institute for Computational Medicine, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA*

[2] *Department of Otolaryngology-Head & Neck Surgery, Johns Hopkins University School of Medicine, Baltimore, MD 212105, USA*

Pre- and peri-lingually deaf adults are benefiting from late cochlear implantation. While much has been written about the emotional experiences, we review auditory plasticity based on 16 months of CI usage by the first author. We suggest that the goal of speech discrimination in quiet via bimodal hearing may accrue from some or all of the following: 1) amplification of low-frequency sounds since infancy, 2) usage of residual hearing via parent-child interaction in auditory training, 3) improved synaptic contact via spike activity from high stimulation rates fused with natural firing from residual hair cells, 4) exposure to singing and music as infants, 5) top-down linguistic processing, 6) reduced cognitive load, 7) episodes of sleep-induced tinnitus-like symptoms after a period of intense auditory exposure, 8) auditory exposure throughout the day, 9) based on inference of imaging scans of 5 oral deaf adults, the distribution of the gray matter cortical thickness of the Heschl's Gyrus (HG) as well as the spatial topography of the acoustic radiation white matter tract from the thalamus to the HG appear to be maintained, and 10) auditory training for bottom-up phoneme processing and auditory working memory.

INTRODUCTION

An increasing number of pre- and peri-lingual oral deaf people get cochlear implants late as adults. This may account for adults age 30-49 years old being the second largest group receiving CI nationally in the United Kingdom (see Fig. 4 in Raine, 2013) and locally at Johns Hopkins Hospital in the past five years. A contributing factor is the decreasing benefit of hearing aids (HAs) due to aging. Evidence suggests that prior use of HAs can provide positive outcomes for pre/peri-lingual deaf late implanted adults (PLDLI) (Caposecco *et al.*, 2012). The purpose of this review is to discuss factors contributing towards the goal of speech discrimination in quiet via bimodal hearing from the perspective of a PLDLI auditory scientist.

LOW-FREQUENCY RESIDUAL HEARING SINCE INFANCY

It has long been observed that successfully mainstreamed deaf adults had usable low-frequency residual hearing (Urbantschitsch and Goldstein, 1898; Bárczi, 1936;

*Corresponding author: tilak@cis.jhu.edu

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Ewing and Ewing, 1938). Then the advent of the transistor resulted in a remarkable example of serendipity with different people making similar observations in the same period (Hardy *et al.*, 1951; Huizing and Pollack, 1951; Wedenberg, 1951; Beebe, 1953; Whetnall, 1956; Guberina, 1957): Deaf children were able to communicate clearly and naturally. These observations were synthesized in the 1970s via the ‘Ling Six Sounds’ as a means to test the potential ability of the deaf child to comprehend sounds (Ling, 2002). The Ling Six Sounds consist of m, oo, ee, ah, sh, and s, which are essential to speech and language development. The first four are in the low-to-mid frequency range and the last two in the moderate-to-high frequency range. Early diagnosis of hearing loss followed by intervention with HAs probably accounts for the maturation of P1 latency in cortical auditory evoked potentials (CAEP) of babies fitted with HA (Nash *et al.*, 2007). This suggests that HA amplification of low-frequency sounds within 20 dB of normal hearing allows a critical part of the ‘speech banana’ to be perceived by the deafened brain. Also the proximity of the parent in communicating with the deaf child is critical in facilitating ‘motherese’ (Brown *et al.*, 2001) especially within the first three years critical for the maturation of the P1 latency (Sharma *et al.*, 2002). This leads to two questions. First, how can low-frequency input with HA enable suprasegmental discrimination in speech production and understanding (Abberton and Fourcin, 1978) and second, why do PLDLI born before the digital HA era take longer to reach ‘reasonable’ speech discrimination levels with CI.

EXPOSURE TO SINGING AND MUSIC AS INFANTS

The spectrum of music is much larger than that of speech, so it is possible that lullabies and nursery rhymes generate low-frequency tones. In turn this may have a positive effect on suprasegmental development seen in CI subjects who began with HA (Most and Peled, 2007). This reinforces the suggestion that access to acoustic hearing, especially low-frequency tones, creates a foundation for music perception accessed later with CI (Hopyan *et al.*, 2012). That music sounds more pleasant (Fuller *et al.*, 2013) may be attributed to the electrical stimulation of the auditory pathways previously developed by acoustic stimulation (Hopyan *et al.*, 2012). There is evidence (Fawkes and Ratnanather, 2009) to show that at one extreme singing improves cadence in deaf children and at the other extreme deaf children are capable of studying and performing music as an academic subject to a very high standard. This leads to the question whether it would be helpful to prime the auditory pathway with HAs via speech and music for a few months prior to CI.

FUSING SYNAPTIC CONTACT AND SPIKE ACTIVITY

HAs can do only so much, i.e., low-frequency amplification to within 20 dB of normal hearing. But the high stimulation rate via CI results in increased synaptic contact and activity at every stage of the auditory pathway from the spiral ganglion cells to the auditory cortex (Kral *et al.*, 2000; Chen *et al.*, 2010). For example, the density of synaptic contacts correlates with the spike activity (O’Neil *et al.*, 2011). So bimodal hearing fuses the natural spike activity from the HA at low frequency with that from the CI at all frequencies to give ‘warmth’ and ‘clarity’, respectively (Crew *et al.*, 2013). It could be argued that the high stimulation rate from the CI induces the natural ‘stochasticity’ properties of the auditory pathway but there is no correlation of stimulation rate with speech perception (Shannon *et al.*, 2011). It is interesting that the parietal cortex activity is significant in deaf adults who have not used HAs (Gilley *et al.*, 2008). This may relate to the initial moments of activation in which the CI patient experiences whole body sensation via the homunculus distribution along the motor cortex. This suggests that the distribution of receptors in the thalamocortical network might be affected by rate-level functions (Metherate *et al.*, 1990). Hence the question whether white-matter tracts to the auditory cortex from the thalamus tolerate high spike activity while tracts to the parietal cortex cannot.

SLEEP-INDUCED PLASTICITY

There have been anecdotal reports that following intense periods of auditory exposure soon after CI activation, ‘whooshing’ brain waves have been experienced either before or at the end of deep REM sleep or both. It is unlikely to be tinnitus as it did not cause distress nor did it appear to be vascular. It is episodic hence being mentioned online by CI subjects. These events could be signs of ‘sleep-induced’ plasticity. It is known that sleep consolidates experience-dependent plasticity (Aton *et al.*, 2013), so do these waves reflect cortical protein synthesis, remodelling of neurons and synapses (Kral, 2013) or the fast propagating waves observed by Reimer *et al.* (2011)? That these episodes do not occur after some time suggests that adaptation has taken place, i.e., it is a ‘positive’ signature of the adapting brain.

TOP-DOWN PROCESSING

PLDLI have already acquired good language so it is not surprising that top-down linguistic processing via contextual analysis is maximised. This top-down cognitive processing mechanism is probably an adaptation of that used in speechreading (Capek *et al.*, 2008). Also the brain is Bayesian, i.e., utilises probabilistic inference (Shannon *et al.*, 1995; Boothroyd, 2010) which suggests that adaptive learning algorithms could be implemented in CI processors. Despite excellent open-set speech recognition scores, the long term challenge is auditory working memory which leads to the question whether this is similar to understanding speech in noise.

REDUCED COGNITIVE LOAD

There is an increase in multi-tasking such as listening to audiobook, podcast, or radio while working on a computer or watching TV; similarly passive listening in meetings is possible. Anecdotal reports of deaf adults who did not use HAs and used signing prior to CI and found it difficult to get ‘over the hump’ in open-set speech suggest that cross-modal plasticity prior to CI may be difficult to unravel. This may be explained by functional MRI studies of deaf adults which showed that those who do not sign process information differently from those who do (Cardin *et al.*, 2013). A key network is the connection between the frontal cortex including Broca’s responsible for executive function and the temporal cortex including association cortex such as the planum temporale (PT) responsible for speech and language processing. The extensive cross-modal plasticity via the takeover of the auditory cortex by visual processing suggests that the acoustic radiation (AR) from the medial geniculate body in the thalamus to the auditory cortex and the optic radiation (OR) from the lateral geniculate body in the thalamus to the visual cortex may cross, overlap, or fuse in deaf adults. It is also challenging to visualise the AR and OR in whole-brain diffusion tensor images (DTI) but Fig. 1 shows it is possible to generate AR via probabilistic based white-matter (WM) fibertracking (Ratnanather *et al.*, 2013) and that spatial topography of the AR is similar in deaf and normal subjects. It is also possible to show that WM tracts from the Heschl's Gyrus (HG) to the posterior region of the PT passes through the OR. This leads to the question whether the CI can "uncouple" the OR and AR.

AUDITORY CORTEX

The thickness of the cortical mantle differs in motor and sensory cortices (Jones, 2004). Figure 2 shows histograms of distances of gray matter (GM) relative to four different cortical surfaces related to hearing, speech, and language from MRI scans from groups of PLDLI adults (prior to CI) and controls. Following Kral and Eggermont (2007), these histograms can be interpreted with respect to upper and deep cortical layers. The differences further from the GM/WM surface could be related to the lack of lateral input that affects bottom-up processing seen in deaf subjects; the similarities closer to the surface could be related to top-down processing in deaf adults with normal CAEP latencies. This leads to the question whether HAs help to prime the auditory pathway and synaptogenesis in the upper layers can be facilitated by auditory training.

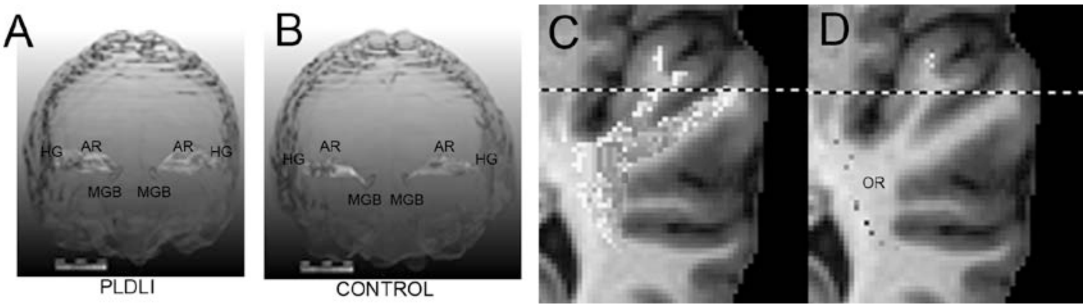


Fig. 1: Acoustic radiation WM tracts from the MGB to the HG in one deaf (A) and one control (B) subject. WM tracts from the HG to the PT (C) and those that pass through the optic radiation (D) terminate in the posterior PT.

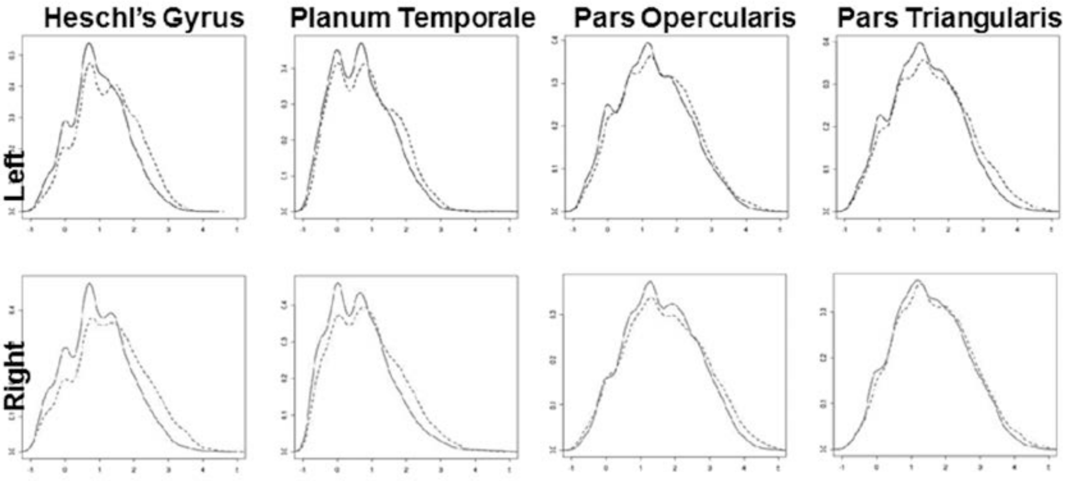


Fig. 2: Normalised distances of GM from cortical surfaces related to hearing, speech, and language. Solid and dashed curves respectively correspond to pooled groups of five age- and gender-matched normal and deaf adults. KS-tests show thinning in deaf group ($p < 0.0001$).

AUDITORY EXPOSURE

Anecdotal reports of low CI usage are difficult to believe. What is the point of the CI if it is not consistently used? (Gordon *et al.*, 2011) First it is suggested that adaptation to noise would be quicker if digital HAs were used prior to CI. It is likely that in the deafened brain, it takes longer to mask noise. So it is important not to let background noise dominate the CI during waking hours. However auditory training in noise is said to improve neural timing (Song *et al.*, 2012). So in the early days, it is helpful to self-test with the Ling Six Sounds iPad app as well as the Virtual Piano software. In comparison with autobiographical accounts which have generally been emotional, perhaps the most substantive account is that by someone forty years after benefiting from auditory training as a child (Beebe, 1953; Younglof, 1997). The HA in the contralateral ear should be used after about eight weeks so that the benefits from bimodal hearing can be accrued. This leads to the question how loudness balance between CI and HA can be optimised (Neuman and Svirsky, 2013).

AUDITORY TRAINING

Auditory training (AT) should begin with bottom-up processing, i.e., acoustic analysis of speech starting with suprasegmental sounds – vowel (V), consonant (C), and consonant-vowel (CV) – and then progressing to CVC and CVCVC. In top-down processing, cognitive analysis is used to extract meaning so it is important not to ‘think’ but to learn to process phonemes. Among the existing AT software, Angel Sound (Fu and Galvin, 2012) is freely available. However there is variation in provision of AT in clinics in the US (Sorkin, 2013) and the UK (Raine, 2013). One solution is to develop adaptive learning tablet applications with feedback which might allow for monthly visits by more patients instead of few weekly visits. In future, music should be incorporated as part of AT as it can improve neural timing (Kraus and Chandrasekaran, 2010).

SUMMARY

For PLDLI adults, it is a work in progress but with new experiences accrued almost on a daily basis. According to Michael Dorman, who was involved in the early days of multichannel CI research, if there is a way to 100% speech recognition in quiet then there are several ways of getting there. Granted that technological developments in CI have begun to mature, it is necessary to develop effective strategies for AT to maximise benefit from the remarkable ability of the brain to adapt to the new auditory input. However, is it reasonable to infer what might happen in PLDLI from anatomical changes in completely deafened animals? MRI, DTI, and functional MRI (fMRI) scans prior to CI could be useful tools for customizing surgical and rehabilitative strategies, but post-operative PET imaging presents challenges. Last but not least, fMRI studies of brain activity via HA amplification are possible without electromagnetic interference from HA (Ratnanather et al., in preparation) and may provide clues for substrates of PLDLI.

ACKNOWLEDEGMENTS

We have benefited from discussions with people too numerous to mention here but JTR would like to dedicate this review to the memory of Dr. Whetnall for starting his parents off on the right track. Data generated in Figs 1-2 were financed by grants from the National Organization for Hearing Research and NIH (P41-EB015909).

REFERENCES

Abberton, E., and Fourcin, A.J. (**1978**). "Intonation and speaker identification," Lang. Speech., **21**, 305-318.

Aton, S.J., Broussard, C., Dumoulin, M., Seibt, J., Watson, A., Coleman, T., and Frank, M.G. (**2013**). "Visual experience and subsequent sleep induce sequential plastic changes in putative inhibitory and excitatory cortical neurons," Proc. Natl. Acad. Sci. USA, **110**, 3101-3106.

Bárczi, G. (**1936**). *Hörerwecken und Hörerziehen* (Josef Rehrl, Salzburg).

Beebe, H. (**1953**). *A guide to help the severely hard of hearing child.* (Basel-Karger, New York).

Boothroyd, A. (**2010**). "Adapting to changed hearing: the potential role of formal training," J. Am. Acad. Audiol., **21**, 601-611.

Brown, P.M., Rickards, F.W., and Bortoli, A. (**2001**). "Structures underpinning pretend play and word production in young hearing children and children with hearing loss," J. Deaf. Stud. Deaf. Educ., **6**, 15-31.

Capek, C.M., Macsweeney, M., Woll, B., Waters, D., McGuire, P.K., David, A.S., Brammer, M.J., and Campbell, R. (**2008**). "Cortical circuits for silent speechreading in deaf and hearing people," Neuropsychologia, **46**, 1233-1241.

Caposecco, A., Hickson, L., and Pedley, K. (**2012**). "Cochlear implant outcomes in adults and adolescents with early-onset hearing loss," Ear Hearing, **33**, 209-220.

Cardin, V., Orfanidou, E., Ronnberg, J., Capek, C.M., Rudner, M., and Woll, B. (**2013**). “Dissociating cognitive and sensory neural plasticity in human superior temporal cortex,” Nat. Commun., **4**, 1473.

Chen, I., Limb, C.J., and Ryugo, D.K. (**2010**). “The effect of cochlear-implant-mediated electrical stimulation on spiral ganglion cells in congenitally deaf white cats,” J. Assoc. Res. Otolaryngol., **11**, 587-603.

Crew, J.D., Galvin, J.J., and Fu, Q.J. (**2013**). “How does electric hearing combine with acoustic hearing for speech and music?” in *CIAP* (Lake Tahoe, NV), p. 160.

Ewing, I.R., and Ewing, A.W.G. (**1938**). *The handicap of deafness* (Longmans, London, New York).

Fawkes, W.G., and Ratnanather, J.T. (**2009**). “Music at the Mary Hare Grammar school for the deaf from 1975 to 1988,” Visions of Research in Music Education, **14**.

Fu, Q.J., and Galvin, J.J. (**2012**). “Auditory training for cochlear implant patients,” in *Auditory Prostheses: New Horizons*. Edited by F.G. Zeng, A.N. Popper, and R.R. Fay (Springer Handbook of Auditory Research), pp. 257-278.

Fuller, C., Mallinckrodt, L., Maat, B., Baskent, D., and Free, R. (**2013**). “Music and quality of life in early-deafened late-implanted adult cochlear implant users,” Otol. Neurotol., **34**, 1041-1047.

Gilley, P.M., Sharma, A., and Dorman, M.F. (**2008**). “Cortical reorganization in children with cochlear implants,” Brain Res., **1239**, 56-65.

Gordon, K.A., Wong, D.D., Valero, J., Jewell, S.F., Yoo, P., and Papsin, B.C. (**2011**). “Use it or lose it? Lessons learned from the developing brains of children who are deaf and use cochlear implants to hear,” Brain Topogr., **24**, 204-219.

Guberina, P. (**1957**). “Verbotonal audiometry; principles & applications,” Ann. Otolaryngol., **74**, 376-377.

Hardy, W.G., Pauls, M.D., and Bordley, J.E. (**1951**). “Modern concepts of rehabilitation of young children with severe hearing impairment,” Acta Otolaryngol., **40**, 80-86.

Hopyan, T., Peretz, I., Chan, L.P., Papsin, B.C., and Gordon, K.A. (**2012**). “Children using cochlear implants capitalize on acoustical hearing for music perception,” Front. Psychol., **3**, 425.

Huizing, H.C., and Pollack, D. (**1951**). “Effects of limited hearing on the development of speech in children under three years of age,” Pediatrics, **8**, 53-59.

Jones, E.G. (**2004**). “Cerebral cortex,” in *Encylopedia of Neuroscience* (Elsevier), pp. 769-773.

Kral, A., Hartmann, R., Tillein, J., Heid, S., and Klinke, R. (**2000**). “Congenital auditory deprivation reduces synaptic activity within the auditory cortex in a layer-specific manner,” Cereb. Cortex, **10**, 714-726.

Kral, A., and Eggermont, J.J. (**2007**). “What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity,” Brain Res. Rev., **56**, 259-269.

Kral, A. (**2013**). “Auditory critical periods: A review from system's perspective,” Neuroscience, **247**, 117-133.

Kraus, N., and Chandrasekaran, B. (**2010**). “Music training for the development of auditory skills,” Nat. Rev. Neurosci., **11**, 599-605.

Ling, D. (**2002**). *Speech and the hearing-impaired child : theory and practice* (Alexander Graham Bell Association for the Deaf and Hard of Hearing, Washington, DC).

Metherate, R., Ashe, J.H., and Weinberger, N.M. (**1990**). "Acetylcholine modifies neuronal acoustic rate-level functions in guinea pig auditory cortex by an action at muscarinic receptors," Synapse, **6**, 364-368.

Most, T., and Peled, M. (**2007**). "Perception of suprasegmental features of speech by children with cochlear implants and children with hearing AIDS," J. Deaf. Stud. Deaf. Educ., **12**, 350-361.

Nash, A., Sharma, A., Martin, K., and Biever, A. (**2007**). "Clinical applications of the P1 cortical auditory evoked potential (CAEP) biomarker," in *A Sound Foundation Through Early Amplification: Proceedings of the Fourth International Conference* (Phonak), pp. 43-50.

Neuman, A.C., and Svirsky, M.A. (**2013**). "Effect of hearing aid bandwidth on speech recognition performance of listeners using a cochlear implant and contralateral hearing aid (bimodal hearing)," Ear Hearing, **34**, 553-561.

O'Neil, J.N., Connelly, C.J., Limb, C.J., and Ryugo, D.K. (**2011**). "Synaptic morphology and the influence of auditory experience," Hear. Res., **279**, 118-130.

Raine, C. (**2013**). "Cochlear implants in the United Kingdom: awareness and utilization," Cochlear Implants Int. Suppl., **14**, S32-S37.

Ratnanather, J.T., Lal, R.M., An, M., Poynton, C.B., Li, M., Jiang, H., Oishi, K., Selemon, L.D., Mori, S., and Miller, M.I. (**2013**). "Cortico-cortical, cortico-striatal and cortico-thalamic white matter fiber tracts generated in the macaque brain via dynamic programming," Brain Connect., **3**, 475-490.

Reimer, A., Hubka, P., Engel, A.K., and Kral, A. (**2011**). "Fast propagating waves within the rodent auditory cortex," Cereb. Cortex, **21**, 166-177.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science, **270**, 303-304.

Shannon, R.V., Cruz, R.J., and Galvin, J.J., 3rd (**2011**). "Effect of stimulation rate on cochlear implant users' phoneme, word and sentence recognition in quiet and in noise," Audiol. Neurootol., **16**, 113-123.

Sharma, A., Dorman, M.F., and Spahr, A.J. (**2002**). "A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation," Ear Hearing, **23**, 532-539.

Song, J.H., Skoe, E., Banai, K., and Kraus, N. (**2012**). "Training to improve hearing speech in noise: biological mechanisms," Cereb. Cortex, **22**, 1180-1190.

Sorkin, D.L. (**2013**). "Cochlear implantation in the world's largest medical device market: utilization and awareness of cochlear implants in the United States," Cochlear Implants Int. Suppl., **14**, S4-S12.

Urbantschitsch, V., and Goldstein, M.A. (**1898**). "The hearing capacity of deaf mutes," The Laryngoscope, **5**, 224-227.

Wedenberg, E. (**1951**). "Auditory training of deaf and hard of hearing children; results from a Swedish series," Acta Otolaryngol. Suppl., **94**, 1-130.

Whetnall, E. (**1956**). "The development of usable (residual) hearing in the deaf child," J. Laryngol. Otol., **70**, 630-647.

Younglof, M. (**1997**). "Mardie's CI Progress" — http://www.listen-up.org/ci/story/mardie.htm.

Using limitations of the auditory system to optimize hearing-aid design

JULIE NEEL WEILE*, MICHAEL NILSSON, AND THOMAS BEHRENS

Clinical Evidence & Communication, Oticon A/S Headquarters, DK-2765 Smørum, Denmark

Recent research has demonstrated that people with hearing impairment have limited ability to take advantage of temporal fine structure information (Strelcyk and Dau, 2009; Hopkins and Moore, 2011). This means that they will not be able to fully utilize auditory cues, such as interaural time differences and detailed pitch perception, which rely on such information. On the other hand, this reduced ability can also be used to improve on certain aspects of hearing-aid functionality. One such area is feedback suppression. Many of the latest hearing-aid introductions feature feedback suppression algorithms which apply a slight frequency shift to de-correlate the hearing-aid output from the input and thus minimize the risk of feedback. This paper will review evidence on temporal fine-structure abilities and relate this to how hearing-aid feedback systems can be designed to achieve a dual goal: to optimize the perceived sound quality of the listener with hearing impairment, whilst minimizing the occurrence of feedback.

BACKGROUND

The human ear processes sound using a number of auditory filters into a series of relatively narrow frequency bands. These bands have good or narrow frequency specificity. When a broadband sound comes in, it is band-pass filtered corresponding to the ‘correct’ position on the basilar membrane, which is tonotopically organized. An incoming signal can be considered as a slowly varying envelope superimposed on a rapid temporal fine structure. Information about the envelope is carried by changes over time in the firing rate of the auditory nerve while information about the temporal fine structure is embedded in the phase-locking pattern. When cochlear hearing loss occurs, the ability to use these fast changes, the temporal fine structure, is believed to decrease (Moore, 2007).

A study by Hopkins and Moore (2007) found that hearing-impaired individuals have trouble utilizing temporal fine structure information compared to normal-hearing individuals. In a group of individuals with moderate cochlear loss they tested complex harmonic tones compared with similar tones with all components shifted by ΔHz. For normal-hearing individuals, a shift like this would be perceived as the shifted tone having higher pitch than the un-shifted one with lower harmonics. Both tones, shifted and un-shifted, had a similar envelope repetition rate of F_0. For

*Corresponding author: jnw@oticon.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

normal-hearing individuals, the smallest detectable shift in frequency was $0.05F_0$. The hearing-impaired group with moderate cochlear loss performed poorer. For most subjects and F_0s the performance was not significantly above chance level even for the maximum shift at $0.5F_0$. Above chance performance was only seen when the hearing loss at the center frequency of the band pass filter was little or none.

The inability to detect shifts in temporal fine structure in individuals with moderate cochlear loss have led to the idea of devising a Temporal Fine Structure (TFS) test (Moore and Sęk, 2009). The underlying hypothesis of the TFS test is that TFS information might be useful in deciding the most appropriate speed of compression for a hearing-impaired individual. The test should be applied quickly and reliably in a clinical situation.

An important note is that even slight or mild hearing losses experience issues with temporal fine structure. Therefore, it is expected that any hearing loss will show a reduced of sensitivity to temporal fine structure information (Ardoint *et al.*, 2010).

ERIKSHOLM RESEACH USING THE TFS1 TEST

Researchers at the Eriksholm Research Centre have explored whether the TFS1 test could be used to unveil differences in normal-hearing and hearing-impaired persons with mild to moderate hearing loss (Hietkamp *et al.*, 2010).

The TFS1 test

The TFS1 test used in the study was similar to the test described in Moore and Sęk (2009)[1]. The test paradigm is based on an A/B comparison of two sequences or intervals of stimuli: one where the F_0 harmonic is held constant in all presentations and one where every other stimulus is shifted by xF_0 (Fig. 1). The task is to select the sequence that contains the shifted stimuli. After each response, the participants get visual feedback whether the response is correct or incorrect.

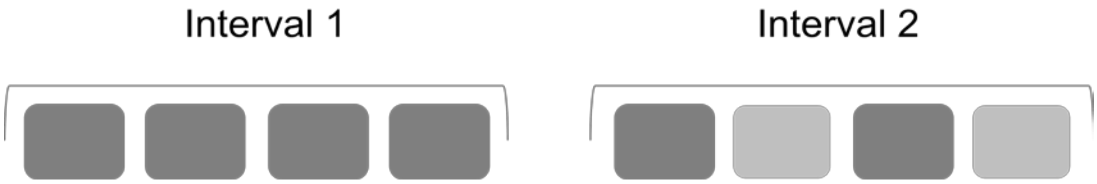


Fig. 1: The task for the participants was to choose the interval with fluctuating stimuli. The order of presentation was randomized for each trial. Test parameters for each test condition are adjustable and include the fundamental frequency (F_0), center frequency (F_c), and harmonic.

[1] Software for the TFS test is available from University of Cambridge’s homepage at http://hearing.psychol.cam.ac.uk/

Test parameters for each test condition are adjustable and include the fundamental frequency (F_0), center frequency (F_c), and harmonic.

The stimuli were band-pass filtered around one of the harmonics of the tone (e.g., the 5th or 11th). The auditory system does not appear to resolve harmonics above the 8th harmonic. This means that all components within the pass band were unresolved when the filter was centred at $11F_0$. Consequently, the excitation patterns were very similar for un-shifted and shifted stimuli. The band-pass filter had a central flat region width a width of $5F_0$ and skirts that decreased by 30 dB/octave. These are relatively shallow slopes which ensure minimal changes in the excitation pattern as components move in and out of the pass band.

The TFS1 test is an adaptive test in the sense that the size of the shift, the *x* in xF_0, is adaptive based on the response from the test person. This means that for individuals who are able to detect the difference in the stimuli, the test is able to calculate a threshold for the amount of perceivable shift in F_0 for a given F_0 and a given harmonic. Thus, the first presentation will include stimuli with the maximum shift. If this is correctly identified, the next presentation will include a smaller shift and so forth. The threshold is then calculated by averaging a specified number of the last reversals. Furthermore, for the study at Eriksholm the reversal rate had a maximum standard deviation (SD) of 0.15 to be deemed reliable. If the SD exceeded this level, another round of testing was included. The test was performed with speech at 20 dB SL and noise at 5 dB SL; the latter was introduced to mask any differences between the stimuli that were not attributed to the difference in temporal fine structure.

Of the hearing-impaired participants many were unable to complete the adaptive procedure. Here a percent correct method was used where the number of correct response at the maximum shift (here $0.5F_0$) was calculated.

Before initiating the test, the participants were trained in the procedure. As the sequences used in the test are of fairly short duration (0.2 s) and with short inter-stimuli pause (0.3 s), the test also requires a certain level of concentration and some short term memory. The training session differed significantly from the test sessions as it only needed to provide a baseline understanding of the test paradigm. For the training session the comparison was between two intervals with the same F_c, but with different F_0s or different repetition rates. Hence, in the training session, the fluctuation was between an interval consisting of four stimuli with an F_c of 1100 Hz and an F_0 of 100 Hz, and an interval that shifted between stimuli with F_0 of 100 Hz and F_0 of 150 Hz. This change in F_0 around the center frequency of 500 Hz changes the repetition rate between the harmonics from 300, 400, 500, 600, and 700 Hz to 100, 350, 500, 650, and 800 Hz. For the test condition, the focus is on the shift of F_0, for the training the focus is on the spread (size) of F_0. In the training session, all participants were able to perceive a difference between the presented sequences, and for all participants a threshold was obtained. The stimuli used for training did not contain a filter, nor was any background noise introduced, leaving greater differences between the stimuli presented in the fluctuating interval.

The baseline for the comparison was five harmonics of F_0 around a center frequency of F_c. Table 1 shows the F_cs and F_0s as well as the maximum shift in F_0 that was used in the study. The 'cleanest' comparison is the one made using the 11th harmonic, cf. its un-resolved nature in the auditory system.

	5th harmonic			11th harmonic		
F_c (Hz)	500	1000	2000	1100	2200	4400
F_0 (Hz)	100	200	400	100	200	400
Max F_0 shift (Hz)	50	100	200	50	100	200

Table 1. The test conditions varied on harmonic, center frequency (F_c), and fundamental frequency (F_0).

All participants were tested using the adaptive procedure, but if they were unable to go below the maximum shift of $0.5F_0$, a percent-correct method was used instead.

Results and conclusions from the study

The results from the study showed that hearing-impaired participants were significantly poorer at detecting the interval with shifted stimuli embedded in it. This was most clear for the test conditions using the 11th harmonic and center frequencies between 1100-4400 Hz. Here the division between hearing-impaired and normal-hearing participants was near binary. Only very few were able to reach a threshold and most scored no better than chance at the maximum shift of $0.5F_0$. The results also suggest that hearing-impaired individuals are better able to perceive differences in the repetition rate of F_0 than differences due to a shift in F_0. Thus, the hearing-impaired listeners seemed more sensitive to changes in envelope than to changes only in the temporal fine structure.

FREQUENCY SHIFTS IN HEARING INSTRUMENT SIGNAL PROCESSING

Using a frequency shift in signal processing in hearing instruments has both advantages and drawbacks. A great advantage is that the frequency shift alongside the feedback cancellation system decreases the susceptibility to entrainment, when external tonality is mistaken for internally generated feedback. A drawback lies in the artefact that may be perceivable when the difference between a shifted and a non-shifted signal is audible. This is most likely to occur in open solutions for individuals with fairly good low frequency hearing. The greater the shift, the more audible it can potentially be. However, greater shifts also provide more efficient de-correlation of hearing-aid output.

These attributes demands some consideration when implementing a frequency shift. In particular, how great a frequency shift must be used and for which inputs are the addition of a frequency shift necessary.

FEEDBACK SHIELD ON THE OTICON INIUM PLATFORM

The Oticon anti-feedback strategy builds on the principles of dynamic phase inversion or feedback cancellation (DFC) for destructive interference between a feedback signal and a cancellation signal produced by the anti-feedback system. In the anti-feedback system, the feedback path is constantly measured. Rapid changes in the feedback path, e.g., when the sound environment is very dynamic or when there are physical movement/changes close to the hearing instruments, warrant rapid and frequent measures of the feedback path. When the feedback path is more stable or only changes slowly, measurements of the feedback path are needed less often.

When the system is whistling or close to whistling, the phase inversion is put to its full use. The feedback path is measured and a mirror or phase-inverted version of the feedback signal is imposed to cancel out the feedback. This may seem simple enough, but the precision of the inverted signal is paramount. The phase inversion must not only be precise, but also fast to cancel feedback out even before it is actually perceivable as feedback or whistling. To be successful in removing the feedback loop, the signal needed to cancel out the feedback must be equally complex.

However, under some conditions or in some environments with high auto-correlation, updating the feedback path is less beneficial for the anti-feedback system because of the risk of disturbing the sound quality. When this is the case, feedback shield maintains the last good feedback path estimate and cancels out feedback based on this estimate. In other situations, it is safe and wise not only to keep updating, but sometimes also to increase the frequency of updates. When doing this, a frequency shift of 10 Hz is enabled to render the system less susceptible to tones in the environment, thus making it easier to correctly identify internally generated feedback loops and not mistakenly attempt to cancel out external tones. The frequency shift works by shifting the entirety of the input signal above a given transition frequency 10 Hz upwards. In shifting the majority of the signal, the envelope is kept intact and only the fine structure is affected.

Because the output of the hearing aid is slightly different from the input due to the frequency shift, potential acoustic leakage will not line up with the input and create a feedback. Thus, processed sound going back to the microphone is more easily distinguished from input from the environment. Using a frequency shift is a very effective method for decreasing system-sensitivity to tonal inputs, thus allowing other parts of the system to update.

Three modes are implemented in the current feedback system to accommodate for the changing/alternating need for using frequent updates, and thus the addition of the frequency shift. The shift between the modes is based on input from two detectors: a howl detector and a tonal detector (Fig. 2).

The main job of the anti-feedback system is to prevent audible feedback or whistling from happening. Thus, the primary task is the detection of potential audible feedback. Based on a calculation of the auto-correlation in the output, the howl detector will determine whether the system will need to adapt and go to a more aggressively updating mode or whether the choice of mode can be based on the input from the tonal detector.

A detection of audible feedback will enable fast updates to the DFC system supported by the 10-Hz frequency shift.

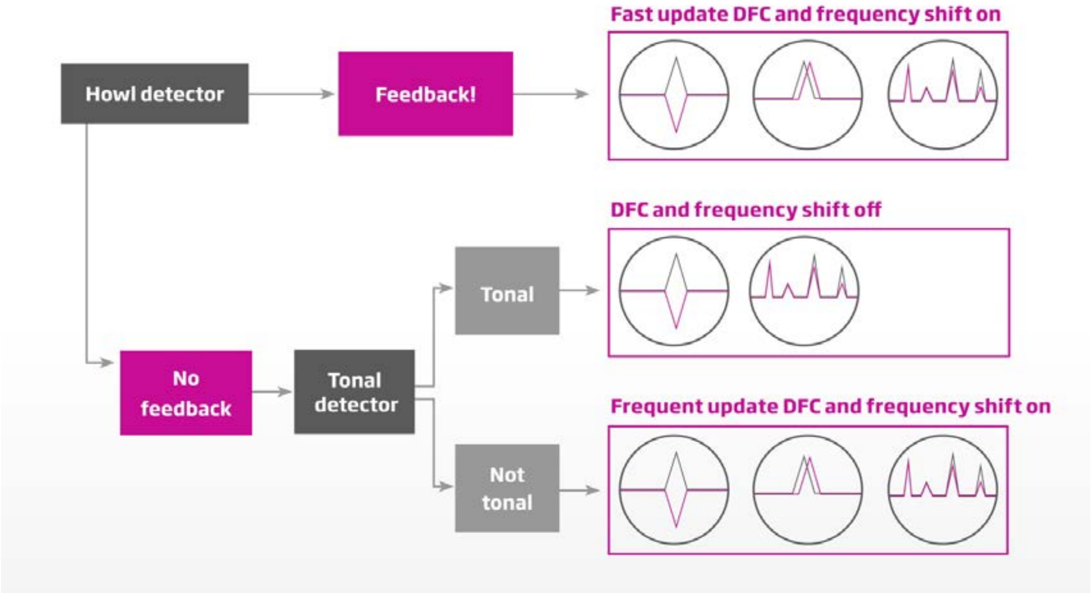

Fast update DFC and frequency shift on
Howl detector
Feedback!
DFC and frequency shift off
Tonal
No feedback
Tonal detector
Frequent update DFC and frequency shift on
Not tonal

Fig. 2: Diagram of how the different detectors and modes are configured in the anti-feedback system.

If audible feedback is not present, the tonal detector determines which mode should be used based on the presence of tonality in the environment. Tonality is characterized by repeatable, harmonic content such as acoustic stimuli like speech and music. This type of content can be mistaken for feedback by the anti-feedback system, so that extra care needs to be asserted to avoid suppressing any valuable information or creating a loop in phase with the input causing additional feedback.

When the content of the input signal is tonal, the system will enable a less aggressive mode. Here the DFC behaves in a stable manner in the sense that the last updated feedback path filter/estimate is applied to the input signal. The assumption is that the acoustic properties around the hearing aid will not have changed and thus this filter will still be applicable. When the DFC is not being updated, the frequency shift is not relevant and will be disabled. When the content of the input signal is *not* tonal, the Inium feedback shield will be in dynamic mode. The risk of degrading either speech or music is no longer present, and therefore, the DFC will allow more frequent updates of the filter according to the subtle changes in the feedback path. The frequency shift is enabled in dynamic mode to ensure robust feedback path estimation.

An internal test compared the previous RISE2 platform to the Inium platform on a feedback 'stress' test of hearing-instrument performance. The test setup consisted of a head-and-torso simulator (HATS) with hearing instruments mounted on the ears and a mechanical arm moving to and from the ear. A test sequence featuring rising pure tones, clicks, classical flute play, and other very tonal sound content was played, processed by the hearing instrument, and recorded in the ear of the HATS.

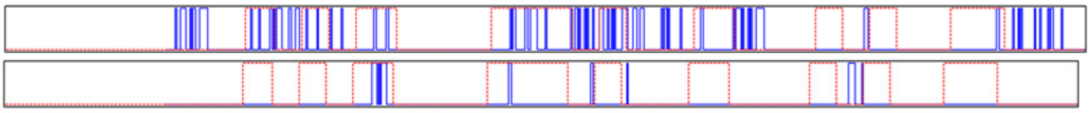


Fig. 3: Results from feedback stress test, RISE2 (top) versus Inium (bottom). The red lines indicate movement of a mechanical arm to and from the ear of the HATS and blue lines indicate identification of audible feedback.

A 'golden-ear' listener evaluated the occurrences of feedback in the recordings with the two instruments. The test showed that the anti-feedback system on the Inium platform reduced the number of audible instances of feedback by 80%.

CONCLUSION

The introduction of a small frequency shift can help improve an anti-feedback system by de-correlating the input without (too) many sacrifices to the requirements/demands for sound quality in hearing-impaired listeners. The implementation of such a shift plays a great role, as research indicates that a shift in the frequency content of the hearing-instrument output may be audible in the form of acoustic artefact and disturbing to the listening experience. The trade-off between the advantages and downsides to the use of the technology warrants diligence when choosing what input can or cannot be added to the signal path. The improvement of the Inium-based feedback shield has yielded great improvements to the frequency of occurrence of feedback in Oticon hearing instruments; a reduction in audible feedback by 80% was seen in an instrument stress test performed by a 'golden-ear' listener.

ACKNOWLEDGEMENTS

We would like to acknowledge the valued help from our colleagues at Oticon A/S Headquarters and at the Eriksholm Research Centre, and in particular Steen Michael Munk, Marianne Philipsson, and Renskje Hietkamp.

REFERENCES

Ardoint, M., Sheft, S., Fleuriot, P., Garnier, S., and Lorenzi, C. (**2010**). "Perception of temporal fine-structure cues in speech with minimal envelope cues for listeners with mild-to-moderate hearing loss," Int. J. Audiol., **49**, 823-831.

Hietkamp, R.K., Andersen, M.R., Kristensen, M.S., Pontoppidan, N.H., and Lunner, T. (**2010**). "The TFS1-test reveals mild hearing loss," American Speech-Language-Hearing Association Convention, Philadelphia, PA.

Hopkins, K., and Moore, B.C.J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am., **122**, 1055-1068.

Hopkins, K., and Moore, B.C.J. (**2011**). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," J. Acoust. Soc. Am., **130**, 334-349.

Moore, B.C.J. (**2007**). "The role of temporal fine structure in normal and hearing impaired hearing," in *Auditory Signal Processing in Hearing-Impaired Listeners*, 1st International Symposium on Auditory and Audiological Research, Helsingør, Denmark.

Moore, B.C.J., and Sęk, A. (**2009**). "Development of a fast method for determining sensitivity to temporal fine structure," Int. J. Audiol., **48**, 161-171.

Strelcyk, O., and T. Dau (**2009**). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," J. Acoust. Soc. Am., **125**, 3328-3345.

Horizontal localization with pinna compensation algorithm and inter-ear coordinated dynamic-range compression

PETRI KORHONEN[*]

Widex, Office of Research in Clinical Amplification (ORCA-USA)

Many hearing-aid users show poorer aided than unaided localization performance even when audibility is accounted for. One source of potential disruption of aided localization include the use of wide dynamic range compression circuits operating independently at each ear in bilateral fittings, which can compromise the interaural-level-difference (ILD) cues used for left-right localization. The natural ILD cues can be restored by coordinating the gain between the two hearing aids wirelessly. Another potential source of disrupted localization include the absence of pinna-shadow when using behind-the-ear (BTE) hearing aids with omnidirectional microphones. A pinna shadow compensation feature that restores the natural attenuation for sounds originating from behind was developed. This study examined the localization performance of hearing-impaired listeners in the horizontal plane when using a BTE hearing aid incorporating inter-ear coordinated compression and a pinna-shadow compensation algorithm. Fifteen listeners who had previously participated in a localization study were recruited. The data demonstrated that the use of the pinna-shadow compensation algorithm improved the localization accuracy over a BTE hearing aid with an omnidirectional microphone. A modest improvement in localization performance was measured for some listeners when using the coordinated inter-ear compression.

INTRODUCTION

The physical presence of pinna attenuates high-frequency sounds that originate from the back and sides by an average of 5 dB from 2 kHz to 8 kHz. This attenuation provides an important acoustic cue for normal-hearing individuals to localize sounds along the median plane. The use of behind-the-ear (BTE) hearing aids with an omnidirectional microphone placed on top of the pinna eliminates the pinna shadow used for front-back localization, because an omnidirectional microphone has the same sensitivity to sounds from all directions. This lack of difference in sensitivity between sounds arriving from the front and the back may reduce front-back localization performance. The absence of the pinna shadow can be corrected so that, despite using a BTE with an omnidirectional microphone, the wearer will still have the 'normal' localization cues. The Digital Pinna (DP) hearing-aid feature was developed to compensate for the difference in input measured between an unaided ear and an aided ear with an omnidirectional BTE hearing aid. The DP algorithm sets the microphone system to a fixed hypercardioid polar pattern above 2000 Hz,

[*]Corresponding author: petri.korhonen@iki.fi

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

while keeping an omnidirectional mode below 2000 Hz. This simulates the natural pinna attenuation for sounds originating from the back.

For sounds arriving from the side of the listener the difference in the acoustic path from the source to the two ears results in a difference in the signal amplitude and phase characteristics at the two ears. This interaural level difference (ILD) and interaural time difference (ITD) provide the cue for left-right localization in the horizontal plane (Blauert, 1997). The use of wide-dynamic-range-compression (WDRC) circuits operating independently at each ear in bilateral fittings may compromise the ILD cues. The WDRC circuit provides more gain to low-level signals, and less gain to high-level signals. Sounds that arrive from the incident side will measure a higher sound pressure level at the microphone opening than the opposite ear because of the head shadow. In this case the WDRC hearing aid applies less gain on the side of the sound source than on the opposite side. As a result, the use of two independently operating WDRC hearing aids at each ear may result in output levels between the two ears such that the natural ILD is not preserved. The coordination of the gain between the two hearing aids can restore the natural ILD. The hearing aid used in the current study included functionality where each input received by one hearing aid was shared wirelessly with the other hearing aid. The gain calculated for the target sound side was used in both hearing aids.

While the inter-ear coordinated compression and digital pinna can restore the natural cues that hearing-aid processing may have distorted, it is possible that listeners may not be able to interpret these new localization cues immediately. In fact, the effect of distorted ILDs on localization has been reported to be rather small in hearing-impaired listeners (Keidser *et al.*, 2006; Musa-Shufani *et al.*, 2006). Maybe the reported small impact of distorted ILDs on localization has been a result of the listeners' inability to fully utilize the ILD cues. For that reason, it would be worthwhile to examine the effects of WDRC on localization with listeners that have participated in auditory localization training.

Keenan (2013) developed and evaluated a training program that focused on localizing sounds in the horizontal plane. The training included both computerized laboratory-based and take-home programs. These programs provided immediate feedback and learning opportunities, and were designed to motivate the participants for greater success by adaptively changing the difficulty of the stimuli by varying the duration of the stimuli, and by exaggerating the pinna-shadow attenuation for sounds that originated from the back. The laboratory-based training utilized a 12-loudspeaker array distributed evenly at 360° separated by 30°. The take-home training used two loudspeakers located at 0° and 180°. The trainee's task was to indicate from which loudspeaker they perceived the stimulus. The trainee was given an opportunity to compare the perception between the correct loudspeaker direction and their indicated loudspeaker direction after each stimulus presentation.

The listeners who have received localization training may be more sensitive to changes caused by the inter-ear coordinated WDRC and the digital pinna. To test this hypothesis, the current study recruited the participants that had received localization training previously in the Keenan (2013) study. The current study examined the horizontal localization performance of these listeners when using a BTE hearing aid incorporating the digital pinna feature and the inter-ear coordinated compression.

METHODS

Subjects

Fifteen participants (7 males and 8 females) with bilateral sensorineural hearing loss were recruited. The averaged pure-tone averages were 48.6 dB HL (standard deviation, SD = 11.8) for the right ear and 50.1 dB HL (SD = 12.0) for the left ear. The symmetry of hearing loss was within 15 dB between ears at any frequency. One subject had a threshold difference of 20 dB at 6000 Hz and another a difference of 25 dB at 8000 Hz. Participants' ages ranged from 28 to 83 yr (mean = 71 yr, SD = 12.9). Ten participants had received one month of home-based training and one month of laboratory-based training as a part of a separate study. Five participants had received six days of laboratory-based training. Participants signed informed consent and were financially compensated for their participation in the study.

Hearing aids

Participants were fitted bilaterally with Widex C4-m-CB BTE hearing aids using custom earmolds. This 15-channel wide-dynamic-range-compression micro-BTE hearing aid uses a relatively long attack time of up to 2 s and a long release time of up to 20 s in each of the 15 channels for most situations. This hearing aid includes a pinna-compensation algorithm called Digital Pinna which was designed as a directional microphone with a hypercardioid pattern above 2000 Hz, which approximates the unaided in-situ directivity below 1.5 kHz and has a directivity index (DI) of 4 dB above 2 kHz. This hearing aid also includes wireless functionality where input received by one hearing aid is shared with the other aid of the bilateral pair wirelessly at a rate of 21 times per second using near field magnetic induction (NFMI). The data exchange coordinates the gain parameters so that the gain at each ear corresponds to the gain calculated for the side of the more intense sound.

Testing

Testing was conducted in a double-walled sound-treated test booth with internal dimensions of 3 × 3 × 2 m (W × L × H). The target stimulus was a three-second female speech sample presented in quiet at a 30 dB sensation level (SL). The stimulus was presented using twelve loudspeakers (KRK-ST6) distributed evenly (30° spacing) on a horizontal plane around the listener (1-m distance) at one-meter height from the floor. The participants were asked to keep their heads fixed towards the loudspeaker at 0°. Four hearing-aid settings were compared: omnidirectional microphone (Omni), omnidirectional microphone with digital pinna (DP), omnidirectional microphone with inter-ear coordinated compression (IE), and omnidirectional microphone with IE and DP (IE+DP). In addition, the unaided performance was measured. The test conditions were counterbalanced across participants. The participants indicated the perceived location of the target by touching a touch-screen computer monitor placed in front of them. The stimulus was presented from each azimuth three times during each test trial.

RESULTS

The current study reported the error characteristics of sound-localization performance using a 'centre of mass' (CoM) method pioneered and presented in detail by Edmondson-Jones *et al.* (2010). The CoM method examines the proportion, direction, and size of errors simultaneously. The CoM analysis is represented visually with a unit circle centered at the origin in the Cartesian coordinate system in the Euclidean plane. The planar Cartesian coordinates for a single response are defined as (sin θ, cos θ) (Fig. 1, left). The CoM is the mean location of all the 'mass' in a group of bodies (Fig. 1, right). Each data point is assumed an equal weight of a unit mass. For a sample of N observations the sample mean is defined as

$$\overline{X} = \left(xCoM, yCoM\right) = \left(\frac{1}{N}\sum_{i=1}^{N}\sin\theta_i, \frac{1}{N}\sum_{i=1}^{N}\cos\theta_i\right)$$

When the responses are perfectly correct the yCoM will be 1 while xCoM will be 0. Thereby, yCoM is a measure of the target accuracy indicating how close the responses are from the perfect responses, and xCoM is a measure of lateral accuracy indicating the response distance from the origin. Applying the standard Multivariate Analysis of Variance (MANOVA) methods, Edmondson-Jones *et al.* (2010) demonstrated that the CoM method performs well in the control of Type I errors and showed its power to detect significant changes in localization responses. Post-hoc methods are also appropriate to investigate the front-back location effects or effects in different quadrants.

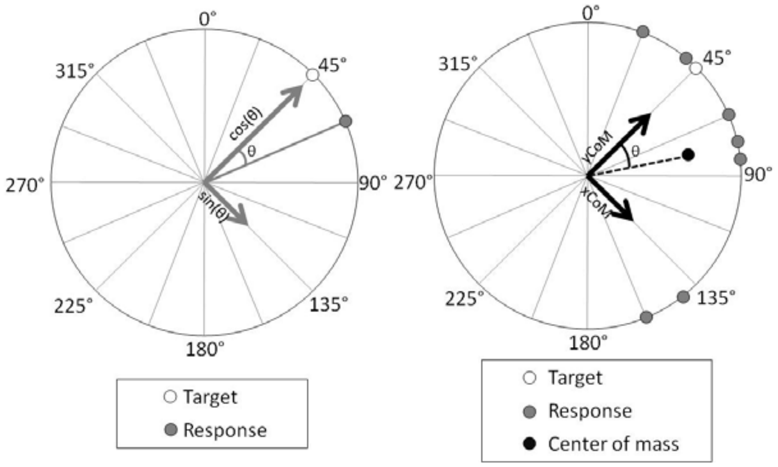


Fig. 1: Visual representation of center of mass (CoM) analysis method for a single response (left), and a group of seven responses (right) for a stimulus at 45°.

Figure 2 shows the averaged localization performance (target-accuracy) for all participants under the five test conditions. Again, the target accuracy of 1 suggests localization performance is perfect. The performance was 0.51 with Omni, 0.58 with IE, 0.68 with DP, with 0.64 IE+DP, and 0.66 unaided. The localization performances between each of the listening conditions were compared pair-wise with one-way ANOVA. The condition pairs with a significant difference are shown with connectors. There was a significant difference (*p* < .05) between the Omni condition and all other conditions. There was no significant difference between unaided condition, and the DP or IE+DP conditions.

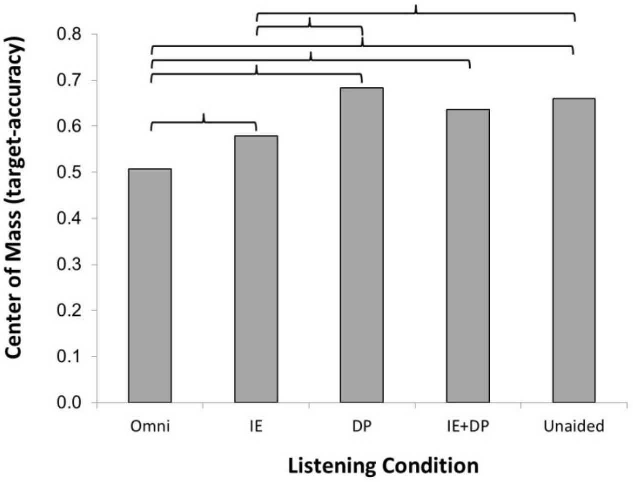


Fig. 2: Averaged localization performance for all participants (N = 15). Test conditions included Omni, IE, DP, IE+DP, and unaided. Comparisons where statistical significance was reached (*p* < .05) are shown with connectors.

Figure 3 compares the unaided and the Omni conditions. Note that the more accurate the performance, the closer the location of the CoM was towards the edge of the unit circle. Statistical significance in the difference in performance between the evaluated conditions was analyzed for front (330°, 0°, 30°), right (60°, 90°, 120°), back (150°, 180°, 210°), and left (240°, 270°, 300°) quadrants separately using one-way ANOVA. The quadrant where there was a significant (*p* < .05) difference between the conditions is indicated with gray shading. With the Omni condition the listeners had difficulties localizing in front-back dimension. A significant difference (*p* < .05) in performance between the unaided and Omni conditions was observed in the back quadrant. This demonstrated that the use of omnidirectional microphone in a BTE hearing aid distorted the natural pinna cue used for front-back localization. No significant difference was observed for the other three quadrants.

Figure 4 compares the Omni and the DP conditions. Digital pinna was designed to correct for the absence of the pinna shadow in a BTE hearing aid with an omnidirectional microphone. In fact, the localization performance was significantly better ($p < .05$) with the DP than with the Omni condition in the front and the back quadrants. No significant difference was observed for the right and left quadrants.

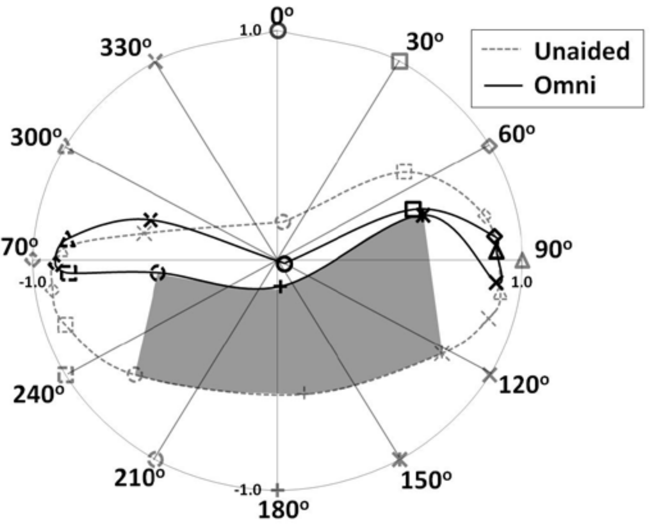


Fig. 3: Localization performance with Unaided (dashed) and Omni (solid). The gray area indicates the quadrant with significant ($p < .05$) difference between the test conditions.

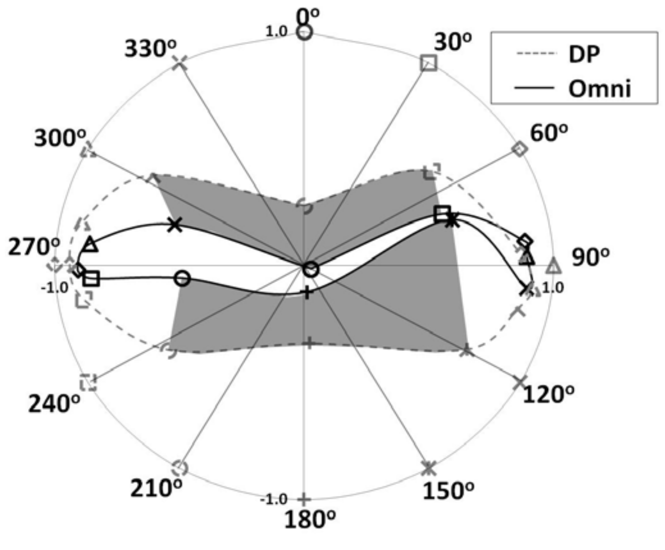


Fig. 4: Localization performance with DP (dashed) and Omni (solid). The gray area indicates the quadrant with significant ($p < .05$) difference between the test conditions.

With both the Omni and the IE conditions the listeners reached a high level of accuracy for sounds arriving from the left or the right, and there was no significant difference between the two conditions for any quadrant. While most participants reached a high level of accuracy for sounds arriving from the sides, individual differences existed. Variation in individual localization performance among participants prompted us to investigate whether the effect of IE compression was dependent on the localization ability. For the participants with the poorest localization performance (yCoM < 0.87) the target accuracy was better with the IE than the Omni condition ($F(1,14) = 5.81$, $p = 0.03$, $\eta^2 = 0.29$, power = 0.6) (Fig. 5).

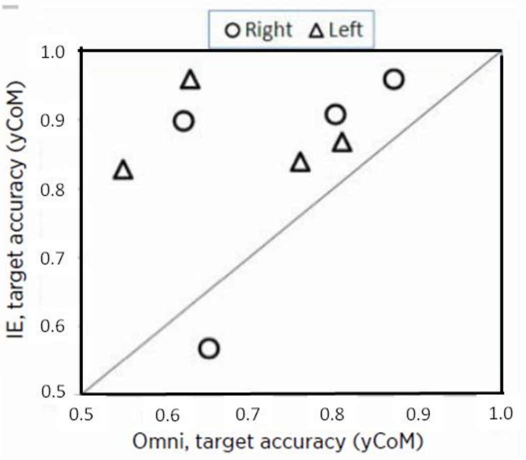


Fig. 5: Scatter-plot comparing individual localization performance between the Omni (x-axis) and IE (y-axis) condition for sounds arriving from the sides (left quadrant: triangles; right quadrant: circles) for the four poorest performers (yCoM < 0.87).

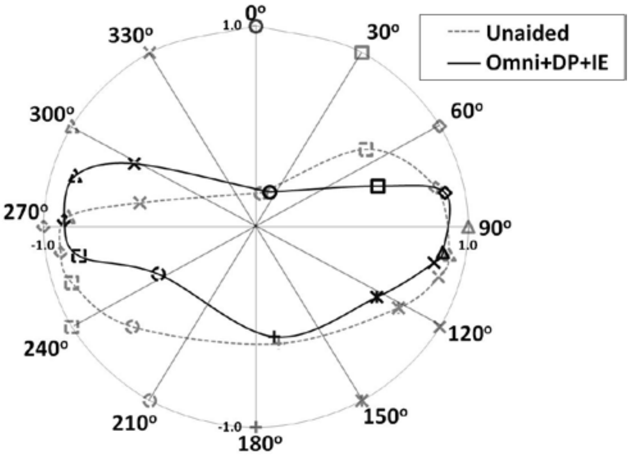


Fig. 6: Localization performance with Unaided (dashed) and IE+DP (solid). The gray area indicates the quadrant with significant (p < .05) difference between the test conditions.

Figure 6 compares unaided performance and the IE+DP conditions. Unlike in the Omni condition, no difference (*p* < .05) was seen in performance between unaided and the IE+DP conditions for any quadrant. In other words, the unaided localization performance was retained when using the digital pinna and inter-ear coordinated compression.

CONCLUSIONS

We demonstrated that the use of the digital pinna feature, as implemented on the hearing aid in the current study, improved front-back localization accuracy in the horizontal plane over a BTE hearing aid with and omnidirectional microphone. We also demonstrated that inter-ear coordinated compression was providing a helpful cue for localization for those listeners who had poorer localization performance for sounds arriving from the sides. The unaided localization performance was better than aided performance when using an omnidirectional microphone alone. However, the use of the digital pinna feature together with inter-ear coordinated compression was successful in restoring the compromised aided performance.

All participants in the current study had received localization training prior to participating in the current study. Also no one wore the hearing aid during the study. We can therefore expect that differences in performance among test conditions reflected the efficacy of the technology and not listener experience or technology alone.

It is worth noting that the effect of the coordinated compression on localization may have been lessened by the slow-acting compression used in the studied hearing aid. In a quiet or in a diffuse sound field, the onset of a sound originating from the side will quickly adjust the gain at each ear in response to changes in the input in a fast-acting WDRC hearing aid. On the other hand, a slow-acting WDRC takes longer to adjust to the final gain setting. Consequently, the natural ILD cues may be interrupted more rapidly in a fast-acting WDRC than a slow-acting WDRC hearing aid, such as the one used in the current study.

REFERENCES

Blauert, J. **(1997)**. *Spatial Hearing* (Cambridge, MA: MIT Press).

Edmondson-Jones, M., Irving, S., Moore, D.R., and Hall, D.A. **(2010)**. “Planar localization analyses: A novel application of a centre of mass approach,” Hear. Res., **267**, 4-11.

Keenan, D. **(2013)**. “An approach to training localization,” Poster presented at Am. Acad. Aud. meeting, AudiologyNow, Anaheim, CA, April 3-6, 2013.

Keidser, G., Rohrseitz, K., Dillon, H., Hamacher, V., Carter, L., Rass, U., and Convery, E. **(2006)**. “The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers,” Int. J. Audiol., **45**, 563-579.

Musa-Shufani, S., Walger, M., von Wedel, H., and Meister, H. **(2006)**. “Influence of dynamic compression on directional hearing in the horizontal plane,” Ear Hearing, **27**, 279-285.

Degradation of spatial sound by the hearing aid

JESPER UDESEN[1,*], TOBIAS PIECHOWIAK[1], FREDRIK GRAN[1], AND ANDREW B. DITTBERNER[2]

[1] *GN ReSound A/S, Lautrupbjerg 7, DK-2750 Ballerup, Denmark*

[2] *GN ReSound North America, 8001 Bloomington Freeway, Bloomington, MN 55420-1036, USA*

It is well known that the hearing aid distorts the spatial cues used to localize sound sources and this has severe consequences for sound localization and for listening in noise. However, it is not clear how the different components in the hearing aid contribute to the degradation of spatial sound. In this study we investigate how the spatial sound is degraded by four hearing aid components: 1) the microphone location, 2) the directionality (beamforming), 3) the compressor, 4) the real ear measurement. Head Related Transfer Functions from an artificial KEMAR head are convolved with appropriate excitation sounds and processed through the respective hearing aid algorithm. The performance metrics under investigation are: 1) interaural level difference (ILD), 2) interaural time difference (ITD), 3) monaural spectral cues. It is found that the main source for ILD degradation is the position of the microphone around the pinna which distorts the ILD by up to 30 dB. It is also found that the real ear measurement compensation severely affects the monaural spectral cues.

INTRODUCTION

It has been known for more than 100 years that the acoustic signals at the ears contain a multitude of information about the spatial nature of any of the sources in the acoustic wave field. This spatial information is encoded in interaural time differences (ITD), interaural level differences (ILD), spectral cues, and reverberation cues (Blauert, 1997). Binaural processing by the brain, when interpreting the spatially encoded information, results in several positive effects: better signal-to-noise ratio (SNR), direction of arrival estimation, depth/distance perception, and synergy between the visual and auditory systems. Therefore, better localization performance will improve sound quality as well as hearing in noise (Hawley *et al.*, 1999).

Even though the benefits of spatial sound are well known, it is not clear how the different components and algorithms of a state-of-the-art hearing aid will distort the spatially encoded information. Previous studies have mainly focused on the localization performance of hearing-impaired test subjects when wearing different types of hearing aids (e.g., Van den Bogaert *et al.* (2006)). Using real hearing aids gives realistic test results but makes it difficult to identify the true sources of any

[*] Corresponding author: judesen@gnresound.com

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

degradation of spatially encoded information.

In this study we investigate how four of the main components of a state-of-the-art hearing aid distort the spatially encoded information. The components are: 1) the spatial position(s) of the microphone(s) on the ear, 2) the hearing aid directionality system, 3) the hearing aid compressor system, 4) the real ear measurement (REM) procedure.

A mathematical model of the algorithm package in a state-of-the-art hearing aid will be used to investigate the degradation of spatially encoded information. The input signal to the model will be encoded with head-related transfer functions (HRTFs) based on the corresponding microphone positions measured on a artificial KEMAR (Knowles Electronics Manikin for Acoustic Research) head. The spatially encoded information will be described by ILD, ITD, and HRTF information and any degradation of these cues will be expressed relatively to the open ear response.

SETUP

The HRTFs were recorded on a KEMAR manikin located in an anechoic room in accordance with ISO 3745 and approved for measurements between 30 Hz and 10 kHz. KEMAR was rotated with a B&K 5960 turntable in steps of 2° covering a full 360° rotation and a KEF Q85S speaker was used to transmit a 5 seconds code length 13 maximum-length-sequence (MLS) signal (Golomb and Gong, 2005). The hardware used for the sound recordings and the play back was a Tucker-Davis-Technologies RX8 sound processor running at a sampling frequency of 48828 Hz. All hardware components were controlled from Matlab.

The microphones used for the recordings were placed at 45 different positions on a small female ear and a large male ear respectively. The microphone positions and the ears can be seen in Fig. 1.

The speaker and microphone transfer functions were removed from the HRTFs offline using a standard deconvolution algorithm implemented in Matlab. Furthermore, the open ear responses were recorded with a 711 coupler.

Fixed directionality (FD) beam forming using a hyper cardioid beam pattern was applied on the following microphone positions $\{[1,11],[11,21],[21,24]\}$. The beam forming filters were derived from the microphone data when they were mounted on KEMAR. The directionality beam forming was implemented using finite-impulse-response (FIR) filters with a length of 101 samples (at f_s=15625 Hz). The compressor algorithm was implemented using a warp band delay line with compressor knee-point at 50 dB.

REM compensation was applied to ensure that the output of the hearing aid had the same amplitude spectrum as the open ear response (the coupler measurement). The REM compensation filter was implemented as a 1501-taps FIR filter which transformed the open ear impulse response at 0° angle into the corresponding head-

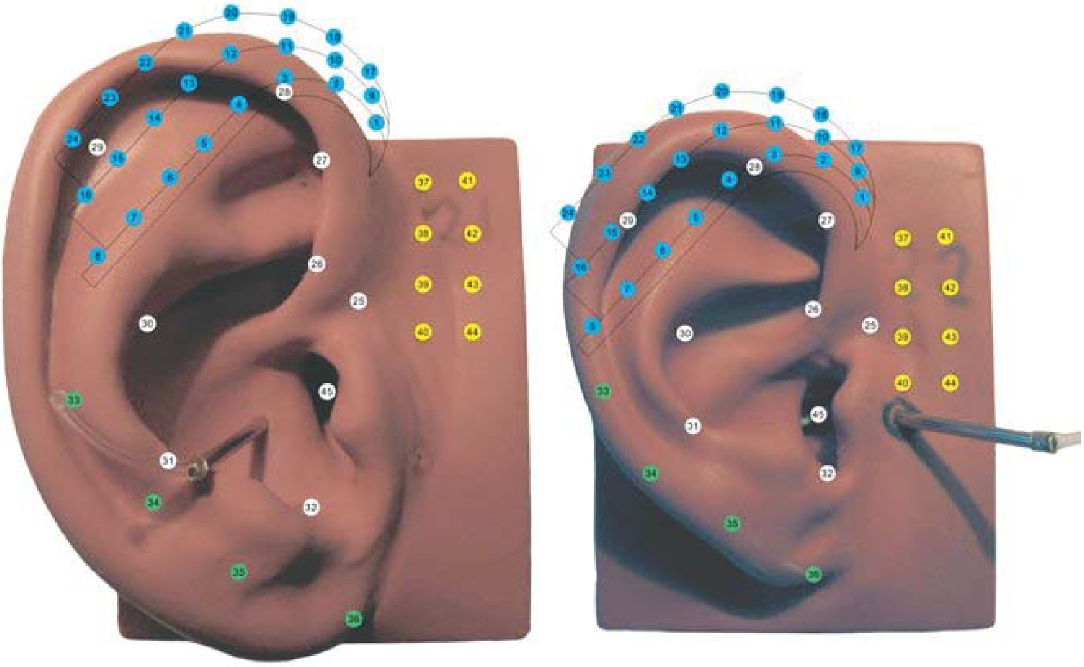


Fig. 1: The two artificial ears used on KEMAR and the 45 microphone positions. The last microphone position is the coupler microphone. All the ‘blue’ positions are located behind the pinna. Position 45 is located at the entrence of the ear canal.

related-impulse-response measured at 0° in a least-square sense. This filter was applied on all the HRTF data for that given microphone position.

RESULTS

The microphone positions

In Fig. 2 the broad band ILD and ITD (ITD only below 2 kHz) for all 46 microphone positions are shown. It is clear that especially ILD is significantly affected by the change in microphone position, and the corresponding ILD error (deviation between the ILD for a given microphone and the corresponding open ear ILD) is as big as 5-8 dB for angles around 90° and 270°. If the ILD of the open ear response was distorted by 5 dB at these angles, it corresponds to an error on the open ear angle estimate of more than 50° (found from visual inspection of Fig. 2). The maximum ITD for the open ear (coupler) is approximately 750 μs which is approximately 100μs less than the maximum ITD values for certain microphone positions. If the open ear ITD is distorted by 100μs this will result in an error of up to 50° (Found from visual inspection of Fig. 2). It should also be noted that the human auditory system is sensitive to ITD changes as small as 13μs (Hartmann, 1999)

In Fig. 3 the frequency dependent ILD for 6 selected microphone positions are shown

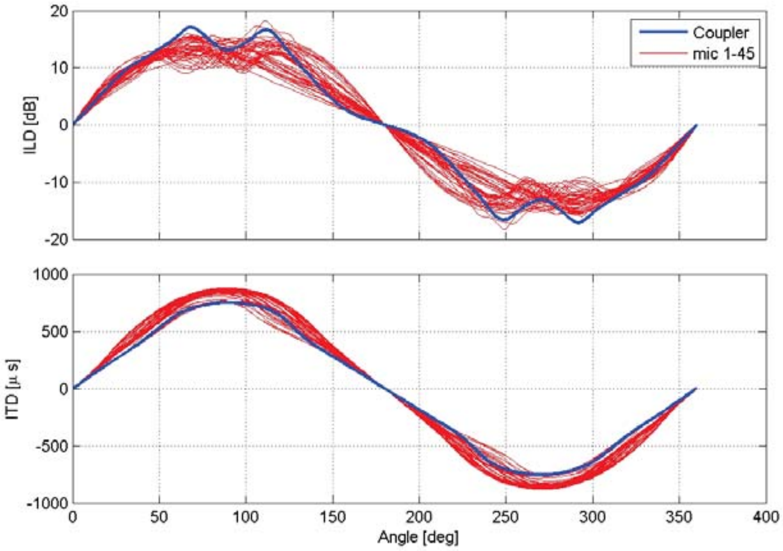


Fig. 2: Top: Broad band ILD (0-10 kHz) for the 45 different microphone positions (red curves) as well as the coupler microphone (thick blue curve). Bottom: The corresponding ITD evaluated between 0-2 kHz using a cross correlation estimator.

where all 6 microphone positions are relevant from a hearing-aid perspective. Based on Fig. 3, it is clear that the ILD error can get much larger at some frequencies than the corresponding broad band ILD error. Especially the microphone positions located behind the pinna have ILD errors of $\sim$ 30 dB at many frequencies. Humans are sensitive to ILD changes of 0.5 dB (Hartmann, 1999) so it is reasonable to assume that an ILD error of 30 dB is noticeable. However, care should be taken here since an increase in ILD at 90° from 40 dB (which is the natural open ear ILD at 5 kHz) to 70 dB will not move the perceived angle of the sound source in space. It is more likely that the listener will experience the sound as internalized. This is often the result when the human auditory system is presented to signals processed through non-personalized HRTFs (Hartmann and Wittenberg, 1996).

Fixed directionality beam forming

Fig. 4 shows the broad band ILD and ITD (ITD only under 2 kHz) as a function of angle for the three different microphone pairs tested for fixed directionality using the large male ear on KEMAR. For comparison also the ILD/ITD for the open ear and for microphone position 1 are shown. The ITD is up to 300 μs higher for the fixed directionality signals compared to the open ear ITD at 260°. This corresponds to an ITD error of $100 \cdot 300\mu\text{s}/750\mu\text{s} = 40\%$. Even though this ITD error is large it is not clear what the perceptual consequences are. The maximum error occurs at angles where the sound source is either directly to the left or to the right of the listener. An

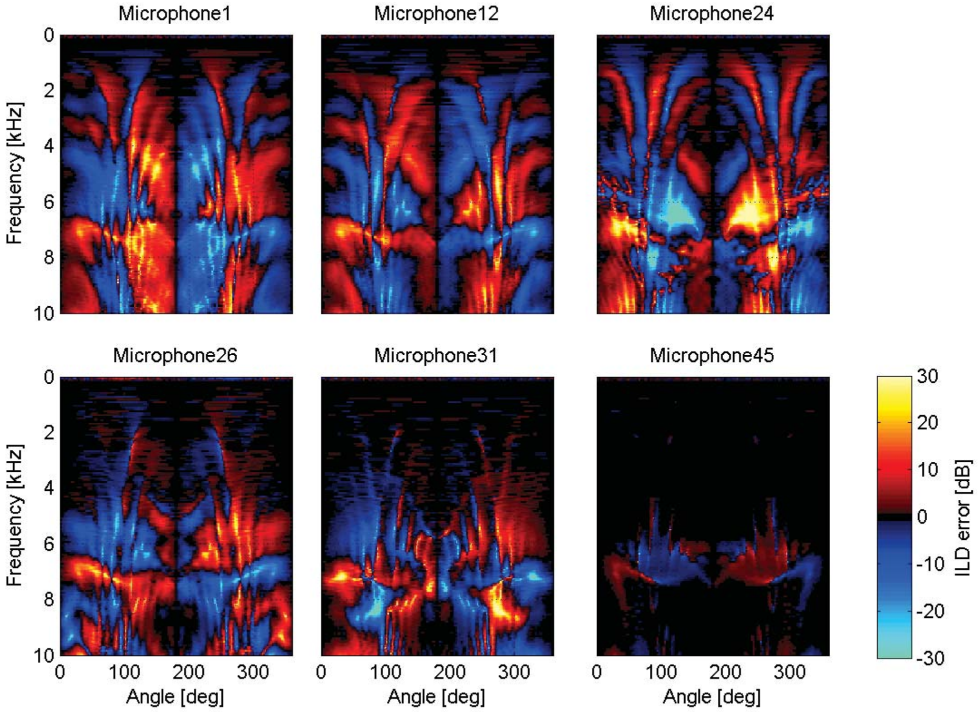


Fig. 3: The ILD error (difference between the open ear ILD and the estimated ILD) for 6 different microphone positions

ITD larger than the natural one at these angles does not move the source location in space, it rather creates a unatural sound impression.

Fig. 4 also shows that fixed directionality distorts the ILD significantly for certain angles. At 120°, where the null in the beam pattern is positioned, the ILD becomes positive for all three tested microphone positions. This means that high-frequency (above 1.5 kHz) signals coming from the right hemisphere at an angle around 120° will be perceived as coming from the left hemisphere. The worst performance is achieved for the microphone pairs (21, 24) which are located on the 'back' of the pinna (pointing backwards). Here the broad band ILD error is 22 dB, which should be compared to the smallest ILD difference of 0.5 dB detectable by humans (Hartmann, 1999).

The compressor

The compressor was tested with four compressor ratios 1, 2, 3, 4, which corresponds to the slope of the input/output gain curve of the hearing aid. The result on broad band ILD and ITD for both male speech and white noise as input signals can be seen in Fig. 5. Here the microphone position 45 was used on the large male ear on KEMAR and SPL was 65 dB (for 0°). The graphs in Fig. 5 show a very clear trend where

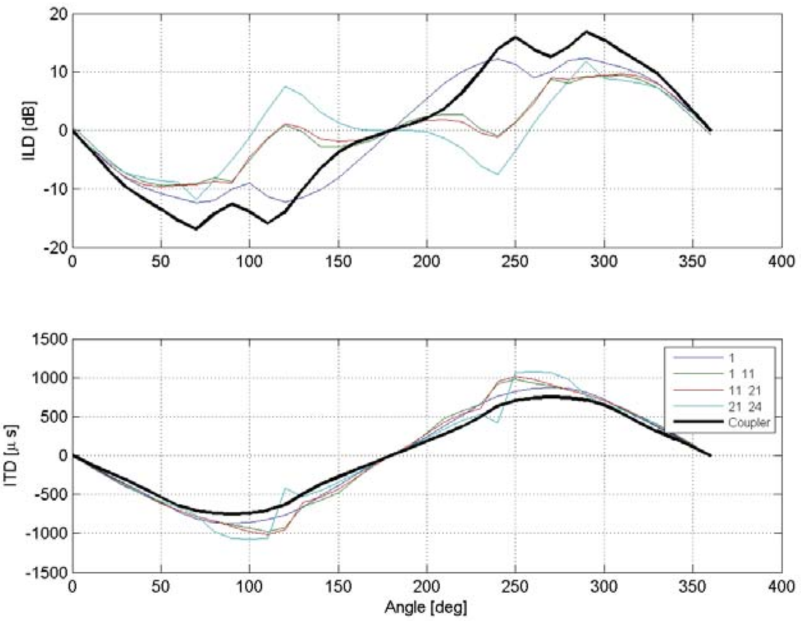


Fig. 4: The ILD and ITD for the fixed directionality beam forming. As reference also the open ear ILD/ITD are shown as well as the ILD/ITD for microphone position 1. Peaks in the ITD plot around 120° is due to the poor signal to noise ratio.

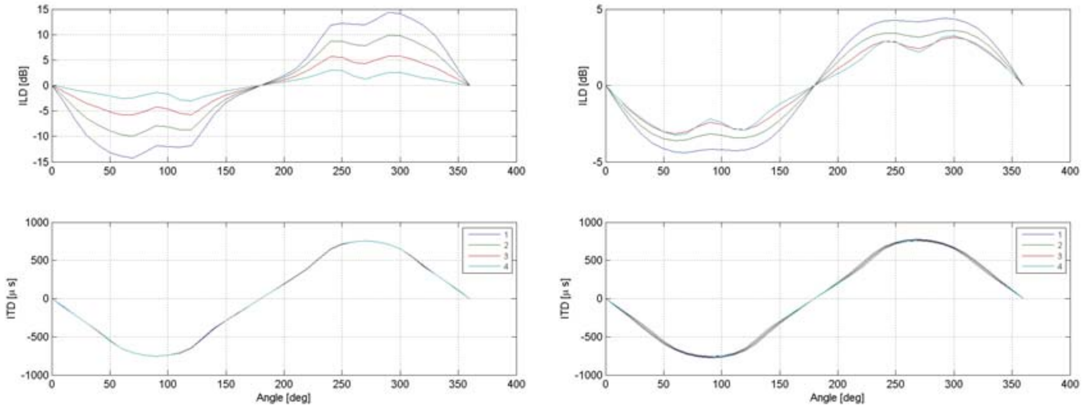


Fig. 5: Left: The broad band ILD and ITD for 4 different compressor ratios when white noise was used as input signal. The microphone at position 45 was used to record the signal on the male ear. Right: same as left but with male voice as input signal.

ILD is decreasing when the compressor ratio is increased. The compressor distortion effect on broad band ILD is up to 12 dB when the input signal is white noise and the uncompressed ILD is 14 dB. The corresponding ILD error is 12 dB/14 dB·100 = 85%. The ILD distortion is significantly less when male voice is used as input signal. Here the ILD error is less than 2 dB and the relative error on the ILD estimate is no more than 2 dB/4 dB·100 = 50%. It should also be noted that Fig. 5 shows that ITD does not change when the compression ratio is increased.

The Real-Ear-Measurement compensation

In Fig. 6 The HRTFs for 6 different microphone positions {1, 12, 24, 26, 45, Coupler} are shown where REM compensation is applied except on the Coupler HRTFs. Fig. 6 shows that REM compensation influences monaural spectral cues which are responsible for front-back localization and externalization of the sound image (Hartmann, 1999). Fig. 6 also shows that there is nearly no difference between the HRTFs for microphone 45 located at the entrance to the ear canal and the Coupler microphone. This result prooves that the ear-canal transfer function is not a function of angle to the external sound source but can be regarded as a stationary FIR filter. According to the basic laws of physics this holds true as long as the diameter of the ear canal is much smaller than the wave length.

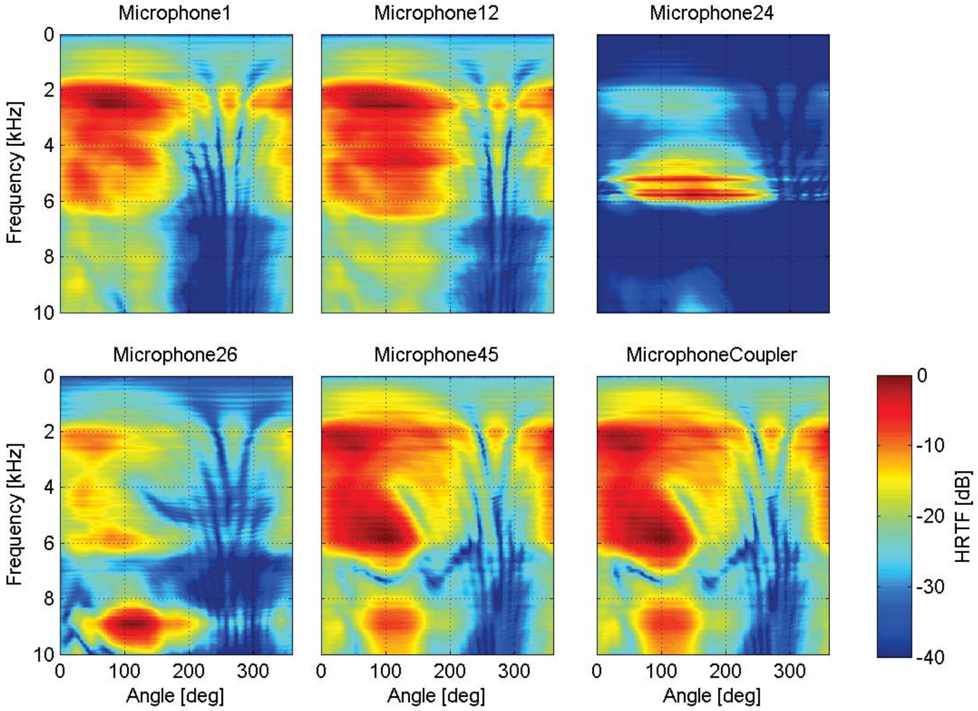


Fig. 6: The HRTFs for microphone positions {1, 12, 24, 26, 45, Coupler} when REM compensation is applied on the large male ear.

When the microphone position moves further away from the entrance to the ear canal the error introduced by applying the REM compensation grows larger. Microphone position 26 shows a significant high-frequency amplification from 90°-180° which most likely will have a huge effect on sound quality. The worst result is obtained with microphone position 24. Here the raw microphone response has a 40-dB narrow dip located around 0° and 5-6 kHz. The REM compensation amplifies this dip by 40 dB at all angles and the result is an HRTF pattern which is significantly different from the open-ear response (except at 0°).

CONCLUSION

It was found that the microphone positions had a significant effect on ILD and ITD. At the entrance to the ear canal the distortion was moderate (less than 10 dB) but behind the pinna microphones introduced ILD errors up to 30 dB at frequencies from 6-8 kHz. Also the ITD error was significant; for some microphone positions it was up to $\sim 100\mu$s. Fixed directionality introduced significant ($\sim$ 20 dB) broad band ILD distortion when sound sources were located around 100°-150°, at other angles the effect was moderate. The compressor had the effect of systematically decreasing ILD. When compression ratio 4 was tested with white noise as input signal the resulting ILD curve had a maximum of 3 dB (compared to 15 dB with no compression). REM compensation did not show any effect on either ILD or ITD but the monaural spectral cues were significantly affected. A positive effect of REM compensation on monaural spectral cues was seen on microphone positions close to the entrance of the ear canal and significant artifacts were introduced when microphones behind the pinna were used.

REFERENCES

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).

Golomb, S.W., and Gong, G. (**2005**). *Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar* (Cambridge University Press).

Hartmann, W.M., and Wittenberg, A. (**1996**). “On the externalization of sound images,” J. Acoust. Soc. Am., **99**, 3678-3688.

Hartmann, W.M. (**1999**). “How we localize sound,” Phys. Today, **52**, 24.

Hawley, M.L., Litovsky, R.Y., and Colburn, H.S. (**1999**). “Speech intelligibility and localization in a multi-source environment,” J. Acoust. Soc. Am., **105**, 3436-3448.

Van den Bogaert, T., Klasen, T.J., Moonen, M., Van Deun, L., and Wouters, J. (**2006**). “Horizontal localization with bilateral hearing aids: Without is better than with,” J. Acoust. Soc. Am., **119**, 515-526.

Evaluation of a frequency lowering hearing instrument algorithm using a non-inferiority test

CHRISTOPHE LESIMPLE, NEIL HOCKLEY*, AND BARBARA SIMON

Bernafon AG, Morgenstrasse 131, 3018 Bern, Switzerland

The primary goal of hearing instrument verification is to demonstrate an improvement on a relevant outcome. It is imprudent to implement an algorithm that improves one outcome while simultaneously degrading another. A traditional test typically uses a superiority hypothesis – H0: New = Conventional and H1: New ≠ Conventional. The absence of statistical significance may be interpreted incorrectly as an absence of clinically relevant differences. An alternative is to start the test with a non-inferiority hypothesis – H0: New < Conventional and H1: New ≥ Conventional. Cross-over designs are often employed because treatment differences are frequently measured within a subject rather than between subjects. Each test period should be long enough for the subject to become acclimatized to each processing change. With these conditions, it is possible to estimate, with the same test, the overall effect of the developed feature and also the period effect. The method of using a cross-over design with a non-inferiority analysis was applied in the testing of a new frequency lowering algorithm. Improved high-frequency functional gain and fricative discrimination was observed. Significant non-inferior SSQ scores between the processing on and off were seen while no period effect was found. These results provide a good approximation of 'real world' acceptance.

INTRODUCTION

Frequency Lowering (FL) algorithms are designed for hearing-impaired people who cannot otherwise obtain benefit from conventional processing (CP) in the high frequencies (HF). The aim of FL processing is to provide improved access to HF cues that would otherwise not be available. Most of the published studies about FL systems are centred on speech recognition and discrimination improvement; however, some of these papers also report the effect of FL systems on sound quality (Simpson *et al.*, 2006; Kuk *et al.*, 2009; Bohnert *et al.*, 2010; Parsa *et al.*, 2013). FL algorithms add artificial signals that may change harmonic ratios, add noise, change timbre, etc., so perceived sound quality may be affected depending on the listener.

Sound quality can be assessed with questionnaires or with perceptual tests. Recent studies, that used questionnaires with different FL algorithms, were unable to show a significant group effect, such as Simpson *et al.* (2006) with the Abbreviated Profile of Hearing Aid Benefit (APHAB) (Cox and Alexander, 1995), Ellis (2012) with the Speech, Spatial and Qualities of Hearing Scale (SSQ) (Gatehouse and Noble, 2004), or Bohnert *et al.* (2010) with self-developed questionnaires. A perceptual test, like the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) design (ITU-R, 2003) used by Parsa *et al.* (2013), investigated subjective ratings with different FL settings and various test stimuli. Test participants rated sound quality

*Corresponding author: nh@bernafon.ch

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

for speech in noise and music samples with different FL settings. The rating difference was not significant between the CP and FL processing for hearing-impaired adults. Conclusions from these studies might be misinterpreted when the authors report no statistically significant difference between the tested conditions for the following reasons:

1. There is a chance that both processing strategies provide more or less the same perceived sound quality.
2. The sample size might be too small to show an existing difference.
3. The outcomes might not be sensitive to the tested FL algorithm. Sensitivity can be affected by floor and/or ceiling effects, or questions that are not relevant to what is being tested.

These studies are superiority trials that are designed to detect differences between treatments (CPMP, 2000). However, a superiority trial cannot be used to conclude that two treatments, or two processing methods in this case, have the same effect. In order to evaluate if two treatments have the same effect, the Committee for Proprietary Medicinal Products (CPMP) (2000) recommends the use of a non-inferiority hypothesis. To the best of the authors' knowledge, this technique has not been used to evaluate hearing instrument (HI) features to date.

Non-inferiority trial in a cross-over design

This non-inferiority trial seeks to find that a new FL algorithm does not perform worse than the reference CP by more than an acceptable amount, i.e., the non-inferiority margin (M_{NI}). Pocock (2003) and D'Agostino *et al.* (2003) present some key issues that need to be addressed when using a non-inferiority design. Therefore, CP should demonstrate superiority over the unaided condition (a) for the same participant population, (b) with an equivalent HI and (c) for the same outcome measure. These authors also state that each processing strategy needs to be tested for a long enough period of time in order for any processing differences to have a realistic opportunity of being observed.

To declare that non-inferiority has been shown, the 95% confidence interval of the difference between both processing systems (FL and CP) should entirely lie above the non-inferiority margin as seen in Fig. 1. Guidelines from the CPMP (2000) also recommend calculating a *p*-value associated with the null hypothesis of inferiority in order to assess the strength of the evidence in favour of non-inferiority.

To reduce the impact of confounding variables and biases (Cox, 2005), a two period cross-over design can be used to assess the processing effect, period effect, and any interaction (Hills and Armitage, 1979). One requirement is that the baseline condition does not change over the two test periods. Thus, it is expected that the test subjects have the same condition at the beginning of each test period. In this design, all participants receive and provide data for both the test FL algorithm and the reference CP. A cross-over design can also be used to test a non-inferiority hypothesis.

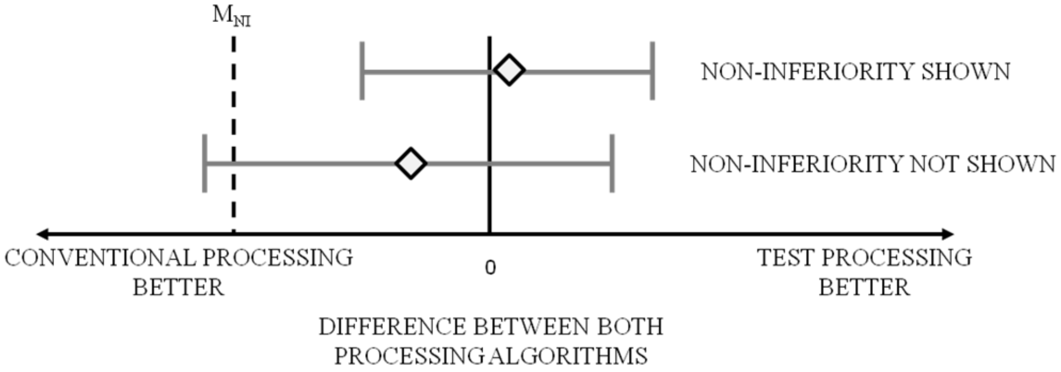

M_{NI}
NON-INFERIORITY SHOWN
NON-INFERIORITY NOT SHOWN
CONVENTIONAL PROCESSING BETTER
0
TEST PROCESSING BETTER
DIFFERENCE BETWEEN BOTH PROCESSING ALGORITHMS

Fig. 1: Schematic presentation of the hypotheses related to a non-inferiority trial. Error bars represent the confidence intervals. In both cases, a superiority hypothesis test would fail to reject the null hypothesis. With a non-inferiority trial in the upper case, it is possible to conclude that the new processing is significantly non-inferior to the reference processing.

Non-inferiority margin determination

The M_{NI} states how close the FL processing must be to the conventional processing. Evidence from previous experiments must show that the CP condition is superior to the unaided one (assay sensitivity) and the test conditions must also be similar for the new trial (constancy assumption). The new processing should also remain superior over the unaided condition by a certain amount (putative comparison). It is reasonable to fix this amount at 50 % of the conventional processing effect over the unaided condition (Jones *et al.*, 1996).

The Speech, Spatial, and Qualities of Hearing Scale (SSQ) questionnaire

The SSQ is a self-report questionnaire divided into three subscales that assess various everyday listening situations. The subject rates their ability to perform in each given listening situation on a scale from 0 to 10 (higher scores always reflect greater ability or less effort). The result is that 'real world' environmental scenarios, with the implemented processing in the HI, can be evaluated.

For this investigation, the qualities subscale of the SSQ was used to investigate various aspects of the perceived sound quality, including: (a) sound quality and naturalness, (b) identification of sound, (c) segregation of sounds, and (d) listening effort. The qualities subscale has shown, in various studies, that CP provides benefit over the unaided condition for adults with mild to severe hearing loss (Noble and Gatehouse, 2006; Jensen *et al.*, 2009; Köbler *et al.*, 2010). The SSQ is available in many languages including German.

Research Question

In this investigation it is hypothesised that the FL algorithm will be judged to provide good sound quality for the hearing-impaired subjects. An improvement over conventional processing is not expected and differences between the FL algorithm and CP should not be clinically relevant. This investigation is designed to show how a non-inferiority test, using the qualities subscale from the SSQ questionnaire, can evaluate the differences in perceived sound quality between the CP and FL processing.

MATERIAL AND METHOD

A cross-over trial with a three-week period was judged to be sufficient for the subjects to become acclimatized to each processing scheme. This time period should be long enough for each subject to experience most of the situations or environments described in the SSQ questionnaire. Two groups were created with two different experimental sequences. The Sequence A group received the FL algorithm in the first period, whereas the Sequence B group received the CP. After three weeks, the processing type was switched so that the Sequence A group received CP and the Sequence B group received the FL algorithm. Group allocation was done using minimization of the following predictive factors: (a) high-frequency hearing loss and (b) participant amplification experience.

Participants

Fourteen subjects between 41 and 79 years of age (average = 64) with a bilateral sensorineural hearing loss took part in this trial. Twelve of them were experienced HI users who had previously used the same compression scheme. The other two subjects were first time users. There were thirteen males and one female. The high-frequency hearing loss was defined by the high-frequency average (HFA) of the air conducted thresholds at 4, 6, and 8 kHz. Ten of the fourteen participants had a HFA that was greater than 70 dB HL. Figure 2 shows the average hearing thresholds for all participants.

Test hearing instrument

The FL algorithm was employed in a commercially available receiver-in-the-ear (RITE) HI. The appropriate acoustic coupling was selected for each subject as recommended by the fitting software. The same instrument was used for both experimental periods. Only the FL algorithm was enabled or disabled during the trial. The gain, compression factors, and automatic features were identical over both periods.

To minimize the placebo effect that is commonly found in HI evaluations (Dawes *et al.*, 2013), the participants were unaware of which algorithm they were trying at any instance in time and of the specific intent of the tested feature.

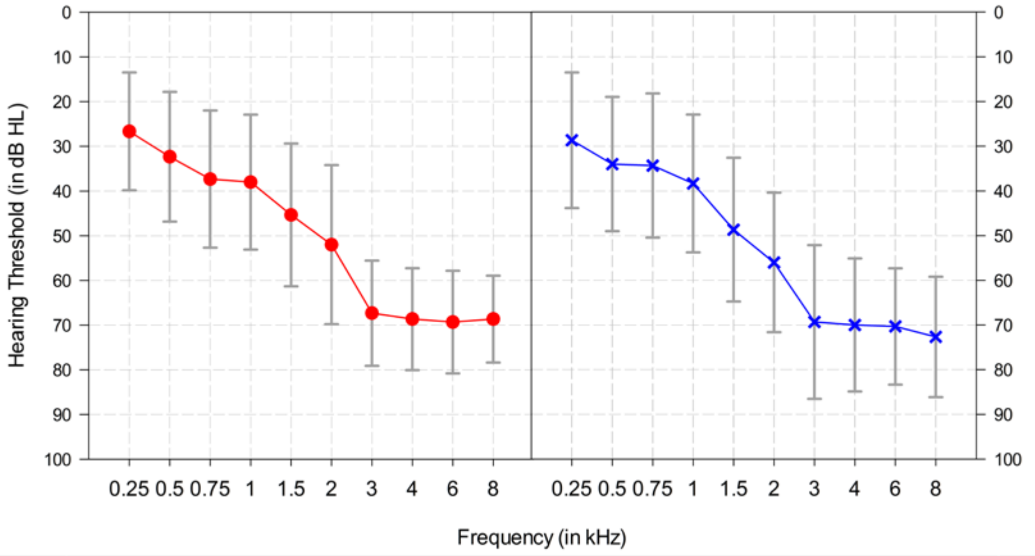


Fig. 2: Average hearing-threshold levels for the fourteen trial participants at each air conducted audiometric frequency. Error bars show the standard deviation.

Non-inferiority determination

The non-inferiority margin determination was based on previous internal pilot investigations. Figure 3 shows what was considered in the determination of the M_{NI}.

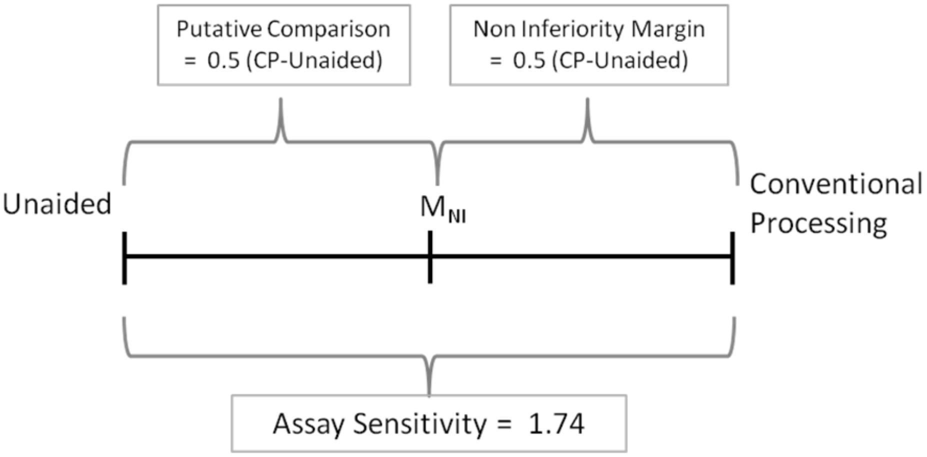


Fig. 3: Considerations in the determination of the M_{NI}. (a) The assay sensitivity will be based on an already known CP effect over the unaided condition. (b) The putative comparison will control the effectiveness of the tested processing.

Based on the observation of a 1.74 score improvement on the SSQ qualities subscale from a previous pilot study, it was appropriate to set the M_{NI} to 50% of the historical effect (improvement from unaided to CP). Due to the fact that the same RITE HIs with the same compression scheme and fitting rationale were used in both trials, it is assumed that the assay sensitivity is obtained and that the constancy assumption is held. The M_{NI} for this trial can be set to a 0.87 SSQ score degradation.

RESULTS

Each participant filled out an SSQ questionnaire after each three-week test period. Mean scores are shown for both processing types and for each qualities subscale attributes in Fig. 4.

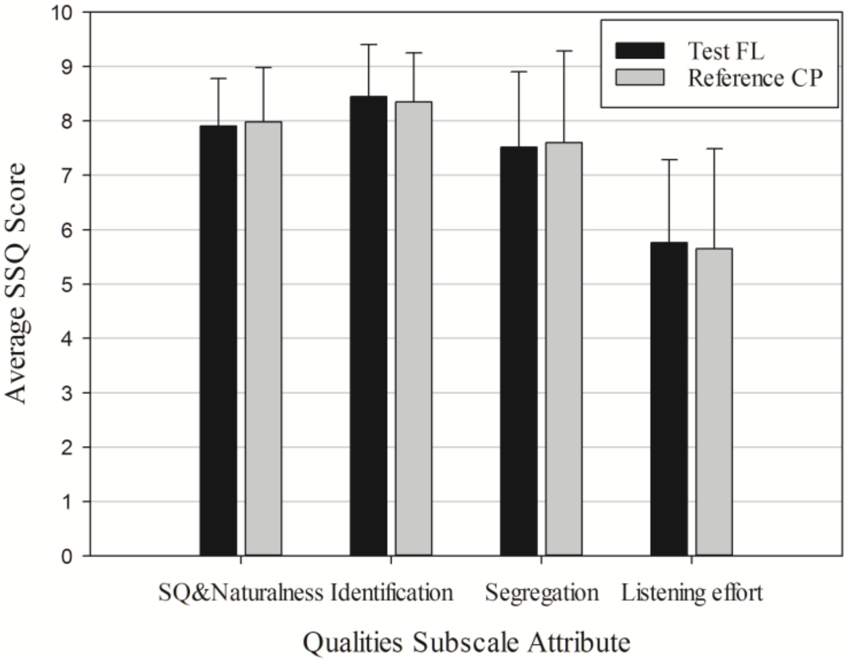


Fig. 4: Average qualities scores from the SSQ within sound quality and naturalness, identification of sound, segregation of sounds, and listening effort (n=14). Results with the HIs with the FL algorithm are in black and the results with CP are in grey. Error bars show one standard deviation.

Based on the rating difference between both processing schemes, it is possible to compute the mean difference and the 95% confidence interval for this outcome. To conclude that FL processing is significantly non-inferior to the CP, the lower boundary of the 95% confidence interval should be higher than the M_{NI}. Under the null hypothesis, the data are distributed with the M_{NI} as means and will follow a *t*-distribution with N-2 degrees of freedom. The *p*-value is derived from these data and all the findings are shown in Table 1.

	Non-Inferiority Test Null hypothesis (H0): Difference ≤ - 0.87				
		95 % CI of the difference			*Reject H0?*
Qualities Attributes	*Difference (T - R)*	*Lower Bound*	*Upper Bound*	*p-value*	*Non-Inferior?*
SQ & Naturalness					
FL-CP	-0.08	-0.43	0.27	0.001	Yes
Identification of Sound					
FL-CP	0.08	-0.31	0.47	0.007	Yes
Segregation of Sounds					
FL-CP	-0.08	-0.43	0.26	0.001	Yes
Listening Effort					
FL-CP	0.11	-0.91	1.13	0.096	No

Table 1: Non-inferiority test results for the different attributes from the qualities subscale of the SSQ. The difference between FL processing and CP imply that positive values are in favour for the tested algorithm.

The qualities subscale outcomes show that the developed FL is significantly non-inferior to the CP for the following attributes: sound quality and naturalness, identification of sound, and segregation of sounds. For the listening effort attribute, the null hypothesis could not be rejected as the confidence-interval lower boundary is smaller than the M_{NI}.

CONCLUSION

Showing client benefit for a newly developed algorithm is the desired outcome for the verification of many HI features. However, when the signal is manipulated, it seems also important to assess how the perceived sound quality might be affected. The use of a non-inferiority hypothesis is probably the only way to show that the sound quality is not significantly degraded with a new algorithm. For the FL algorithm evaluated in this investigation, it was possible to address this concern by using the qualities subscale from the SSQ questionnaire with a non-inferiority test. Three out of the four attributes from the qualities subscale were significantly non-inferior. Based on these outcomes it is expected that this FL algorithm will provide a perceived sound quality that is comparable to the CP and that this will result in good acceptance by the HI wearer.

REFERENCES

Bohnert, A., Nyffeler, M., and Keilmann, A. (**2010**). "Advantages of a non-linear frequency compression algorithm in noise," Eur. Arch. Otorhinolaryngol., **267**, 1045-1053.

Cox, R.M. (**2005**). “Evidence-based practice in provision of amplification,” J Am Acad Audiol. 2005, **16**, 419-38.

Cox, R.M., and Alexander, G.C. (**1995**). “The abbreviated profile of hearing aid benefit,” Ear. Hearing, **16**, 176-186.

CPMP: Committee for Proprietary Medicinal Products (**2000**). “Points to consider on switching between superiority and non-inferiority,” European Medicines Agency (EMEA), CPMP/EWP/482/99.

D'Agostino, R.B. Sr., Massaro, J.M., and Sullivan, L.M. (**2003**). “Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics,” Stat. Med., **22,** 169-186.

Dawes, P., Hopkins, R., and Munro, K.J. (**2013**). “Placebo effects in hearing-aid trials are reliable,” Int. J. Audiol., **52**, 472-477.

Ellis, R.J. (**2012**) “Benefit and predictors of outcome from frequency compression hearing aid use,” PhD thesis, University of Manchester, UK.

Gatehouse, S., and Noble, W. (**2004**). “The speech, spatial and qualities of hearing scale (SSQ),” Int. J. Audiol., **43**, 85-99.

Hills, M., and Armitage, P. (**1979**). “The two-period cross-over clinical trial,” Br. J. Clin. Pharmacol., **8**, 7-20.

ITU-R (**2003**). *Recommendation BS.1534: Method for the subjective assessment of intermediate quality levels of coding systems.* International Telecommunications Union, Geneva, Switzerland.

Jensen, N.S., Akeroyd, M.A., Noble, W., and Naylor, G. (**2009**). “The Speech, Spatial and Qualities of Hearing scale (SSQ) as a benefit measure,” NCRAR conference on *The Ear-Brain System: Approaches to the Study and Treatment of Hearing Loss*, Portland, October 2009 (poster).

Jones, B., Jarvis, P., Lewis, J.A., and Ebbutt, A.F. (**1996**). “Trials to assess equivalence: the importance of rigorous methods,” Brit. Med. J., **313**, 36-39.

Köbler, S., Lindblad, A.C., Olofsson, A., and Hagerman, B. (**2010**). “Successful and unsuccessful users of bilateral amplification: differences and similarities in binaural performance,” Int. J. Audiol., **49**, 613-627.

Kuk, F., Keenan, D., Korhonen, P., and Lau, CC. (**2009**). “Efficacy of linear frequency transposition on consonant identification in quiet and in noise,” J. Am. Acad. Audiol., **20**, 465-479.

Noble, W., and Gatehouse, S. (**2006**). “Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial, and Qualities of Hearing Scale (SSQ),” Int. J. Audiol., **45**, 172-181.

Parsa, V., Scollie, S., Glista, D., and Seelisch, A. (**2013**). “Nonlinear frequency compression: effects on sound quality ratings of speech and music,” Trends Amplif., **17**, 54-68.

Pocock, S.J. (**2003**). “The pros and cons of noninferiority trials,” Fundam. Clin. Pharmacol., **17**, 483-490.

Simpson, A., Hersbach, A.A., and McDermott, H.J. (**2006**). “Frequency-compression outcomes in listeners with steeply sloping audiograms,” Int. J. Audiol., **45**, 619-629.

Analyzing the effects on the internal signal-to-noise ratio for bilateral hearing-aid systems configured for asymmetric processing

FREDRIK GRAN[1,*], JESPER UDESEN[1], TOBIAS PIECHOWIAK[1],
AND ANDREW B. DITTBERNER[2]

[1] *GN ReSound A/S, Lautrupbjerg 7, DK-2750 Ballerup, Denmark*

[2] *GN ReSound North America, 8001 Bloomington Freeway, Bloomington, MN 55420-1036, USA*

This paper investigates how bilateral hearing-aid systems configured to perform asymmetric processing affect the internal signal-to-noise ratio (SNR) in the auditory system. Here, an asymmetric hearing-instrument (HI) system is characterized by directional noise reduction in the instrument in one ear whereas the contra-lateral device is adjusted for omni mode processing. The Equalization and Cancellation model is used to evaluate the internal SNR of the auditory system. Two reference conditions were also created, a system with directionality in both HI, and one with omni-mode processing in both HI. A speaker was placed to the front, and another speaker was placed at the side. In the first experiment, the target was assumed to be in the front direction and the noise was assumed to be coming from the side. Here, it was shown that the asymmetric system provided the same SNR as the system with directionality in both HI. The noise and target positions were interchanged and the experiment was repeated. In this case, the asymmetric system provided similar SNR as the system with omni-mode processing in both HI, which for this test condition provided a better SNR than the system with directionality in both HI.

INTRODUCTION

Directional hearing-aid systems have been shown to improve speech intelligibility in noisy conditions (Ricketts and Dittberner, 2002). Directionality algorithms and/or technologies aim at preserving signals originating from the look direction (0 degrees) whilst suppressing sources from all other directions. In digital dual-microphone systems this is typically done by placing a null in a direction where the masker is assumed to be located. An inherent aspect of this processing strategy is that the listener loses sensitivity to sources to the side and in the back as compared to single microphone systems (omni-mode processing). Asymmetric processing schemes (omni mode in one ear and directional technology in the contralateral ear) have been shown to provide similar speech understanding performance for hearing-impaired subjects as when applying symmetrically configured hearing aids programmed to

*Corresponding author: fgran@gnresound.com

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

provide directionality towards the front (Bentler *et al.*, 2004; Cord *et al.*, 2005). In both these studies, the target signal was originating from the front and the masker signals were originating from other directions, playing mutually uncorrelated speech-shaped noise. Both papers showed no significant difference between the asymmetric processing condition and the symmetric directionality condition. In Hornsby and Ricketts (2007), several target and masker configurations were investigated, 1) target in front and five masker sources evenly distributed in a circle around the listener, 2) target to the front and five masker sources evenly distributed between 50°-130° and 3) target at 90° and three masker sources evenly distributed between 45°-135°. The purpose was to mimic listening in diffuse noise with configuration 1, listening to a target in front with interferers coming predominantly from the left in configuration 2, and trying to concentrate on a talker to the right with the majority of the masker energy coming from the left in condition 3. The hearing aids were programmed for symmetric omni-mode processing, symmetric directionality processing, and asymmetric processing. Both symmetric directionality and asymmetric processing showed a benefit on speech reception thresholds (SRT) compared to symmetric omni-mode processing for conditions 1 and 2, whereas a performance degradation was observed for condition 3. The SRT degradation was found to be smaller for the asymmetric processing compared to the symmetric directionality processing when the hearing instrument programmed for omni-mode processing faced away from the masker sources.

The purpose of this paper was to investigate if binaural listening models can predict these phenomena. In particular, the Equalization and Cancellation (EC) model (Kock, 1950; Durlach, 1960, 1963) was used to model the binaural signal-to-noise ratio (SNR) of the auditory system for different target and masker conditions as well as different hearing-aid processing configurations. The EC model was proposed to model the binaural masking level differences (BMLD) of detecting tones in noise for dichotic vs diotic signal presentation. This model was later modified and used to explain several data sets for more complicated listening experiments, such as modeling speech-intelligibility improvement for speech masked by a single noise source in an anechoic space (Zurek, 1992), speech-intelligibility improvements in multi-talker speech-shaped interference in an anechoic space (Culling *et al.*, 2004), speech-intelligibility tasks in anechoic and diffuse conditions, both for hearing-impaired and normal-hearing listeners (Beutelmann and Brand, 2006). In Wan *et al.* (2010), an extended version of the EC model was used to explain the data sets acquired in Hawley *et al.* (2004).

HEARING-AID TECHNOLOGY AND PROCESSING

Data acquisition and measurement equipment

The experiments involved measuring hearing-instrument-related impulse responses (HRIR) on KEMAR. In this paper the HRIRs were measured on a KEMAR manikin in the horizontal plane with an angular resolution of 2 degrees. An anechoic room was used for the HRIR measurements. The room was in accordance with ISO 3745. The distance from the speaker to the rotation axis of KEMAR was 1.5 m. The speaker used in all experiments was a KEF Q85S (serial number: 740107G). The phase was inverted by connecting ($-$) on the speaker to (+) on the ROTEL RB-1050 power amplifier. The recoded microphone signals were convolved with the inverse of the speaker impulse response before further processing. All measurements were performed at a sampling frequency of 48828 Hz using a Tucker Davis RX8 multiprocessor controlled by MATLAB R2010b, The MathWorks Inc., Natick, MA. The signal presented through the speaker was a maximum length sequence (MLS) signal (Proakis and Salehi, 1994). In the anechoic room the code length was $(2^{11}-1) = 2047$ samples. This corresponds to an acoustic distance of 14.2 m. It was found that the room reflections were below the noise floor at this distance. The corresponding intensity for the speaker signal at KEMAR's position (without KEMAR present) was 74 dB SPL. The HRIRs were measured on a pair of modified receiver-in-the-ear hearing aids where the front and the rear microphone signals were accessible.

Processing modes

Omni-mode processing was created by simply extracting the front-microphone signal from the hearing instrument. The directionality processing was created using filter and sum beamforming by placing a null at 180°. The filters in the beamformer had 21 taps. Three different hearing-aid processing configurations were tested:

- **Bilateral omni mode** was created by using the omni signal in both hearing aids.
- **Bilateral directionality mode** was created by using the beamformer output in both hearing instruments.
- **Asymmetric mode** was generated by choosing the omni signal in the right hearing instrument and the beamformer output in the left hearing instrument.

The directivity patterns for the left (black solid line) and right (gray solid line) hearing instruments can be seen in Fig. 1 (1 kHz) and Fig. 2 (4 kHz). The left plot shows the bilateral omni mode processing configuration, the middle plot shows the configuration where both hearing instruments are programmed to perform beamforming and the right plot shows the asymmetric processing configuration where one instrument performs omni-mode processing and the other performs beamforming.

SIMULATION SETUP

Four different simulations were created: 1) target at 0° and masker at 120°, 2) target at 120° and masker at 0°, 3) target at 0° and masker at $-120°$, 4) target at $-120°$ and masker at 0°. Binaural HRIRs were then created for the three different processing configurations and the resulting impulse responses were processed by the EC model. Let $A_q(f)$ be the spectrum of a realization of the target component after the EC process

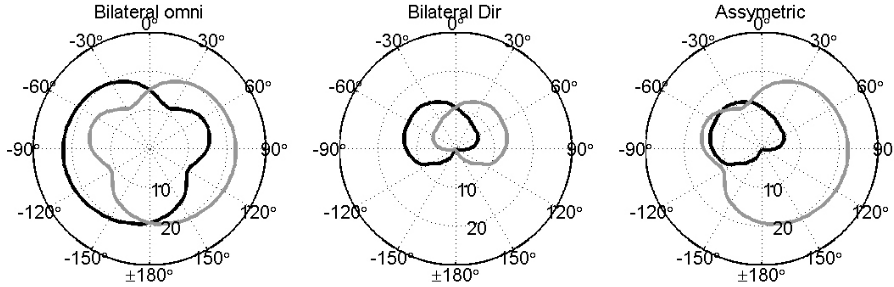


Fig. 1: Directivity patterns for the left (black solid line) and right (gray solid line) hearing instruments for the frequency of 1 kHz.

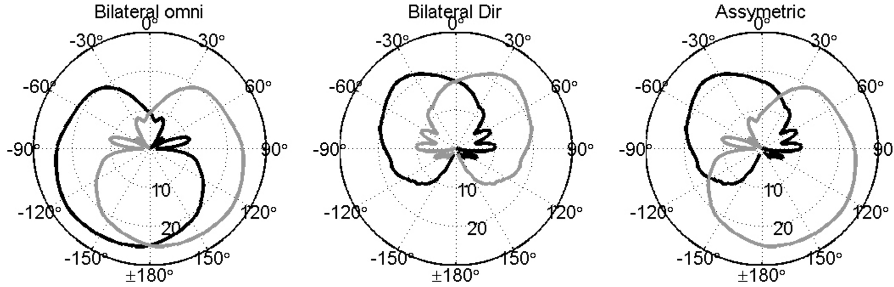


Fig. 2: Directivity patterns for the left (black solid line) and right (gray solid line) hearing instruments for the frequency of 4 kHz.

and let the corresponding masker spectrum after EC processing be given by $B_q(f)$. The binaural signal-to-noise ratio was then estimated as

$$\mathrm{SNR}(f) = \frac{\sum_{q=0}^{Q-1} \left|A_q(f)\right|^2}{\sum_{q=0}^{Q-1} \left|B_q(f)\right|^2}, \qquad \text{(Eq. 1)}$$

In this paper $Q = 10000$ realizations were used.

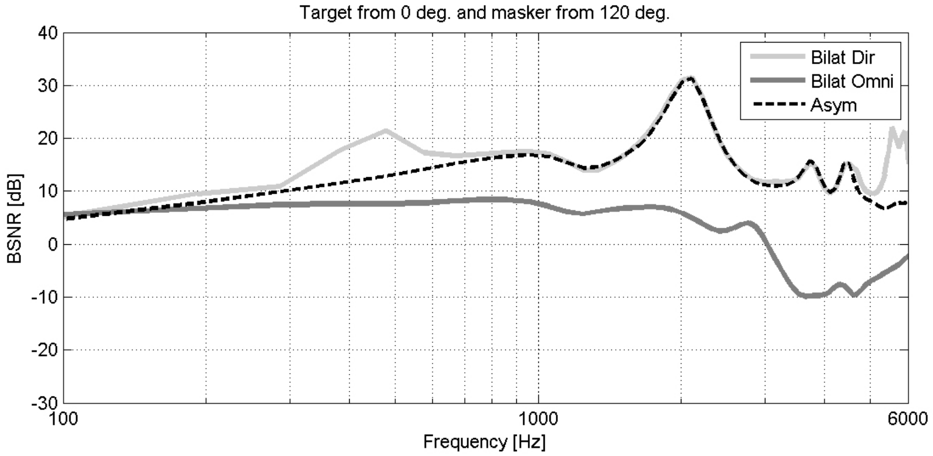


Fig. 3: Binaural SNR estimated by the EC model with a target presented from 0 degrees and a masker from 120 degrees. The bilateral omni mode is given by the dark gray curve, the bilateral directionality is given by the light gray curve, and the asymmetric configuration is seen in the dashed black plot.

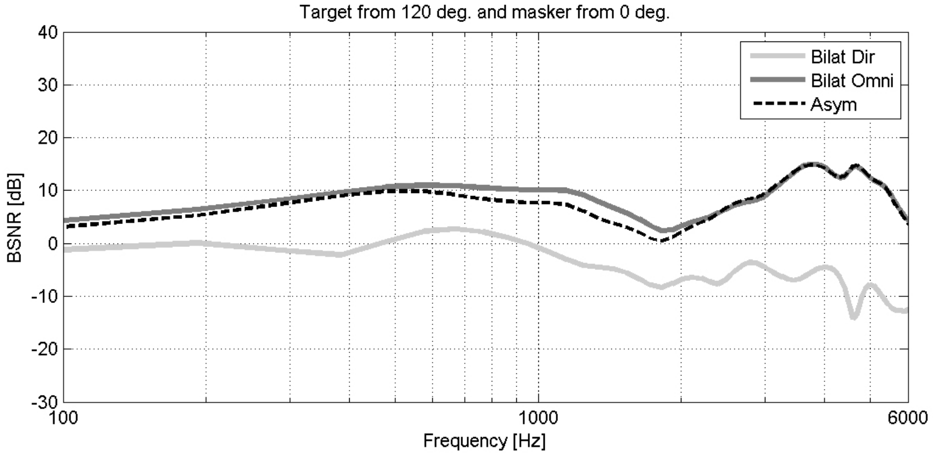


Fig. 4: Binaural SNR estimated by the EC model with a target presented from 120 degrees and a masker from 0 degrees. The bilateral omni mode is given by the dark gray curve, the bilateral directionality is given by the light gray curve and the asymmetric configuration is seen in the dashed black plot.

SIMULATION RESULTS

The binaural SNR predicted by the EC model for the four different simulations is given in Figs. 3-6. In all figures, the bilateral omni-mode results are given by the solid dark gray curve, the bilateral directionality mode results are given by the light gray curve, and the results for the asymmetric processing are given by the dashed black curve. In Fig. 3, the target and masker are configured so that the target is in front of the listener (0°) and the masker is to the side (120°). In the asymmetric processing mode, the masker faces the hearing aid which is programmed to perform omni-mode processing. The bilateral directionality mode has better SNR than the bilateral omni configuration. This is to be expected, since the directionality algorithm suppresses sources from from the rear and to the side. The asymmetric configuration, however, seems to have similar performance as the bilateral directionality mode.

In Fig. 4, the target and masker positions are interchanged compared to Fig. 3. The masker is now assumed to be positioned in the front and the target is placed to the side. For the asymmetric processing mode, the target signal is now facing the hearing aid which is programmed to perform omni-mode processing. The bilateral directionality mode has worse SNR than the bilateral omni configuration across all frequencies. Again, this is to be expected, since the has better sensitivity to the side and to the rear compared to the directionality algorithm. The asymmetric configuration, however, now seems to have similar performance as the bilateral omni mode. Comparing the results in Fig. 3 and Fig. 4, it seems as if the auditory system is able to use the processing mode which gives the best SNR for the target of interest when presented with asymmetric beampatterns.

In Fig. 5, the masker is assumed to be positioned to the left of the listener ($-120°$) and the target is in front of the listener. For the asymmetric processing mode, the masker signal is now facing the hearing aid which is programmed to perform directionality-mode processing. The bilateral directionality mode has again better SNR than the bilateral omni configuration across all frequencies. The results for the asymmetric configuration now seem to be a bit more mixed as compared to the results in Fig. 3. Up to approximately 500 Hz, the asymmetric configuration has similar SNR as the bilateral directionality mode. Above this frequency, performance seems to degrade and resembles the performance given by the bilateral omni-mode results.

In Fig. 6, the target is assumed to be positioned to the left of the listener ($-120°$) and the masker is in front of the listener. For the asymmetric processing mode, the target signal is now facing the hearing aid which is programmed to perform directionality-mode processing. The bilateral directionality mode has worse SNR than the bilateral omni configuration across all frequencies. The results for the asymmetric configuration seem to resemble the performance given by the bilateral directionality-mode results. In this target/masker setup, there seems to be no advantage of asymmetric processing. If one analyzes the beampatterns in Fig. 1 and Fig. 2, it is seen that the target position of $-120°$ is particularly unfavorable for the asymmetric configuration, as the beampatterns of both the left and the right hearing instrument display very low sensitivity in this region.

DISCUSSION

Modeling binaural listening performance with asymmetric beampatterns yields two major conclusions: If the target and masker are configured so that one of the sources is

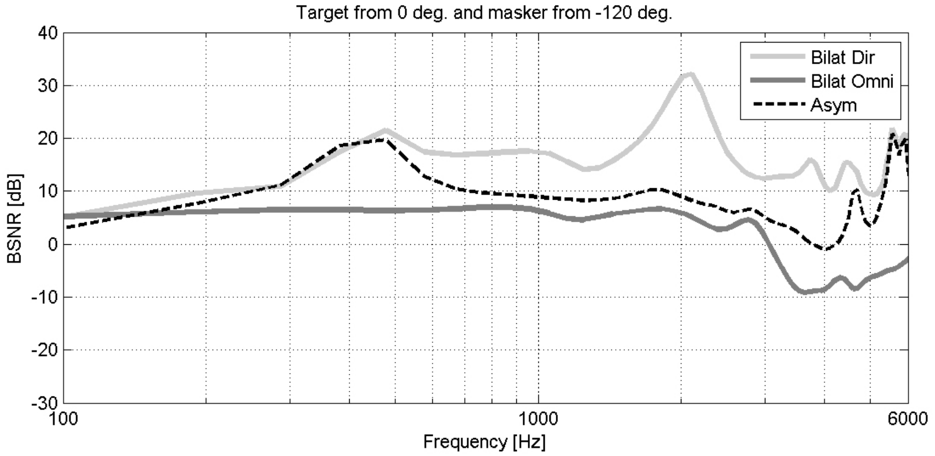


Fig. 5: Binaural SNR estimated by the EC model with a target presented from 0 degrees and a masker from 120 degrees. The bilateral omni mode is given by the dark gray curve, the bilateral directionality is given by the light gray curve and the asymmetric configuration is seen in the dashed black plot.

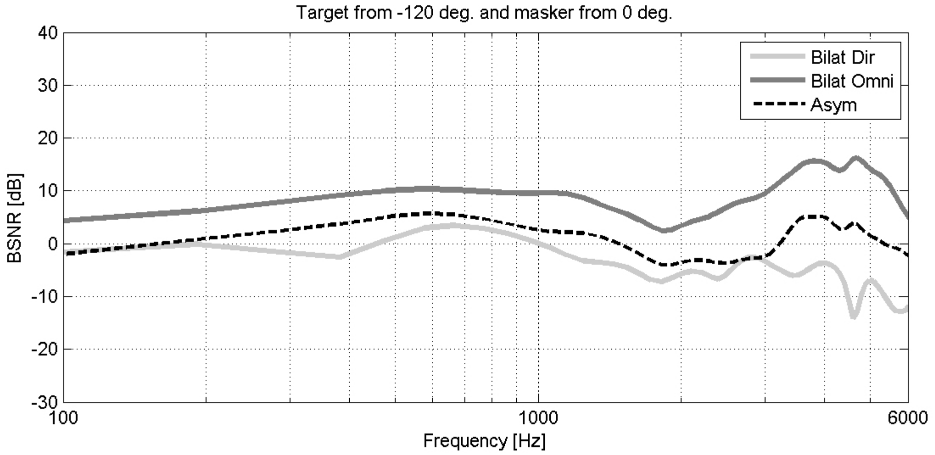


Fig. 6: Binaural SNR estimated by the EC model with a target presented from 0 degrees and a masker from 120 degrees. The bilateral omni mode is given by the dark gray curve, the bilateral directionality is given by the light gray curve and the asymmetric configuration is seen in the dashed black plot.

in front of the listener and the other source is placed facing the hearing aid performing omni-mode processing, it seems as if the listener can get the same performance as the bilateral directionality mode when listening to the source in front. However, if attention is turned to the source to the side, the listener achieves similar performance as with the bilateral omni-mode configuration. In this particular target/masker configuration, the model predicts that the asymmetric processing mode would yield good speech understanding to the front as well as to the side. When one source is placed in the front and the other source is placed facing the hearing instrument programmed for directionality, the results are more mixed. When trying to focus on the source in front, a listening benefit is seen up to approximately 500 Hz. When trying to focus on the source to the side the asymmetric processing only displays small improvement as compare to the bilateral directionality configuration. This suggests that if this listening situation occurs, the hearing-aid system should switch so that the hearing aid facing the interferer performs omni-mode processing.

REFERENCES

Bentler, R.A., Egge, J.L., Tubbs, J.L., and Dittberner, A.B. (**2004**). “Quantification of directional benefit across different polar response patterns,” J. Am. Acad. Audiol., **15**, 649-659.

Beutelmann, R., and Brand, T. (**2006**). “Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing impaired listeners,” J. Acoust. Soc. Am., **120**, 331-342.

Cord, M., Surr, R., Walden, B., and Dittberner, A.B. (**2005**). “Asymmetric directional microphone fittings,” American Academy of Audiology 17th Annual Convention, Washington D.C.

Culling, J.F., Hawley, M.L., and Litovsky, R.Y. (**2004**). “The role of head induced interaural time and level differences in the speech reception threshold for multiple interferring sound sources,” J. Acoust. Soc. Am., **116**, 1057-1065.

Durlach, N.I. (**1960**). “Note on the equalization and cancellation theory of binaural masking level differences,” J. Acoust. Soc. Am., **32**, 1075-1076.

Durlach, N.I. (**1963**). “Equalization and cancellation theory of binaural masking level differences,” J. Acoust. Soc. Am., **35**, 1206-1218.

Hawley, M.L., Litovsky, R.Y., and Culling, J.F. (**2004**). “The benfit of binaural hearing in a cocktail party: Effect of location and type of interferer,” J. Acoust. Soc. Am., **115**, 833-843.

Hornsby, B.W.Y., and Ricketts, T.A. (**2007**). “Effects of noise source configuration on directional benefit using symmetric and asymmetric directional hearing aid fittings,” Ear Hearing, **28**, 177-186.

Kock, W.E. (**1950**). “Binaural localization and masking,” J. Acoust. Soc. Am., **22**, 801-804.

Proakis, J.G., and Salehi, M. (**1994**). *Communcation Systems Engineering* (Prentice Hall).

Ricketts, T., and Dittberner, A.B. (**2002**). “Directionality amplification for improved signal-to-noise ratio: Strategies, measurements and limitations,” in *Hearing Aids: Standards, Options and Limitations*, 2nd ed. Edited by M. Valente, pp. 274-346.

Wan, R.W., Durlach, N.I., and Colburn, H.S. (**2010**). “Application of an extended equalization-cancellation model to speech intelligibility with spatially distributed maskers,” J. Acoust. Soc. Am., **128**, 3678-3690.

Zurek, P.M. (**1992**). “Binaural advantages and directional effects in speech intelligibility,” in *Acoustical Factors affecting Hearing Aid Performance*, 2nd ed. Edited by G.A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255-276.

Comparison between the equalization and cancellation model and state of the art beamforming techniques

FREDRIK GRAN[1,*], J ESPER UDESEN[1],
T OBIAS PIECHOWIAK[1], AND ANDREW B. DITTBERNER[2]

[1] *GN ReSound A/S, Lautrupbjerg 7, DK-2750 Ballerup, Denmark*

[2] *GN ReSound North America, 8001 Bloomington Freeway, Bloomington, MN 55420-1036, USA*

This paper investigates the performance of a selection of state-of-the-art array signal-processing techniques for the purpose of predicting the binaural listening experiments from the equalization and cancellation (EC) paper by Durlach written in 1963. Two different array signal-processing techniques are analyzed, 1) filter and sum beamforming (FS), and 2) minimum variance distortionless response (MVDR) beamforming. The theoretical properties of these beamformers for the specific situation of prediction of binaural masking level differences are analyzed in conjunction with the EC model. Also, the performance of the different beamformers on the data sets in the Durlach paper from 1963 is compared to the EC model.

INTRODUCTION

Some of the earliest work on binaural listening effects date back to the duplex theory presented by Lord Rayleigh (1876, 1907), where interaural time and level differences (ITDs and ILDs) characterized the localization of sound sources. Over four decades later, it was shown (Cherry, 1953) that the benefit of listening with two ears compared to monaural listening is especially pronounced in complex listening scenarios with several competing talkers. The binaural listening advantage in these adverse circumstances, also referred to as the ‘cocktail party problem’, was extensively studied in the fifties and sixties (Cherry and Taylor, 1954; Cherry and Sayers, 1956; Leaky and Cherry, 1957; Sayers and Cherry, 1957; Cherry and Bowles, 1960) where a cross correlation model was used to explain the binaural listening effect.

The Equalization and Cancellation (EC) model was proposed to model the binaural masking level differences (BMLD) of detecting tones in noise for dichotic vs diotic (Kock, 1950; Durlach, 1960, 1963) signal presentation.

This model was later modified and used to explain several data sets for more complicated listening experiments, such as modeling speech-intelligibility improvement for speech masked by a single noise source in an anechoic space (Zurek, 1992), speech-intelligibility improvements in multi-talker speech-shaped interference in an anechoic space (Culling *et al.*, 2004), speech-intelligibility tasks in anechoic and diffuse

[*] Corresponding author: fgran@gnresound.com

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

conditions, both for hearing-impaired and normal-hearing listeners (Beutelmann and Brand, 2006). In Wan *et al.* (2010), an extended version of the EC model was used to explain the data sets acquired in Hawley *et al.* (2004).

In Durlach (1963), the EC model was compared to array processing, where the model tries to put a null at the location of the masker to suppress this component as much as possible. The purpose of this paper is to investigate how more generic beamforming techniques compare to the EC model, both from a theoretical stand point, but also in terms of predictive performance on the original data sets. In particular, fixed filter and sum beamforming is investigated (Johnson and Dudgeon, 1993), as well as the minimum variance distortionless response (MVDR) beamforming technique (Capon *et al.*, 1967).

GENERAL MODEL

The general data model assumes a binaural signal set consisting of a mixture of two signals, one representing the target and one the masker. The short-term spectrum of the target is denoted $X(f,t)$ and the corresponding spectrum of the masker is denoted $Y(f,t)$. Then the binaural signal set can be written as:

$$\underbrace{\begin{pmatrix} S_{\mathrm{l}}(f,t) \\ S_{\mathrm{r}}(f,t) \end{pmatrix}}_{\mathbf{s}(f,t)} = \underbrace{\begin{pmatrix} A_{\mathrm{l}}(f) \\ A_{\mathrm{r}}(f) \end{pmatrix}}_{\mathbf{a}(f)} X(f,t) + \underbrace{\begin{pmatrix} B_{\mathrm{l}}(f) \\ B_{\mathrm{r}}(f) \end{pmatrix}}_{\mathbf{b}(f)} Y(f,t), \qquad \text{(Eq. 1)}$$

where $S_{\mathrm{l}}(f)$ and $S_{\mathrm{r}}(f)$ are the signal mixtures, $A_{\mathrm{l}}(f)$ and $A_{\mathrm{r}}(f)$ are the left and right acoustical transfer functions for the target, respectively, and $B_{\mathrm{l}}(f)$ and $B_{\mathrm{r}}(f)$ are the left and right acoustical transfer functions for the masker, respectively. Note that the assumption here is that these transfer functions do not change over time. Furthermore, it is assumed that $X(f,t)$ and $Y(f,t)$ are independent stochastic variables and spectrally white. In this paper, the binaural signal is estimated via a beamforming approach where

$$\mathbf{b}(f,t) = \mathbf{w}^H(f)\mathbf{s}(f,t), \qquad \text{(Eq. 2)}$$

where $\mathbf{b}$ is the binaural spectrum and $\mathbf{w}^H$ is the complex conjugate transpose of the coefficients used to combine the right- and left-ear signals. The coefficients are defined as:

$$\mathbf{w}(f) = \mathbf{M}(f)\mathbf{h}(f), \qquad \text{(Eq. 3)}$$

where $\mathbf{M}$ can be interpreted as a process that models amplitude and timing jitters and is defined by:

$$\mathbf{M}(f) = \begin{pmatrix} (1-\varepsilon_1)e^{-j2\pi f\delta_1} & 0 \\ 0 & (1-\varepsilon_2)e^{-j2\pi f\delta_2} \end{pmatrix}, \qquad \text{(Eq. 4)}$$

where ε_1 and ε_2 are independent Gaussian-distributed variables with zero mean and a variance of 0.25, δ_1 and δ_2 are independent Gaussian-distributed variables with zero mean and a variance of 105 μs and $\mathbf{h}$ are the beamforming coefficients applied to minimize the masker and enhance the target. The experienced reader immediately realizes that if one chooses the beamforming coefficients to be:

$$\mathbf{h}_{\mathrm{EC}}(f) = \begin{matrix} B_r(f) & -B_l(f) \end{matrix} \Big)^T \qquad \text{(Eq. 5)}$$

the equalization and cancellation model follows from Eq. 2 and $(\cdot)^T$ is the transpose of $(\cdot)$.

ARRAY SIGNAL PROCESSING TECHNIQUES

In this section the two beamformers are derived for the condition described in Eq. 1.

Filter and Sum beamformer

The Filter and Sum (FS) beamformer is an array signal-detection technique developed for optimal signal detection in white Gaussian-distributed noise in the maximum likelihood sense (Johnson and Dudgeon, 1993). The beamforming coefficients would in this case be:

$$\mathbf{h}_{\mathrm{FS}}(f) = \begin{matrix} A_l(f) & A_r(f) \end{matrix} \Big)^T = \mathbf{a}(f), \qquad \text{(Eq. 6)}$$

Minimum Variance Distortionless Response beamformer

In the Minimum Variance Distortionless Response (MVDR) beamformer, the strategy is to suppress all noise sources as much as possible while maintaining the signal of interest. If the model in Eq. 1 is used this can be expressed mathematically as:

$$\begin{gathered} \mathbf{h}_{\mathrm{MVDR}}(f) = \arg\min_{\mathbf{h}} \mathbf{h}^H \mathbf{R}_{\mathbf{ss}}(f)\mathbf{h} \\ \text{subject to } \mathbf{h}^H \mathbf{a}(f) = 1 \end{gathered} \qquad \text{(Eq. 7)}$$

where

$$\mathbf{R}_{\mathbf{ss}}(f) = E\left[\mathbf{s}(f,t)\mathbf{s}^H(f,t)\right] \qquad \text{(Eq. 8)}$$

is the spatial auto correlation matrix of $\mathbf{s}(f,t)$ and E is the expectancy operator.

SIMULATION SETUP

The binaural signal-to-noise ratio (SNR) was evaluated using stochastic simulations. Once a given experimental setup E had been determined (i.e., determining $\mathbf{a}$ and $\mathbf{b}$) and a given set of beamforming coefficients $\mathbf{h}$ had been chosen, the binaural SNR was estimated as:

$$\mathrm{SNR}_E(\mathbf{h},f) = \frac{\sum_{q=0}^{Q-1} \left|\mathbf{w}_q^H(\mathbf{h},f)\mathbf{a}(f)\right|^2}{\sum_{q=0}^{Q-1} \left|\mathbf{w}_q^H(\mathbf{h},f)\mathbf{b}(f)\right|^2}, \qquad \text{(Eq. 9)}$$

where $\mathbf{w}_q$ is the qth realization of the stochastic process defined by Eq. 3 and Q is the total number of realizations used in the simulation. If the beamforming coefficients are chosen so that $\mathbf{h} = \mathbf{h}_{\mathrm{EC}}$, this SNR estimate is actually equivalent to the variable denoted the EC factor in Durlach's paper from 1963, because the spectral amplitudes of the target and masker are the same and uniform over frequency. In this paper $Q =$ 10000 realizations were used, as this was found to be sufficient to generate a good approximation of the results presented in Durlach (1963). The ratio of binaural SNR between two different experimental conditions E and E' is then given by

$$R_{E/E'}(\mathbf{h}, f) = \frac{\mathrm{SNR}_E(\mathbf{h}, f)}{\mathrm{SNR}_{E'}(\mathbf{h}, f)}. \qquad \text{(Eq. 10)}$$

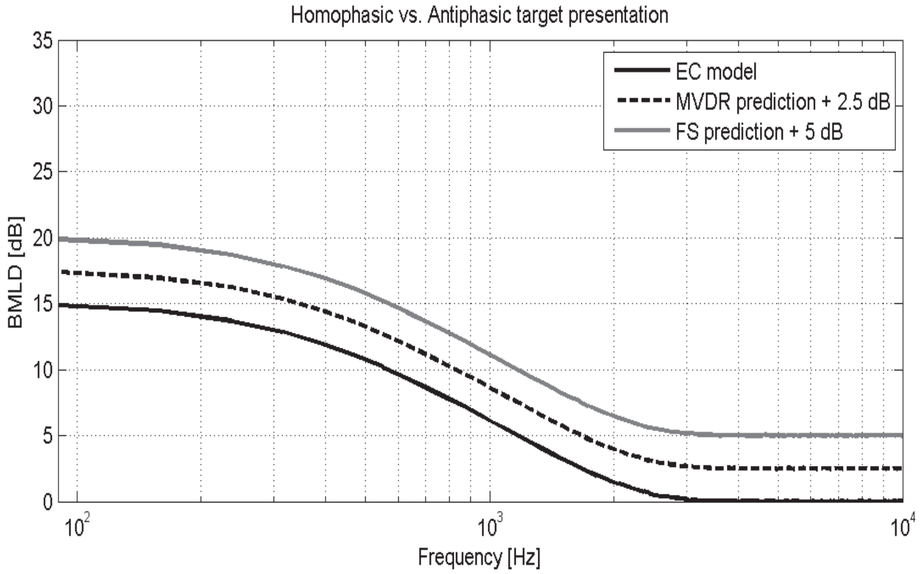


Fig. 1: BMLD for the antiphasic signal presentation compared to the homophasic signal presentation, where the masker is homophasic in both cases. The traditional EC model is given by the solid black curve, the MVDR prediction offset by 2.5 dB is given by the dashed black curve, and the FS beamformer prediction offset by 5 dB is given by the gray curve. Both beamformers are capable of accurately predicting the BMLD according to the original EC model.

SIMULATION RESULTS

In this section a selection of the experimental setups from Durlach (1963) will be reproduced and the BMLD predictions of the different beamformers are compared to the corresponding BMLD prediction of the EC model.

Antiphasic vs homophasic as a function of frequency

In the first simulation, the SNR for antiphasic target-signal presentation was compared to the homophasic target-signal condition. The masker was in both cases homophasic:

$$\text{Condition A} \quad : \quad \mathbf{a}(f) = \quad 1 \quad -1 \;)^T, \mathbf{b}(f) = \quad 1 \quad 1 \;)^T \qquad \text{(Eq. 11)}$$

$$\text{Condition H} \quad : \quad \mathbf{a}(f) = \quad 1 \quad 1 \;)^T, \mathbf{b}(f) = \quad 1 \quad 1 \;)^T \qquad \text{(Eq. 12)}$$

In Fig. 1, the quantity $R_{\text{A/H}}(f)$ is plotted as a function of frequency. The traditional EC model is given by the solid black curve, the MVDR prediction offset by 2.5 dB is given by the dashed black curve and the FS beamformer prediction offset by 5 dB is given by the gray curve. Both beamformers are capable of accurately predicting the BMLD according to the original EC model.

Variations in the interaural time delays of the signal and noise

The following section describes various conditions where the interaural delay is varied for the masker or the target. The first condition describes a situation where the delay of the target is varied:

$$\begin{aligned} \text{Condition DT}: \mathbf{a}(f) &= \quad 1 \quad e^{-j2\pi f\tau} \;)^T, \\ \mathbf{b}(f) &= \quad 1 \quad 1 \;)^T. \end{aligned} \qquad \text{(Eq. 13)}$$

In Fig. 2, $R_{\text{DT/H}}(\tau)$ is shown where the frequency is $f = 167$ Hz and τ is varied between -3 and 3 ms. The traditional EC model is given by the solid black curve, the MVDR prediction offset by 2.5 dB is given by the dashed black curve, and the FS beamformer is given by the gray curve. All beamformers have the correct predictions for ±3 ms and 0 ms. The filter and sum beamformer has the wrong shape in between these points and seems to be a shifted and inverted version of the EC model. The MVDR, however, seems capable of accurately predicting the BMLD.

In Fig. 3, the situation is reversed and the delay of the masker is varied:

$$\begin{aligned} \text{Condition DM}: \mathbf{a}(f) &= \quad 1 \quad 1 \;)^T, \\ \mathbf{b}(f) &= \quad 1 \quad e^{-j2\pi f\tau} \;)^T. \end{aligned} \qquad \text{(Eq. 14)}$$

$R_{\text{DM/H}}(\tau)$ is shown where the frequency is $f = 500$ Hz and τ is varied between 0 and 4 ms. The traditional EC model is given by the solid black curve, the MVDR prediction offset by 2.5 dB is given by the dashed black curve, and the FS beamformer prediction is given by the gray curve. All beamformers have the correct predictions for 0 ms, 2 ms, and 4 ms. All predictions seem periodic, however, the filter and sum beamformer again has the wrong shape in between 0, 2, 4 ms compared to the EC model, whereas the MVDR accurately predicts the BMLD.

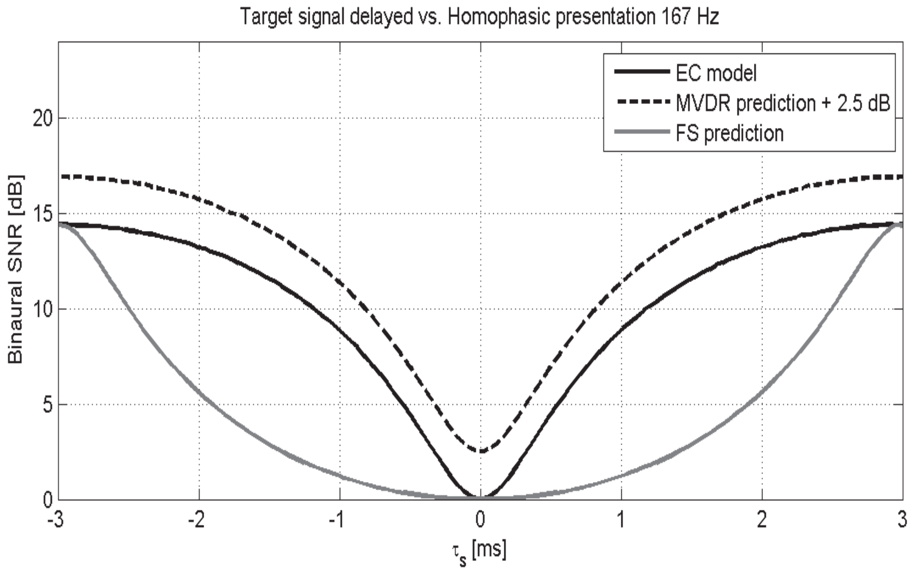


Fig. 2: BMLD for interaurally time-delayed target condition compared to homophasic target presentation. The center frequency was 167 Hz. The traditional EC model is given by the solid black curve, the MVDR prediction offset by 2.5 dB is given by the dashed black curve, and the FS beamformer prediction is given by the gray curve.

DISCUSSION

In this paper two different beamformers were analyzed for the purpose of predicting BMLDs: the filter and sum beamformer (FS) and the minimum variance distortionless response (MVDR). The work spawned from a statement in Durlach (1963) where the EC model was compared to a null-pointing array. Analogous to this, adaptive beamforming techniques automatically adjust the nulls of the array to correspond to the directions of the interferers. The mathematical details of the processing both for the static FS beamformer and for the adaptive MVDR showed large discrepancies in the beamforming coefficients compared to the EC model. However, when applying the beamformers to the examples in the original paper, it was shown that the MVDR was able to accurately predict the BMLD given by the EC model, whereas the static FS beamformer only accounted for the correct BMLD in the condition with the target signal in anti-phase and the masker signal in phase in the two ears. The MVDR has the advantage over the EC model that it does not need any a priori knowledge of the acoustic transfer function between the masker and the listener; instead, it only requires information about the target. This can simplify the use of the model when investigating complex listening environments with multiple interferers from different directions and/or diffuse-noise listening conditions.

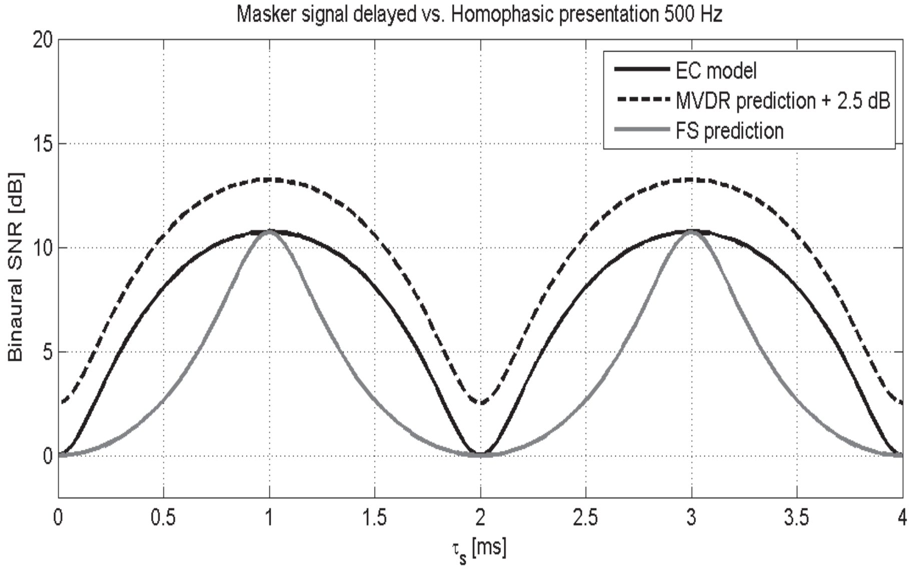


Fig. 3: BMLD for interaurally time-delayed masker condition compared to homophasic masker presentation, where the target is homophasic in both cases. The center frequency was 500 Hz. The traditional EC model is given by the solid black curve, the MVDR prediction offset by 2.5 dB is given by the dashed black curve, and the FS beamformer prediction is given by the gray curve.

REFERENCES

Beutelmann, R., and Brand, T. (**2006**). “Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing impaired listeners,” J. Acoust. Soc. Am., **120**, 331-342.

Capon, J., Greenfield, R., and Kolker, R. (**1967**). “Multidimensional maximum-likelihood processing of a large aperture seismic array,” P. IEEE, **55**, 192-211.

Cherry, E.C. (**1953**). “Some experiments on the recognition of speech with one and two ears,” J. Acoust. Soc. Am., **25**, 975-979.

Cherry, E.C., and Bowles, J.A. (**1960**). “Contribution to a study of the cocktail party problem,” J. Acoust. Soc. Am., **32**, 884.

Cherry, E.C., and Sayers, B.M.A. (**1956**). “Human cross-correlator - a technique for measuring certain parameters of speech perception,” J. Acoust. Soc. Am., **28**, 889-895.

Cherry, E.C., and Taylor, W.K. (**1954**). “Some further experiments upon the recognition of speech, with one and with two ears,” J. Acoust. Soc. Am., **26**, 554-559.

Culling, J.F., Hawley, M.L., and Litovsky, R.Y. (**2004**). “The role of head induced interaural time and level differences in the speech reception threshold for multiple interferring sound sources,” J. Acoust. Soc. Am., **116**, 1057-1065.

Durlach, N.I. (**1960**). "Note on the equalization and cancellation theory of binaural masking level differences," J. Acoust. Soc. Am., **32**, 1075-1076.

Durlach, N.I. (**1963**). "Equalization and cancellation theory of binaural masking level differences," J. Acoust. Soc. Am., **35**, 1206-1218.

Hawley, M.L., Litovsky, R.Y., and Culling, J.F. (**2004**). "The benfit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am., **115**, 833-843.

Johnson, D.H., and Dudgeon, D.E. (**1993**). *Array Signal Processing* (Prentice Hall, Englewood Cliffs, NJ).

Kock, W.E. (**1950**). "Binaural localization and masking," J. Acoust. Soc. Am., **22**, 801-804.

Leaky, D.M., and Cherry, E.C. (**1957**). "Influence of noise upon the equivalence of intensity differences and small time delays in two-loudspeaker systems," J. Acoust. Soc. Am., **29**, 284-286.

Rayleigh, L. (**1876**). "Our perception of the direction of sound," Nature, **14**, 32-33.

Rayleigh, L. (**1907**). "On our perception of sound direction," Phil. Mag., **6**, 213-242.

Sayers, B.M., and Cherry, E.C. (**1957**). "Mechanism of binaural fusion in the hearing of speech," J. Acoust. Soc. Am., **28**, 973-987.

Wan, R.W., Durlach, N.I., and Colburn, H.S. (**2010**). "Application of an extended equalization-cancellation model to speech intelligibility with spatially distributed maskers," J. Acoust. Soc. Am., **128**, 3678-3690.

Zurek, P.M. (**1992**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors affecting Hearing Aid Performance*, 2nd ed. Edited by G.A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255-276.

Profiling hearing-aid sound

MORTEN L. JEPSEN* AND CHRISTIAN NORUP

Audiological Research and Innovation, Widex A/S, DK-3540 Lynge, Denmark

Assessment of audio quality has a strong tradition within concert hall acoustics, music reproduction and telecommunication, and some of the associated methods have recently been applied to hearing aid sound (Simonsen and Legarth, 2010). Many assessment methods have been developed and evaluated, and one of the most valuable methods is the use of assessment panels consisting of trained listeners (e.g., Legarth *et al.*, 2012). Considerations about sound quality are an integral part of hearing-aid development as hearing-aid gain strategies and processing modify the sound by applying, e.g., frequency-dependent gain and dynamic-range compression, in order to compensate for consequences of hearing impairment. Hearing-aid manufacturers use different processing principles and different signal-processing technology to obtain this compensation. In the present study, the aim was to obtain the sound-attribute profile for Widex devices and compare this to profiles of devices from other manufacturers, as well as an earlier Widex device. The listening panel comprised listeners with hearing impairment and was provided by DELTA SenseLab. The sound preference of the listening panel was also measured in a variety of acoustic scenarios focusing on speech and music conditions. It was found that the sound profiles of the different manufacturer devices were different and that this may be explained by differences in processing principles and technology.

INTRODUCTION

The aim of the signal processing in modern hearing aids is to provide the optimal gain and feature strategy to allow the hearing impaired to hear similarly to normal hearing listeners without compromising sound quality with regards to comfort and naturalness. Different manufacturers apply different principles and technology to reach these goals. Effectively, this means that the overall perceived sound quality of devices from different manufacturers can be quite different. The present study aims at quantifying subjectively perceived sound quality using a number of sound attributes, and identifying a manufacturer-specific *sound signature*. It is hypothesized that sound quality profiles are different across hearing-aid manufacturers and may also be similar across series within one manufacturer.

Sound quality assessment has been used for general product sound evaluation in many applications. The methods and analysis tools originate from a broader field of sensory evaluation, which have proved to be very influential in, e.g., the food industry. It is common to use a panel of trained assessors, who have a common

*Corresponding author: moje@widex.com

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

language and proven sensitivity to small quality differences. Sound reproduction systems for listening to music and entertainment have historically also had very fruitful use of quality assessment. In assessment methods the goal is often to establish a set of meaningful sound quality attributes and explore how these are related to preference or perceived 'good' quality (e.g., Gabrielsson and Sjögren, 1979). Sound quality assessment in the hearing-aid industry deviates from other consumer industries since listeners with hearing impairment may perceive sound quality differently from listeners with normal hearing. It is, however, unclear whether a group of listeners that are uniform with regards to pure-tone audiometry will have similar sound quality perception. In the present study, a perceptual sound quality evaluation was conducted, using a panel with trained listeners with homogeneous sensorineural hearing impairment.

METHODS

Assessor panel

The panel of listeners was developed by DELTA SenseLab. Their panel consists of a number of listeners with moderate sloping hearing loss (N3) (Legarth *et al.*, 2012). An assessor can only be included in the panel based on a satisfying evaluation of his ability to reproduce data in identical conditions and show sensitivity to changes in sound quality. Eleven listeners participated in this study. The execution of the experiment and the administration of the listeners were handled by DELTA SenseLab, who were paid for their services.

Attributes and acoustic scenarios

In order to show various aspects of sound quality, a number of sound samples were used for different test scenarios, comprising babble, female speech, male speech, pop music, classical music, traffic and nature. Overall aspects of preference and attributes were evaluated by the listening panel for all these scenarios. In the first part of the assessment, the panel was asked to rate sound samples according to preference, while in the second part, more detailed information about the perceived sound quality was obtained. The panel evaluated the sound samples according to six attributes of sound quality, namely 'naturalness', 'fullness', 'loudness', 'sharpness', 'distortion', and 'tube sound'. These six attributes were chosen on the basis of earlier experiences of DELTA SenseLab.

Devices

The present article presents the results obtained with the assessor panel for four hearing-aid devices: (A) a current Widex device, (B) a Widex device from year 2005, (C) a current competitor device #1 and (B) a current competitor device #2. All devices were set up with proprietary fitting rationale recommended by the manufacturer. In conditions of music signals, a recommended music/entertainment program was chosen.

Recording setup

The stimuli that were presented to the subjects were recorded at Widex facilities according to guidelines of DELTA. Recordings were made with devices mounted on the KEMAR and sound was played back using a 5.1 loudspeaker setup, using multichannel sound files provided by DELTA SenseLab. Closed BTE moulds, with 1.5-mm venting recommended for the given hearing loss, and standard hearing-aid tubing were used. The premise was that the devices should be exposed to the same stimuli presented in the same setup. The aided sound was recorded by use of the coupler microphone. The resulting sound files were then sent to DELTA and scrambled, such that the recorded devices were blinded to their test leaders and assessors.

RESULTS

Preference

In this section, the rating of preference is presented, in terms of 'overall device preference', 'device preference for male speech in quiet' and 'device preference for classical music'. The data are presented as the mean device ratings across listeners on a 100-point scale, and the error bars reflect 95% confidence intervals. Figure 1 shows the preference ratings overall and for the two specific conditions.

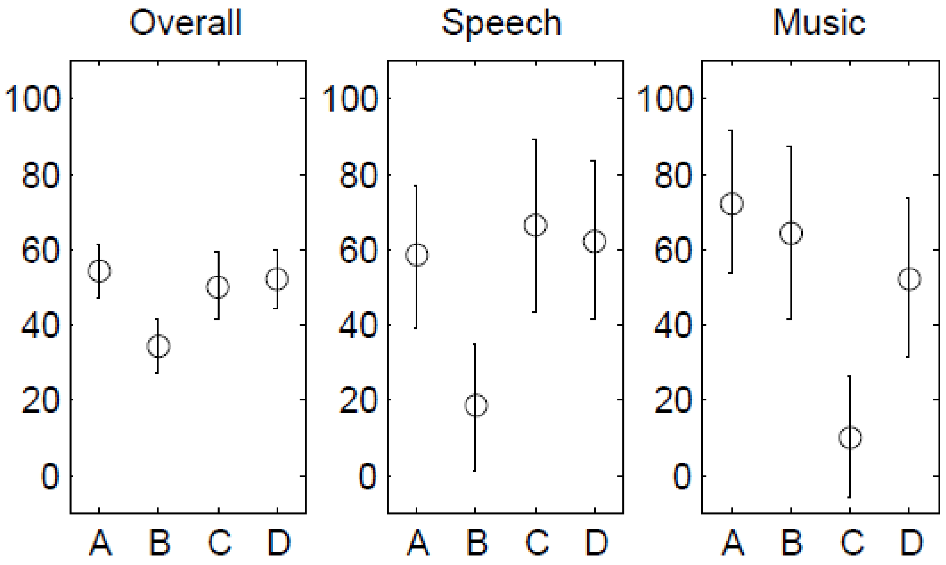


Fig. 1: Preference ratings for 'overall', 'speech in quiet' and 'classical music'.

In overall preference, devices A, C, and D are not significantly different, while device B is rated significantly poorer than the three others. From the middle panel it can be seen that devices A, C, and D achieve similar ratings for male speech, while device B is rated significantly lower with regards to preference. The confidence intervals are larger due to the lower number of responses that the mean data are based on. The right panel shows the data from the condition where classical music was the stimulus. Here the pattern is different, as devices A and B are rated higher than C and D, while the devices A, B, and D are rated significantly higher than C.

Profiles

Many factors of subjective perception are involved in the determination of preference. So in order to obtain a higher resolution of the preference rating, six sound quality attributes were tested. Figure 2 shows the ratings of 'naturalness' of the four devices, averaged across scenarios. This is shown as an example, and similar data were obtained for each of the six test attributes. There is a trend in that device A has the most natural sound. Device A also had the highest overall preference rating. Interestingly, device B had a relatively low rating on overall preference, yet reaching a rather good rating on naturalness. This illustrates that preference does include and combine elements from a set of sound quality attributes.

To show the subjective rating from all six attributes at once, Fig. 3 shows 'spiderweb plots' where each point in the hexagon indicates the rated value of each attribute. By connecting the data points in this graphical representation, the sound profile of a test device may be visualized, and thereby compared for the four devices.

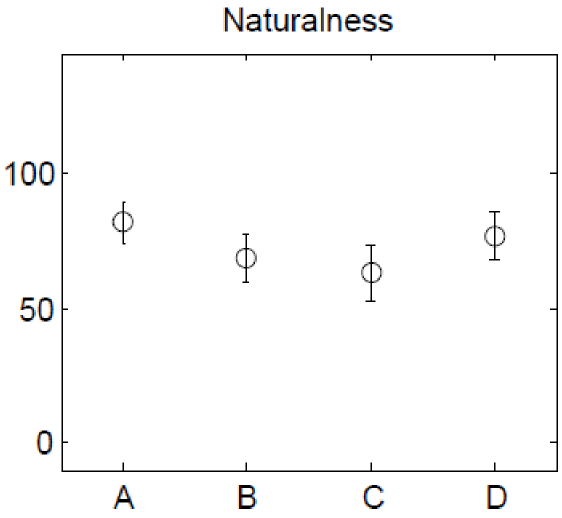


Fig. 2: Results for the average rating of 'naturalness'.

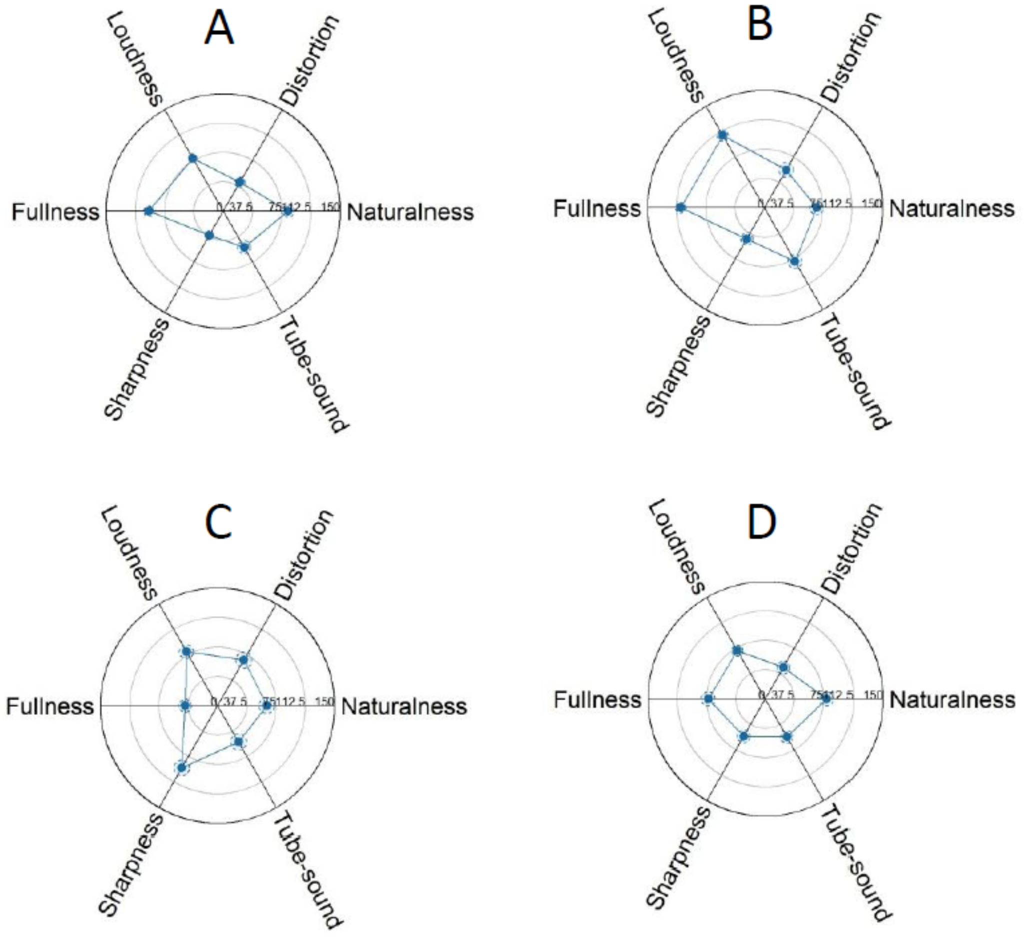


Fig. 3: Sound-attribute profiles of the four test devices. The radius of the symbol represents 95% level of confidence.

The timbre characteristics represented by ‘fullness’ and ‘sharpness’ show differences in the products. Device D shows a medium level of fullness and sharpness, whereas device C has much treble and little bass, leading to a sharp sound. Devices A and B are both rated high on fullness which indicates a sound with a strong bass reproduction. The individual product characteristics can be described from the profile plots (Fig. 3):

A. Very full (bassy) sound with little ‘sharpness’ and low level of ‘distortion’.
B. The most bassy (high on fullness) device, and also significantly louder than any other devices. It has a high rating on ‘tube sound’ which could be related to by the high loudness rating.
C. A sharp/thin sound with some ‘distortion’. Lowest rating on ‘naturalness’.
D. Medium fullness and sharpness. Average ‘loudness’, ‘distortion’, and ‘tube sound’.

Looking at the shapes of the profiles, it appears that devices A and B have a similar tilted rectangular shape, even though device B has higher values in 'loudness', 'fullness', and 'tube sound'. These tilted rectangles are somewhat different from the shapes of devices C and D. This is interesting, as these devices are from the same manufacturer, namely Widex. This would suggest that the shape identifies the Widex sound signature, since they are both Widex devices. Devices C and D come from other manufacturers and clearly have different sound profiles.

DISCUSSION AND CONCLUSIONS

It was possible to identify a Widex sound signature based on the panel's evaluation. Whether this sound profile is optimal and most preferable is not given as such. However, for particular attributes there is an intuitive link between good sound quality and preference. An attribute like 'naturalness' seems like something that should be as high as possible to be preferred, while 'distortion' and 'tube sound' can be associated with poor sound quality and should be minimized. The remaining three attributes, 'loudness', 'sharpness', and 'fullness', do not have clear relation to poor or good sound. With the present data, it is not possible to conclude anything about a clear relation between the tested attributes and preference. However, if more data were available, it could be possible to create a map from attribute rating to preference, using methods of factor analysis or principal components.

The hearing-aid sound is thought to be strongly associated with the manufacturers' fitting rationales and underlying audiological principles. It is assumed that the manufacturer responsible for device C provides a fitting rationale with more high-frequency gain compared to devices A, B, and D, leading to a high rating on sharpness. The Widex devices (A and B) have relatively more 'fullness'.

Technological aspects other than amplification rationales may have an impact on the perception of 'distortion' and 'tube sound'. It is likely that modern digital signal processing algorithms introduce distortion, but factors determined by audiological principles, such as compression speed, can also have a large impact on perceived 'distortion', as well as 'tube sound'. Furthermore, the hearing-aid acoustics related to venting can also contribute to especially the perception of tube sound due to the direct sound path.

In conclusion, the availability of sound quality profiles allow for the formulation of specific goals for sound quality of future devices, while the evaluation methods used here may be used to quantitatively test whether the goals have been achieved.

REFERENCES

Gabrielsson, A., and Sjögren, H. (**1979**), "Perceived sound quality of sound-reproducing systems," J. Acoust. Soc. Am., **65**, 1019-1033.

Legarth, S., Simonsen, C.S., Dyrlund, O., Bramsløw, L., and Jespersen, C.T. (**2012**). "Establishing and qualifying a hearing impaired expert listener panel," Poster at the International Hearing Aid Research Conference.

Simonsen, C.S., and Legarth, S.V. (**2010**). "A procedure for sound quality evaluation of hearing aids," Hearing Review, **17**, 32-37.

Modeling potential distributions inside the cochlea caused by electrical stimulation

ANJA CHILIAN[1,2,*], ANDRÁS KÁTAI[1], TAMÁS HARCZOS[1,3], AND PETER HUSAR[1,2]

[1] *Fraunhofer Institute for Digital Media Technology IDMT, D-98693 Ilmenau, Germany*

[2] *Institute of Biomedical Engineering and Informatics, Faculty of Computer Science and Automation, Ilmenau University of Technology, D-98693 Ilmenau, Germany*

[3] *Institute for Media Technology, Faculty of Electrical Engineering and Information Technology, Ilmenau University of Technology, D-98693 Ilmenau, Germany*

During the last decades the average speech intelligibility of cochlear-implant (CI) users has steadily been improved. Nevertheless, problems still occur especially in complex listening situations. One reason for that is the inaccurate signal transmission between CI electrodes and stimulated nerve cells. To develop new methods overcoming this problem, models are required that provide insight into the processes of electrical stimulation inside the complex geometry of the cochlea. This paper presents a detailed model of the electrically stimulated cochlea. The model consists of a virtual three-dimensional representation of the most important structures of the human cochlea. It serves as a basis for the volume conductor model, which was developed using finite element method. It allows for computation of the electrical potentials inside the modeled structures caused by current applied to the CI electrodes. The presented model was used to compare current spread for different electrode positions and configurations. The results show that the model can represent characteristic differences in spatial selectivity and hence be a help in realizing spatially more focused electrical stimulation.

INTRODUCTION

A cochlear implant (CI) is an electronic device to provide a sensation of sound to patients with severe to profound hearing loss. It bypasses damaged parts of the ear by electrical stimulation of the auditory nerve. Due to advances in technology and signal processing, most CI users reach good speech intelligibility in quiet environments. However, complex listening situations remain challenging. One factor contributing to this problem is the electrode-neuron interface. Current applied to the CI electrodes spreads along the fluid-filled cochlea. Therefore, different electrodes can excite overlapping populations of auditory neurons, which leads to channel interactions.

To improve signal transmission between CI electrodes and stimulated nerve cells, deeper knowledge about the processes of electrical stimulation inside the complex

*Corresponding author: anja.chilian@idmt.fraunhofer.de

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

geometry of the cochlea is required. However, experimental investigations are not practicable due to the small dimensions of the cochlea. Models are a more feasible option. For example, they can be used to investigate the electrical potentials generated by a CI.

In this paper, we present a detailed three-dimensional model of the electrically stimulated human cochlea. The model allows for computation of electrical potentials inside the cochlear structures caused by current applied to CI electrodes. We use this model to compare simulated potential distributions for various electrode positions and configurations.

METHODS

3-D model of the human cochlea

In order to create a model of the electrically stimulated cochlea, a representation of the cochlear geometry is necessary. For this purpose, a 3-D model incorporating all important structures of the human cochlea was developed. To obtain a realistic and detailed representation, histological sections of the human temporal bone served as a basis for modeling.

To create the model, the contours of all important anatomical structures were defined in various cross sections through the cochlear turns. Then the defined contours in the different cutting planes were connected to each other to create solids. Figure 1 shows one of the defined cross sections. Structures that were not modeled were considered to be too small to influence the calculated electrical potentials significantly.

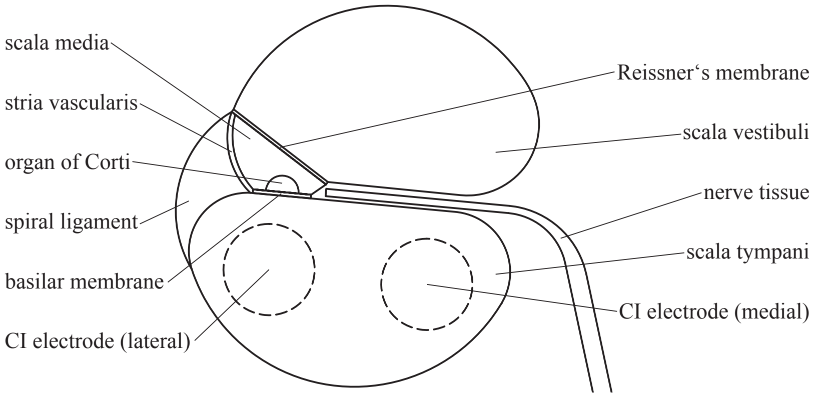


Fig. 1: Cross section through the second cochlear turn containing all modeled anatomical structures of the cochlea and the nerve tissue. The two circles indicate the electrode positions that were modeled: a lateral placement near the outer wall and a medial placement near the modiolus.

The resulting virtual model, which is shown in Fig. 2, was completely embedded into a cylinder with a diameter of 20 mm and a height of 11 mm. This cylinder models bone tissue. In this way, it could be ensured that all gaps were filled.

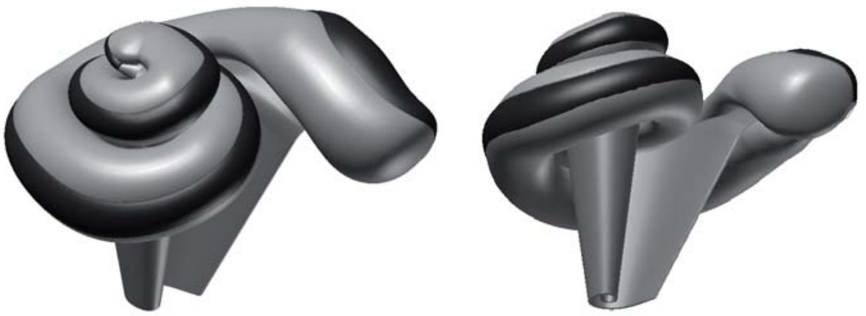


Fig. 2: Illustration of the created 3-D model of the human cochlea in top view (left) and side view (right). The surrounding bone tissue has been removed for better visualization. The following structures are visible: spiral ligament, nerve tissue, scala vestibuli, and scala tympani.

In addition to the anatomical structures, the implanted CI electrodes were included into the model. Shape and size of the electrodes comply approximately with the Cochlear^{TM} Nucleus® full-band straight electrode. Therefore, 22 banded electrode contacts on a cylindrical electrode carrier were constructed. For the sake of convenience, tapering of the electrode carrier was disregarded. The electrode contacts have a diameter of 0.5 mm and are evenly spaced over a length of 17 mm. The resulting electrode distance is 0.75 mm. The diameter of the electrode carrier is 0.55 mm. Furthermore, the modeled electrode array was placed in two different positions. As shown in Fig. 3, a medial placement close to the modiolus and a lateral placement near the outer wall of the cochlea were modeled.

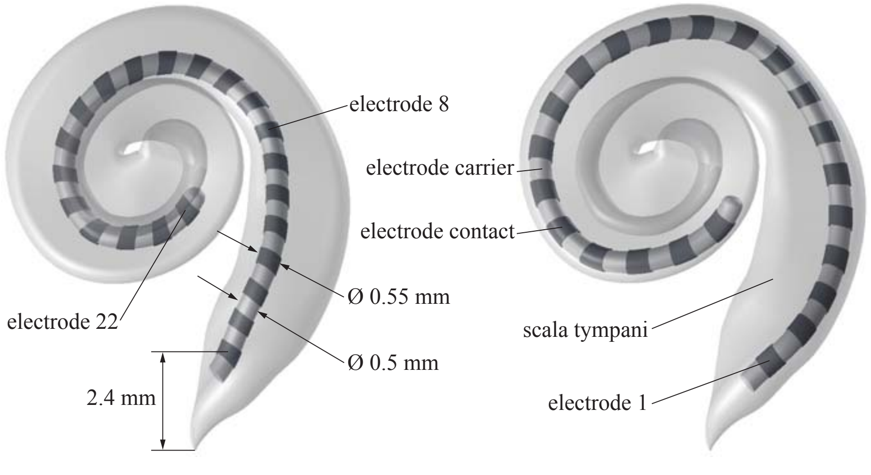


Fig. 3: Visualization of medial (left) and lateral (right) electrode placements inside the scala tympani.

Calculation of electrical potentials

Based on the 3-D model of the human cochlea, a volume conductor model was developed. It was used to calculate the electrical potential distribution inside the cochlear structures as a result of electrical stimulation. Current applied to a CI electrode spreads through the cochlear tissues, which are volume conductors. Hence, a volume conduction problem has to be solved to determine resulting electrical potentials. Because of the electrical properties of the cochlear tissues, the volume conduction problem can be approximated as a quasi-static problem, which is described by Poisson's equation:

$$\nabla^2 \varphi = -\frac{I_s}{\sigma} \qquad \text{(Eq. 1)}$$

where φ is the electrical potential, I_s is the volume current source, and σ is the electrical conductivity. Due to the complex geometry of the cochlea, numerical methods are necessary to solve Eq. 1. For this purpose, the finite element method (FEM) was used.

The FEM model was created using COMSOL Multiphysics®. At first, the created 3-D geometry was imported into the software. Afterwards, the electrical conductivity σ was defined for all modeled structures. All materials were approximated by pure resistances and the values were based on data published by various authors, who developed similar volume conductor models of the cochlea. Table 1 summarizes the values we used in the FEM model.

modeled structure	σ in $\frac{\text{S}}{\text{mm}}$	reference
scala tympani	1.43	Finley *et al.* (1990); Frijns *et al.* (1995)
scala vestibuli	1.43	Finley *et al.* (1990); Frijns *et al.* (1995)
scala media	1.67	Finley *et al.* (1990); Frijns *et al.* (1995)
basilar membrane*	0.0625	Frijns *et al.* (1995); Strelioff (1973)
Reissner's membrane*	0.00098	Frijns *et al.* (1995); Strelioff (1973)
organ of Corti	0.012	Frijns *et al.* (1995); Strelioff (1973)
stria vascularis	0.0053	Frijns *et al.* (1995); Strelioff (1973)
spiral ligament	1.67	Frijns *et al.* (1995); Strelioff (1973)
nerve tissue	0.3	Frijns *et al.* (1995)
bone tissue	0.156	Frijns *et al.* (1995); Suesserman (1992)
electrode carrier	10^{-15}	Tognola *et al.* (2007)
electrode contact	10^{6}	Tognola *et al.* (2007)

Table 1: Specific electrical conductivities σ of the modeled structures in the volume conductor model. For structures marked with an asterisk (*) upscaled values are given (see below).

The modeled solids had to be discretized into smaller tetrahedra. In total, the generated mesh consists of approximately one million elements. To prevent problems with meshing, we upscaled the thicknesses of the basilar membrane and Reissner's membrane by factors of 5 and 10, respectively. To compensate for that, conductivity values of these tissues were also upscaled (see Table 1). This method is in line with that used by Frijns *et al.* (1995).

FEM simulations were performed for two different stimulation protocols. For monopolar electrode configuration the current was applied to one electrode contact. The outer boundaries of the bone cylinder served as ground. For bipolar electrode configuration the current was applied to one electrode and the same current with opposite sign was applied to a neighboring electrode.

RESULTS

Effect of electrode position

Using FEM simulation the potential distribution in all modeled structures can be obtained. To investigate the effect of electrode position, electrical potentials were calculated for monopolar stimulation of electrode 8 with a current of 0.852 mA. Figure 4 compares resulting potential distributions for medial and lateral electrode placements. It shows equipotential lines in a mid-modiolar cross section through the active electrode.

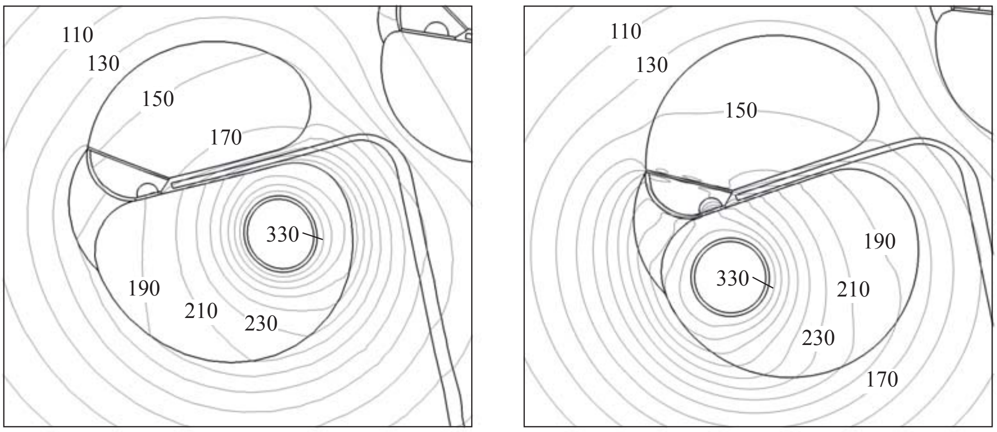


Fig. 4: Equipotential lines in a cross section through the basal cochlear turn for medial (left) and lateral (right) electrode placements caused by monopolar stimulation (0.852 mA) of electrode 8. The numbers indicate the electrical potential in mV. Equipotential lines are spaced by 20 mV. Black lines represent the contours of the modeled cochlear structures.

For the medially placed electrode Fig. 4 reveals higher potential variations in the nerve tissue than for the laterally placed electrode. One reason for that is the smaller electrode to nerve fiber distance for medial placement. Furthermore, current mainly flows along the highly conductive scala tympani. Surrounding tissues like basilar membrane, stria vascularis, organ of Corti, and bone obstruct the current flow, because of their lower electrical conductivities (see Table 1). This is indicated by the accumulation of equipotential lines in these tissues. By contrast, conductivity of the spiral ligament is comparable to that of the scala tympani. Hence, current leaks from the scala tympani through the outer wall, particularly for the lateral electrode placement. As a result, higher current intensities are necessary to excite neurons using laterally placed electrodes.

Effect of electrode configuration

The potential distribution in the nerve tissue is of primary interest because it gives some indication of possible neural excitation. Hence, resulting electrical potentials on the surface of the nerve tissue are shown in Fig. 5 to compare different electrode configurations. It illustrates equipotential lines for monopolar and bipolar stimulation applying a current of 0.852 mA. Electrode 8 served as the active electrode and was placed in medial position.

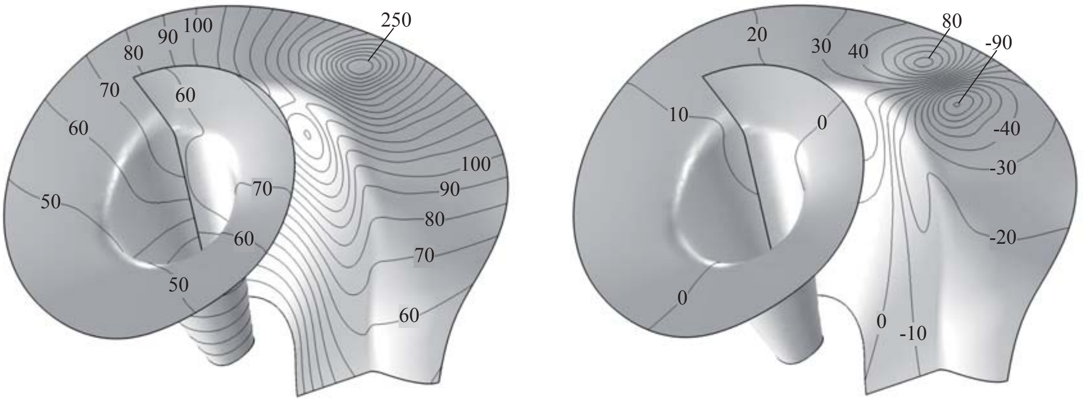


Fig. 5: Equipotential lines on the surface of the nerve tissue for monopolar (left) and bipolar (right) stimulation (0.852 mA) of electrode 8 (medial placement). The numbers indicate the electrical potential in mV. Equipotential lines are spaced by 10 mV.

In Fig. 5 it can be seen that the electrical potential reaches its maximum near the active electrode and falls off with increasing distance. For bipolar electrode configuration an additional potential minimum is visible near the neighboring return electrode and a zero potential line occurs between both electrodes.

By comparison, monopolar stimulation causes a relatively wide spatial distribution of the electrical potential, whereas potential variations are more localized with bipolar stimulation. Furthermore, monopolar configuration induces higher electrical potential values than bipolar configuration. To better illustrate this effect, the electrical potential was calculated along a spiral path on the surface of the nerve tissue near the basilar membrane. The results for all modeled electrode positions and configurations are shown in Fig. 6.

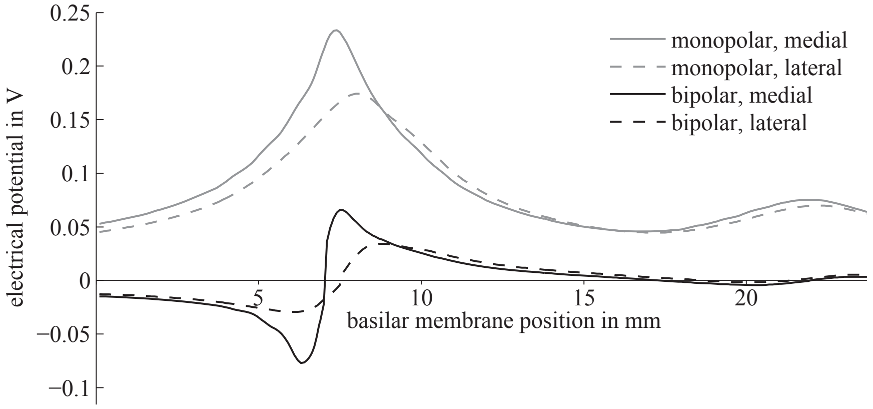


Fig. 6: Course of the electrical potential along a spiral path on the surface of the nerve tissue (close to the basilar membrane) for different electrode configurations and positions. Electrode 8 served as the active electrode.

It is visible that electrode position and configuration influence both spatial distribution and amplitude of the electrical potential in the cochlear tissues. Monopolar stimulation of the medial electrode causes the highest potential values. For laterally placed electrodes the spatial distribution gets wider. Bipolar stimulation produces more localized potentials, but the amplitudes are much smaller in comparison to monopolar stimulation.

Furthermore, Fig. 6 reveals an additional increase of potential values for positions above 20 mm along the spiral path. This position corresponds to nerve fibers one turn above the stimulated electrode. These fibers also have a relatively small distance to the active electrode and may be excited by higher current levels. This effect is called cross-turn or ectopic stimulation and can only be simulated by three-dimensional models of the cochlea.

CONCLUSIONS

In this paper, we have presented a detailed and realistic model of the implanted human cochlea. This model was used to calculate potential distributions inside the cochlear structures. Characteristic differences in spatial selectivity were shown for various electrode configurations and positions. These results are in good agreement with previous findings in the literature, e.g., Briaire and Frijns (2000) and Tognola *et al.* (2007). Hence, the model can be used to investigate different aspects of electrical stimulation.

There are many possible applications of the model. For example, effects of various electrode designs or stimulation protocols on resulting electrical potentials can be evaluated. In this way, the model could be a help in realizing spatially more focused electrical stimulation and consequently reducing channel interactions.

However, in order to infer from the calculated potential distributions about neural excitation, a nerve fiber model is essential. Therefore, further work will concentrate on extensions of the model, to additionally simulate nerve fiber responses. Thus it would be possible to evaluate influences on the spread of neural excitation.

REFERENCES

Briaire, J.J., and Frijns, J.H. (**2000**). “Field patterns in a 3D tapered spiral model of the electrically stimulated cochlea,” Hear. Res., **148**, 18-30.

Finley, C.C., Wilson, B.S., and White, M.W. (**1990**). “Models of neural responsiveness to electrical stimulation,” in *Cochlear Implants: Models of the Electrically Stimulated Ear*. Edited by J.M. Miller and F.A. Spelman (Springer, New York), ISBN: 9783540970330, pp. 55-96.

Frijns, J.H., de Snoo, S.L., and Schoonhoven, R. (**1995**). “Potential distributions and neural excitation patterns in a rotationally symmetric model of the electrically stimulated cochlea,” Hear. Res., **87**, 170-186.

Strelioff, D. (**1973**). “A Computer Simulation of the generation and distribution of cochlear potentials,” J. Acoust. Soc. Am., **54**, 620-629.

Suesserman, M. (**1992**). *Noninvasive microelectrode measurement technique for performing quantitative, in vivo measurements of inner ear tissue impedances*, Ph.D. thesis, University of Washington.

Tognola, G., Pesatori, A., Norgia, M., Parazzini, M., Di Luca, R., Ravazzani, P., Burdo, S., Grandori, F., and Svelto, C. (**2007**). “Numerical modeling and experimental measurements of the electric potential generated by cochlear implants in physiological tissues,” IEEE Trans. Instrum. Meas., **56**, 187-193.

Making use of auditory models for better mimicking of normal hearing processes with cochlear implants: first results with the SAM coding strategy

TAMÁS HARCZOS[1,2,*], ANJA CHILIAN[1,3], ANDRAS KATAI[1], FRANK KLEFENZ[1], IZET BALJIĆ[4], PETER VOIGT[5], AND PETER HUSAR[1,3]

[1] *Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany*

[2] *Electronic Media Technology Lab, Faculty of Electrical Engineering and Information Technology, Ilmenau University of Technology, Ilmenau, Germany*

[3] *Institute of Biomedical Engineering and Informatics, Faculty of Computer Science and Automation, Ilmenau University of Technology, Ilmenau, Germany*

[4] *HELIOS Hospital Erfurt, Department of Otolaryngology, Erfurt, Germany*

[5] *Cochlear-Implant Rehabilitationszentrum Thüringen, Erfurt, Germany*

Stimulation based on auditory modeling, or SAM, is a new speech-processing strategy for cochlear implants that we developed recently at Fraunhofer IDMT. SAM incorporates active cochlear filtering along with the mechanoelectrical transduction of the inner hair cells, so that several psychoacoustic phenomena are accounted for inherently. SAM was tested with a group of five CI users: We investigated speech perception in quiet and in the presence of noise or reverberation, pitch discrimination abilities (for pure tones and sung vowels), and consonant discrimination. We also asked for subjective quality rating for speech and music snippets. Tests were repeated with the everyday strategy of the implantees and results were compared. This paper presents the test results in detail and compares outcomes with those of the previously published simulation studies. Results are encouraging, although more tests would be needed to increase statistical significance.

INTRODUCTION

Increased processing speeds make applications using auditory models that mimic some properties of the human ear viable. The idea of using models of the human auditory system in cochlear implants (CIs) is not new (see Wilson *et al.*, 2010), but still fairly uncharted. In Harczos *et al.* (2013) we presented a novel sound-processing strategy, SAM (Stimulation based on Auditory Modeling), which was based on hydromechanical and neurophysiological models of the human ear and could be employed in auditory prostheses.

SAM incorporates active cochlear filtering (basilar membrane and outer hair cells) along with the mechanoelectrical transduction of the inner hair cells, so that travel-

*Corresponding author: hzs@idmt.fraunhofer.de

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

ling-wave delays and several psychoacoustic phenomena are accounted for inherently. The produced stimulation patterns differ greatly from that of the wide-spread Cochlear ACE™ (advanced combination encoder) strategy. Although the computation of SAM requires considerably more operations than that of ACE, the current C/C++ implementation of SAM can run in real-time on a state-of-the-art desktop computer.

At ISAAR 2011 we showed the outline of the algorithm along with first simulation results concerning speech reception thresholds (Harczos *et al.*, 2012b) and horizontal-plane localization abilities using SAM (Harczos *et al.*, 2012a). We also presented a real-time visualization of the strategy and a vocoder algorithm making SAM stimuli audible (Chilian *et al.*, 2012). In the meantime we did first tests with CI users to explore benefits with SAM. In this paper, we present these results.

METHODS

Participants

Five post-lingually deafened adult CI users participated in the study. They were all native speakers of German and had at least two years of CI experience at the commencement of the study. Every subject had a Nucleus® Freedom™ implant with a Contour Advance™ electrode together with a Freedom™ sound processor from Cochlear®. More detailed demographic information is presented below.

Subject	**Age** (yr)	**Deaf** (yr)	**CI** (yr)	**Most probable cause of hearing impairment**	**Laterali-zation**	***CSR*** (pps)	***N***
S1	37	3	4	Circulatory disorder	Bimodal	900	11
S2	70	1	5	Genetic	Unilateral	900	9
S3	69	15	2	Diphtheria	Bimodal	900	8
S4	50	1	5	Genetic / Traumatic	Bilateral	1200	10
S5	27	3	13	Meningitis	Bilateral	900	8

Table 1: Detailed demographic information. The ACE parameters *CSR* and *N* mean channel stimulation rate and number of spectral peaks, respectively.

Assessment procedure

Cochlear-implant users' performance was measured in various ways with a number of tests as listed below.

(1) Testing of speech intelligibility in quiet with the *Freiburg monosyllabic test* (see Hahlbrock, 1953). Corresponding results did not appear to be particularly meaningful and were not listed in this paper.

(2) Testing of speech intelligibility in speech-shaped noise with the *Oldenburg sentence test* (*OLSA*, see Wagener *et al.*, 1999).

(3) Testing of speech intelligibility in simulated reverberant environments using clean but reverberated OLSA sentences with four distinct magnitudes of reverberation. Sentences were played back and the subject was asked to repeat them. The percentage of correctly repeated words was computed.

(4) Testing of pitch discrimination thresholds for pure tones and sung vowels in an adaptive three-interval three-alternative forced-choice ('3I-3AFC', or 'odd-one-out') procedure using the 1-up 2-down paradigm (see Levitt, 1971). Without having to tell which tone was lower or higher in frequency, the subject was asked to identify the tone that was different in pitch. The discrimination threshold was then computed for each test-tone type.

(5) Testing of discrimination ability of consonant pairs (b/p, m/n, n/l, and k/t) using minimal-pair words. In each trial, either two similarly sounding (e.g., bark/park) or two identical words were played back in a sequence. The subject was asked to tell if the two words were the same or not. The percentage of correct answers was computed for each consonant pair.

(6) Subjective quality rating of speech and music via direct comparison of snippets processed with either SAM or ACE.

Reverberation conditions as used in assessment method (3) are listed in Table 2. RT_{60} and STI mean reverberation time and speech transmission index (Steeneken and Houtgast, 1980), respectively. The former is the time required for the sound level to decrease by 60 dB, while the latter is a measure of speech transmission quality. STI is a well-established objective measurement predictor of how well a listener may understand speech using the given transmission channel. STI values may vary between 0 (bad) and 1 (excellent). STI values presented in Table 2 were calculated for the same (randomly selected) OLSA sentence for 65 dB SPL presentation level.

	Reverb-1	**Reverb-2**	**Reverb-3**	**Reverb-4**
Simulated environment	living room	empty office	train station	stairwell (concrete walls)
RT_{60}	935 ms	1440 ms	1380 ms	2700 ms
STI	0.988	0.897	0.745	0.467

Table 2: Summary of reverberation conditions and related parameters.

Conditions as used for the pitch discrimination tests described in assessment method (4) are listed in Table 3. Each presented sequence consisted of three tones (two identical reference and an odd one, each 600 ms long) separated by a 400-ms pause. The spectral distance between the differing and the reference tones was varied adaptively with a quantization of one semitone. Frequencies (or fundamental frequencies in the case of sung vowels) of the tones, as expressed in notes, were determined to be symmetrical around the centre of the valid range for the given test variant (see Table 3). The initial distance was six semitones. The intensity of each tone was randomized by +/− 3 dB to reduce any unwanted effects of loudness variations on the subjects' ranking of pitches. Subjects were instructed to ignore loudness variations, if they perceived any. The task was to identify the tone that was different in pitch.

	Pure tones (C5)	Pure tones (C6)	Pure tones (C7)	Female sung "A" and "I"	Male sung "A" and "I"
Range	C4 (262 Hz) – C6 (1046 Hz)	C5 (523 Hz) – C7 (2093 Hz)	C6 (1046 Hz) – C8 (4186 Hz)	C4 (262 Hz) – F5 (698 Hz)	G2 (98 Hz) – A#3 (233 Hz)
Centre of range	C5 (523 Hz)	C6 (1046 Hz)	C7 (2093 Hz)	G#4 (415 Hz)	D#3 (156 Hz)

Table 3: Summary of conditions for pitch discrimination tests.

Within five sessions (each 2 × 45 minutes plus breaks) all tests were conducted with each participant using both the ACE and the SAM strategies. The latest individual clinical map was used with ACE. With SAM a new map was created and fitted for each CI user. Subjects were provided an excerpt of 6 to 10 minutes of an audio book prior to testing with SAM to get accustomed to the new strategy.

Except for the duration of fitting and initial practice with SAM, subjects were blinded to the choice of processing strategy used in any test.

Stimulation

The Nucleus Implant Communicator (NIC) version 2 from Cochlear® (see Irwin, 2006; Swanson and Mauch, 2006) was employed to directly stream the stimuli from the PC to the CI. All computer programs developed and used during this study were able to process sounds by both the ACE and the SAM strategy. This way, the switch between the strategies was easy for the operator and without attracting subjects' attention.

RESULTS

Speech intelligibility

The standard OLSA test revealed that implant users S4 and S5, being already high-performers with the ACE strategy (i.e., OLSA SRT < 0 dB), could not benefit from switching to the SAM strategy in terms of speech intelligibility in speech-shaped noise. For the other three subjects (having about 10-15 dB worse speech reception thresholds using ACE than S4 and S5) the switch to SAM manifested itself in better SRTs (on average 2.44 dB better).

Results based on the reverberant OLSA corpus showed similar trends: SAM showed no benefit in S4 and S5, while the other three subjects achieved slightly better scores on average. For detailed results, please see Table 4.

Speech intelligibility test type	S1		S2		S3		S4		S5	
	SAM	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM	ACE
OLSA (Standard)	1.9 dB	4.9 dB	5.2 dB	6.3 dB	7.6 dB	10.9 dB	-3.6 dB	-5.9 dB	-2.3 dB	-4.1 dB
OLSA (Reverb-1)	92 %	85 %	87 %	87 %	80 %	55 %	97 %	100 %	100 %	100 %
OLSA (Reverb-2)	80 %	83 %	82 %	76 %	49 %	52 %	100 %	100 %	100 %	100 %
OLSA (Reverb-3)	73 %	70 %	70 %	56 %	37 %	36 %	93 %	100 %	94 %	100 %
OLSA (Reverb-4)	65 %	20 %	9 %	16 %	4 %	18 %	60 %	71 %	93 %	100 %

Table 4: Results of speech-intelligibility tests with the OLSA corpus. The first row shows speech reception thresholds (in dB SNR) measured with the standard OLSA test procedure (speech-shaped noise). The other rows show percentage of correctly identified words of reverberant OLSA sentences at four fixed reverberation magnitudes. Cells with grey background colour denote cases where the ACE strategy performed better.

Pitch discrimination

Since the SAM strategy was designed to provide a considerable amount more temporal pitch information than ACE does, cochlear-implant users were expected to perform better (in terms of pitch-discrimination performance) with SAM than with ACE. Test results showed that this expectation was reasonable: except for isolated cases, all tests delivered much better scores with the proposed new signal-processing strategy.

Signal type in pitch test	S1		S2		S3		S4		S5	
	SAM	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM	ACE
Pure tones (C5)	2.3	8.5	4.6	2.0	2.2	3.9	1.4	2.3	1.5	2.5
Pure tones (C6)	3.3	8.7	3.3	2.5	1.8	2.5	1.5	1.7	1.3	1.0
Pure tones (C7)	2.8	4.1	1.5	2.7	1.5	3.0	3.0	3.5	1.8	2.3
Female sung A	10.3	6.0	6.4	5.0	6.2	7.4	4.3	6.6	5.9	7.5
Female sung I	7.8	10.7	2.5	3.3	3.4	6.5	2.0	3.8	1.8	4.0
Male sung A	6.0	6.5	6.0	12.5	3.5	6.4	6.1	7.4	6.2	6.3
Male sung I	4.5	7.7	7.7	13.5	4.8	10.4	4.8	6.6	6.0	7.0

Table 5: Pitch-discrimination thresholds (in semitones) measured using the adaptive 3-AFC procedure (with 1-up 2-down rule) that targeted 70.7% ($p = 1/\sqrt{2}$) correct discrimination level. Cells with grey background colour denote cases where the ACE strategy performed better.

Table 5 shows discrimination thresholds (in semitones) of all subjects for various signal types. Tests with pure tones seem to be much easier for all subjects: The number of semitones (ST) for the 70.7% discrimination threshold averages to 2.83 ($\sigma = 1.8$), while the same measure for the sung vowels yields 6.28 ST ($\sigma = 2.57$). Vowel ‘I’ sung by the male singer proved to be the most difficult signal: Subjects needed a pitch difference of 7.29 ST ($\sigma = 2.81$) on average (i.e., an interval larger than a perfect fifth!) to correctly identify the difference (with $p = 1/\sqrt{2}$).

The results listed in Table 5 clearly indicate that the tested CI listeners can utilize the additional temporal information provided by the new strategy. The benefit with SAM averages to 1.16 ST ($\sigma = 2.17$), 1.02 ST ($\sigma = 2.27$), and 2.86 ST ($\sigma = 2.36$) for pure tones, female sung vowels, and male sung vowels, respectively.

Consonant discrimination

Results of the consonant-discrimination tests did not deliver clear trends, as shown in Table 6. CI users’ performance seems to be at about the same level with both strategies.

Consonant pair	S1		S2		S3		S4		S5	
	SAM	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM	ACE
b / p	100 %	87 %	100 %	100 %	93 %	87 %	73 %	67 %	100 %	100 %
m / n	27 %	33 %	40 %	67 %	73 %	40 %	87 %	80 %	72 %	67 %
n / l	67 %	40 %	67 %	60 %	53 %	67 %	80 %	80 %	87 %	60 %
k / t	73 %	93 %	80 %	80 %	87 %	87 %	80 %	80 %	80 %	100 %

Table 6: Percentages of correct answers in the consonant pairs test. Cells with a grey background denote cases where the ACE strategy performed better.

Subjective quality

At first sight, subjective quality ratings yielded mixed strategy preferences (see Table 7). However, the preferences of the two bimodal users (S1 and S3) of the test group were remarkable. Since these subjects still had a more or less natural contralateral auditory perception to compare with (hearing aid in the contralateral ear), results would suggest that stimulation via SAM elicits more natural sensation.

Signal type	S1	S2	S3	S4	S5
Speech	SAM better	ACE better	SAM better	ACE better	SAM better
Music	SAM better	no preference	SAM better	no preference	no preference

Table 7: Results of subjective quality rating after direct comparisons. Cells with a grey background denote cases where the ACE strategy was preferred.

DISCUSSION

SAM is a novel speech-processing strategy for implantable auditory prostheses that we have tested in a pilot study with five cochlear-implant users. Even though the number of testees was very low (and hence the variance of the results considerably high), we were able determine some trends of benefit with SAM: Better speech reception thresholds in speech-shaped noise of CI users performing poorly with ACE, and much better pitch-discrimination performance of all testees were the most prominent quantifiable results. These were also predicted by the simulation study published in Harczos *et al.* (2012b).

Another important outcome was that bimodal users rated the quality of sensation higher with SAM than with ACE. This indicates that the firing patterns of the auditory nerve elicited by the SAM stimulation are more similar to the natural ones. Investigations with an acoustical simulation tool (see Chilian *et al.*, 2012) also showed strong preference for SAM (over ACE) in normal-hearing subjects.

Tests of the presented study have also shown that no subject was stressed or disturbed by SAM. Furthermore, knowing that a successful switch from one CI strategy to another may take weeks or months, the fact that all participants understood speech immediately after switching to SAM was an astonishing outcome by itself.

Unfortunately, the NIC v2 tool provided by Cochlear® did not support continuous real-time streaming, which had two important implications. First, there was an unavoidable delay (ranging from seconds to minutes, depending on the duration of the test signal) between sending a stimulus signal from the PC and perceiving it via the CI. Second, to be able to communicate with the CI users, their everyday processor (using ACE) needed to be placed back and turned on again, which might have interfered with the learning processes involved in extracting information from the SAM stimulation patterns.

Preparations are currently underway in our lab to be able to provide a longer uninterrupted habituation and testing period with SAM. Furthermore, we plan to run a longer study including at least 20 CI users to yield more statistically relevant results.

Finally, as the simulation study in Harczos *et al.* (2012a) indicates huge improvements in horizontal plane localization with binaural SAM configurations over ACE, this issue should also be investigated with cochlear-implant users.

ACKNOWLEDGMENTS

We thank volunteer subjects for participating in the pilot study and the staff of the Cochlear-Implant Rehabilitationszentrum Thüringen for their support.

The authors would also like to thank Thomas Vandahl and Akos Kitta for their help with programming, Gero Szepannek for his help with statistics, and Cochlear Limited for access to implant communication software and hardware. We are also grateful for the talks with Horst Hessel, Thomas Hocke, and Jan Poppeliers.

This research was supported in part by grant B514-09020 of the Thuringian Ministry of Education, Science and Culture, Germany, and by Cochlear Ltd.

REFERENCES

Chilian, A., Braun, E., and Harczos, T. (**2012**). "Acoustic simulation of cochlear implant hearing," in *Speech perception and auditory disorders*. 3rd International Symposium on Auditory and Audiological Research. Nyborg, Denmark. Edited by T. Dau, M.L. Jepsen, T. Poulsen, and J. C.-Dalsgaard. ISBN: 978-87-990013-3-0. (The Danavox Jubilee Foundation, Copenhagen), pp. 425-432.

Hahlbrock, K.-H. (**1953**). "Über Sprachaudiometrie und neue Wörterteste," Arch. Ohren Nasen Kehlkopfheilkd., **162**, 394-431.

Harczos, T., Chilian, A., and Katai, A. (**2012a**). "Horizontal-plane localization with bilateral cochlear implants using the SAM strategy," in *Speech perception and auditory disorders*. 3rd International Symposium on Auditory and Audiological Research. Nyborg, Denmark. Edited by T. Dau, M.L. Jepsen, T. Poulsen, and J. C.-Dalsgaard. ISBN: 978-87-990013-3-0. (The Danavox Jubilee Foundation, Copenhagen), pp. 339-345.

Harczos, T., Fredelake, S., Hohmann, V., and Kollmeier, B. (**2012b**). "Comparative evaluation of cochlear implant coding strategies via a model of the human auditory speech processing," in *Speech perception and auditory disorders*. 3rd International Symposium on Auditory and Audiological Research. Nyborg, Denmark. Edited by T. Dau, M.L. Jepsen, T. Poulsen, and J. C.-Dalsgaard. ISBN: 978-87-990013-3-0. (The Danavox Jubilee Foundation, Copenhagen), pp. 331-338.

Harczos, T., Chilian, A., and Husar, P. (**2013**). "Making use of auditory models for better mimicking of normal hearing processes with cochlear implants: the SAM coding strategy," IEEE Trans. Biomed. Circuits Syst., **7**, 414-425.

Irwin, C. (**2006**) "NIC v2 Software Interface Specification E11318RD (Technical Report)," Lane Cove NSW, Australia, Cochlear Ltd.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am., **49**, 467-477.

Steeneken, H.J.M., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am., **67**, 318-326.

Swanson, B.A., and Mauch, H. (**2006**). "Nucleus Matlab Toolbox 4.20 software user manual," Lane Cove NSW, Australia, Cochlear Ltd.

Wagener, K., Kühnel, V., and Kollmeier, B. (**1999**). "Entwicklung und Evaluation eines Satztests in deutscher Sprache I: Design des Oldenburger Satztests (Development and evaluation of a sentence test in German language I: Design of the Oldenburg sentence test)," Z. Audiol., **38**, 4-15.

Wilson, B.S., Lopez-Poveda, E.A., and Schatzer, R. (**2010**). "Use of auditory models in developing coding strategies for cochlear implants." in *Computational Models of the Auditory System*. Edited by R. Meddis, R.R., Fay, E.A., Lopez-Poveda, and A.N. Popper (Springer Science+Business Media LLC, Boston, Massachusetts, USA), pp. 237-260.

Across-electrode processing in CI users: a strongly etiology dependent task

STEFAN ZIRN[1,2,*], J OHN-MARTIN HEMPEL[1]
M ARIA SCHUSTER[1], AND WERNER HEMMERT[2]

[1] *Department of Otorhinolaryngology, Head and Neck Surgery, Ludwig-Maximilians-University Munich, Germany*

[2] *IMETUM, Bio-Inspired Information Processing, Technische Universität München, Germany*

To investigate across-electrode processing in cochlear-implant (CI) users, we established an experimental setup that allows measuring comodulation masking release (CMR) using controlled electrical stimulation of auditory nerve fibers. In this paper we present results of a flanking-band type of CMR experiment with uncorrelated (UC) vs. comodulated (CM) masker components. To deal with the large current spread in electrical stimulation that may introduce additional masking especially in the UC condition, we now compare two different electrode configurations: proximate vs. remote alignments of flanking bands in reference to the on-signal band. Results of 18 test subjects revealed no significant difference between CMR[UC-CM] magnitudes across these two conditions (p = 0.3), whereas outcomes varied strongly across test subjects. To highlight different groups of performers, a hierarchical cluster analysis was conducted. N = 5 CI users showed no or even negative CMR. The majority of N = 9 CI users exhibited positive and significant CMR (around 3 dB). Finally, a subset of N = 4 CI users showed considerable CMR magnitudes (6-10 dB). Etiology was a good indicator for the remaining individual CMR capabilities.

INTRODUCTION

The normal-hearing (NH) auditory system provides elaborated strategies to segregate different sounds with overlapping spectra occurring at the same time, usually an unsolvable task for cochlear-implant (CI) users. An important neural mechanism in this context is across-frequency processing: There is good evidence that the auditory system is able to make comparisons across the outputs of auditory filters (Moore, 2012). Many natural sounds exhibit highly-correlated temporal envelope fluctuations in different frequency bands. Common amplitude fluctuation across-frequency facilitates comodulation masking release (CMR) and may also contribute to auditory grouping (Bregman, 1990). CMR illustrates the fact that detectability of a sinusoidal signal masked by a narrow-band masker can be markedly improved by simultaneously presenting additional maskers at frequencies remote from the signal

[*]Corresponding author: stefan.zirn@med.uni-muenchen.de

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

frequency, provided the envelope fluctuations across frequencies are coherent (Hall *et al.*, 1984).

Two different types of CMR measurements are established in acoustic experiments: band-widening and flanking-band type of CMR experiments (for a review, see Verhey *et al.* (2003)). We concentrate on the latter type (see methods section).

Recent stimulation strategies in cochlear implants are often based on continuous interleaved sampling (CIS): In simple terms, the signal first goes through a set of bandpass filters which divide the acoustic waveform into different frequency channels. The envelopes of each channel are then detected by rectification and low-pass filtering according to a Hilbert transform. Current pulses are generated with amplitudes proportional to the envelopes of each frequency band and transmitted to multiple intracochlear electrodes. In CIS strategies stimulation is usually realized sequentially and not simultaneously across channels. The pulse rate is usually constant.

As CMR is sensitive to low-frequency level fluctuations represented by the temporal envelope of the signals (e.g., Epp *et al.* (2009)) and such low-frequency envelope fluctuations are usually well perceived by CI users (Shannon, 1992), our assumption was that CMR in CI users may exist. In a former publication we addressed this issue (Zirn *et al.*, 2013). There we described how we stimulated relatively apical electrodes with fixed distance between the flanking- and on-signal band-electrodes. As a result we could show that approx. 30% of CI users (7/21) were able to benefit from correlations in a masker in terms of facilitated target detection. The pattern of masked detection thresholds across test subjects, revealing peripheral masking due to current spread, cannot explain the whole effect: The difference of detection thresholds between the uncorrelated and the comodulated condition resulted from a lower detection threshold in the comodulated condition in the subjects with considerable CMR magnitudes. Peripheral masking due to current spread would provoke more masking energy in the on-signal band in the uncorrelated condition and therefore higher detection thresholds in this condition.

To embrace this issue from another perspective we now compare the results of two different configurations of active electrodes: proximate and remote flanking bands in reference to an on-signal masker. This is explained in more detail in the next section.

METHODS

A CMR flanking-band experiment was adapted to electrically-induced hearing. We orientated ourselves to a typical acoustic type of flanking-band CMR experiment. Here, the masker consists of several narrow-band maskers; one at the signal frequency and one or more narrow-band maskers spectrally separated from signal frequency. The masker components are amplitude-modulated either uncorrelated or correlated (comodulated) and the difference of masked detection thresholds of the embedded target sinusoidal signal determines CMR. The underlying definition that can be investigated with this setup is the uncorrelated (UC)-comodulated (CM) CMR[UC-

CM] definition. The masker component at signal frequency is termed on-signal band (OSB) and the other components spectrally remote to the OSB are called flanking bands (FB). To achieve relatively high CMR magnitudes up to 12-13 dB in normal-hearing listeners (Epp *et al.*, 2009) four FBs are often used.

Adapted to electrically induced hearing we stimulated across five intracochlear electrodes. The medial electrode (#14) contained OSB and target. FBs were streamed to proximate or remote four electrodes (see Table 1).

Condition	Electrode configuration
Proximate flanking band electrodes	18, 16, 12, 10
Remote ¨ ¨ ¨	22, 20, 8, 6

Table 1: Addressed electrodes in different test conditions

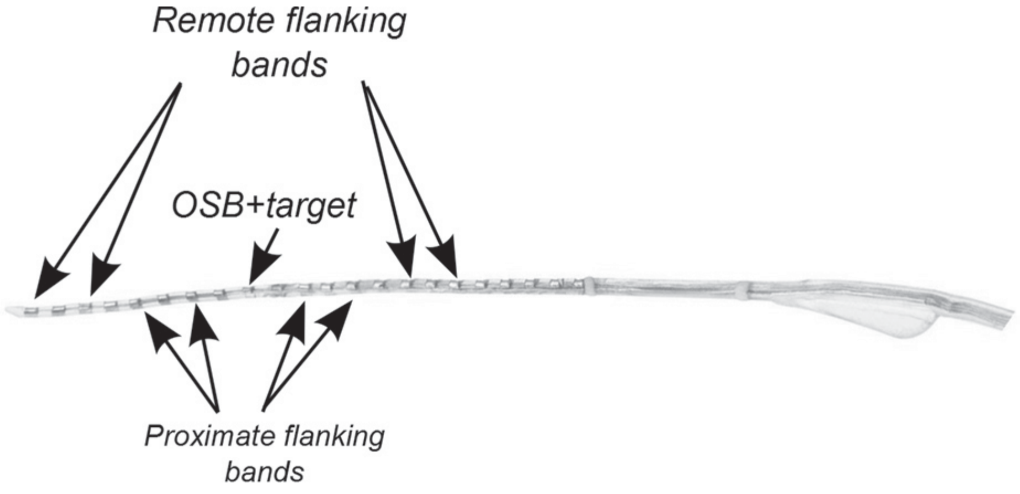


Fig. 1: Electrode configuration in different test conditions. The implant shown is a CI422 by courtesy of Cochlear Ltd.

The required addition of two uncorrelated signals (OSB plus sinusoidal target signal) was conducted in with the original signals (with carrier frequency; pointwise addition with constructive and destructive interference depending on phase relationships of OSB and target; center frequencies orientated at the usual frequency table of the fitting software – Custom Sound Version 3.2, Cochlear Ltd. with 22 active channels). For determination of the overall sound pressure level when adding two non-coherent sounds, see Eq. 1.

$$L_1 + L_2 = 10\log_{10}(10^{L_1/10} + 10^{L_2/10}) \ \text{[dB]} \qquad \text{(Eq. 1)}$$

After determination of the envelope using a Hilbert transform, the signal was then used to modulate a biphasic pulse train and streamed to electrode #14. Additionally four flanking bands (either uncorrelated (Fig. 1) or comodulated (Fig. 2) to the OSB; all biphasic current pulse trains) were presented to proximate or remote electrodes (see Fig. 2).

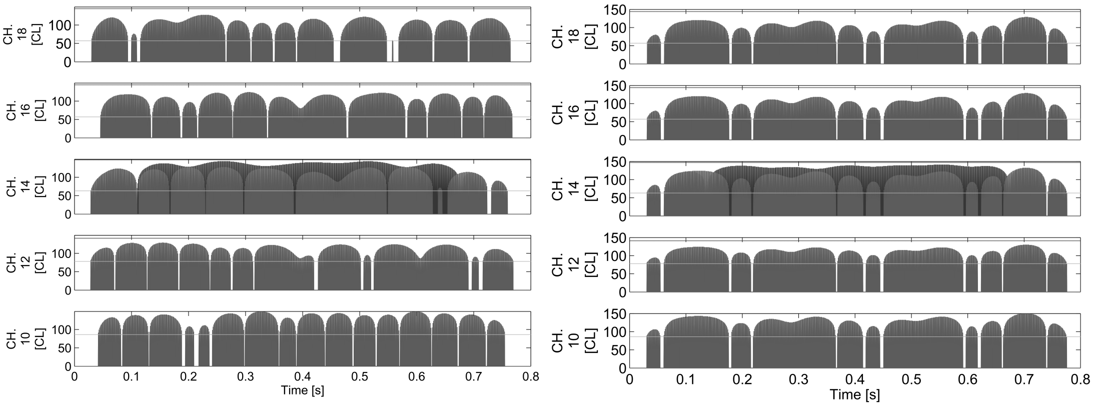


Fig. 2: Superimposed stimulation sequences with and without target (+10 dB signal-to-noise ratio) in the proximate uncorrelated condition (left) and comodulated condition (right). Shown are the positive phases of biphasic current pulse trains that are streamed to five CI electrodes. CL: Cochlear current levels. Horizontal lines indicate electrode-specific current levels at individual hearing threshold levels (T-levels – lower lines) and most comfortable levels (C-levels – upper lines).

Duration of the target was 0.6 sec, OSB duration 0.8 sec. The target was temporally centered in the OSB. All stimuli (also FBs) were ramped up and down at signal onset and offset (100-ms $\cos^2$ ramps).

Procedure

A three-interval, three-alternative forced-choice procedure with adaptive signal-level adjustment was used to determine the masked threshold of the target. CI users had to indicate which of the intervals contained the signal. A graphical user interface with visual feedback was therefore used. The signal level was adjusted according to a two-down, one-up rule to estimate the 70.7% point of the psychometric function (Levitt, 1971). The initial step size was 8 dB. After every second reversal the step size was halved until a step size of 1 dB was reached. The run was then continued for another four reversals. From the level at these last four reversals, the mean was calculated and used as an estimate of the threshold. The final threshold estimate was taken as the mean over two threshold estimates.

Stimulation Hardware

Streaming of stimuli was achieved using the Nucleus Implant Communicator (NIC) and the Nucleus Matlab Toolbox from Cochlear Ltd. Envelopes were inserted in the frequency-time matrix and processed with the following steps of the Advanced Combination Encoder stimulation strategy of Cochlear with 5 maxima, 900 pulses per channel per s, 25-µs pulse width, monopolar stimulation.

Participants

We included 18 test subjects that were provided with cochlear implants from Cochlear Ltd. unilaterally (N = 12) or bilaterally (N = 6). In case of a bilateral CI user, the ear with better performance was selected for the experiment (based on results in Freiburg monosyllables and Oldenburger Sentence test in steady-state interference). Mean age of participants was 55 yrs ± 15. Cochlear implants were types CI24R, CI24RE, CI422, or CI512. All of them are fully compatible with NIC streaming.

RESULTS

Mean masked detection thresholds are shown in Fig. 3. Across all test subjects, a highly-significant release form masking occurred in the proximate condition (3.2 dB ± 0.8, Wilcoxon signed-rank test: $p = 0.006$) and significant magnitudes in the remote condition (4.2 dB ± 0.3, $p = 0.02$).

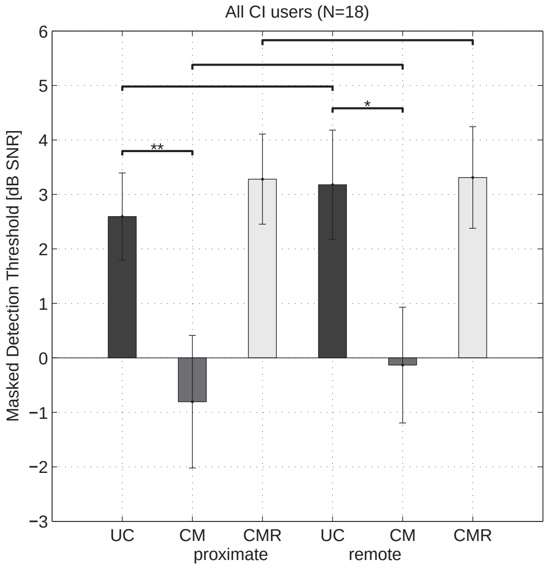


Fig. 3: Mean masked detection thresholds and CMR magnitudes in the proximate (left) and remote (right) condition. Error bars depict one standard error of the mean.

The difference of CMR magnitudes in the proximate vs. remote condition was not significant (Wilcoxon signed-rank test: $p = 0.3$). The same holds for differences of underlying masked detection thresholds in the UC proximate vs. remote condition (p=0.1) and CM proximate vs. remote condition ($p = 0.7$) condition. To deal with the large inter-individual variability across test subjects, a hierarchical cluster analysis into three clusters was calculated. The three clusters revealed groups of CI users that performed very differently. A group of N = 5 test subjects showed no release of masking or even negative values (cluster 1). The majority (N = 9) showed considerable CMR magnitudes of approx. 3 dB (in the proximate as well as in the remote condition – cluster 2). A subgroup of N = 4 CI users showed larger mean CMR magnitudes with better detection thresholds.

DISCUSSION

CMR magnitudes in the proximate flanking-band condition correspond to that found on more apical electrodes in an earlier similar test setup (Zirn *et al.*, 2013). The position of flanking bands (proximate or remote) had minor impact on CMR outcomes (Wilcoxon signed-rank test: p = 0.3). A mainly peripheral explanation for the measured effect as a consequence of masking due to current spread is therefore unlikely. Furthermore, beating between the carrier frequencies of two masker bands cannot occur in constant rate envelope-based electrical stimulation. This finding corroborates our notion that a subset of CI users is able to effectively make comparisons across the stimulation sites. CMR magnitudes were dependent on etiology:

Etiology	**CMR[UC-CM] prox**
Progressive	0.7 dB ± 1.3 (N = 7)
Congenital	4.3 dB ± 1 (N = 6)
Acute hearing loss	5.8 dB (N = 1) (N = 6)
Otitis media	6 dB ± 0.5 (N = 3) (N = 6)
Noise trauma	10 dB (N = 1)

Table 2: Etiologies and corresponding CMR magnitudes

Our hypothesis: Across-electrode processing can be impaired by long-term hearing loss and/or specific etiologies that implicate retro-cochlear impairments.

Results are only valid for test-specific stimuli with direct controlled stimulation. Amplitude comodulation across electrodes is altered by CI signal processing when stimulated acoustically.

An aspect that so far cannot be addressed based on the available data-set is the influence of individual C-levels, dynamic range, or spread of spatial excitation measured with electrically-evoked compound action potentials. The large inter-individual variability of results makes a clear statement in this context difficult. We therefore try to increase the number of test subjects.

ACKNOWLEDGMENTS

This work was supported by Cochlear Ltd., Dr. H. Hessel (IIR-73; CMR in CI users CRDL) and the German Federal Ministry of Education and Research within the Munich Bernstein Center of Computational Neuroscience (reference number 01GQ1004B).

REFERENCES

Bregman, A.S. (**1990**). *Auditory scene analysis: the perceptual organization of sound* (Cambridge, MA: MIT Press).

Epp, B., and Verhey, J.L. (**2009**). "Superposition of masking releases," J. Comput. Neurosci., **26**, 393-407.

Hall, J., Haggard, M., and Fernandes, M. (**1984**). "Detection in noise by spectrotemporal pattern analysis," J. Acoust. Soc. Am., **76**, 50-56.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am., **49**, 467-477.

Moore, B.C.J. (**2012**). *An Introduction to the Psychology of Hearing, 6th ed.* (Cambridge: Emerald Group Pub).

Shannon, R.V. (**1992**). "Temporal modulation transfer functions in patients with cochlear implants," J. Acoust. Soc. Am., **91**, 2156-2164.

Verhey, J.L., Pressnitzer, D., and Winter, I.M. (**2003**). "The psychophysics and physiology of comodulation masking release," Exp. Brain Res., **153**, 405-417.

Zirn, S., Hempel, J.M., Schuster, M., and Hemmert, W. (**2013**). "Comodulation masking release induced by controlled electrical stimulation of auditory nerve fibers," Hear. Res., **296**, 60-66.

Interaural bimodal pitch matching with two-formant vowels

François Guérit[1,*], Josef Chalupper[2], Sébastien Santurette[1],
Iris Arweiler[2] and Torsten Dau[1]

[1] *Centre for Applied Hearing Research, Technical University of Denmark, DK-2800 Lyngby, Denmark*

[2] *Advanced Bionics European Research Center GmbH, D-30625 Hanover, Germany*

For bimodal patients, with a hearing aid (HA) in one ear and a cochlear implant (CI) in the opposite ear, usually a default frequency-to-electrode map is used in the CI. This assumes that the human brain can adapt to interaural place-pitch mismatches. This 'one-size-fits-all' method might be partly responsible for the large variability of individual bimodal benefit. Therefore, knowledge about the location of the electrode array is an important prerequisite for optimum fitting. Theoretically, the electrode location can be determined from CT scans. However, these are often not available in audiological practice. Behavioral pitch matching between the two ears has also been suggested, but has been shown to be tedious and unreliable. Here, an alternative method using two-formant vowels was developed and tested with a vocoder system simulating different CI insertion depths. The hypothesis was that patients may more easily identify vowels than perform a classical pitch-matching task. A spectral shift is inferred by comparing vowel spaces, measured by presenting the first formant in the HA and the second either in the HA or the CI. Results suggest that pitch mismatches can be derived from such vowel spaces. In order to take auditory adaptation in individual patients into account, the method is tested with CI patients with contralateral residual hearing.

INTRODUCTION

In the last years, an increased number of patients having residual contralateral hearing received a cochlear implant (CI). This population is therefore combining the neural excitation coming from the CI and that from the ear stimulated acoustically. However, due to the variability in electrode placement in the cochlea and in cochlear duct length among patients, it is difficult to activate nerve fibers with the same frequency-to-place map as in the contralateral ear. Typically, a standard frequency-to-electrode allocation is used across subjects for the clinical fitting, assuming that the brain can adapt to a mismatch. The evolution of speech perception over time after implantation supports the theory of accommodation to a frequency shift (e.g., Skinner *et al.*, 2002). However, a complete adaptation might not be possible in the case of large mismatches. Rosen *et al.* (1999) showed that even after a long-term training period with a vocoder

*Corresponding author: fguerit@elektro.dtu.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

system simulating a 6.5-mm basalwards shift, speech recognition was worse than for the unshifted condition. More recently, Siciliano *et al.* (2010) used a 6-channel vocoder and presented odd channels in the right ear, shifted 6 mm basally, while keeping the even channels unshifted in the left ear. After 10 hours of training, subjects showed poorer speech perception in this condition than when presented with the three unshifted channels only, suggesting that they did not benefit from combining the mismatched maps.

The above findings suggest that the electrode-array location is crucial for adequate fitting and optimal benefit from the CI. Although electrode location can theoretically be determined from computed-tomography (CT) scans, these are often unavailable in audiological practice and require an additional dose of radiation. For patients having residual hearing in the opposite ear, behavioral pitch-matching has been suggested but is rather difficult because of the different percepts elicited by the implant and the acoustic stimulation. Carlyon *et al.* (2010) also showed that results of behavioral pitch-matching are strongly influenced by nonsensory biases and that the method is tedious and time-consuming. Here, based on the ability to fuse vowel formants across ears (Carlson *et al.*, 1975), an alternative method using two-formant vowels was developed and tested. This method is thought to be clinic-friendly, using stimuli that the CI users are dealing with in their everyday life.

The question addressed in the present study is the following: Can the second formant (F2) of a two-formant vowel be used as a pitch-matching stimulus by presenting it either on the aided/normal-hearing side or on the implanted side? If the implant is perfectly fitted, the perceived vowel distributions should not depend on the ear to which F2 is presented, when fixing the first formant (F1) on the acoustic side. In the presence of an interaural mismatch, vowel distributions should show differences when presenting F2 to the acoustic vs the electric side. To test this hypothesis, an experiment with normal-hearing (NH) listeners using a vocoder system and simulated interaural mismatches was implemented. In order to take auditory adaptation into account, as well as the difficulties regarding the fusion between electric and acoustic percepts, the method was also tested with bimodal (BM) and single-sided-deaf (SSD) CI users.

METHODS

Subjects

Eight NH listeners were tested, all of them native German speakers. Their hearing thresholds were below 20 dB HL at all audiometric frequencies, and the mean age was 25.4 years, ranging from 22 to 30 years.

Eleven implant users were tested in the ENT department of the Unfallkrankenhaus (UKB) in Berlin, and were all native German speakers. Five BM and six SSD implant users took part in the experiment. The mean age was 55.6 years, ranging from 33 to 78 years. The subjects were post-lingually deafened and had a a similar experience with their implant (mean: 19.9 months, SD: 2.1 months). All were equipped with

Advanced Bionics electrode arrays and processors.

Stimuli and equipment

Two-formant vowels were generated using a Matlab-based Klatt synthesizer (Klatt, 1980), and embedded into the consonants /t/ and /k/. The duration of the vowels was slightly longer than normal (≈350 ms) for ease of recognition in CI users. The stimuli were presented at 60 dB SPL. F1 was set at 250 Hz and 400 Hz, and F2 between 600 Hz and 2200 Hz in 200 Hz steps. With these settings, six different German vowels could be elicited when progressively increasing F2 with fixed F1: [u:]/[y:]/[i:] with F1 at 250 Hz and [o:]/[ø:]/[e:] with F1 at 400 Hz.

A monaural (F1 and F2 in the left channel) and a dichotic (F1 in the left and F2 in the right channel) version were created for each stimulus. For the study with NH listeners, the right channel was vocoded using a vocoder mimicking Advanced Bionics CI processing (Litvak *et al.*, 2007). 16-channel noise excitation was used for this vocoder, with noise bands having 25 dB/octave of attenuation. Three different settings were used: 'Voc1' (perfect fitting), 'Voc2' (slight basal shift, ≈ 0.45 octave), and 'Voc3' (larger mismatch, ≈ 0.85 octave). For the NH listeners, Sennheiser HDA 200 headphones were used, ensuring a good interaural attenuation (Brännström and Lantz, 2010). Test procedures were implemented in Matlab and all testings were conducted in a double-walled sound-attenuating listening booth.

For the implant users, the right channel was connected to the implant processor, using the Advanced Bionics Direct Connect® system. Subjects were seated in a booth, and the left channel was connected to a loudspeaker, placed 1 meter to the left or right side of the subjects, to stimulate their non-implanted ear. Subjects indicated their responses orally to the audiologist in charge of the experiment, who was using the custom Matlab-based interface outside the booth.

Procedure for NH listeners

NH subjects were forced to categorize each stimulus using one of six possibilities, chosen to match with the frequency range of the stimuli (Table 1). They could listen to each stimulus up to three times if needed. The different combinations of F1 and F2 resulted in two blocks of 18 stimuli each: a monaural and a dichotic block.

The first part of the test was performed using the monaural stimuli and organized as follows: (1) two repetitions of the stimulus block were presented for training only, (2) five repetitions were recorded ($5 \times 18 = 90$ presentations). All stimuli were presented in a random order, and subjects were aware of the number of remaining presentations.

After this first test, the subjects were trained to fuse stimuli that were non-vocoded on one side and vocoded on the other. This was done by listening to 8 minutes of an audio-book, from which the right channel had been vocoded and the left channel lowpass-filtered at 500 Hz to mimic a typical audiogram of bimodal listeners. Subjects were asked to listen carefully to both sides, with the aim to train them to combine the non-vocoded and vocoded percepts. After this training, nine dichotic sub-tests (three for each vocoder setting, presented in a random order) were administered, following the same protocol as for the monaural test: (1) two repetitions of the dichotic stimulus block were presented for training only, (2) five repetitions of the block were recorded.

Possible choice	TUK	TÜK	TIK	TOK	TÖK	TEK
Phonetic equivalent	[u:]	[y:]	[i:]	[o:]	[ø:]	[e:]
Typical F1 [Hz]	320	301	309	415	393	393
Typical F2 [Hz]	689	1569	1986	683	1388	2010

Table 1: Possible vowel choices for the NH subjects during the categorization task. Phonetic equivalent as well as typical F1 and F2 values (Strange *et al.*, 2004) are indicated. 250 Hz was chosen rather than 300 Hz for F1 when synthesizing the vowels to make sure that subjects would differentiate stimuli having two different F1.

Procedure for implant users

The same categorization task was used, but to reduce the duration of the experiment, only stimuli with F1 at 250 Hz were presented. Accordingly, only 'TUK', 'TÜK', and 'TIK' were possible responses during the task. The experiment was divided into two sub-tests, the first one with the monaural stimulus set, and the second one with the dichotic set. For each sub-test, the stimulus set was repeated twice for training only, and then 10 repetitions were recorded, all stimuli being randomly presented.

RESULTS

NH listeners

Figure 1 shows the vowel categorization results for the 8 NH listeners. In the left panel (A/E), results of the 'monaural' test are plotted. For F1 = 250 Hz as well as for F1 = 400 Hz, changing F2 from 600 Hz to 2200 Hz evokes clearly different vowels: [u:]/[o:] for F2≈800 Hz; [y:]/[ø:] for F2≈1500 Hz; [i:]/[e:] for F2≈2000 Hz. These patterns are consistent with previously reported North-German vowel maps (e.g., Strange *et al.*, 2004).

When presenting F2 to the right ear vocoded without any mismatch ('Voc1'), the three vowel distributions are broader (panels B and F in Fig. 1). This was expected, as the noise-vocoder creates a spread of excitation. However, the distributions still reflect the three different vowels centered at similar values of F2 to without the vocoder. For example, the mid-F2 vowel (*black curve*) has its distribution centered around 1400 Hz ('TÜK') and 1600 Hz ('TÖK') for both conditions.

When simulating a shift with the vocoder ('Voc2' and 'Voc3'), vowel distributions are affected, as seen in panels C, D, G, and H in Fig. 1. The low-F2 vowels (TUK and TOK) progressively disappear. Shifting the vocoder basally assigns channels to

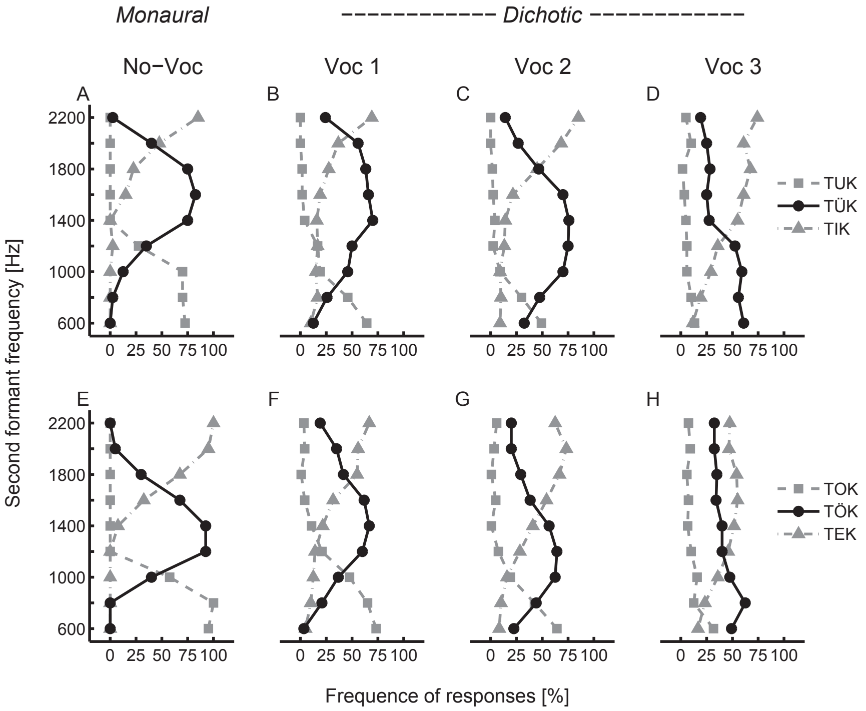


Fig. 1: Mean results (N = 8) of the categorization test for the NH listeners. The number of occurrences (in %) for each vowel is indicated as a function of the frequency of F2. *Top panel:* F1 is fixed at 250 Hz, therefore only the occurrence of the choices TUK, TÜK, and TIK is shown. *Bottom panel:* F1 is fixed at 400 Hz, only the occurrence of the choices TOK, TÖK, and TEK is shown. *Left panel (A/E):* F1 and F2 are presented in the left channel. *Mid-left panel (B/F):* F1 is in the left channel while F2 is in the right channel, processed with an unshifted vocoder. *Mid-right panel (C/G):* F1 is in the left channel while F2 is in the right channel, processed with a slightly shifted vocoder. *Right panel (D/H):* F1 is in the left channel while F2 is in the right channel, processed with a more pronouncedly shifted vocoder.

higher place-frequencies. Therefore, F2 frequencies at 600 Hz in the original signal are shifted, evoking vowels having a higher F2 frequency. In a similar way, the high-F2 vowels (TIK and TEK) are more and more represented, and the mid-F2 vowels (TÜK and TÖK) have their distribution shifted downwards in frequency using this representation.

To assess the simulated shift quantitatively, the F2 distribution of the mid-F2 vowels (categories TÜK and TÖK) are fitted by means of a Gaussian distribution. Fitted center frequencies (mean of the Gaussian distribution) are shown in Fig. 2. The expected center frequencies (dashed gray lines in Fig. 2) are calculated using the

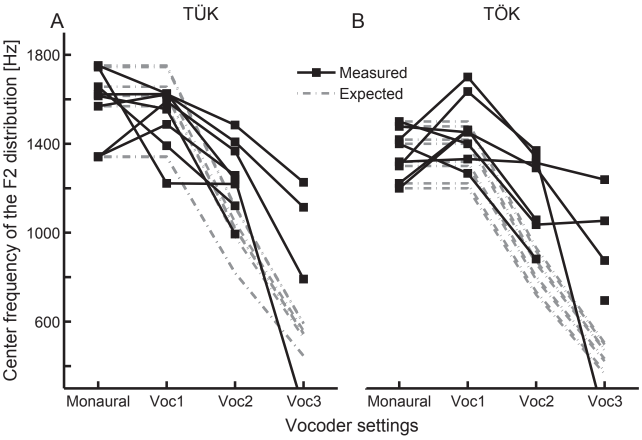


Fig. 2: Fitted center frequencies for individual NH listeners' (N = 8) mid-F2 vowel distributions. (*A*) Fitted center frequencies of the category 'TÜK' (F1 = 250 Hz). (*B*) Fitted center frequencies of the category 'TÖK' (F1 = 400 Hz). For panels (*A*) and *B*), a Gaussian fit was applied for the F2 distribution, and the center is plotted (black squares) for the different conditions. Expected centers for each individual were calculated from the results of the 'monaural' condition and the vocoder settings, and are indicated in dashed gray lines. Center frequencies reaching the frequency limits (100-4000 Hz) of the fitting procedure were removed.

vocoder settings and the fitted center frequency of the 'monaural' condition of each subject. Even though there is a trend of these fitted center frequencies to follow the expected shift from the vocoder, variability is high across subjects, especially for the largest mismatch ('Voc3'). Moreover, for the larger mismatch, some subjects showed a rather flat distribution, indicating a difficulty to fuse the two percepts: No effect of changing F2 indicates that they based their response on F1 only.

CI listeners

Vowel distributions for the monaural condition for both the SSD and BM implant users were very similar to the NH listeners' distributions: the three categories (TUK, TÜK, and TIK) were similarly distributed over the F2 frequency range. An example of one subject's monaural distribution is shown in Fig. 3 (panel A). Assuming that the brain would adapt to mismatches, similar vowel maps would be expected when presenting the second formant either in the implanted or non-implanted ear, as shown for NH listeners in Fig. 1. This was only observed for one of the eleven subjects (panel B in Fig. 3). For the other subjects, various patterns could be observed, and three of them are shown in panels C to E. Some subjects showed a pattern resembling a basal shift (C), others showed a rather flat distribution (D), and one subject even never perceived the mid-F2 vowel (E). This variability was seen for both groups (SSD and BM) and does not imply that these subjects have a mismatch, as discussed later.

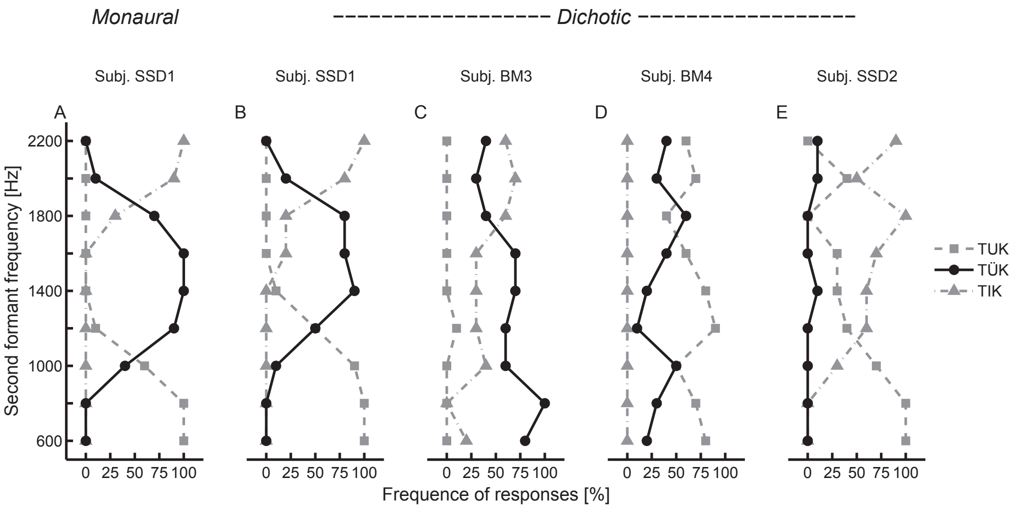


Fig. 3: Five examples of individual CI listeners categorization results. (*A*) 'Monaural' condition results of one subject. The mid-F2 vowel is highlighted in black. (*B*) 'Dichotic' results where the subject has a similar distribution to the 'monaural' condition. (*C*) 'Dichotic' results resembling a basal shift of the electrode array. (*D*) 'Dichotic' results where the subject showed uniform categorization. (*E*) 'Dichotic' results where the subject almost never perceived the mid-F2 vowel.

DISCUSSION AND CONCLUSIONS

NH listeners were able to fuse formants of two-formant vowels when presenting them dichotically with F2 vocoded. Fusion was challenging with the two different percepts, but this was overcome by a careful training and description of the test. The effect of simulating a shift could be seen in the vowel distributions. The low-F2 vowels ([u:] and [o:]) were less represented as the shift was increased. Estimates of the shifts from the mid-F2 vowels ([y:] and [ø:]) were overall smaller than their theoretical value, with high across-subjects variability, and might not represent the best way to estimate a shift. Overall, the NH listeners' results suggest that this new procedure could be a tool to indicate the existence of a mismatch, but that it remains challenging to evaluate this mismatch quantitatively.

Vowel distributions could be derived for all CI users in the monaural acoustic condition, indicating an ability to perform the task reliably. Despite this, large individual differences were observed for dichotic bimodal stimulation, with listeners showing either basal or apical shifts, or generally-poor vowel discrimination. This could be due to the difficulty to fuse percepts more than to possible mismatches. Indeed, for some NH subjects having difficulty to fuse non-vocoded and vocoded percepts, similar distributions could be seen, where the subjects would focus mainly on F1. This was overcome for NH subjects by training them to fuse percepts before categorizing two-formant vowels. Adequate training should be investigated for CI patients in order to obtain vowel distributions based on the fusion of both formants.

CT-scan insertion depth evaluation should be compared to the vowel distributions of the CI patients to look for a possible correlation and shed light on the large variability observed. Moreover, speech perception results using either the CI stimulation only, the non-implanted side only, or both, will be collected for the tested patients. It might be interesting to look at a potential effect of having a dominant ear or a good combination of information across ears. As a general conclusion, the two-formant task is reliable and straight-forward in NH listeners and has potential to detect a mismatch in bimodal CI patients. However, it is difficult to obtain a quantitative estimate of the mismatch with this method and fusion issues should be overcome.

REFERENCES

Brännström, K.J., and Lantz, J. (**2010**). "Interaural attenuation for Sennheiser HDA 200 circumaural earphones", Int. J. Audiol., **49**, 467-471.

Carlson, R., Fant, G., and Granström, B. (**1975**). "Two-formant models, pitch and vowel perception", in *Auditory analysis and perception of speech*. Edited by G. Fant, pp 55-82.

Carlyon, R.P., Macherey, O., Frijns, J.H.M., Axon, P.R., Kalkman, R.K., Boyle, P., Baguley, D.M., Briggs, J., Deeks, J.M., Briaire, J.J., Barreau, X., and Dauman, R. (**2010**). "Pitch comparisons between electrical stimulation of a cochlear implant and acoustic stimuli presented to a normal-hearing contralateral ear", J. Assoc. Res. Oto., **11**, 625-640.

Klatt, D.H. (**1980**). "Software for a cascade/parallel formant synthesizer", J. Acoust. Soc. Am., **67**, 971-995.

Litvak, L.M., Spahr, A.J., Saoji, A.A., and Fridman, G.Y. (**2007**). "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners", J. Acoust. Soc. Am., **122**, 982-991.

Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants", J. Acoust. Soc. Am., **106**, 3629-3636.

Siciliano, C.M., Faulkner, A., Rosen, S., and Mair, K. (**2010**). "Resistance to learning binaurally mismatched frequency to place maps: implications for bilateral stimulation with cochlear implants", J. Acoust. Soc. Am., **127**, 1645-1660.

Skinner, M.W., Ketten, D.R., Holden, L.K., Harding, G.W., Smith, P.G., Gates, G.A., Neely, J.G., Kletzker, G.R., Brunsden, B., and Blocker, B. (**2002**). "CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients", J. Assoc. Res. Oto., **3**, 332-350.

Strange, W., Bohn, O.-S., Trent, S.A., and Nishi, K. (**2004**). "Acoustic and perceptual similarity of North German and American English vowels", J. Acoust. Soc. Am., **115**, 1791-1807.

Assessment, modeling, and compensation of inner and outer hair cell damage

STEFFEN KORTLANG* AND STEPHAN D. EWERT

Medizinische Physik, Universität Oldenburg and Cluster of Excellence 'Hearing4all', Oldenburg, Germany

Reduced temporal fine structure (TFS) sensitivity is proposed to accompany cochlear hearing loss even if audibility and loudness perception are compensated for by hearing aids, or can be present in elderly listeners with unremarkable audiometric thresholds. In both cases, inner hair cell (IHC) damage or neuronal degeneration of subsequent stages can be assumed to play a role. To investigate psychoacoustic measures for assessment of IHC loss, random frequency modulation (FM) detection thresholds in quiet and in background noise were collected for six young normal-hearing (NH) listeners, six older NH listeners, and eleven HI listeners. Two possible detection mechanisms based on phase-locking and amplitude modulation (AM) were assessed in a probabilistic, 'spiking' auditory model [Meddis, J Acoust Soc Am 119, 406 (2006)]. IHC and outer hair cell (OHC) damage were incorporated and adapted to predict the psychoacoustic data. The resulting hearing-impaired (HI) model was then used to simulate the auditory nerve (AN) response in aided conditions with an improved model-based dynamic compression algorithm [based on Ewert and Grimm, ISAAR, 393 (2011)]. Comparison to simulated normal-hearing AN responses revealed partial compensation of OHC damage while IHC damage resulted in supra-threshold 'internal noise' which might contribute to the limited benefit from compensation strategies in hearing aids.

INTRODUCTION

Even if audibility and loudness perception are restored by dynamic compression strategies in hearing aids, supra-threshold processing deficits may persist. Recently, it has been shown that sound exposure can lead to a permanent impairment of auditory-nerve (AN) fibers with low spontaneous rate (LSR) in the absence of elevated audiometric thresholds (Kujawa and Libermann, 2009). Such a degeneration of AN fibers or losses of synaptic elements in the inner hair cells (IHC) might reduce the redundancy of neural coding (Henry and Heinz, 2012), acting as a source of 'internal noise' in the signal representation. Particularly, the usability of temporal fine structure (TFS) information in the signal might be reduced as consequence of IHC damage. TFS sensitivity was shown to decline with hearing loss and age (e.g., Hopkins and Moore, 2011).

As a measure of TFS sensitivity, low-rate frequency modulation (FM) detection thresholds are proposed here and assessed in three different subject groups. FM

*Corresponding author: ....

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

detection requires accurate coding of temporal and spectral cues. However, FM detection tasks are assumed to involve less central or complex stages such as higher-level language processes which are involved in, e.g., speech perception tasks. FM detection therefore appears suited as a measure of early (peripheral) damage in auditory perception. Here it is hypothesized that auditory deafferentiation results in an undersampled representation of the signal at the level of the AN (see also Lopez-Poveda and Barrios, 2013), whereas a loss of outer hair cells (OHC) primarily results in a loss of compression and broader auditory-filter bandwidths. The effects of these two independent impairments are simulated in an FM detection model. By reducing the amount of AN fibers in the model, the neural coding fidelity is diminished ('internal noise'), while filter broadening as a consequence of OHC damage increases the effect of 'external noise' in conditions with noise maskers. The goal of this study is to better understand and to predict perceptual consequences of IHC and OHC damage with regard to FM detection that may also contribute to poor speech-in-noise performance.

RANDOM FREQUENCY MODULATION DETECTION THRESHOLDS

Method

Random frequency modulation detection thresholds (RFMDTs) were examined in six normal-hearing young listeners (NH-Y**)**, six normal-hearing older listeners (NH-O), and eleven hearing-impaired listeners (HI) with sloping sensorineural hearing loss. In an adaptive three-interval, three-alternative, forced-choice (3-AFC) procedure, the FM interval had to be detected against a pure-tone reference at 500 Hz, 2 kHz, and 6 kHz. The random frequency modulation (RFM) tones were generated by imposing a bandpass noise (1-4 Hz) as instantaneous frequency deviation to the pure tone's frequency (f_c). The frequency modulation depth was expressed as root-mean-square (RMS) deviation of the instantaneous frequency from f_c. To reduce amplitude modulation (AM) based detection cues, an additional 1-4-Hz bandpass-noise AM was applied with an RMS modulation depth of −12 dB. After a training run, four threshold runs were performed. To ensure a comfortable level and comparable loudness among all subjects, NH listeners were measured at 65 dB HL, while signals were presented to the HI group at the level of medium loudness (L_{25}) obtained from categorical loudness scaling (CLS; Brand and Hohmann, 2002). To test the impact of external noise on FM detection, Gaussian white noise with 5-ERB bandwidth centred around f_c was added at signal-to-noise ratios (SNRs) of +3 dB and −3 dB in two further conditions. The signals were 500 ms in duration including 25-ms Hann ramps. Further details are provided in Ewert *et al.* (2013).

Results and discussion

Average RFMDTs for the different conditions, test frequencies, and listener groups are shown in Fig. 1. Empirical results are represented by the bars of different shades of grey indicating +/− one standard deviation. Detection thresholds increase with additional external noise and significantly differ between listener groups. In line with literature, RFMDTs at medium loudness were elevated for the NH-O (grey) and HI (dark grey) group (e.g., Strelcyk *et al.*, 2009). However, it is apparent that the differences between the listener groups are less systematic at 6 kHz, where a larger spread per group is observed. This aspect is particularly interesting given that, in this frequency region, audiometric thresholds, loudness growth (characterized by CLS), and filter bandwiths (as estimated in an additional notched noise measurement) differed most between the NH-Y and HI group. Thus, OHC-related processing deficits may play a smaller role and RFMDTs appear to be a meaningful choice to examine IHC-related impairment. In particular, in the absence of elevated audiometric thresholds, altered loudness growth, or increased filter bandwidth for the NH-O listeners, results indicate independent supra-threshold processing disorders reducing the temporal coding fidelity. A three-way repeated-measures analysis of variance (ANOVA) with factors subject group, centre frequency, and condition confirmed that all factors were significant ($p < 0.001$).

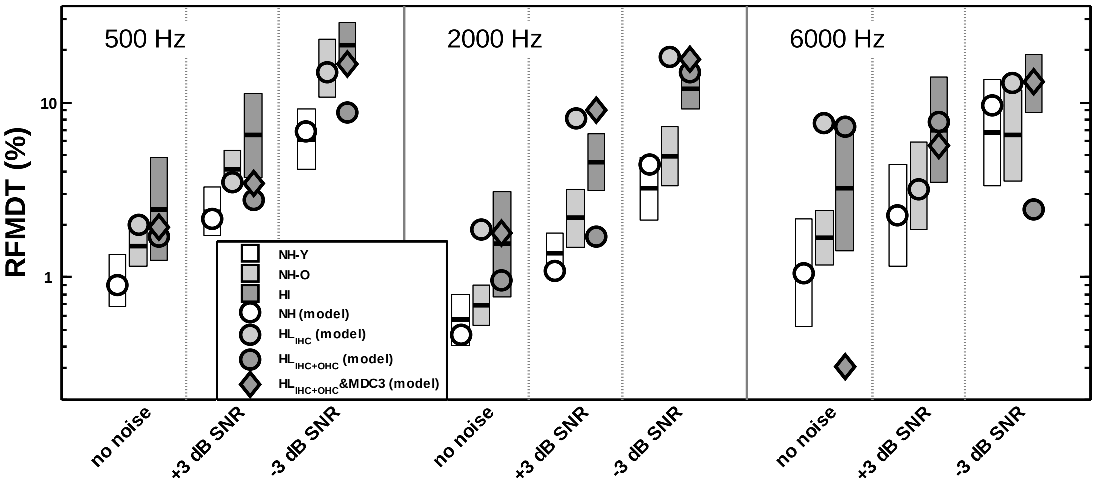


Fig. 1: RFMDTs as a function of test frequency and noise condition. Within each listener group, the thick horizontal line indicates the geometric mean. Model results are indicated by the different symbols (see legend).

MODELING

The RFMDTs presented above can be assumed to rely on two independent detection cues: At small modulation and carrier frequencies, distinctness from a pure sinusoid is thought to rely on gentle fluctuations in the timing of neural spikes (Strelcyk *et al.*, 2009), i.e., TFS cues. At higher rates, FM detection is thought to be based primarily on a place mechanism due to FM-induced AM (‘FM-to-AM conversion’). Both cues were considered here at the level of the AN, including timing (spike phaselocking, TFS) and level (spike density, AM) information.

Model structure

The auditory model (MAP) by Meddis (2006; MAP1_14g release) was used to simulate AN responses. In the model, stimuli are filtered with a linear bandpass to model the outer and middle ear and then subjected to the dual-resonance-non-linear (DRNL) filter (Meddis *et al.*, 2001) accounting for non-linear basilar membrane (BM) processing. The DRNL output is then converted to membrane potentials in the IHC stage, to generate either an AN spiking probability or probabilistic AN spike responses including refractory effects and adaptation. To keep interpretation as simple as possible, feedback paths available in the model (acoustic and medial olivocochlear reflex) were deactivated here. As an example of the MAP model output, Fig. 2 shows the mean AN spiking probability of a high-spontaneous-rate (HSR) fiber as a function of time and characteristic frequency (CF) for the three intervals of the 3-AFC measurement (2 kHz). Here, the target was in the middle interval containing an FM of 5%-RMS modulation depth. It is apparent that the FM introduces additional AM to the AN activity. For illustration purposes, the pattern at the 2-kHz channel is marked with the grey plane.

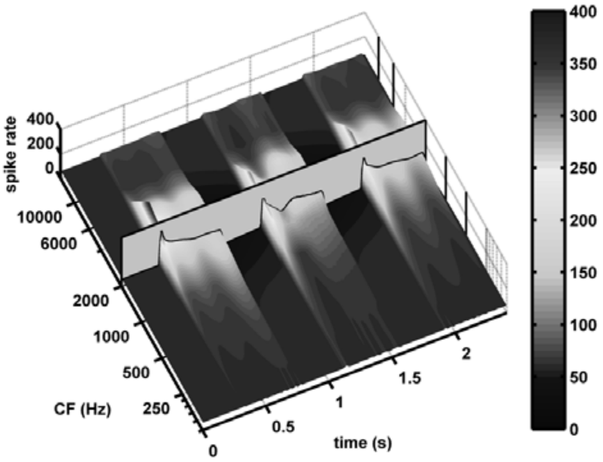


Fig. 2: Mean spike rate (spike probability output of MAP) for a HSR fiber in response to three intervals of a 3-AFC measurement.

In the further modeling, the ‘spiking’ mode was used to generate spike trains at BM characteristic frequencies of 125 Hz up to 16 kHz in octave steps, including the intermediate frequencies (187.5 Hz, 375 Hz, 750 Hz, …). As illustrated in Fig. 3, the FM target and unmodulated reference signal were passed through the model. Spike trains of 100 AN fibers were simulated for each auditory channel. As the more sensitive HSR fibers outnumber the LSR fibers by about 4 to 1 (e.g., Schnupp, 2011), they were separated into 80 HSR and 20 LSR fibers. The spike patterns were cut to match the steady-state part of the signal (450 ms excluding 25 ms on- and offset ramps). For the two above-mentioned detection mechanisms, two independent paths were modeled: i) To account for the AM cue (upper pathway in Fig. 3), post-stimulus time histograms (PSTH) were formed by summing the output of the 100 fibers. These PSTHs form a similar pattern to the spiking probability used in Fig. 2, but underly stochastic fluctuations due to the spiking process. A 4th-order zero-phase bandpass filter (1-8 Hz) was applied to the PSTHs to extract low-rate AM information. Finally, for each BM channel, the AM cue was calculated as the variance of the bandpass-filtered PSTH (PSTH-VAR) and transformed to log domain. The FM tone usually shows higher variance than the pure tone. ii) To calculate the TFS cue (lower pathway in Fig. 3), first-order inter-spike-interval (ISI) histograms were calculated to examine the distribution of the observed times between spikes merged over all AN fibers. Phase-locking to f_c is represented by the local maxima in the distribution separated by one period ($1/f_c$) of the signal. The strength of phase-locking was quantified by the vector strength of the ISI histogram (ISI-VS) to f_c. Here, the FM tone usually shows lower values of ISI-VS caused by smearing of the maxima.

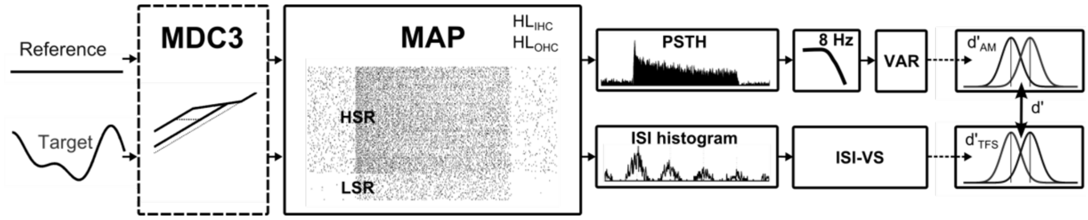


Fig. 3: Block diagram of the model to predict RFMDTs including AM and TFS information extraction in the upper and lower pathway, respectively.

The calculations were repeated N times to estimate mean and standard deviation for PSTH-VAR and ISI-VS, from which two detectabilty measures d'_{AM} and d'_{TFS} (Cohen's *d*) were calculated as RMS over the individual d's in each BM channel. The final detectability measure *d'* was calculated by combination of d'_{AM} and d'_{TFS} using a weighting factor α:

$$d' = \sqrt{\alpha \cdot d'_{AM}{}^{2} + (1-\alpha) \cdot d'_{TFS}{}^{2}} \qquad \text{(Eq. 1)}$$

Method

Different model versions were used to predict the RFMDT data. The NH model (representing group NH-Y) used the standard settings proposed by Meddis (2006) despite the above-mentioned changes with 100 AN fibers for each BM channel. Similar to Lopez-Poveda and Barrios (2013), IHC loss (HL_{IHC}) was modeled as a reduction of AN fibers from 100 to 10 (reducing the spike rate by a factor of 10, equivalent to 20 dB linear attenuation). This reduction of AN fibers results in decreased *d'* values (equivalent to internal noise reducing the acuity of spike pattern analysis). The stimulus level for the two models was 65 dB HL as in the experiment. The HI group was modeled as a combination of IHC and OHC loss ($HL_{IHC+OHC}$). Here, a reduction of gain (gain loss, GL) in the non-linear pathway of the DRNL stage (see, e.g., Jepsen and Dau, 2011) was additionally introduced. GL was estimated for all subjects of the HI group from audiometric thresholds (HL) and the lower slope of the loudness function (m_{low}) derived from the CLS measurement, as suggested in Ewert and Grimm (2011). Finally, GL was averaged across all HI subjects resulting in values ranging from, e.g., 18 dB at 500 Hz, 30 dB at 2 kHz, and 32 dB at 6 kHz. For the HI group simulations, the mean stimulus level of all HI listeners in the FM detection measurement was used (77, 78, and 79 dB HL for 500 Hz, 2 kHz, and 6 kHz, respectively).

RFMDTs were simulated by estimating *d'* values for N = 100 repetitions for all experimental conditions and 20 different RMS FM depths, separated equally on log space between 0.2 and 36%. Third-order polynomial functions were fitted to the *d'* functions in a least-squares sense. For all model versions, a single *d'* threshold per frequency was selected, so that the NH model accounted best for the data of the NH-Y group. A weighting factor of $\alpha = 0.27$ was chosen.

Comparison of predicted thresholds and data

Predicted RFMDTs are shown as the black symbols in Fig. 1. It is apparent that the NH model (white circles) can reproduce the main trends of the NH-Y data. At high frequencies, the $\mathrm{HL}_{\mathrm{IHC}}$ model (light grey circles) tends to overestimate the performance of the $\mathrm{HI}_{\mathrm{IHC}}$ group, showing the need for frequency-dependant IHC loss estimates. Additional OHC damage (dark grey circles) did mainly influence the results at the higher two frequencies and in the presence of external noise. For the noise conditions, broadening of auditory filters as a consequence of OHC damage (gain loss) in the model should reduce the salience of the TFS cue as more (external) noise energy falls into the filter. However, this TFS cue is dominant at low frequencies where the applied gain loss was low and the overall effect of OHC damage was thus low in the model simulations. In some cases, OHC damage led to lower RFMDTs in the model simulations, contradicting data. This might be related to changes in the model's AN pattern for the higher signal levels applied the HI group simulations.

MODEL-BASED DYNAMIC COMPRESSION

The effect of dynamic compression on RFMDTs was assessed using the physiologically-motivated model-based dynamic compression algorithm (MDC3, Fig. 4) which is based on the algorithm of Ewert and Grimm (2011) and Ewert *et al.* (2013). The input signal was analysed in a 4th-order Gammatone filterbank (30 bands, one ERB wide). In each filter channel, the instantaneous level (L_{inst}), the instantaneous frequency (F_{inst}), and a smoothed (50-ms 1st-order lowpass) broad-band, 'long-term' level (L_{lt}), estimated over five adjacent frequency bands, were calculated. A model for BM compression for NH and HI subjects was computed in real-time, based on a combination of L_{lt} and L_{inst}. Off-frequency component suppression was realized using the F_{inst} estimate. The difference between the modeled NH and individual HI BM-I/O function was applied as gain per frequency band. The output signal of the algorithm was generated by a 2nd-order Gammatone resynthesis filterbank including delay compensation between the channels (Hohmann, 2002). In comparison to Ewert *et al.* (2013), the main new stage was the 'HI broadband compensation' (see Fig. 4) prior to resynthesis. It estimates the intensities in the normal and impaired system, taking into account filter widening. A level correction is applied based on the effect of widened filters, resulting in a slightly reduced gain for broadband input signals.

The compressor was fitted for the $\mathrm{HL}_{\mathrm{IHC+OHC}}$ model using the same GL estimates as described above. Reference and target signals with an input level of 65 dB HL were processed. Model results for simulated aided RFMDTs are shown in Fig. 1 (dark grey diamonds). In some of the external-noise conditions dynamic compression led to higher RFMDTs, most likely due to reduction of AM cues. Overall the results are not clear-cut, however, it is obvious that dynamic compression is not suited to significantly improve thresholds of the $\mathrm{HL}_{\mathrm{IHC+OHC}}$ model compared to the unaided condition (dark grey circles).

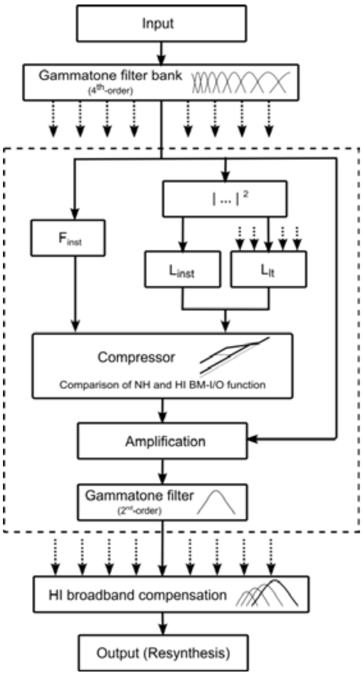


Fig. 4: Block diagram of the model-based dynamic compression algorithm MDC3. Details are described in the text.

GENERAL DISCUSSION

The proposed model for the simulation of RFMDTs is able to reproduce NH data within good accuracy. Neither the model of $\mathrm{HL}_{\mathrm{OHC}}$, nor dynamic compression (as can be found in most hearing aids) showed consistent effects in the model simulations. Reduced TFS sensitivity was empirically found in both HI and elderly NH listeners. This could be accounted for by modeled IHC damage which resulted in supra-threshold ‘internal noise’. The effect of increased external noise on the AN representation as a consequence of broadened filters in case of OHC damage could not be mimicked by the model. A possible confound comes from ‘unrealistic’ changes in the model’s AN representation depending on the signal level. It is obvious that the salience of both cues available in the model, TFS and AM, could not be improved by dynamic compression, which might explain limited benefit of hearing aids when audibility is not the sole problem. Taken together, a first promising step towards a framework for (aided) performance prediction based on stochastic AN responses with individually adjustable IHC and OHC loss was suggested. While the model is able to mimick supra-threshold processing deficits in HI listeners in the TFS and AM domain, further work is required to, e.g., determine frequency-dependent IHC-loss estimates and to assess the effect of signal level on the model's AN representation.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research BMBF, research unit 'Modellbasierte Hörgeräte' (13EZ1127D), and the German research funding organization DFG (SFB TRR 31).

REFERENCES

Brand, T., and Hohmann, V. (**2002**). "An adaptive procedure for categorical loudness scaling," J. Acoust. Soc. Am., **112**, 1597-1604.

Ewert, S.D., and Grimm, G. (**2011**). "Model-based hearing aid gain prescription rule" in Proceedings of ISAAR 2011: *Speech Perception and Auditory Disorders*. Nyborg, Denmark, pp. 393-400.

Ewert, S.D., Kortlang, S., and Hohmann, V. (**2013**). "A Model-based hearing aid: Psychoacoustics, models and algorithms," ASA, ICA 2013 Montréal, **19**, 1.

Henry, K., and Heinz, M.G. (**2012**). "Diminished temporal coding with sensorineur-al hearing loss emerges in background noise," Nat. Neurosci., **15**, 1362-1364.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a Gammatone filterbank," Acta Acustica, **88**, 433-442.

Hopkins, K., and Moore, B.C.J. (**2011**). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech pereception in noise," J. Acoust. Soc. Am., **130**, 334-349.

Jepsen, M., and Dau, T. (**2011**). "Characterizing auditory processing and perception in individual listeners with sensorineural hearing loss," J. Acoust. Soc. Am., **129**, 262-281.

Kujawa, S.G., and Liberman, M.C. (**2009**). "Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss," J. Neurosci., **29**, 14077-14085.

Kortlang, S., Mauermann, M., Kollmeier, B., and Ewert, S.D. (**2012**). "Characterization of IHC loss and its relevance to hearing aid gain prescription rules," Poster at IHCON, Lake Tahoe, USA.

Lopez-Poveda, E.A., and Barrios, P. (**2013**). "Perception of stochastically undersampled sound waveforms: A model of auditory deafferentiation," Front. Neurosci., **7**, 124.

Meddis, R. (**2001**). "A computational algorithm for computing auditory frequency selectivity," J. Acoust. Soc. Am., **109**, 2852-2861.

Meddis, R. (**2006**). "Auditory-nerve first-spike latency and auditory absolute threshold: a computer model," J. Acoust. Soc. Am., **119**, 406-417.

Schnupp, J., Nelken, I., and King, A. (**2011**). *Auditory Neuroscience – Making Sense of Sound*, MIT Press.

Strelyck, O., and Dau, T. (**2009**). "Relations between frequency selectivity, temporal fine-strucutre processing and speech reception in impaired hearing," J. Acoust. Soc. Am., **125**, 3328-3345.

A simplified measurement method of TMTF for hearing-impaired listeners

TAKASHI MORIMOTO[1,*], TAKESHI NAKAICHI[1], KOUTA HARADA[1], YASUHIDE OKAMOTO[2], AYAKO KANNO[2], SHO KANZAKI[3], AND KAORU OGAWA[3]

[1] *RION Co. Ltd., Tokyo, Japan*

[2] *Inagi Municipal Hospital, Tokyo, Japan*

[3] *Keio University Hospital, Tokyo, Japan*

It is difficult to understand speech for listeners with reduced temporal resolution. To measure the auditory index of temporal resolution in clinical diagnosis, a novel measurement method was proposed. It is called a simplified measurement method of temporal modulation transfer function (S-TMTF). This method is based on measurement of temporal modulation transfer function (TMTF). The novelty of S-TMTF lies in the use of only two thresholds for estimation of peak sensitivity and 3-dB cutoff frequency. One is a threshold of modulation depth and the other is a threshold of modulation frequency. In this study, to evaluate the practicability and accuracy of peak sensitivity and 3-dB cutoff frequency, both S-TMTF and TMTF were measured for normal-hearing and hearing-impaired subjects. Results of S-TMTF were significantly correlated with that of TMTF and the measurement time of S-TMTF could be shortened to one fourth of the time for TMTF. Furthermore, the measurement time will be shortened by using the method of limits. S-TMTF would be applied for clinical diagnosis of hearing impairment.

INTRODUCTION

It is well known that temporal resolution is reduced for hearing-impaired listeners. Narne and Vanaja (2009) said that it is difficult to understand speech for listeners with reduced temporal resolution. To measure the auditory index of temporal resolution, gap detection threshold (GDT) and temporal modulation transfer function (TMTF) are often used in psychoacoustical experiments (Shen and Richards, 2013).

GDT is the threshold of time for detecting a silent interval embedded between two noise bursts. Penner (1977) reported that, for normal-hearing subjects, the threshold of time is usually approximately 3 ms but that it is larger for the hearing impaired. TMTF is the threshold of modulation depth as a function of modulation frequency. Usually, seven thresholds of modulation depth are measured for estimation of two parameters. TMTF can express sensitivity to modulation depth and detection ability of fast modulation frequency (Formby and Muir, 1988; Eddins, 1993). These abilities

*Corresponding author: t-morimoto@rion.co.jp

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

are called the peak sensitivity and 3-dB cutoff frequency, respectively. Bacon and Viemeister (1985) reported that the thresholds of modulation depth for hearing-impaired listeners were increased more than for normal-hearing listeners. If TMTF is measured in clinical diagnosis, the auditory characteristic for hearing-impaired might be described more precisely. Furthermore, this information might denote hearing-aid fitting parameters or be applied for new hearing prostheses. Therefore, applying TMTF in clinical diagnosis is desired. It is difficult to apply for clinical diagnosis since the measurement of TMTF is more time consuming than that of GDT.

In this paper, for shortened measurement time, a simplified measurement method of TMTF (S-TMTF) is proposed. This method needs two measurement points, one is a threshold of modulation depth at a lower modulation frequency, and the other is a threshold of modulation frequency for which a signal with sufficient modulation depth is used. For investigation of practicability and accuracy of S-TMTF, the measurement time and two parameters of S-TMTF were compared with TMTF for normal-hearing and hearing-impaired subjects.

OUTLINE OF CONVENTIONAL TMTF

The measurement of TMTF used a sinusoidal-amplitude-modulated broadband noise. Modulation depth thresholds were measured as a function of modulation frequency. Figure 1 shows experimental data of TMTF for normal-hearing and hearing-impaired listeners. The ordinate indicates modulation depth thresholds and the abscissa shows modulation frequency. The modulation depth threshold is often measured using modulation frequencies of 8, 16, 32, 64, 128, 256, 512 Hz (Eddins, 1993; Shen and Richards, 2013). Thresholds are expressed in decibels as $20\log_{10}(m)$, where m is the modulation-depth parameter. When the modulation-depth parameter m is 1.0, the signal has a modulation depth of 100%, i.e., the modulation depth is expressed as 0 dB, the carrier level falls to zero and rises to twice its non-modulated level. Additionally, when m is 0.5 and 0.1, the modulation depths is expressed as -6 dB (50%) and -20 dB (10%), respectively. Modulation-depth thresholds are fairly constant from a 8-Hz modulation frequency to about 50 Hz. Beyond 50-Hz modulation frequency, the threshold increases more slowly at a rate of about 3 dB per octave of modulation frequency (Bacon and Viemeister, 1985). For a hearing-impaired listener, modulation-depth thresholds increase less than that of normal-hearing listeners, as shown in Fig. 1. This difference expresses a degraded ability of temporal resolution (Zeng *et al.*, 1999). These thresholds are very well fitted with a first-order Butterworth filter. Formby and Muir (1988) and Eddins (1993) modeled the TMTF using a function of the form

$$\phi(f_m) = L_{ps} - 10\log_{10}(1/(1+(f_m/f_c)^2)) \qquad \text{(Eq. 1)}$$

where $\phi(f_m)$ denotes the modeled TMTF, L_{ps} denotes peak sensitivity (dB, 20 log m), f_c denotes the 3-dB cutoff frequency, f_m denotes modulation frequency. For normal-hearing listeners, typically L_{ps} is approximately -24 dB and f_c is approximately 140 Hz (Shen and Richards, 2013).

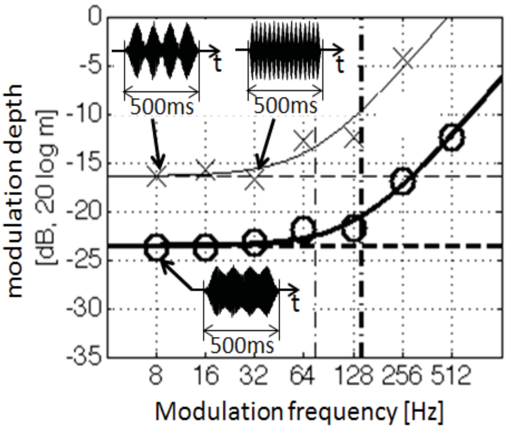


Fig. 1: Example of TMTF. The ordinate denotes modulation depth and the abscissa denotes modulation frequency. The circles and heavy lines denote the results of normal-hearing listeners and the x-marks and narrow lines denote the results of hearing-impaired listeners. Solid lines denote the model of TMTF, dashed lines denote L_{ps}, and chain lines denote f_c. The waveform of the sinusoidal-amplitude-modulated noise is indicated for different conditions.

A PROPOSED SIMPLIFIED MEASUREMENT METHOD OF TMTF

For shortened measurement time, a simplified measurement method of measurement of TMTF (S-TMTF) is proposed. S-TMTF is based on the conventional method of TMTF. The characteristic of S-TMTF lies in the use of only two thresholds for estimating both L_{ps} and f_c. One is a modulation-depth threshold at a lower modulation frequency and the other is a modulation-frequency threshold at fixed modulation depth. This method is fast and is almost as accurate as a conventional TMTF measurement. A procedure of S-TMTF is shown as follows:

Step 1 : A modulation depth threshold $\phi(\alpha)$ is measured, where α is an arbitrary modulation-frequency value. Usually, α is a lower modulation frequency (e.g., 8 Hz). The measured $\phi(\alpha)$ is considered to be L_{ps} in S-TMTF.

Step 2 : Modulation frequency f_{m1} is measured at the modulation depth which is set at $\phi(\alpha)+\beta$, where β is an arbitrary value. The value of $\phi(\alpha)+\beta$ should be set at an audible modulation depth at a lower modulation frequency.

Step 3 : f_c is estimated from substituting L_{ps} and f_{m1} into Eq. 2.

$$f_c = f_{m1}(10^{-L_{ps}/20} - 1)^{-1/2} \qquad \text{(Eq. 2)}$$

				Frequency (Hz)						
Sub.	Age	Sex	Ear	125	250	500	1000	2000	4000	8000
							(dB HL)			
HI-A	82	F	L	45	45	45	55	65	70	80
HI-B	66	F	L	70	85	80	75	80	75	85
HI-C	68	F	R	50	60	55	55	60	65	75
HI-D	72	F	R	65	70	70	75	80	75	100
HI-E	84	F	L	60	65	60	55	60	75	85
HI-F	71	F	R	60	55	55	55	65	55	85
HI-G	79	M	R	50	65	55	40	55	55	75
HI-H	85	F	R	60	60	55	60	65	75	85
HI-I	86	F	L	70	70	70	65	75	70	85
HI-J	83	M	L	70	70	65	65	70	65	70
HI-K	82	M	R	35	45	50	80	95	90	80
HI-L	79	F	R	30	40	50	55	60	65	75
HI-M	75	M	L	55	45	45	55	70	70	85

Table 1: Profiles and hearing thresholds for tested ears of individual hearing-impaired subjects.

EXPERIMENT

L_{ps} and f_c were measured by both TMTF and S-TMTF. Results were compared and the measurement practicability and accuracy of S-TMTF was evaluated.

Subjects

16 normal-hearing subjects and 13 hearing-impaired subjects participated. The age of normal-hearing subjects ranged from 25 to 43 years. All subjects had hearing thresholds better than 15 dB HL at all audiometric frequencies in the tested ear. The age of hearing-impaired subjects ranged from 66 to 86 years. Table 1 shows the individual profiles and hearing thresholds for hearing-impaired subjects.

Stimuli and equipment

A broadband noise (20-14000 Hz) was generated and controlled digitally. The duration of the noise was 500 ms, including 2.5 ms rise/fall cosine ramps. The noise was presented from a personal computer with a 16-bit digital-to-analog converter (Roland QUAD-CAPTURE) to the subject's tested ear via supra-aural headphones. Sony MDR-V6 headphones and Sennheiser HD 380PRO headphones were used for normal-hearing and hearing-impaired listeners, respectively. The presented level was fixed at 60 dB SPL for normal-hearing and fixed at 20 dB SL for hearing-impaired subjects.

Procedure of modulation-depth threshold for S-TMTF and TMTF

For S-TMTF, the modulation-depth threshold was measured at 8 Hz, i.e., the value of α was set at 8 Hz. For TMTF, the thresholds of modulation depth were measured at seven modulation frequencies. These frequencies were set at 8, 16, 32, 64, 128, 256, and 512 Hz. In this study, the modulation-depth threshold at 8 Hz was reused as the threshold in the S-TMTF method. Detection thresholds were obtained using an adaptive, three-interval, three-alternative, forced-choice procedure (3I, 3AFC), with a two-down and one-up rule tracking the 70.7% point on the psychometric function (Levitt, 1971). Listeners did not receive any feedback concerning the correct interval after each trial. To shorten measurement time, a detection-threshold task was carried out once, i.e., there was no repeated measurement. For modulation-depth thresholds for S-TMTF and TMTF, the modulation depth was started at 0 dB. Twelve reversals were obtained in a given task, and the threshold estimation for the task was taken as the mean value of the last eight reversals. The step-size was set to 4 dB at the first four reversals and 2 dB thereafter.

Procedure of modulation-frequency threshold for S-TMTF

The modulation-frequency threshold was measured at modulation depth $\phi(\alpha)+\beta$, where β was the absolute value of $L_{ps}/2$. Detection thresholds were obtained with the same procedure as the modulation-depth threshold. The modulation frequency was started at 8 Hz. The step-size was set to 2 octaves at the first four turnarounds and 1 octave thereafter.

RESULTS

The measurement time of S-TMTF was approximately 10 minutes and the measurement time of TMTF was approximately 40 minutes for each subject. The measurement time of S-TMTF was thus shortened to one-fourth of the time of TMTF. Figure 2(a) shows the correlation diagram of f_c and (b) shows L_{ps} estimated from S-TMTF and TMTF for normal-hearing and hearing-impaired subjects. The ordinate denotes values estimated from S-TMTF and the abscissa shows values estimated from TMTF. Correlation coefficients for f_c and L_{ps} were 0.89 ($p < 0.01$) and 0.92 ($p < 0.01$), respectively. Both parameters estimated from S-TMTF were significantly correlated with TMTF. For f_c, differences between S-TMTF and TMTF were within the smaller step-size for all subjects. For L_{ps}, differences between S-TMTF and TMTF were within the smaller step-size for 25 out of 29 subjects.

DISCUSSION

The measurement time of S-TMTF was 10 minutes and was shortened to one-fourth of the time of TMTF for normal-hearing and hearing-impaired subjects. This result shows that it is possible to measure in clinical diagnosis. However, there are other clinical measurements such as an audiogram and speech-intelligibility tests at the same consultation session. It is desirable to shorten the measurement time of S-TMTF. In a

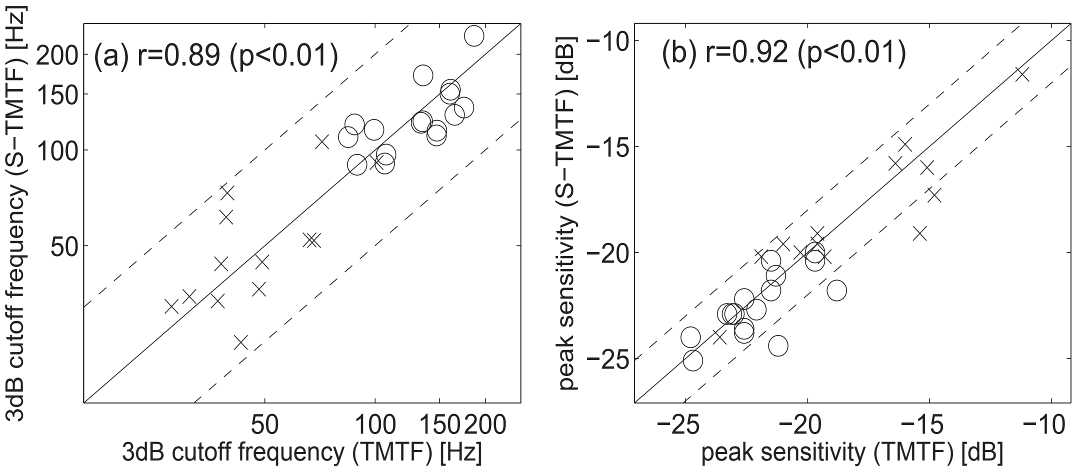


Fig. 2: Correlation diagrams between S-TMTF and TMTF for normal-hearing and hearing-impaired subjects. (a) Correlation diagram of f_c. (b) Correlation diagram of L_{ps}. The ordinate denotes values estimated from S-TMTF and the abscissa denotes values estimated from TMTF. The circles denote results of normal-hearing subjects and the x-marks denote results of hearing-impaired subjects. Dashed lines denote smaller step-size ranges for each measurement.

further investigation, detection thresholds will be obtained using the method of limits rather than the 3AFC for shortened measurement time.

The f_c and L_{ps} estimated from S-TMTF were significantly and highly correlated with TMTF. The accuracy of S-TMTF was confirmed because the differences in f_c between S-TMTF and TMTF were within the smaller step-size for all subjects, and the differences in L_{ps} were within the smaller step-size for 25 subjects out of 29 subjects (86%). On the other hand, differences in L_{ps} between the two methods were out of the smaller step-size range for 4 subjects out of 29 (2 normal-hearing and 2 hearing-impaired subjects). On results of remeasurement, differences were within the smaller step-size for the mentioned subjects with normal hearing. The thresholds may need to be measured twice or more to determine more accurate results, as reported by Bacon and Viemeister (1985). The sufficient performance, however, was obtained from only one determination, as mentioned above.

SUMMARY

To measure the auditory index of temporal resolution in clinical diagnosis, a simplified measurement method of temporal modulation transfer function (S-TMTF) was proposed. S-TMTF needs only two thresholds, one is a threshold of modulation depth at 8 Hz and the other is a threshold of modulation frequency for which a signal with sufficient modulation depth was used. To evaluate the measurement practicability and accuracy of S-TMTF, f_c and L_{ps} were measured using S-TMTF and TMTF. The results showed that (1) the measurement time of S-TMTF was 10 minutes which is one-fourth of the time of TMTF, (2) two parameters estimated from S-TMTF were significantly correlated with TMTF for normal-hearing and hearing-impaired subjects. S-TMTF was fast and was almost as accurate as conventional TMTF. S-TMTF would be applied for clinical diagnosis of hearing impairment.

REFERENCES

Bacon, S.P., and Viemeister, N.K. (**1985**). "Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners," Audiology, **24**, 117-134.

Eddins, D.A. (**1993**). "Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region," J. Acoust. Soc. Am., **93**, 470-479.

Formby, C., and Muir, K. (**1988**). "Modulation and gap detection for broadband and filtered noise signals," J. Acoust. Soc. Am., **84**, 545-550.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am., **49**, 467-477.

Narne, V.K., and Vanaja, C.S. (**2009**). "Preception of speech with envelope enhancement in indivduals with auditory neuropathy and simulated loss of temporal modulation processing," Int. J. Audiol., **48**, 700-707.

Penner, M.J. (**1977**). "Detection of temporal gaps in noise as a measure of the decay of auditory sensation," J. Acoust. Soc. Am., **61**, 552-557.

Shen, Y., and Richards, V.M. (**2013**). "Temporal modulation transfer function for efficient assessment of auditory temporal resolution," J. Acoust. Soc. Am., **133**, 1031-1041.

Zeng, F.G., Oba, S., Grade, S., Sininger, Y., and Starr, A. (**1999**). "Temporal and speech processing deficits in auditory neuropathy," Neuroreport, **10**, 3429-3435.

Dichotic listening: A predictor of speech-in-noise perception in older hearing-impaired adults?

LIMOR LAVIE*, KAREN BANAI, AND JOSEPH ATTIAS

Department of Communication Sciences and Disorders, University of Haifa, Israel

The objective of the study was to examine the relations between two auditory processes, dichotic listening and speech perception in noise. Both involve listening to competing signals and significantly decline with age. Dichotic listening and speech identification in multitalker noise were tested in 36 elderly participants with symmetric mild-to-moderate hearing loss. High negative correlations between the SNR levels in which 50% and 30% of the words were correctly identified and the dichotic scores were found. These correlations were attributed to the dichotic score in the non-dominant ear. Our data suggest that dichotic listening, a major processing deficit in hearing-impaired older adults, could potentially serve as a reliable predictor of speech-in-noise perception in this population.

INTRODUCTION

The most common complaint of elderly hearing-impaired individuals is the difficulties in understanding speech in the presence of background noise. These difficulties are more prominent in the presence of competing speech, of either one speaker or (to a larger extent) in the presence of multi-talker babble noise (Divenyi and Haupt, 1997; Schneider and Pichora-Fuller, 2001). These have been attributed to age-related peripheral hearing loss (Humes and Roberts, 1990; Humes, 1996; Killion, 1997), as well as age-related cognitive decline (e.g., Gordon-Salant and Fitzgibbons, 1993; Pichora-Fuller *et al.*, 1995; Martin and Jerger, 2005; Schneider *et al.*, 2005; Tun *et al.*, 2002; Wingfield *et al.*, 2005; Humes *et al.*, 2007), changes in central auditory processes (e.g., Schneider *et al.*, 1994; Frisina and Frisina, 1997; Divenyi and Haupt, 1997; Strouse *et al.*, 1998; Snell and Frisina, 2000; Frisina, 2001; Gordon-Salant and Fitzgibbons, 2001; Schneider and Pichora-Fuller, 2001; Divenyi *et al.*, 2005; Martin and Jerger, 2005), or the combination of cognitive, central, and peripheral causes (Martin and Jerger, 2005).

Dichotic listening, like speech perception in multi-talker noise, is a challenging listening situation because listeners are required to cope with competing speech signals. Studies on dichotic listening provide evidence for age-related changes in central auditory processing. An overall decline in dichotic scores was reported, together with enlarged right-ear advantage (REA) for speech signals due to large reduction in the left-ear dichotic scores (left-ear deficit, LED) (Jerger *et al.*, 1994; Jerger *et al.*, 1995; Noffsinger *et al.*, 1996; Wilson and Jaffe, 1996; Strouse and Wilson, 1999; Strouse *et al.*, 2000; Hallgren *et al.*, 2001; Roup *et al.*, 2006). In

*Corresponding author: lavielimor@gmail.com

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

addition, Dos-Santos *et al.* (2008a,b) showed a decline in REA during dichotic listening in the presence of noise due to higher dichotic scores in the left ear and lower right-ear scores.

It was claimed that the substantial LED in verbal dichotic tasks and the significant right-ear deficit (RED) in non-verbal tasks may have a considerable impact on older adults' ability to use binaural information effectively, including the information which is used for speech identification in noise (Jerger *et al.*, 1995; Strouse-Carter *et al.*, 2001). It was further suggested that dichotic listening and speech perception in noise may be related to each other, since both involve listening to competing signals (Martin and Jerger, 2005), and both decline with age. Moreover, Givens *et al.* (1998) reported a significant correlation between dichotic listening scores and hearing-aid satisfaction. Thus, the aim of the current study was to examine the relations between these auditory processes, and to investigate whether the perceptual difficulties of older hearing-impaired adults in complex listening environments can be predicted using a relatively simple dichotic listening test.

METHODS

Participants: A group of 36 participants who never used hearing aids, ages 64-88 years (mean age in years ± s.d., 76.3 ± 5.9; median, 77 years), 20 men and 16 women, were recruited from an audiology clinic. All participants had a symmetric sensory hearing loss of 30-70 dB at 0.5-4 kHz, with flat or mild-moderate slope audiograms and symmetric speech-recognition scores (PB-50, 86.22 ± 11.64, 85.44 ± 10.77, right ear, left ear, respectively). They were cognitively fit (mini mental state examination, 27.9 ± 1.4; inclusion score: ≥ 24; digit span standard score, 9.4 ± 2.2; inclusion score: ≥ 6). 35 participants were right-handed and one participant left-handed, as tested with the Edinburgh Dexterity questionnaire (Oldfield, 1971).

The participants underwent dichotic monosyllabic words test and tests of speech identification in noise.

Dichotic tests: 6 dichotic lists, each with 25 pairs of phonetically-balanced monosyllabic words, adopted from the Hebrew speech-discrimination test (PB-50), were recorded and digitally normalized for length and intensity using Nuendo 3.2.0 audio software. The test stimuli were presented at each participant's most comfortable level (MCL) through calibrated TDH 39 earphones and a MAICO MA 52 audiometer in a sound-proof room, such that one word of each pair was presented to the right ear, while the other word was presented simultaneously to the left ear. Each pair of dichotic words was preceded with a carrier phrase which was played simultaneously to both ears: "please repeat…". A four-second silent interval was inserted after every dichotic pair to enable the participants to repeat the test words and the experimenter to write them down. Each correctly identified word was scored 4%, and the dichotic score was the sum of the correct scores in each ear. In addition, the total dichotic score was calculated as the sum scores in the two ears.

Speech in noise: 26 lists of 20 bi-syllabic recorded words, based on the Hebrew SRT word lists were used together with multi-talker babble noise which was comprised of

4 Hebrew speakers (2 men and 2 women), all recorded, normalized, and mixed using Nuendo 3.2.0 audio software. The test words were presented at each participant's individual MCL from a loudspeaker located one meter in front of the listener (azimuth 0°) and the babble noise were presented simultaneously from two loudspeakers located one meter from the listener at azimuths +45° and −45°. The word lists and babble noise were presented using MAICO MA 52 audiometer in descending signal-to-noise ratios (SNRs). SNRs were adjusted by changing the level of the noise while keeping the level of the words constant. Levels of SNRs for which each participant recognized 30% and 50% of the words were eventually identified. The 50% level is reported because it is commonly used for threshold estimation (e.g., HINT, Nilsson *et al.*, 1994; QuickSIN, Killion *et al.*, 2004). The 30% level was selected because the average SNR at which this level of performance was achieved in the pre-test (−0.44 dB) represents an SNR for common daily environments (e.g., subway or aircraft, Schneider *et al.*, 2002).

RESULTS

Dichotic listening: Very low dichotic scores were found in both ears. The average scores were 58.8% ± 17.9 in the dominant ear and 37.8% ± 19.7 in the non-dominant ear, thus the average REA was 21%.

Speech identification in noise: the average SNR level, required to reach the 50% level of word identification, was +1.25 dB ± 2.8, and the average SNR level required to reach the 30% level of word identification was −0.44 dB ± 2.3.

Both tests were characterized by large variability among the participants: The dichotic scores ranged from 16% to 88% in the dominant ear, and from 4% to 76% in the non-dominant ear; in the speech identification in noise tests there were differences of up to 10 dB in SNR levels between the participants.

We calculated the Pearson correlations between dichotic performance and achievements in speech identification in noise. High negative correlations were observed between the total dichotic score and the SNR levels at which 50% and 30% of the words were correctly identified ($r = -0.710$, $p < 0.001$ at both levels, see Fig. 1), indicating that in general, listeners with better dichotic scores tended to have better speech identification in noise and vice versa.

In addition, high negative correlation was found between the non-dominant ears' dichotic scores and both SNR levels at which 50% ($r = -0.707$; $p < 0.001$) and 30% ($r = -0.708$, $p < 0.001$) of the words were correctly identified (Fig. 2). Similar correlations were found for the dominant ear ($r = -0.597$, $p < 0.001$ and −0.596, $p < 0.001$, respectively). Nevertheless, when regression models were used to predict the SNR levels required to achieve 50% or 30% word identification, the dichotic score in the non-dominant ear was a significant predictor (50%: $\beta = -.56$, $t = -3.36$, $p = .002$; 30%: $\beta = -.57$, $t = -3.38$, $p = .002$), but those in the dominant ear had no additional contribution (50%: $\beta = .20$, $t = -1.19$, $p = .24$; 30%: $\beta = -.16$, $t = -1.17$, $p = .25$). This latter analysis suggests that the correlations between the total dichotic scores and speech-in-noise identification may result from disruption of dichotic listening and not simply from reduced speech perception in both ears under competing signal conditions.

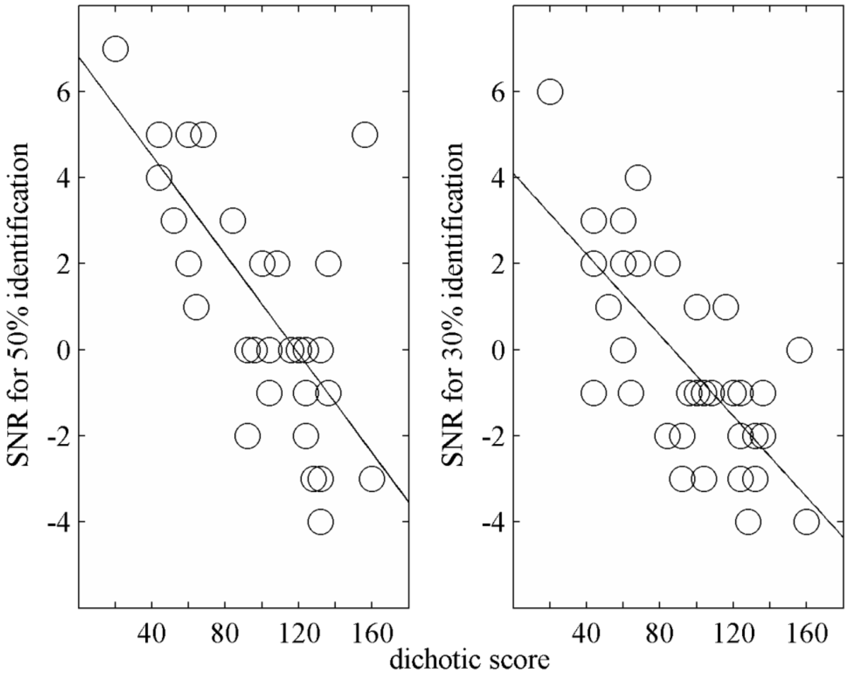


Fig. 1: The relationship between speech identification in noise (50% identification level on the left panel, 30% identification on the right panel) and the total dichotic score. Individual data are shown in circles. Lines show the linear fit between SNRs and dichotic scores.

DISCUSSION

Low achievements were observed in the dichotic listening test, with extremely low scores in the non-dominant ears (LED), making an average REA of 21%. These results are in line with previous studies (e.g., Strouse *et al.*, 2000; Hallgren *et al.*, 2001; Roup *et al.*, 2006), demonstrating the deficits in processing dichotic words in older adults, as opposed to young normal-hearing listeners that typically score 90%-100% in dichotic listening tests with minimal discrepancies (about 2%) between the ears (Roup *et al.*, 2006).

In the speech-in-noise tests high SNR levels were required to identify both 50% and 30% of the words, demonstrating the difficulties hearing-impaired older adults may face in common daily environments (Frisina and Frisina, 1997; Gordon-Salant, 2005; Martin and Jerger, 2005).

There was a large variability in the dichotic scores and in speech identification in noise. Inter-subject variability is one of the main characteristics of the elderly population in general, and particularly of hearing-impaired older adults. This variance is apparent in various hearing and auditory-processing tests and has a major role in the differences among individuals in hearing-aid acclimatization and satisfaction (Humes and Nelson, 1991; Gordon-Salant and Sherlock, 1992; Humes *et al.*, 1994; Roup *et al.*, 2006).

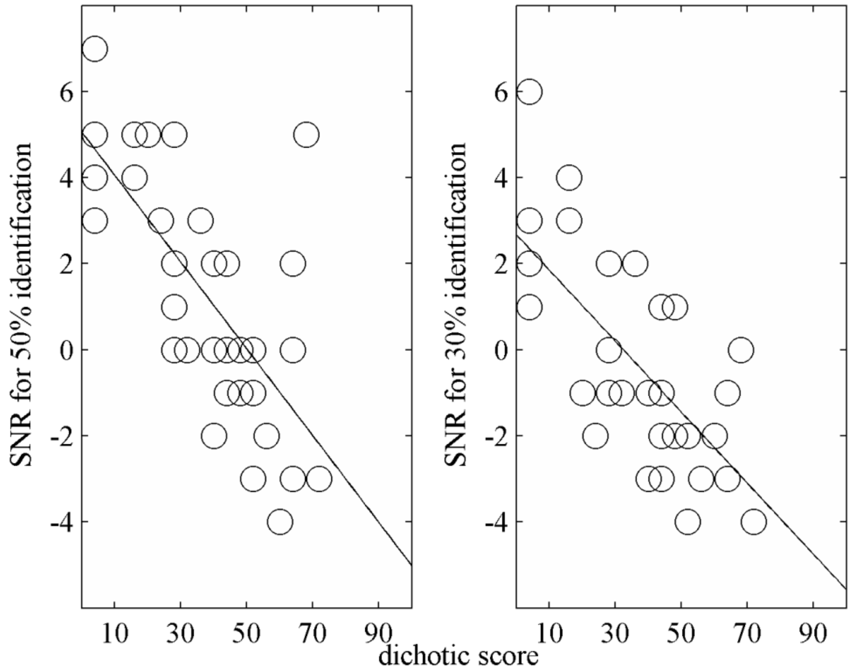


Fig. 2: The relationship between speech identification in noise and the non-dominant ear dichotic score. For further details see Fig. 1.

Low dichotic scores with large LED and low scores in speech identification in multi-talker babble noise both reflect deficits in the ability of the auditory system to process competing speech stimuli. Indeed, we found high correlations between speech-in-noise scores and dichotic scores of the non-dominant ear, the dominant ear, and the total dichotic scores: Participants who had lower dichotic achievements tended to require better signal-to-noise ratios to identify the test words. These results support the claims made by Jerger *et al.* (1995) and Strouse-Carter *et al.* (2001) that deterioration in dichotic listening may be significant in elderly people's capability to separate target speech from competing spatially-separated speech stimuli. Moreover, deterioration in dichotic listening may result in lower satisfaction with hearing aids (Chmiel and Jerger, 1996; Givens *et al.*, 1998).

Speech in noise is routinely measured by no more than 25% of audiologists in the course of hearing-aid fitting (Weinstein, 2013). Dichotic listening is easy and fast to evaluate. Our results suggest that dichotic listening tests may be a good predictor for individuals' abilities to understand speech in the presence of multi-talker noise, and thus can serve as an available and reliable tool in hearing rehabilitation counseling.

REFERENCES

Chmiel, R., and Jerger, J. (**1996**). "Hearing aid use, central auditory disorder, and hearing handicap in elderly persons," J. Am. Acad. Audiol., **7**, 190-202.

Divenyi, P.L., and Haupt, K.M. (**1997**). "Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. I. Age and lateral asymmetry effects," Ear Hearing, **18**, 42-61.

Divenyi, P.L., Stark, P.B., and Haupt, K.M. (**2005**). "Decline of speech understanding and auditory thresholds in the elderly," J. Acoust. Soc. Am., **118**, 1089-1100.

Dos Santos, S.S., Specht, K., Hämäläinen, H., and Hugdahl, K. (**2008a**). "The effects of background noise on dichotic listening to consonant-vowel syllables," Brain Lang., **107**, 11-15.

Dos Santos, S.S., Specht, K., Hämäläinen, H., and Hugdahl, K. (**2008b**). "The effects of different intensity levels of background noise on dichotic listening to consonant-vowel syllables" Scand. J. Psychol., **49**, 305-310.

Frisina, D.R., and Frisina, R.D. (**1997**). "Speech recognition in noise and presbycusis: relations to possible neural mechanisms," Hear. Res., **106**, 95-104.

Frisina, D.R. (**2001**). "Possible neurochemical and neuroanatomical bases of age-related hearing loss-presbycusis," Semin. Hearing, **22**, 213-226.

Givens, G.D., Arnold, T., and Hume, W.G. (**1998**). "Auditory processing skills and hearing aid satisfaction in a sample of older adults," Percept. Motor Skill., **86**, 795-801.

Gordon-Salant, S., and Sherlock, L.P.G. (**1992**). "Performance with an adaptive frequency response hearing aid in a sample of elderly hearing-impaired listeners," Ear Hearing, **13**, 255-262.

Gordon-Salant, S., and Fitzgibbons, P.J. **(1993)**. "Temporal factors and speech recognition performance in young and elderly listeners," J. Speech Hear. Res., **36**, 1276-1285.

Gordon-Salant, S., and Fitzgibbons, P.J. (**2001**). "Sources of age-related recognition difficulty for time-compressed speech," J. Speech Lang. Hear. Res., **44**, 709-719.

Gordon-Salant, S. (**2005**). "Hearing loss and aging: new research findings and clinical implications," J. Rehabil. Res. Dev., **42**, 9-23.

Hallgren, M., Larsby, B., Lyxell, B., and Arlinger, S. (**2001**). "Cognitive effects in dichotic speech testing in elderly persons," Ear Hearing, **22**, 120-129.

Humes, L.E., and Roberts, L. (**1990**). "Speech recognition difficulties of the hearing-impaired elderly: the contribution of audibility," J. Speech Hear. Res., **33**, 726-735.

Humes, L.E., and Nelson, K.J. (**1991**). “Recognition of synthetic speech by hearing-impaired elderly listeners,” J. Speech Hear. Res., **34**, 1180-1184.

Humes, L.E., Watson, B.U., Christensen, L.A., Cokely, C.G., Halling, D.C., and Lee, L. (**1994**). “Factors associated with individual differences in clinical measure of speech recognition among the elderly,” J. Speech Hear. Res., **37**, 465-474.

Humes, L.E. (**1996**). “Speech understanding in the elderly,” J. Am. Acad. Audiol., **7**, 161-167.

Humes, L.E., Burk, M.H., Coughlin, M.P., Busey, T.A., and Strauser, L.E. (**2007**). “Auditory speech recognition and visual text recognition in younger and older adults: similarities and differences between modalities and the effects of presentation rate,” J. Speech Lang. Hear. Res., **50**, 283-303.

Killion, M.C. (**1997**). “SNR loss: “I can hear what people say, but I can't understand them””, Hearing Review, **4**, 8-14.

Killion, M.C., Niquette, P.A., Gudmundsen, G.I., Revit, L.J., and Banerjee, S. (**2004**). “Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners,” J. Acoust. Soc. Am., **116**, 2395-2405.

Jerger, J., Chmiel, R., Allen, J., and Wilson, A. (**1994**). “Effects of age and gender on dichotic sentence identification,” Ear Hearing, **15**, 274-286.

Jerger, J., Alford, B., Lew, H., Rivera, V., and Chmiel, R. (**1995**). “Dichotic listening, event related potentials, and interhemispheric transfer in the elderly,” Ear Hearing, **16**, 482-498.

Martin, J.S., and Jerger, J.F. (**2005**). “Some effects of aging on central auditory processing,” J. Rehabil. Res. Dev., **42**, 25-44.

Nilsson, M., Soli, S.D., and Sullivan, J.A. (**1994**). “Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise,” J. Acoust. Soc. Am., **95**, 1085-1099.

Noffsinger, D., Martinez, C.D., and Andrews, M. (**1996**). “VA-CD data from elderly subjects,” J. Am. Acad. Audiol., **7**, 49-56.

Oldfield, R.C. (**1971**). “The assessment and analysis of handeness: The Edinburgh inventory,” Neuropsycholοlgia, **9**, 97-113.

Pichora-Fuller, M.K., Schneider, B.A., and Daneman, M. (**1995**). “How young and old adults listen and remember speech in noise,” J. Acoust. Soc. Am., **97**, 593-608.

Roup, C.M., Wiley, T.L., and Wilson, R.H. (**2006**). “Dichotic word recognition in young and older adults,” J. Am. Acad. Audiol., **17**, 230-240.

Schneider, B.A., Pichora-Fuller, M.K., Kowalchuk, D. and Lamb, M. (**1994**). Gap detection and the precedence effect in young and old adults. The Journal of the Acoustical Society of America. 95(2): 980-991

Schneider, B.A., and Pichora-Fuller, M.K. (**2001**). “Age-related changes in temporal processing: Implications for speech perception,” Semin. Hearing, **22**, 227-240.

Schneider, B.A., Daneman, M., and Pichora-Fuller, M.K. (**2002**). “Listening in aging adults: from discourse comprehension to psychoacoustics,” Can. J. Exp. Psychol., **56**, 139-152.

Schneider, B.A., Daneman, M., and Murphy, D.R. (**2005**). “Speech comprehension difficulties in older adults: cognitive slowing or age related changes in hearing?” Psychol. Aging, **20**, 261-271.

Snell, K.B., and Frisina, D.R. (**2000**). “Relationships among age-related differences in gap detection and word recognition,” J. Acoust. Soc. Am., **107**, 1615-1626.

Strouse, A., Ashmead, D.H., Ohde, R.N., and Grantham, D.W. (**1998**). “Temporal processing in the aging auditory system,” J. Acoust. Soc. Am., **104**, 2385-2399.

Strouse, A., and Wilson, R.H. (**1999**). “Stimulus length with dichotic digit recognition,” J. Am. Acad. Audiol., **10**, 219-229.

Strouse, A., Wilson, R.H., and Brush, N. (**2000**). “Effect of order bias on the recognition of dichotic digits in young and elderly listeners,” Audiology, **39**, 93-101.

Strouse-Carter, A., Noe, C.M., and Wilson, R.H. (**2001**). “Listeners who prefer monaural to binaural hearing aids,” J. Am. Acad. Audiol., **12**, 261-272.

Tun, P.A., O’Kane, G., and Wingfield, A. (**2002**). “Distraction by competing speech in young and older adult listeners,” Psychol. Aging, 17, 453-467.

Weinstein, B.E. (**2013**). *Geriatric audiology*, 2nd edition (Thieme, New-York).

Wilson, R.H., and Jaffe, M.S. (**1996**). “Interaction of age, ear and stimulus complexity on dichotic digit recognition,” J. Am. Acad. Audiol., **7**, 358-364.

Wingfield, A., Tun, P.A., and McCoy, S.L. (**2005**). “Hearing loss in older adulthood. What it is and how it interacts with cognitive performance,” Curr. Dir. Psychol. Sci., **14**, 144-148.

Model-based loudness compensation for broad- and narrow-band signals

DIRK OETTING[1,2], STEPHAN D. EWERT[2,*], VOLKER HOHMANN[2], AND JENS-E. APPELL[1]

[1] *Project Group Hearing, Speech and Audio Technology, Fraunhofer IDMT and Cluster of Excellence "Hearing4all", Oldenburg, Germany*

[2] *Medizinische Physik, Universität Oldenburg and Cluster of Excellence "Hearing4all", Oldenburg, Germany*

A fundamental problem when attempting to restore loudness perception in hearing-impaired listeners are differences in the loudness perception of narrow- and broad-band signals when compared to normal-hearing listeners. Here, a multi-channel dynamic compression algorithm is presented where the signal-to-masking ratio (SMR) is used to modify the channel gain function. Result 1: The evaluation of this approach using a loudness model showed that the loudness perception of hearing-impaired listeners can be restored to the loudness perception of a normal-hearing listener for signals with different bandwidths. Result 2: Inconsistencies between the individual measured loudness function using the categorical loudness scaling procedure and the model predictions were found. The available model parameters, being i) hearing threshold level, ii) outer, and iii) inner hair-cell loss, were not sufficient to fit the model to the individual narrow-band loudness perceptions.

INTRODUCTION

Loudness perception of hearing-impaired (HI) listeners differs from the loudness perception of normal-hearing (NH) listeners. Typically, HI listeners show increased hearing threshold levels (HTL) whereas uncomfortable loudness levels (UCL) remain at the same level as in NH listeners (Bentler and Cooley, 2001). Therefore, to restore loudness perception in HI listeners, a compression algorithm is required which applies the appropriate gain for signals with low amplitudes and reduces the gain for signals with high amplitudes. The individual narrow-band loudness perception can be measured using categorical loudness scaling (CLS; Brand and Hohmann, 2002). The result of the CLS procedure is a loudness function which maps the signal level to the perceived loudness category. Level-dependent gain functions for restoring narrow-band loudness perception with a compression algorithm can be derived when comparing the measured loudness function with the average NH loudness function for the same signal (compare Fig. 3c). It is known that the gain required to restore the narrow-band loudness perception in a multi-channel dynamic compression algorithm leads to overly high gains for broad-band signals (Latzel *et al.*, 2004), resulting in too high loudness impressions. Using both signal types in a current loudness model clarifies why different gain values for narrow- and broad-band signals are required.

[*] Corresponding author: stephan.ewert@uni-oldenburg.de

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Figure 1a shows different channel gain functions required for restoring the specific loudness perception using the loudness model of Chen *et al.* (2011) and the model of Moore and Glasberg (1997). The specific loudness of Bark channel no. 8 (920-1080 Hz) was calculated for a NH listener and a HI listener having a 50-dB flat hearing loss with standard model parameter settings of 80% outer (OHC), and 20% inner hair-cell (IHC) loss. The channel gain in dB required to restore specific loudness in this channel to normal was calculated for a 1/3-octave, low-noise noise stimulus (LNN) centred at 1 kHz, and for a stationary speech-shaped noise (IFnoise) generated from the international speech test signal (ISTS; Holube *et al.*, 2010) as a function of input level. The estimated gain for restoring the narrow-band loudness function differs by more than 10 dB between both loudness models. The difference between the narrow- and broad-band gain function required to restore the specific loudness is shown in Fig. 1b. The model of Chen *et al.* (2011) estimates a gain reduction of up to 9 dB for the broad-band signal for medium signal levels. Using the model of Moore and Glasberg (1997), only 2-4 dB of gain reduction is predicted for high signal levels.

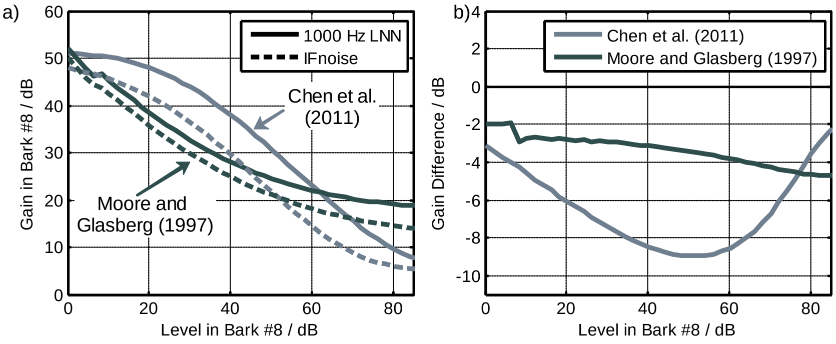


Fig. 1: a) Example of model calculations for the different channel gain functions required in a multi-channel dynamic compression algorithm to restore specific loudness at 1 kHz for a narrow-band noise and broad-band noise (IFnoise). b) Difference between the gain functions from a) showing that both models estimate different channel gain reductions to restore normal loudness perception in HI when a broad-band signal is presented.

It can be concluded from the model calculation that a multi-channel dynamic compression algorithm that analyses and processes the input signal independently in each frequency channel is not capable of applying the correct gain for restoring loudness of both narrow-band and broad-band sounds. To restore loudness using a multi-channel dynamic compression algorithm some measure of the actual signal's bandwidths is required to control the compressive gain functions depending on the signal's bandwidth. In this study, the signal-to-masking ratio (SMR) calculated in each processing channel as an estimator of the signal's bandwidth is introduced. The calculation of the SMR and its integration in a compression algorithm is described in the next section. To demonstrate the properties of the suggested approach, the loudness model of Chen *et al.* (2011) is used in the following.

DYNAMIC COMPRESSION ALGORITHM

A multi-channel compression algorithm was implemented in the frequency domain using an overlap-add (FFT, sampling rate 22 kHz, frame length 408 samples) processing scheme. The signal level is calculated for each of the 24 Bark-spaced channels formed by adding up the squared magnitudes of the corresponding FFT-bins. According to the excitation pattern calculation proposed by Moore and Glasberg (1997), the masking of each Bark-channel on the neighbouring Bark-channels is calculated. Instead of the quite complex calculation scheme proposed by Moore and Glasberg (1997) a more efficient approximate method to calculate the masking patterns was used. Based on the channel levels and the masking slopes, the SMR for each channel as an estimator of the signal's bandwidth was calculated. As shown in the left figure of Fig. 2, we define the SMR to be the difference in dB between the channel level and the maximum masking level caused by all other channels (dashed lines). In the left panel of Fig. 2 the SMR is about 10 dB, which corresponds to the value of the upward masking slopes (10 dB/Bark at medium overall level) for signals having the same level in each Bark band (i.e., uniform-exciting noise, UEN; Fastl and Zwicker, 2007).

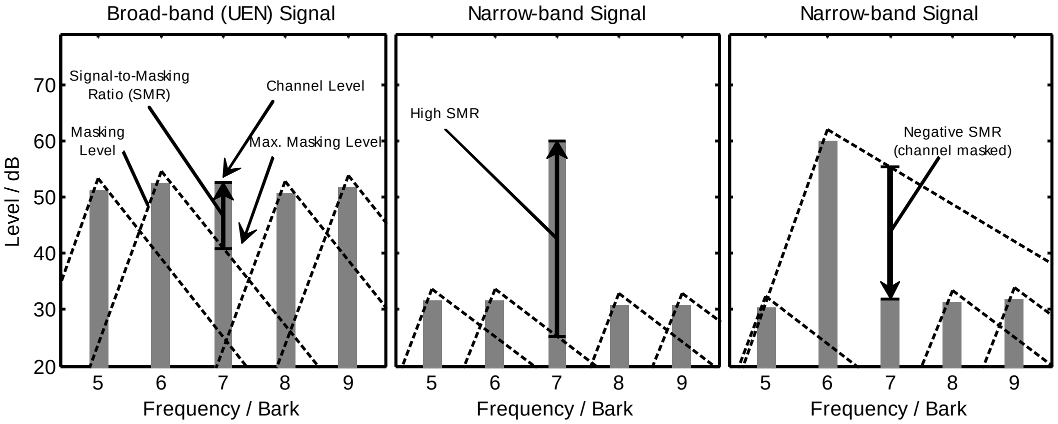


Fig. 2: Calculation of the SMR. The SMR is the difference in dB between the channel level and the maximum masking level. Left panel: A broad-band signal with equal channel levels produces a SMR of about 10 dB at medium input levels. Middle panel: A narrow-band signal produces a high on-frequency SMR value, and (right panel) negative SMR values in the neighbouring channels.

The middle and right panels of Fig. 2 show the calculation of the SMR for a narrow-band signal. Most of the signal energy falls into channel #7 and the masking levels of the other channels are well below the signal level. Accordingly, the SMR value in channel #7 is high. If the narrow-band signal falls into channel #6 as shown in the right panel, the masking level of channel #6 towards channel #7 is much higher than the channel level in channel #7. Hence, the SMR in channel #7 is negative, meaning that this channel will not be perceived and thus should not be amplified by the compressor. In summary, high SMR values indicate a narrow-band signal prominent in the respective Bark-channel, low SMR values indicate a broad-band signal, and negative SMR values indicate signal components that are not perceived.

In the next step the SMR was integrated in a dynamic compression scheme as a major control parameter to adapt the channel gain based on the signal's bandwidth parametrically. This is achieved by modifying the channel level which is used as the input level to the channel gain function. The channel's gain functions are initially set to restore the narrow-band loudness perception. This corresponds to the 'LoudFit' fitting rationale by Herzke and Hohmann (2005) which will serve in this work as a comparison fitting rationale. Figure 3a shows examples for the SMR-dependent modification of the channel level. For high SMR values (SMR > 20 dB) it is assumed that the signal is narrow-band and therefore no further modification to the channel level is applied. In contrast, broad-band signals lead to low SMR values and the estimated 'effective' channel level is increased with decreasing SMR. The 'effective' channel level is used as input to the channel gain function. Since the SMR-dependent modification is always positive, this modification leads to a reduced amount of gain due to the steeper loudness function of HI listeners. This is illustrated in Fig. 3c for the loudness functions derived from categorical loudness scaling. The amount of SMR-dependent level modification is adapted to the individual loudness perception.

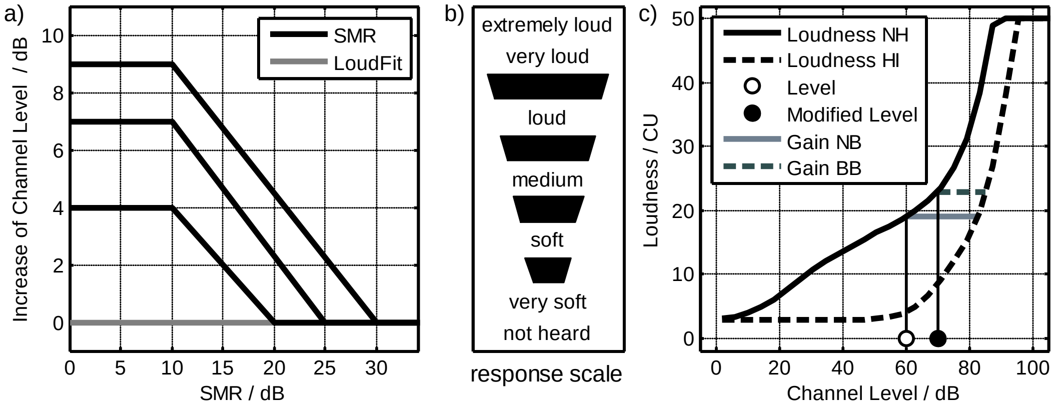


Fig. 3: a) Increase of channel level depending on the SMR. The amount of level increase can be individually adjusted (4, 7, and 10 dB in this example). b) Response scale of the CLS procedure. c) Simulated narrow-band loudness functions of a NH and a HI listener for CLS. According to the 'LoudFit' fitting procedure, the gain to restore the loudness of a 60 dB narrow-band signal ('Gain NB') is approximately 23 dB. For a broad-band signal the channel level is increased ('effective' level), hence the applied gain ('Gain BB') is reduced from approx. 23 to 15 dB for a broadband signal having the same channel level as a narrow-band signal.

EVALUATION OF THE SMR-BASED COMPRESSION ALGORITHM

To evaluate the basic properties of the proposed algorithm, the loudness model of Chen *et al.* (2011) was used and two hearing losses where simulated: a flat 50-dB hearing loss (FHL) and a sloping haring loss (SHL; audiograms are shown in the middle panel of Fig. 4 and 5). The model's standard parameters of 80% OHC loss and 20% IHC loss were used for the FHL und the SHL. To compare the results with results of the CLS procedure, the transformation from the model output in sone to the CLS response scale in categorical units (CU) according to Heeren *et al.* (2013) was used. Uniform-exciting noise (UEN), i.e., noise that produces the same channel level in each bark band (Fastl and Zwicker, 2007) was used as a test signal. The UEN was limited to bandwidths of 1, 5, and 15 Bark centred at 8.5 Bark (1000 Hz) and is accordingly referred to as UEN1, UEN5, and UEN15. The aim was to restore the NH loudness function for all test signals and hearing losses using the SMR algorithm.

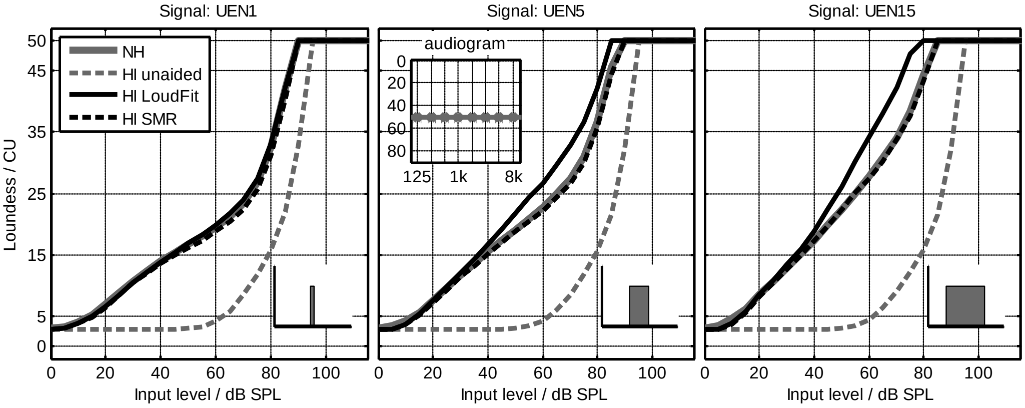


Fig. 4: Loudness functions for a NH and a HI listener with a flat 50-dB hearing loss for UEN with a bandwidth of 1, 5, and 15 Bark, left to right panels, respectively. The black solid line and the black dashed line show the aided loudness function using the fitting rationale 'LoudFit' and the proposed SMR approach. It can be observed that the LoudFit rationale leads to a too high loudness sensation for broad-band signals (i.e., the UEN5 and UEN15 signal shown in the middle and right panel, respectively), whereas the SMR algorithm is able restore the HI loudness perception to NH perception for all signals independent of bandwidth.

In Fig. 4 and 5 the unaided (grey dashed line) and aided loudness (black curves) functions for the modelled FHL and SHL are shown. Normal-hearing loudness perception (indicated by the solid grey curve) was the target for the two aided conditions tested: aided according to the 'LoudFit' rationale (black line) and aided with the SMR-algorithm (black dashed line). Figure 4 and 5 show that the 'LoudFit' fitting rationale as well as the SMR algorithm were able to compensate the loudness perception for the narrow-band signal UEN1 (left panels). For broader signals like UEN5 and UEN15 (middle and right panel, respectively) the 'LoudFit' procedure applied too much gain resulting in a higher loudness perception for the HI when compared to NH, especially at medium levels.

In contrast, the SMR algorithm was able to correctly reduce the gain for broad-band signals and in consequence to restore HI's loudness perception to normal independent of bandwidth. Very similar results were obtained in Fig. 5 for the SHL simulations. Again, too much gain was applied for broad-band signals when signals were processed according to the 'LoudFit' rationale, whereas the SMR algorithm again restored normal loudness for all signals. Here, the SMR-dependent level modification was adjusted to be 4 dB for signals having a low SMR value (lowest function in Fig. 3a for both hearing losses tested. However, depending on the individual hearing loss or preference other modifications might be required. To find those individual settings of the SMR modification, information about the individual perception of broad-band signals is required. This information can be derived from individualized loudness models (as it is done for two types of hearing-impaired here) or by CLS measurements of broad-band signals carried out in addition to the standard clinical procedure of CLS measures with narrow-band signals.

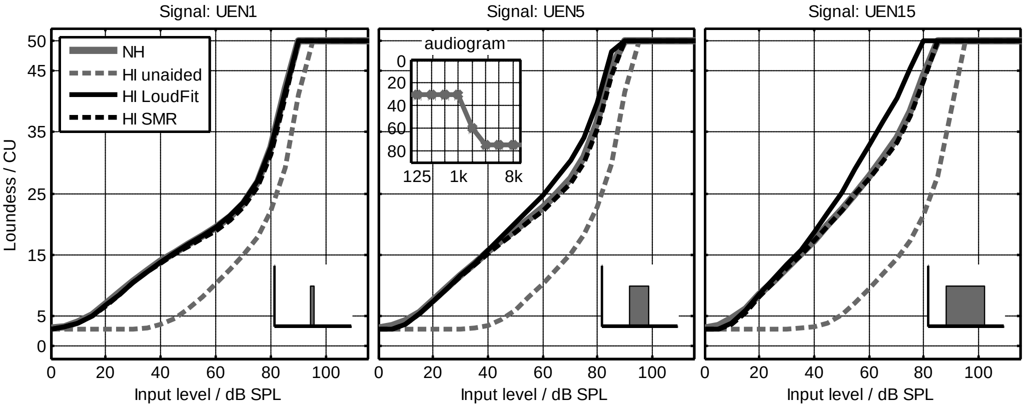


Fig. 5: same as Fig. 4 but for a listener with a sloping hearing loss (SHL).

INDIVIDULIZATION OF THE LOUDNESS MODEL

As pointed out above, the individual adjustment of the SMR algorithm requires information about the individual loudness perception of narrow-band signals over frequency to derive frequency-dependent gain functions as well as information about the perception of broad-band signals to adjust the SMR-based modification. While the CLS measurement with narrow-band signals becomes more and more clinical practice, information about the perception of broad-band sounds is typically not collected. To close this gap, a loudness model could be used, which can be individualized based on data from CLS measurements with narrow-band stimuli to predict the individual broad-band loudness perception. The loudness model of Chen *et al.* (2011) for NH and HI listeners provides the parameters IHC-related and OHC-related hearing loss to individualize the model predictions, whereas the sum of both losses equals approximately the total hearing loss for hearing losses below 60 dB HL (Chen and Hu, 2013).

Following the idea to adjust the loudness model to predict individual CLS data measured with narrow-band stimuli, the model of Chen *et al.* (2011) was adjusted using the IHC and OHC parameters. The result is shown in Fig. 6. Grey curves show variations of the IHC and OHC parameters, whereas thick curves show loudness scaling data measured in two HI listeners having the same hearing threshold of 45 dB HL at 2 kHz but different loudness perception above threshold. A systematic variation of OHC/IHC loss configurations under the constraint of a fixed audiometric threshold (grey lines in Fig. 6) showed that the loudness model was not able to predict the loudness perception of the two HI listeners at all input levels. Therefore, it does not seem meaningful to use the loudness model of Chen *et al.* (2011) to model individual loudness perception of broad-band signals as it is needed for individual adjustment of the SMR algorithm.

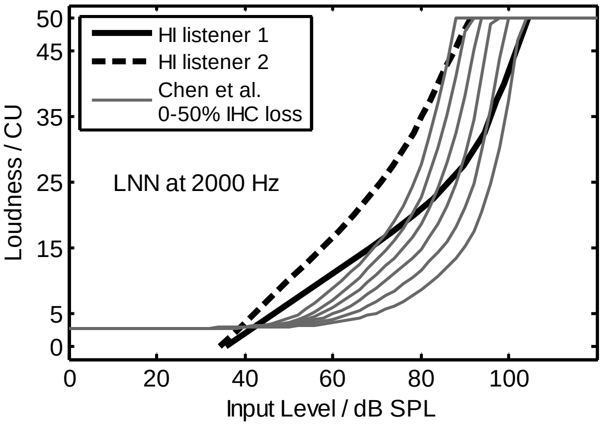


Fig. 6: Loudness functions of two HI listeners derived from CLS measurements with a narrow-band noise signal centred at 2 kHz. Both listeners had a hearing threshold of 45 dB HL. The modelled loudness perception (grey lines) does not match the measured loudness function independent of the possible parameter configuration.

SUMMARY AND CONCLUSION

A multi-channel dynamic compression algorithm for restoring the loudness perception of HI listeners was proposed, which is capable to restore normal loudness perception for narrow-band and broad-band stimuli. The algorithm uses the signal-to-masking ratio (SMR) to modify the level in each processing channel. This ‘effective’ channel level is then used to determine the channel gain. The evaluation of the algorithm with two simulated HI listeners (flat and sloping hearing loss) using the recent loudness model by Chen *et al.* (2011) showed that the SMR approach is able to restore loudness perception for narrow- and broad-band signals. Thereby, the SMR algorithm requires information about the individual loudness perception of broad-band signals. A first approach to gather this information from predictions of the loudness model of Chen *et al.* (2011) failed, because the model could not be adjusted to correctly predict loudness perception for narrow-band signals. The model predictions in the lower loudness domain (between 'very soft' and 'soft') for the HI listeners were too low for all possible model parameter configurations when compared to the measured loudness in CLS.

As a conclusion, the SMR-algorithm requires to be adjusted using additional CLS measurements with broad-band signals. Further evaluations of the SMR algorithm using CLS of different everyday signals will show if the individually-adjusted SMR-algorithm restores the loudness perception to normal for narrow- and broad-band signals. Further improvements of recent loudness models to predict the individual loudness perception of a single HI is required for future research and fitting of model-based algorithms which individually restore loudness for a variety of stimuli.

AKNOWLEDGMENTS

This work was supported by the DFG Cluster of Excellence EXC 1077/1 'Hearing4all'.

REFERENCES

Bentler R.A., and Cooley L.J. (**2001**). "An examination of several characteristics that affect the prediction of OSPL90 in hearing aids," Ear Hearing, **22**, 58-64.

Brand T., and Hohmann, V. (**2002**). "An adaptive procedure for categorical loudness scaling," J. Acoust. Soc. Am., **112**, 1597-1604.

Chen, Z., Hu, G., Glasberg, B.R., and Moore, B.C.J. (**2011**). "A new model for calculating auditory excitation patterns and loudness for cases of cochlear hearing loss," Hear. Res., **282**, 69-80.

Chen, Z., and Hu, G. (**2013**). "CHENFIT-AMP, a nonlinear fitting and amplification strategy for cochlear hearing loss," IEEE T. Bio-Med. Eng., **60**, 3226-3237.

Fastl, H., and Zwicker, E. (**2007**). *Psychoacoustics: Facts and Models* (Springer, Berlin), Third Ed.

Heeren, W., Hohmann, V., Appell, J.E., and Verhey, J.L. (**2013**). "Relation between loudness in categorical units and loudness in phons and sones," J. Acoust. Soc. Am., **133**, EL314-EL319.

Herzke, T., and Hohmann, V. (**2005**). "Effects of instantaneous multiband dynamic compression on speech intelligibility," EURASIP J. App. Sig. P., **18**, 3034-3043.

Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (**2010**). "Development and analysis of an International Speech Test Signal (ISTS)," Int. J. Audiol., **49**, 891-903.

Latzel, M., Margolf-Hackl, S., Denkert, J., and Kießling, J. (**2004**). "Präskriptive Hörgeräteanpassung auf Basis von NAL-NL1 im Vergleich mit einem lautheitsbasierten Verfahren," DGA 7. Jahrestagung.

Moore, B.C.J., and Glasberg, B.R. (**1997**). "A model of loudness perception applied to cochlear hearing loss," Audit. Neurosci., **3**, 289-311.

Effects of NALR on consonant-vowel perception

CHRISTOPH SCHEIDIGER* AND JONT B. ALLEN

Human Speech Recognition Group, University of Illinois, Urbana, IL, USA

Consonant vowel (CV) identification experiments in masking noise with 16 hearing-impaired (HI) ears at two different gain conditions, i.e., flat-gain (FG) and spectral correction (National Acoustic Laboratory Revised prescriptive procedure, NALR), were administered (Han, 2011). In both gain conditions, listeners were directed to adjust the presentation level to their most comfortable loudness (MCL). MCL testing runs contrary to the common approach of adjusting the presentation level, depending on the pure tone thresholds (PTTs) and the long term average speech spectrum (LTASS) (Posner and Ventry, 1977; Zurek and Delhorne, 1987). The results, however, prove that for speech testing MCL is justified. A more rigorous definition for audibility based on entropy in recognition experiments is provided. Furthermore, the effectiveness of NALR for CV perception is investigated. The average error went down from 20.1% ($\sigma = 3.7$) to 16.3% ($\sigma = 2.8$). For 50.5% of the token[1]-ear pairs (TEPs) the error and entropy both went down, while for 15.1% of the TEPs the entropy and error went up with NALR. In order to evaluate statistically siginificant effects of NALR, the confusion matrix data were clustered, and the number of ears which switched clusters when NALR was applied were investigated. In addition, the subjects' confusions under both conditions were studied and compared to the confusions of other HI and normal-hearing (NH) subjects.

INTRODUCTION

The goal of this research is to better understand speech perception in hearing-impaired ears. The human speech recognition (HSR) group at the University of Illinois at Urbana-Champaign takes the approach to look at CV recognition tasks of NH as well as HI subjects. CVs are chosen, as opposed to words, phrases, or sentences in order to reduce the influence of higher-order (context) processing in the auditory pathway, which permits the control of differences in cognitive abilities (e.g., memory, semantics) (Miller *et al.*, 1951). A second goal of this paper is to address what audibility means in speech perception experiments and to determine how it can be verified. Lastly, we will show that, despite its high variability, speech as a test for hearing loss and hearing-aid evaluation can deliver more detailed insights than the commonly used pure tones, which is in contrast to what Walden *et al.* (1983) and Zurek and Delhorne (1987) suggested.

*Corresponding author: csche@elektro.dtu.dk

[1] In this document a token is defined as a recorded sound (i.e., CV). One consonant (e.g., /p/) can have many tokens.

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

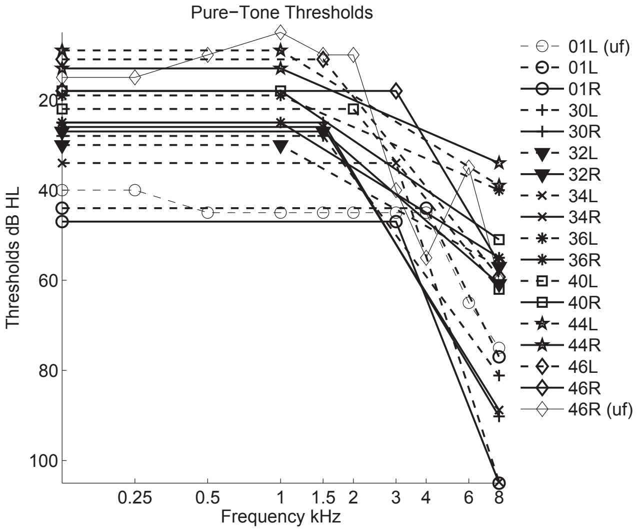


Fig. 1: PTTs for the sixteen ears

HI ear	Age	PTA	MCL	
			FG	NALR
44L	65	10	82	77
44R	65	15	78	77
46L	67	8.3	82	85
46R	67	16.6	82	86
40L	79	21.6	79, 81	
40R	79	23.3	80	80
36L	72	26.6	68	75
36R	72	28.3	70	75
30L	66	30	80	79
30R	66	26.6	80	79
32L	74	35	79	81
32R	74	26.6	77	78
34L	84	31.6	84	85
34R	84	28.3	82	85
02L	82	45	83	88
02R	82	46.6	82	89
(μ,σ)	(74,7)	(29,15)	(79,4)	(81,5)

Table 1: Subjects' age, PTAs and MCLs (dB SPL)

METHODS

The two conditions (i.e., FG and NALR) were administered as separate experiments (Han, 2011). Each of the 8 subjects (16 HI ears) passed a middle-ear examination and their hearing thresholds were measured before each experiment. All 16 ears had mild-to-moderate hearing loss. Fig. 1 shows the fitted PTTs according to Trevino (2013).

The CV syllables consisted of 14 consonants (6 stops, 6 fricatives, and 2 nasals) followed by /a/. Two talkers (1 male and 1 female) were selected per consonant. The tokens were chosen from those for which there was less than 3% error at SNRs < -2 dB in previous NH experiments. The male tokens for /f, s, ʒ, n/ + /a/ were removed from the analysis, because they had to be changed between the two conditions, leaving 12 CVs for comparison (24 tokens). The tokens used for the experiments are well characterized: the perceptual cues have been identified by the 3DDS method (Li *et al.*, 2012) and the CMs at 6 SNRs were previously determined in both white noise and speech weighted noise.

The subjects were able to adjust the presentation level at any time during the experiment, however, as seen in Table 1 only 40L made use of this option. All of the subjects had one practice session before they began the experiment. Syllable presentation was randomized over consonants, speakers, and SNRs (12, 6, and 0 dB, plus quiet). For each condition, SNR and subject, a token was presented between 5 and 10 times (depending on the error); this resulted in 800-1000 trials per subject.

RESULTS

The resulting *confusion matrices* (CM) were analyzed using the following tools.

Entropy

Information theory and entropy were introduced by Shannon (1948). Miller and Nicely (1955) were the first to apply an entropy analysis to speech confusion data. Entropy, a measure of the randomness of a response, is defined as the expected value of the information ($\log(1/p_i)$), the CM row sum is $\sum_i p_i = 1$, where $i = 1 \dots 14$ (14 is the number of possible responses):

$$\mathscr{H}(\mathbf{p}) = \sum_{i=1}^{I} p_i \log_2\left(\frac{1}{p_i}\right). \qquad \text{(Eq. 1)}$$

Audibility: Posner and Ventry (1977) found that subjects perform below their maximum speech discrimination abilities if tested under MCL conditions. The data, however, suggest that most tokens were fully audible to all the subjects under both conditions. We suggest that calculating the entropy in quiet is a more meaningful audibility measure for CV identification experiments than LTASS and PTTs. Low entropy implies consistency, which is a strong test of audibility, even if the error is high (cf. Fig. 2 (a)).

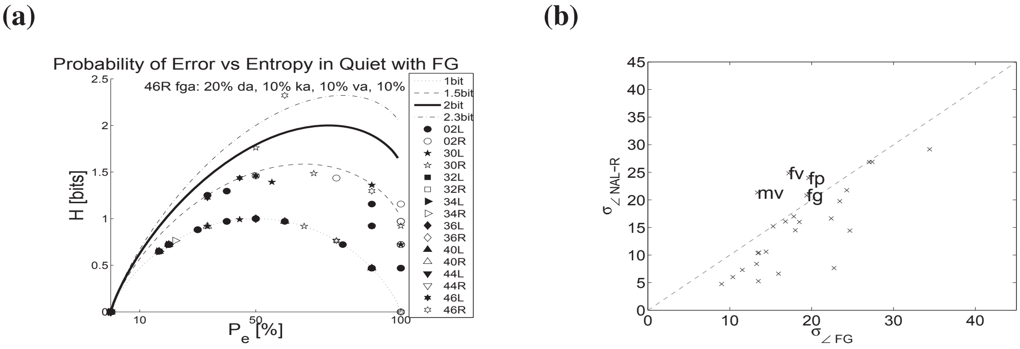


Fig. 2: **(a)** P_e vs $\mathscr{H}$ plot for all subjects and tokens in quiet (FG condition). Entropy is low, even though in many cases the error is high, thus audibility is not an issue. The 2-bit curve is a reasonable audibility threshold based on the Miller and Nicely (1955) confusion groups. According to this definition only the female token of /gɑ/ is not audible for subject 46R. All the other sounds are audible for all subjects. **(b)** The standard deviation of the angles between the correct response and the ears decreases with NALR in all but the four labeled cases (fg, fp,fv, mv) subjects responded more consistently, (mv = male /va/ token).

Effects of NALR: 24 tokens can be compared between the two experiments and 16 ears. This results in 384 cases, when collapsed over SNR. Those cases can be categorized according to how the entropy and error changed from the FG to the NALR experiment. Most of the cases (50.5%) are the cases where NALR decreased both the entropy and error. The second largest group is the one where NALR increased both the entropy and error (15.1%). The other two categories only contain the few remaining TEPs (cf. Fig. 3).

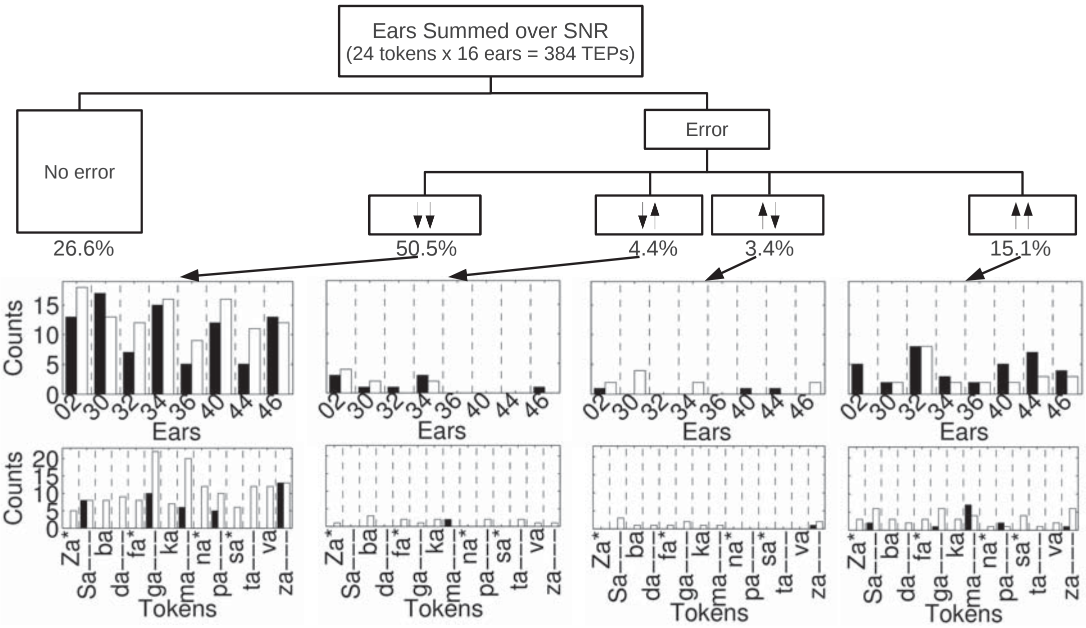


Fig. 3: Categorization of the CV perception data for the 24 tokens and 16 listeners collapsed over four SNRs. For 102 (26.6%) of the 384 TEPs there was zero error in both conditions. The remaining 282 TEPs are grouped into one of 4 major categories, in the category labels the first arrow indicates what affect NALR had on the entropy, the second one indicates what happen to the error with NALR: (↓↓) (50.5%), (↓↑) and (↑↓) are small categories (4.4% and 3.4%)(↑↑) (15.1%). The histograms display the listener (top) and token (bottom) distributions. They show many of the TEPs in one category belong to a particular ear or token. The black bars represent the left ear and the male token, respectively, whereas the white bar represents the right ear and the female token, respectively. The * indicates the male token was excluded for the analysis (e.g., Za*).

Direction cosine

Every confusion matrix defines a vector space, where each row is a vector in that space. In order to find the distance between two tokens (rows), a norm must be defined: we chose a metric called the *Hellinger Distance* (HD), which uses the square roots of the probability vectors **p**. Via *Schwartz's inequality*, it is possible to calculate an angle θ_{lm} between any two tokens in the vector space. The angle is a measure of how different two response vectors are. The HD can also be used to measure the difference between the two experiments or between a listener's response and the correct answer.

The HD seems to be an underutilized measure for the analysis of CMs.

$$cos\ \theta_{lk} = \mathbf{p}_l \cdot \mathbf{p}_m = \sum_{i=1}^{I} \sqrt{p_{l,i}}\sqrt{p_{m,i}} \qquad \text{(Eq. 2)}$$

Confusions: The angle between the correct answer and the response is a measure of the change in confusions. The mean angle ($\mu_{\angle}$) and the standard deviation ($\sigma_{\angle}$) of the angles for one token are expected to decrease if the ears become more accurate in their response or if they become more consistent in their answers, respectively.

NALR has a significant impact on the standard deviation: a *paired t-test* results in $\alpha = 0.05 > p = 0.013$; in addition, the means of the two conditions are significantly different ($p = 2.0 \times 10^{-7}$). From the scatter plot in Figure 2 (b) one can see that the variance of the angles ($\sigma_{\angle}$) goes down with NALR in all but 3 cases: fv (i.e., female /va/), fp and fg. The mean angle ($\mu_{\angle}$) goes down with NALR for all 24 tokens.

K-means clustering

Once normed vector space is defined, the elements in this space may be clustered. For each of the 24 tokens, there are $2 \times 4 \times 14 = 112$ (2 conditions, 4 SNRs and 14 ears) data points in the fourteen dimensional space. The *k-means* algorithm is then used to group the data points into $K = 4$ clusters, with each cluster represented by its cluster centroid $\mathbf{c}_k$, $k = 1, \ldots, K$. The $\mathbf{c}_k$s are then sorted according to their entropy (Eq. 1).

Classifying NALR: By comparing the centroid ($\mathbf{c}_k$) assignments of two points of a subject at a given SNR – representing the two different gain conditions – it is possible to investigate the impact of NALR. For all tokens, $\mathbf{c}_1$ (smallest entropy) represents the centroid of the points closest to the correct answer. Subjects that go from a higher entropy cluster ($\mathbf{c}_2$,$\mathbf{c}_4$,$\mathbf{c}_4$) to $\mathbf{c}_1$ at a given SNR because of NALR, are considered cases where NALR worked. These pairs are assigned to the category "Best" (B: $\mathbf{c}_x \rightarrow \mathbf{c}_1$, $x = 2, 3, 4$). Points that leave $\mathbf{c}_1$ because of NALR are cases where NALR failed, thus categorized as "Worst" (W: $\mathbf{c}_1 \rightarrow \mathbf{c}_y$, $y = 2, 3, 4$). Pairs of points that stay in the same cluster are classified as "Neutral" (N: $\mathbf{c}_z \rightarrow \mathbf{c}_z$, $z = 1, 2, 3, 4$). The points that change cluster but do not leave or go to $\mathbf{c}_1$ are either classified as "Improved" (I) or "Degraded" (D) depending on whether they changed to a lower or higher entropy cluster (I: $\mathbf{c}_x \rightarrow \mathbf{c}_y$ and D: $\mathbf{c}_y \rightarrow \mathbf{c}_x$, $x < y$).

In the k-means analysis, listeners' responses are not collapsed over SNR, but they are grouped according to proximity in the vector space. Restricting to $K = 4$ clusters helps to come to statistically more meaningful results. If the responses of the same listener at the same SNR in the two experiments differ only a little, they will be grouped into the same cluster and insignificant changes are thus eliminated. When examining all 1568 cases ($4 \times 14 \times 24 = 1344$), one can see that 191 cases (14.2%) fall into the "B" category and that in 76 cases (5.68%), NALR failed ("W" category). The "N" category contains 68.7% of the cases, "I" 9.2%, and "D" 2.3%.

Token	List	Conf (+ /ɑ/)	NALR	Ears /8		$\bar{P}_e$ (%)	
	/16			FG	NALR	FG	NALR
f109gɑ	14	/d, v, b, f/	↓ˆ	0	0	46.9	36.1
m112bɑ	13	/v, f, p/	↓	3	0	41.5	28.5
f101bɑ	13	/d, g, v/	↓	3	1	37.9	35.9
f103mɑ	12	/v, n/	↑↓	2	2	26.7	18.8
f106zɑ	10	/Z, v/	=	3	3	34.4	28
f109fɑ	10	/s/	↓	1	1	31.4	18.9
m118zɑ	10	/Z, s/	↓	1	2	30.6	11.7
f101nɑ	10	/m, v/	=	1	1	17.3	5.8
f103kɑ	9	/t/	↓	2	2	26.1	27.6
f103ʃɑ	9	/s, z/	↓	0	0	9	9.5
f105ʒɑ	8	/z, S, g/	↑	2	1	40.1	32.6
f103pɑ	8	/t, k/	↓	1	0	23	20.8
f101vɑ	7	/m, f/	↓	2	1	27.4	20.3
m111gɑ	7	/d/	↓	0	0	21.5	21.4
f103sɑ	6	/f, Z/	↓ˆ	1	3	19.9	12.2
m118pɑ	6	/t/	↓ˆ	3	1	10.9	4.1
m120sɑ	5	/z/	↓	2	0	25.7	37.3
f105dɑ	4	/t/	↓	1	1	9.8	2.3
m111kɑ	3	/t/	↓	1	1	16.3	2.3
m118mɑ	3	/n/	=	0	0	9.2	2.6
f108tɑ	3	none	↓	3	1	8.7	1.6
m112tɑ	2	none	↓	2	1	5	1.3
m118ʃɑ	2	/Z, z/	↓	1	0	4.5	1
m118dɑ	1	/t/	↓	0	0	6.3	0.8

Table 2: The *List* column shows how many of the 16 ears have enough error to be taken into account for further analysis. The *NALR* column shows what happened to the entropy: ↓ down, ↑ up. The symbol ˆ indicates that NALR reduced the entropy, yet it still remained high; "=" indicates no significant change. The *Ears* column shows how many out of the 8 listeners have ears that perform differently. $\bar{P}_e$ shows the average error.

Comparison to NH subjects

Table 2 shows split up by token (i) how many ears have a sufficient number of errors in order to be considered significant, (ii) what the main confusions are for both experiments (are they consistent across ears?), (iii) how the entropy of the listeners change with NALR and (iv) for how many subjects the two ears are remarkably different (as measured by angle between the responses), (v) what the average error is for the token.

For each token it is interesting to know (i) how many ears have a sufficient number of errors in order to be considered significant, (ii) what the average error is for the token, (iii) what the main confusions are for both experiments (are they consistent across ears?), (iv) if the confusions that were made in the NALR experiment were expected (same Miller and Nicely confusion groups, expected from the normal-hearing 3DDS data of the particular token), (v) how the entropy of the listeners change with NALR, and (vi) for how many subjects the two ears are remarkably different (as measured by angle between the responses). The results for all 24 tokens are summarized in Table 2.

CONCLUSIONS

Audibility

Despite the uncommon approach of measuring CV confusions at MCL, the data demonstrates, based on the low entropy in quiet, that audibility was not an issue. Audibility is not rigorously defined. Given the results of our CV recognition experiments, we propose the use of entropy as means of defining audibility as opposed to PTA and LTASS. The following reasons further support this proposal:

1. The LTASS is irrelevant when it comes to CV perception, because CV cues are found to be bursts or frequency edges (Li, 2010; Li *et al.*, 2012), whereas the long-time speech spectrum is dominated by vowels.
2. CV perception is binary: the acoustic speech cue either can be heard or cannot be heard (Singh and Allen, 2012).
3. PTTs do not characterize the audibility of acoustic speech cues as indicated by the 3DDS method (Li, 2010). PTTs for example are an inadequate predictor of the audibility of a plosive burst, which can be much more intense than the LTASS in a critical band over a few centi-seconds (Wright, 2004).

From the reasoning stated above, it follows that a sound with 100% error ($\mathscr{H} = 0$ bit) must be audible. This is plausible since the ear must be listening to some signal properties, otherwise it would not be so consistent. On the other hand, a listener who responds randomly across all 14 consonants has $P_e = 0.93$ and $\mathscr{H} = 3.8$bits, indicating the listener cannot hear the signal. The average size of the Miller and Nicely (1955) confusion groups (/p, t, k/; /b, d, g/; /f, θ, s, ʃ/; /v, ð, z, ʒ/; /m, n/) is 3, therefore a response with 3 equally likely responses can be taken as an audibility threshold. The subject is most likely guessing when confusions outside of a known confusion group appear. In Fig´. 2 (a) the 2-bit curve representing the audibility threshold is plotted thicker. Only one point (ear 46R female /gɑ/) lies above the line, for all the other ears and tokens audibility can be assumed not to be the problem.

Effects of NALR

NALR generally, decreases the entropy (see *NALR* column in Table 2, Fig. 3 and also, the k-means result). This means the responses with NALR, show on average smaller confusion groups. The ears become more consistent in their responses, based on the decreasing standard deviation $\sigma_{\angle}$, which means the angles in the 14 dimensional space between the responses and the correct answer become more similar for all ears. In addition, the responses become closer to the correct answer, since the mean angle ($\mu_\angle$) of all ears per token decreases with NALR. Therefore, NALR not only decreases the randomness of the answers but also causes the ears to agree more on a token basis. This gives hope for training of listeners with their specific problems, since they all seem to agree on the signal they hear. Given the presented data, we have demonstrated the effectiveness of NALR using a speech test instead of pure tone tests. This suggests that a carefully constructed speech test can be used as a diagnostic tool: From the results listed in Table 2, we know all listeners for whom CV tokens cause problems and therefore can get detailed information about their hearing loss. Carefully characterized CVs can be used to find specific problems in HI subjects, that PTTs cannot.

REFERENCES

Han, W. (**2011**). *Methods for robust characterization of consonant perception in hearing-impaired listeners*. PhD thesis, University of Illinois.

Li, F. (**2010**). *Perceptual cues of consonant sounds and impact of sensorineural hearing loss on speech perception*. PhD thesis, University of Illinois at Urbana-Champaign.

Li, F., Trevino, A., Menon, A., and Allen, J.B. (**2012**). “A psychoacoustic method for studying the necessary and sufficient perceptual cues of American English fricative consonants in noise,” J. Acoust. Soc. Am., *132*, 2663-2675.

Miller, G.A., Heise, G.A., and Lichten, W. (**1951**). “The intelligibility of speech as a function of the context of the test materials,” J. Exp. Psychol., **41**, 329-335.

Miller, G.A., and Nicely, P.E. (**1955**). “An analysis of perceptual confusions among some English consonants,” J. Acoust. Soc. Am., **27**, 338-352.

Posner, J., and Ventry, I.M. (**1977**). “Relationships between comfortable loudness levels for speech and speech discrimination in sensorineural hearing loss,” J. Speech Hear. Disord., **42**, 370-375.

Shannon, C.E. (**1948**). “A mathematical theory of communication,” Bell Syst. Tech. J., **27**, 379-423.

Singh, R., and Allen, J.B. (**2012**). “The influence of stop consonants’ perceptual features on the Articulation Index model, J. Acoust. Soc. Am., **131**, 3051-3068.

Trevino, A. (**2013**). *Techniques for understanding hearing impaired perception of consonant cues*. PhD thesis, University of Illinois at Urbana-Champaign.

Walden, B.E., Holum-Hardegen, L.L., Crowley, J.M., Schwartz, D.M., and Williams, D.L. (**1983**). “Test of the assumptions underlying comparative hearing aid evaluations,” J. Speech Hear. Disord., **48**, 264-273.

Wright, R. (**2004**). “A review of perceptual cues and cue robustness,” in *Phonetically-Based Phonology*. Edited by B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press), pp. 34-57.

Zurek, P., and Delhorne, L. (**1987**). “Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment,” J. Acoust. Soc. Am., **82**, 1548-1559.

Systematic groupings in hearing-impaired consonant perception

Andrea C. Trevino* and Jont B. Allen

Beckman Institute, University of Illinois at Urbana-Champaign, IL, USA

Auditory training programs are currently being explored as a method of improving hearing-impaired (HI) speech perception; precise knowledge of a patient's individual differences in speech perception allows one to more accurately diagnose how a training program should be implemented. Remapping or variations in the weighting of acoustic cues, due to auditory plasticity, can be examined with the detailed confusion analyses that we have developed at UIUC. We show an analysis of the responses of 17 ears with sensorineural hearing loss to consonant-vowel stimuli, composed of 14 English consonants followed by the vowel /ɑ/, presented in quiet and speech-shaped noise. Although the tested tokens are noise-robust and unambiguous for normal-hearing listeners, the subtle natural variations in signal properties can lead to systematic differences for HI listeners. Specifically, our recent findings have shown token-dependent individual variability in error and confusion groups for HI listeners. A clustering analysis of the confusion data shows that HI listeners fall into specific groups. Many of the token-dependent confusions that define these groups can also be observed for normal-hearing listeners, under higher noise levels or filtering conditions. These HI-listener groups correspond to different acoustic-cue weighting schemes, highlighting where auditory training should be most effective.

INTRODUCTION

One of the primary goals of auditory training techniques is improving the consonant recognition of listeners with sensorineural hearing loss. Training has been shown to be effective treatment in terms of both consonant and word recognition; the work of Boothroyd and Nittrouer (1988) and Bronkhorst *et al.* (1993) generalizes these results by demonstrating how the perception of individual phones and low-context syllables predicts the perception of words and sentences. Although significant improvements can be observed from both analytic and synthetic training (Sweetow and Palmer, 2005), the effects are difficult to measure and are most easily observed for listeners with the most pre-training recognition error (Walden *et al.*, 1981). Analysis of the effects of training tends to focus on discrimination ability and overall error; the effects on consonant confusions would provide an additional dimension to the analysis, often without the collection of additional data.

In general, auditory training methodologies do not focus on the listener-specific

*Corresponding author: atrevin2@illinois.edu

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

consonant recognition deficiencies (i.e., individual differences) that are present prior to the training period. Although an identical, overarching approach is desirable when initially assessing the efficacy of a training scheme, it may not be the most beneficial for providing treatment to the patient population. Our previous works (Trevino and Allen, 2013a,b) have shown that patients with mild-to-moderate hearing loss have consonant recognition errors that are usually limited to a small subset of test consonant-vowel tokens. This indicates that, for maximum efficacy and efficiency, a targeted approach is necessary in the implementation of training programs. In addition, we have explored the significant effects of talker variability on HI perception, particularly across tokens of the same consonant (i.e., within-consonant perceptual differences). These within-consonant differences, again, highlight the need for a targeted, patient-specific approach, as well as the importance of considering token variability in the analysis of perceptual data.

The confusion matrix has been the fundamental basis for analyzing consonant recognition data for over 50 years (Miller and Nicely, 1955). In this paper, we introduce a technique, k-means clustering based on the Hellinger distance, for analyzing similarity of consonant confusions. This analysis is performed on a token-by-token basis, as recommended in the conclusions of our previous works on within-consonant HI perceptual differences (Trevino and Allen, 2013a,b). A more precise understanding of how HI listeners are using the acoustic cues that are available to them provides a detailed diagnosis, which could be used to refine the implementation of auditory training programs.

METHODS

Subjects

Nine subjects with sensorineural hearing loss were recruited for this study from the Urbana-Champaign, IL community. All subjects reported American English as their first language and were paid to participate. Typanometric measures showed no middle-ear pathologies (type A tympanogram). The ages of eight HI subjects ranged from 65 to 84; one HI subject (14R) was 25 years old. Based on the pure-tone thresholds, all ears had $>$ 20 dB of hearing loss (HL) for at least one frequency in the range 0.25-4 kHz.

The majority of the ears in our study have slight-to-moderate hearing loss with high-frequency sloping configurations. One HI ear (14R), has an inverted high-frequency loss, with the most hearing loss $<$ 2 kHz and a threshold within the normal range at 8 kHz. For further listener details, including level of hearing loss, age, and most comfortable level, see Trevino and Allen (2013a,b).

Speech materials

All stimuli used in this study were selected from the Linguistic Data Consortium Database (LDC-2005S22). Speech was sampled at 16 kHz. Fourteen naturally-spoken

American English consonants (/p, t, k, f, s, ʃ, b, d, g, v, z, ʒ, m, n/) were used as the test stimuli. Each consonant was spoken in an isolated consonant-vowel (CV) context, with the vowel /ɑ/. Two tokens were selected (1 male and 1 female talker) for each consonant, resulting in a total of 28 test tokens (14 consonants × 2 talkers = 28 tokens). The term *token* is used throughout this work to refer to a single CV speech sample from one talker.

The 28 test tokens were selected based on their NH perceptual scores in quiet and speech-weighted noise. To ensure that tokens were unambiguous and robust to noise, each token was selected based on a criterion of $\leq 3.1\%$ error for a population of 16 NH listeners, calculated by combining results in quiet and -2 dB signal-to-noise ratio (SNR) of noise (i.e., no more than 1 error over a total N=32, per token) (Phatak and Allen, 2007). Such tokens are representative of the LDC database; Singh and Allen (2012) shows, for the majority of tokens, a ceiling effect for NH listeners ≥ -2 dB SNR. One token of /fɑ/ (male talker, label m112) was damaged during the preparation of the tokens, thus it has not been included in this analysis.

The stimuli were presented with flat gain at the *most comfortable level* (MCL) for each individual HI ear. For the majority of the HI ears the MCL was approximately 80±4 dB SPL; only two subjects did not choose an MCL within this range (36L/R chose 68/70 dB SPL and 14R chose 89 dB SPL).

Experimental procedure

The speech was presented at 4 SNRs (0, 6, 12 dB, and quiet) using speech-weighted noise, generated as described by Phatak and Allen (2007). Presentations were randomized over consonant, talker, and SNR. The total number of presentations for each consonant ranged from N = 40-80 for each HI ear (total N = 5-10 over two adaptive phases × 2 tokens × 4 SNRs). The Vysochanskiï–Petunin inequality was used to verify that the number of trials was sufficient to determine correct perception within a 95% confidence interval, as described in the appendix of Singh and Allen (2012).

All of the data-collection sessions were conducted with the subject seated in a single-walled, sound-proof booth. The speech was presented monoaurally via an Etymotic ER-3 insert earphone. The contralateral ear was not masked or occluded. The subject chose their MCL (for non-test speech samples) before testing began. A practice session, with different tokens from those in the test set, was run first in order to familiarize the subject with the testing paradigm and to confirm their MCL setting. After hearing a single presentation of a token, the subject would choose from the 14 possible consonant responses by clicking one of 14 CV-labeled buttons on the graphical user interface, with the option of up to 2 additional token repetitions, to improve accuracy. Additional experimental details are provided in Han (2011) and Trevino and Allen (2013a,b).

Data analysis

The variability of naturally-spoken acoustic cues can lead to HI within-consonant differences in both error and consonant confusions (Trevino and Allen, 2013a,b); therefore, calculations at the token level are necessary in any analysis that attempts to understand how a HI listener is using and interpreting the acoustic cues that are available to them. In this paper, the data are analyzed at the token level, with individual data points for the HI ears.

The Hellinger distance is a metric for computing the distance between two probability distributions. The probability distributions that we compare in this paper are the ones defined by each row of a confusion matrix. In the case of this experiment, there are 14 possible consonant responses. This vector of probabilities can be considered as a point in 14-dimensional space, where each dimension corresponds to each possible consonant response. Distances between confusion results are computed within this 14-dimensional space; the distances provide a measure of consonant-confusion similarity, which can be used to compare HI ears, SNRs, or tokens.

We will show that the squared Hellinger distance is equivalent to 1 minus the direction cosine, when computed from the square root of probabilities. This relationship allows us to use widely-known algorithms that employ 1 minus the direction cosine, such as spherical k-means clustering, to analyze the data. Let $P_{r|s}(snr, HI)$ be the probability of the consonant response r for a fixed stimulus s, SNR, and HI ear; the probabilities for all possible responses for a fixed stimulus would be a row of the confusion matrix. A data point in the 14-dimensional space, $\mathbf{x}$, is then defined as $x_i = \sqrt{P_{r_i|s}(snr, HI)}$, $i = 1, 2, 3, \ldots 14$. Since the vector is composed of probabilities that sum to 1, the points lie on the unit sphere, $||\mathbf{x}|| = 1$. Let $\mathbf{x}, \mathbf{y}$ be two data points in the 14-dimensional space. We define the notation for an inner product as

$$< \mathbf{x}, \mathbf{y} >= \sum_i x_i y_i$$

and the norm as

$$< \mathbf{x}, \mathbf{x} >= ||\mathbf{x}||^2 = \sum_i x_i^2.$$

Then the square of the Hellinger distance

$$\begin{aligned} H^2(\mathbf{x}, \mathbf{y}) &= \frac{1}{2}||\mathbf{x} - \mathbf{y}||^2 = \frac{1}{2}(||\mathbf{x}||^2 - 2 < \mathbf{x}, \mathbf{y} > + ||\mathbf{y}||^2) \\ &= 1 - < \mathbf{x}, \mathbf{y} >= 1 - ||\mathbf{x}||||\mathbf{y}||cos(\Theta_{xy}) = 1 - cos(\Theta_{xy}). \end{aligned}$$

Thus, the spherical k-means algorithm, which forms groups based on $1 - cos(\Theta_{xy})$ between points distributed on the unit sphere, produces results that also minimize the Hellinger distance. The spherical k-means clustering algorithm is implemented in MATLAB, with the *kmeans()* function. For each token, one of the clusters is always composed of the data points where HI listeners correctly perceived the consonant; the remaining clusters are composed of the data with varying degrees of error. Therefore, assuming there are errors, the minimum possible K for a token is 2.

Additionally, the angle between the HI listener response x and the plane representing the 'primary' confusion groups can be calculated. With this implementation, HI-listener data that contain varying degrees of the same primary confusions would show zero distance between the points; non-zero distances would indicate the degree of deviation from the primary confusion group.

The k-means algorithm groups HI-listener data that are similar in terms of the confusions. The size and number of clusters is a function of the diversity of hearing impairment across listeners in the study (i.e., there is no fixed prior), therefore, a k-means implementation which does not assign a prior probability to each cluster models the experimental setup more realistically than a Gaussian Mixture Model (GMM). The G-means algorithm (Hamerly and Elkan, 2004) was added to the implementation in order to automatically select the number of means, K, based on an Anderson-Darling test of statistical significance.

RESULTS

For a fixed consonant token, HI listeners vary widely in both the degree of error and the SNR threshold at which errors begin to occur. Despite this individual variability, we have observed that different HI ears tend to have similar token-dependent confusions once an error is made (Trevino and Allen, 2013b). If HI listeners generally share a similar confusion group for a particular token, then an auditory training scheme that corrects for this confusion should be effective for a broad population of patients. In order to explore the extent of the similarities across HI listeners, we use the spherical k-means clustering algorithm to group the listeners based on confusions. The data at all tested SNRs is used together in the k-means clustering analysis, since the different severities of hearing impairment across the many listeners leads to errors at different SNRs.

Each cluster identified by the k-means algorithm is composed of listeners with similar consonant confusions. The number of clusters, K, for each token is determined by the G-means algorithm, which selects K iteratively based on a statistical test of the cluster distributions (Hamerly and Elkan, 2004). As a result of incorporating the statistical test, the number of resulting clusters K is the amount of significantly different confusion groups that are present in our data. For example, the case of two resulting clusters, $K = 2$, indicates that all of the listener data are distributed within the cluster of correct-response data points and a second cluster defined by a single confusion group. From the results in Table 1, we see that 17 out of the 27 tokens have $K \leq 3$, indicating that all of the HI data for these tokens fall into one of 3 confusion-based clusters. 22 out of 27 tokens have $K \leq 4$. This small number of clusters for the majority of tokens indicates that, generally, only a few token-dependent confusion groups are present in the HI data.

For each cluster, the primary confusions that define the k$^{\text{th}}$ mean, along with the number N of data points within each cluster, are included in Table 1. Results for the cluster of ‘correct’ responses (i.e., the cluster of data with no more than 1 error over 5-10 trials) are also included. From the results in Table 1, we see that the confusions that define the clusters can vary across tokens of the same consonant. For example, /d, g, v/ confusions are present for the female /bɑ/ token, while only /v/ confusions dominate the responses for the male /bɑ/ token. In addition, the large number of data points, N, in the ‘correct’ clusters of all tokens indicates that the mild-to-moderate HI listeners in this study did not have widespread errors. These are observations that have been made previously in Trevino and Allen (2013b); this analysis shows how these observations can also be made from the results of k-means clustering.

The extent of the similarity across listener responses can be quantified by the angle between the points in the spherical vector space. These angles can be expressed as direction cosines or Hellinger distances, as described in the Methods section, and can range from 0°to 90°. The angle Θ_{x,μ_k} between a data point $\mathbf{x}$ in the k$^{\text{th}}$ cluster and the k$^{\text{th}}$ cluster mean μ_k provides a measure of how well each mean represents the overall group of data points. The average of this measure, $\widehat{\Theta}_{x,\mu_k}$, is analogous to the variance within each cluster. Results for $\widehat{\Theta}_{x,\mu_k}$ are shown in Table 1. For reference, when each data point $\mathbf{x}$ is the result of 5-10 presentations, as ours are, an angle of 18°-27°lies between a vector of correct responses and a vector with a single incorrect response. Overall, the clusters defined by a larger number of primary confusions have larger $\widehat{\Theta}_{x,\mu_k}$ values. Systematic groupings of HI data in terms of consonant confusions is observed for all the tested tokens.

DISCUSSION

Our past studies (Trevino and Allen, 2013a,b) have found that HI listeners with mild-to-moderate hearing loss make errors with only a small subset ($< 25\%$) of listener-dependent consonant tokens at low noise levels, although the error for these tokens can be as high as chance performance. In addition, we observed significant individual variability across HI ears in terms of the degree of error and which sounds are perceived in error, despite similar hearing thresholds. These findings verify the need for an individualized approach when implementing an auditory training program. Based on our data, an individualized auditory training program would, ideally, first identify the sounds/acoustic cues that a HI listener has difficulty with in quiet and low-levels of noise, in order to focus the training appropriately. In addition, this initial test would provide a precise outcome measure after the training is completed. A test that identifies a HI listener’s difficulties in terms of identifying and interpreting acoustic cues would be ideal when prescribing such a training program. A context-free, high-entropy (i.e., large response set), consonant identification task paired with a token-level analysis allows one to identify the specific acoustic cue-processing difficulties of each HI individual.

We have introduced k-means clustering as a flexible tool for analyzing confusion

CV	**k$^{\text{th}}$ Mean (N)**	$\widehat{\Theta}_{x,\mu_k}$		**CV**	**k$^{\text{th}}$ Mean (N)**	$\widehat{\Theta}_{x,\mu_k}$
bɑ$_{F101}$	k_1 : correct (39)	12°		**bɑ**$_{M112}$	k_1 : correct (32)	15°
$K=2$	k_2 : b, d, g, v (29)	36°		$K=4$	k_2 : b, v (21)	27°
					k_3 : b, v (9)	19°
dɑ$_{F105}$	k_1 : correct (61)	10°		**dɑ**$_{M118}$	k_1 : correct (61)	10°
$K=3$				$K=2$	k_2 : d, g, t (7)	25°
fɑ$_{F109}$	k_1 : correct (39)	14°		-		
$K=2$	k_2 : f, s, v (29)	34°				
gɑ$_{F109}$	k_1 : correct (35)	8°		**gɑ**$_{M111}$	k_1 : correct (54)	10°
$K=2$	k_2 : b, d, f, g, v (33)	48°		$K=4$		
kɑ$_{F103}$	k_1 : correct (50)	11°		**kɑ**$_{M111}$	k_1 : correct (56)	9°
$K=3$	k_2 : k, p, t (11)	25°		$K=2$	k_2 : k, t (12)	23°
	k_3 : t (7)	22°				
mɑ$_{F103}$	k_1 : correct (46)	11°		**mɑ**$_{M118}$	k_1 : correct (61)	9°
$K=3$	k_2 : m, v (12)	28°		$K=2$	k_2 : m, n, v (7)	16°
	k_3 : m, n (10)	26°				
nɑ$_{F101}$	k_1 : correct (52)	10°		**nɑ**$_{M118}$	k_1 : correct (43)	12°
$K=4$	k_2 : m, n (9)	25°		$K=4$	k_2 : m, n (15)	4°
pɑ$_{F103}$	k_1 : correct (59)	13°		**pɑ**$_{M118}$	k_1 : correct (61)	12°
$K=6$				$K=2$	k_2 : f, p, t, z (7)	35°
sɑ$_{F103}$	k_1 : correct (55)	11°		**sɑ**$_{M120}$	k_1 : correct (45)	11°
$K=3$	k_2 : s, ʒ, z (7)	26°		$K=5$	k_2 : s, z (11)	10°
tɑ$_{F108}$	k_1 : correct (61)	6°		**tɑ**$_{M112}$	k_1 : correct (62)	6°
$K=2$	k_2 : f, p, s, t (7)	40°		$K=2$		
vɑ$_{F101}$	k_1 : correct (43)	11°		**vɑ**$_{M118}$	k_1 : correct (29)	14°
$K=3$	k_2 : f, v (15)	32°		$K=7$	k_2 : p, v (12)	25°
	k_3 : b, d, m, n, v (10)	38°			k_3 : m, n, v (11)	28°
ʃɑ$_{F103}$	k_1 : correct (60)	7°		**ʃɑ**$_{M118}$	k_1 : correct (65)	6°
$K=2$	k_2 : s, ʃ, z (8)	24°		$K=2$		
ʒɑ$_{F105}$	k_1 : correct (42)	11°		**ʒɑ**$_{M107}$	k_1 : correct (36)	13°
$K=4$	k_2 : z (16)	18°		$K=3$	k_2 : g, ʒ, z (17)	32°
					k_3 : v, ʒ, z (15)	38°
zɑ$_{F106}$	k_1 : correct (35)	14°		**zɑ**$_{M118}$	k_1 : correct (38)	14°
$K=7$	k_2 : ʒ, z (11)	9°		$K=6$	k_2 : ʒ, z (11)	18°
	k_3 : s, ʒ, z (8)	19°			k_3 : v, ʒ, z (9)	20°

Table 1: Clustering results for 27 CV tokens. Talker gender and identification number are indicated by the CV subscript. The resulting total number of clusters K is included in the CV column. Each row shows the data for a single cluster; to focus on clusters with similar listeners, clusters with less than 6 data points are omitted. The main confusions comprising the k^{th} cluster means ($> 5\%$) are listed under k^{th} Mean (N), with N being the number of data points within each cluster (out of 68 total). Similarities across HI ears within a cluster are quantified by the average angle between the members of each cluster and the k$^{\text{th}}$ mean, $\widehat{\Theta}_{x,\mu_k}$.

matrix data. Such a clustering analysis can be conducted without averaging across tokens, consonants, SNRs or HI ears. The k-means clusters of HI data correspond to different acoustic cue-weighting schemes and indicate where auditory correction or training may be useful. Although there are many individual differences across HI listeners, the small number of resulting clusters from the analysis of our data shows that the listeners are processing and interpreting the acoustic cues that are present in speech similarly. These results suggest that, once the sounds that are difficult for a HI listener are diagnosed by a speech test, a common cue-correction scheme can be effective for a broad population of listeners.

REFERENCES

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am., **84**, 101-114.

Bronkhorst, A.W., Bosman, A.J., and Smoorenburg, G.F. (**1993**). "A model for context effects in speech recognition," J. Acoust. Soc. Am., **93**, 499-509.

Hamerly, G., and Elkan, C. (**2004**). "Learning the k in k-means," Adv. Neur. In., **16**, 281-288.

Han, W. (**2011**). *Methods for robust characterization of consonant perception in hearing-impaired listeners*. PhD thesis, University of Illinois, Urbana-Champaign.

Miller, G.A., and Nicely, P.E. (**1955**). "An analysis of perceptual confusions among some english consonants," J. Acoust. Soc. Am., **27**, 338-352.

Phatak, S.A., and Allen, J.B. (**2007**). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am., **121**, 2312-2326.

Singh, R., and Allen, J.B. (**2012**). "The influence of stop consonants' perceptual features on the articulation index model," J. Acoust. Soc. Am., **131**, 3051-3068.

Sweetow, R., and Palmer, C.V. (**2005**). "Efficacy of individual auditory training in adults: a systematic review of the evidence," J. Am. Acad. Audiol., **16**, 494-504.

Trevino, A., and Allen, J.B. (**2013a**). "Individual variability of hearing-impaired consonant perception," in *Seminars in Hearing*, Vol. 34 (Thieme Medical Publishers) pp. 74-85.

Trevino, A., and Allen, J.B. (**2013b**). "Within-consonant perceptual differences in the hearing impaired ear," J. Acoust. Soc. Am., **134**, 607-617.

Walden, B.E., Erdman, S.A., Montgomery, A.A., Schwartz, D.M., and Prosek, R.A. (**1981**). "Some effects of training on speech recognition by hearing-impaired adults," J. Speech Lang. Hear. Res., **24**, 207-216.

The benefit of cochlear-implant users' head orientation to speech intelligibility in noise

JACQUES A. GRANGE* AND JOHN F. CULLING

Cardiff University, School of Psychology, Cardiff CF103AT, United Kingdom

Speech reception thresholds (SRTs) in noise improve when the speech and noise sources are spatially separated. This spatial release from masking (SRM) is usually investigated in fixed-head situations. We studied free-head situations in audio and audio-visual conditions. We compared normally-hearing and cochlear-implant (CI) users' spontaneous and directed head-orientation strategies when attending to speech in noise with a progressively declining signal-to-noise ratio. SRM-model predictions suggested benefits of head orientation away from the target speech that we hypothesized would motivate head rotation. As signal-to-noise ratio declined, observed head tracks differed greatly between listeners. Audio-visual presentation reduced the amount of head rotation. When directed, listeners made more effective use of head rotation. Audio and audio-visual SRTs were acquired at fixed, 0, and 30 deg head orientations with respect to the target speech. At the most favourable 30-deg head orientation, SRM reached 8 and 6 dB for NH listeners and CI users respectively. Lip-reading yielded improvements of 3 and 5 dB on average across conditions. CI users confirmed that training in optimizing both their position and head orientation with respect to target speaker and noise source position in a social setting was both currently missing and likely valuable.

INTRODUCTION

Bilateral cochlear implantation provides service users with several benefits over unilateral implantation. In addition to sound-source localization being made possible to some extent, Van Hoesel and Tyler (2003) showed that bilateral cochlear-implant users (BCIs) benefit from improved speech intelligibility in noise (SpIN) when speech and noise sources are spatially separated. However most studies to date have considered such spatial release from masking (SRM) in a fixed-head situation (e.g., Van Hoesel and Tyler, 2003; Litovsky *et al.*, 2006; Loizou *et al.*, 2009). Furthermore and with few exceptions, most examined SRM by comparing speech co-located with noise in front of the listener with speech in front and noise azimuthally separated by 90 deg to the left or to the right, configurations known to not make optimum use of the head-shadow effect due to the bright spots located at ±90 deg. Our model of SRM (Jelfs *et al.*, 2011) predicted the spatial configuration providing the maximum benefits of bilateral over unilateral implantation, later confirmed by Culling *et al.* (2012) with normally hearing-listeners (NHs) and cochlear-implant users (CIs). The SRM model could also be used to predict how

*Corresponding author: grangeja@cf.ac.uk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

head orientation away from facing the speech could yield improved SpIN. From the conclusions of Culling *et al.* (2012) one could guide CIs with respect to their optimal seating strategy. Seating options in a restaurant could however be limited and the only degree of freedom left would be head orientation.

In a first experiment we examined in a sound-deadened room whether the model predictions translated to CIs adopting effective spontaneous free-head orientation strategies. We also established whether simple guidance could immediately make a large difference in the lowest speech-to-noise ratio (SNR) they could successfully reach. A baseline was established with NHs. Trials were conducted in audio-only or audio-visual (AV) conditions to measure the impact of lip-reading.

In another two experiments with the same participants, in the same room and spatial configurations, fixed-head speech reception thresholds (SRTs) were measured as well as their improvement by a modest, 30 deg head orientation away from the speech-facing direction. This enabled direct validation of the model predictions without compromising lip-reading, 30 deg being thought as an acceptable gaze angle for lip-reading purposes. In the BCIs case we also measured summation and squelch (Schleich *et al.*, 2004).

MODEL OF SRM

The model of SRM originally introduced by Lavandier and Culling (2010) takes as input speech-shaped noise, which has been convolved with binaural-room-impulse-response recordings to create a reverberant speech-shaped noise target and interferer.

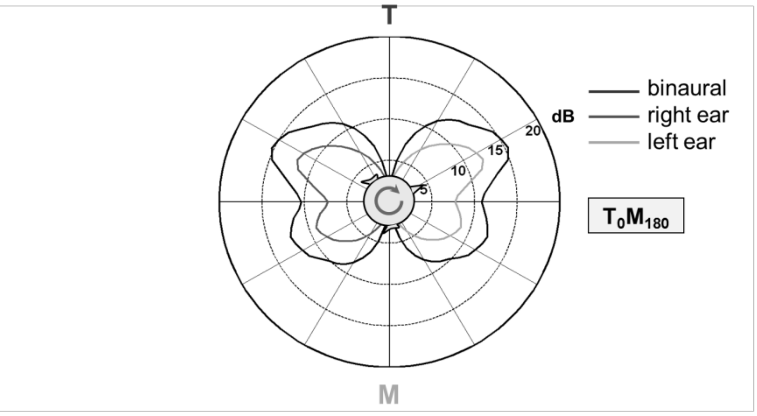


Fig. 1: Predicted head-orientation benefit (anechoic condition) with target in front and masker at 180 deg.

A first path calculates the expected binaural advantage due to binaural unmasking using equalization-cancellation theory to predict the binaural masking level difference. A second path predicts the benefits of better-ear listening or head-shadow effect. Combined, the two paths account for the two cues associated with SRM (Bronkhorst and Plomp, 1988). The two outputs are simply added to generate an SRM prediction. Jelfs *et al.* (2011) refined the model by enhancing its computational efficiency. Culling *et al.* (2012) made use of the model to predict SRM for NHs and CIs with one speech-shaped interfering noise. Figure 1 illustrates the model predictions as a function of head orientation in anechoic conditions. The two inner curves are predictions for left or right ear alone. Because bilateral CI users do not benefit from binaural unmasking, the head-orientation benefit they would experience is the outermost these two curves, i.e., the benefits of better-ear listening. The outer curve is a NH prediction which includes binaural unmasking. Up to 12 and 16 dB benefit is predicted at ±60 degrees head orientation for CIs and NHs respectively.

MATERIALS AND METHODS

Participants

12 NH participants aged from 18 to 22 (age mean: 20) were recruited from the undergraduate Cardiff University population. 9 CIs aged 35 to 72 (age mean: 62) participated, of which 5 were bilateral CIs and 4 were unilateral CIs. CIs were recruited through the UK National CI User Association and used implants from a mix of manufacturers (Cochlear, MedEl, and Advanced Bionics). All CIs had had their last implant fitted 2+ years before being tested.

Laboratory setup

Two sound-deadened rooms were used and acoustically matched. As schematically shown in Fig. 2, 4 × Cambridge Audio Minx loudspeakers were arranged at cardinal positions around a circle of radius 1.3 m centered on the listener's head, driven by a 6-channel Auna solid-state amplifier and an ESI MAYA44+ I/O sound card. A 17-inch screen was positioned below the front speaker, through which the target speech was always presented. A shaving mirror was used to assist listeners in adopting the correct head orientations during the SRT runs. The RT60 of the rooms was derived from impulse measurements to be circa 60 ms. A webcam fitted on the ceiling above the listener's head enabled covert video recording and subsequent extraction of participant head-tracks.

Stimulus presentation and preparation

The speech and noise were presented either directly by Matlab and Playrec or through the VideoLAN player. A set of 320 high-predictability-SPIN-sentence audio-visual clips were recorded to measure the impact of lip-reading. The reading of sections of The Wonderful Wizard of Oz by L. Frank Baum was video-recorded for the free-head task, a material chosen for its predictability. 570 sentences from the Harvard IEEE sentence corpus were used for more precise audio SRT measurements. All audio material was sampled or re-sampled at 44.1 kHz and rms-normalized.

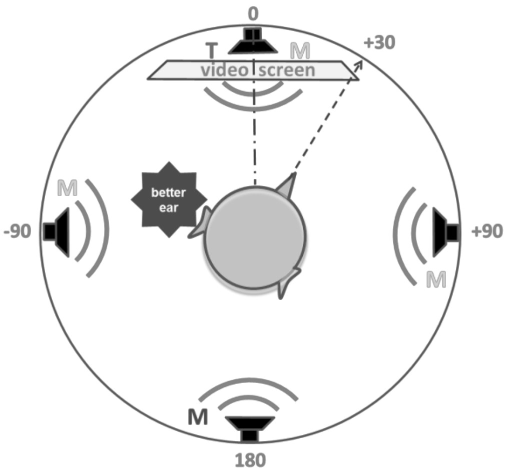


Fig. 2: Highlighted (darker masker or target markers M or T) are the T_0M_{180} spatial configuration and $H_{30}M_{180}$ global configuration.

Spatial configurations

The model predicted that maximum SRM gain could be obtained with target at 0 deg and masker at 180 deg azimuths respectively (the T_0M_{180} configuration). Informed by a prior NH study and given most previous studies tested for SRM with speech in front and masker at ±90 deg, tests were conducted in 16 combinations of spatial configuration, head orientation, and presentation modality: spatial configuration with target and masker at 0 deg vs. target at 0 deg and masker at ±90 deg or 180 deg; head either facing the speech or with the head rotated by ±30 deg; audio or AV. A 90 deg masker separation or 30 deg head turn favoured the listener's best performing ear for speech perception in noise. The model predicted that a favourable 30 deg head turn would provide either the bulk of the attainable audio SRM in the T_0M_{180} configuration or the maximum attainable SRM in the favourable T_0M_{90} configuration. The T_0M_{90} configuration would allow us to correlate our new audio-only data with the Culling *et al.* (2012) study and other prior studies. The T_0M_{180} would maximize benefit of head rotation according to model predictions. T_0M_0 acted as a reference for all other SRT data or as control in the free-head experiment. Spatial configuration and head angle were combined for simplicity into a global configuration code such as the $H_{30}M_{180}$ configuration highlighted in Fig. 2.

RESULTS AND DISCUSSION

Experiment 1: Free-head listening task

The central plot of Fig. 3 is the outcome of the undirected, audio-only condition. SNR at source (proportional to time) is presented radially and head orientation azimuthally. To the left, the change of behaviour reflects the effect of lip-reading; to the right how much or how quickly a listener can learn to make use of head orientation. Arrows highlight the range of optimum orientations. For the sake of brevity only the T_0M_{180} UCIs plots are included here. Each line corresponds to a given participant's head track; a circle is positioned at the SNR/time corresponding to the last 3-5 words correctly understood by the participant, corresponding to a self-reported SRT-50 for their final head angle.

Contrary to 45% of NH listeners who made spontaneous use of head orientation, only 10% of CIs appeared to turn their heads (undirected, audio-only) and the presence of the visual cues totally eradicated head turns. Once directed, most CIs achieved a very significant improvement, most reaching the optimum SRM orientation(s). No CI turned their heads when the speaker's face was visible and all reached between 5 and 10 dB more intelligibility when directed. 90% of CIs reached the very best performance by combining head orientation and lip-reading in the AV directed task. Most reached 20-25 deg away from the speech direction.

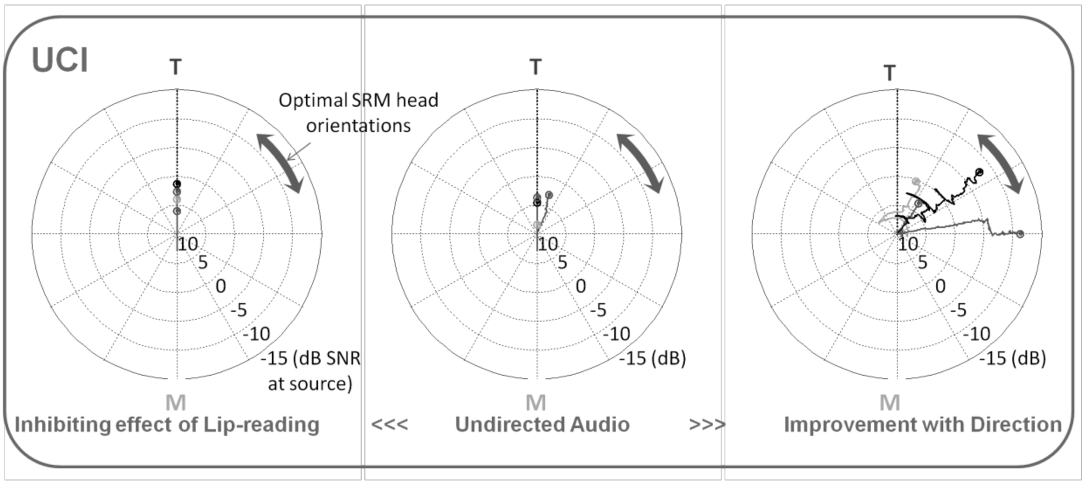


Fig. 3: UCI example of head orientation tracks.

Experiment 2: Gain of lip-reading

In each spatial configuration and making use of the SPIN sentence material, the SRTs acquired in audio-only were subtracted from SRTs acquired in the AV mode. Figure 4 shows the benefit that lip-reading provided over and above SRM benefits. NH's and CI's lip-reading gains were respectively found to be 3 and 5 dB on average across conditions. Moreover and most importantly, lip-reading was confirmed to be beneficial in all configurations and a 30 deg head turn did not significantly and adversely affect lip-reading.

Experiment 3: SRM, summation, and squelch findings

Figure 5 presents on the right hand side the SRM results obtained for NHs and BCIs. Dotted and dashed lines are the model predictions for each group, continuous lines and symbols are means across participants. The error bars reflect the standard error of the means.

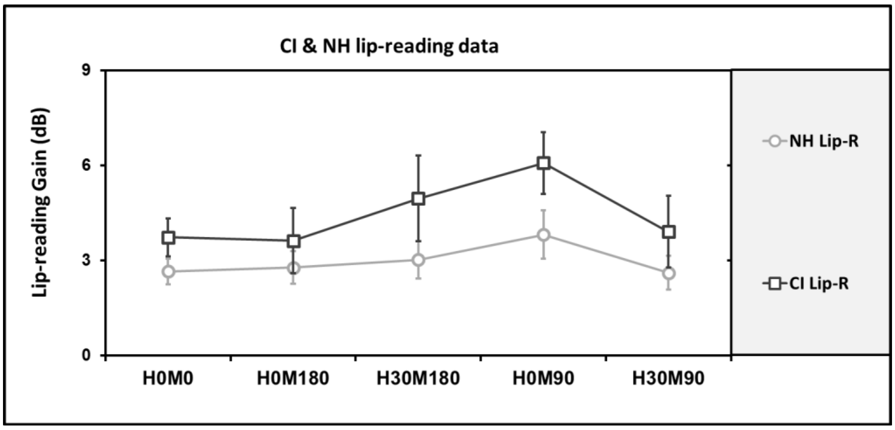


Fig. 4: NHs and CIs lip-reading gain.

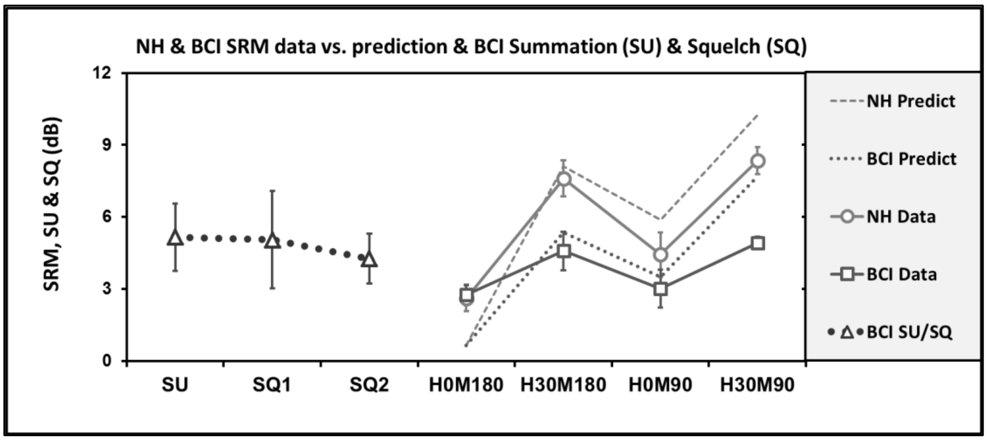


Fig. 5: NHs SRM; BCIs SRM, summation and squelch.

The predicted benefit of head rotation from H_0M_{180} to $H_{30}M_{180}$ is noticeably larger for NHs (7.4 dB) than for BCIs (4.3 dB) and so solely due to binaural unmasking. NHs and BCIs all benefitted from a 30-deg head turn in both spatial configurations by up to 4.5 dB and 1.9 dB for NHs and BCIs, respectively. The lower than predicted CI 30-deg head-turn gain may be attributable to the variety of microphone positions in the various implants used by our participants; the further away from the KEMAR manikin's microphone position in the ear canal, the larger the effect. The discrepancy is indeed expected to be attributable to variations in head-shadow effect. On the left hand side of Fig. 5, summation and squelch data are displayed. These are the SRT improvement between the best ear/implant performance (second implant disabled) and performance with both cochlear implants on. The summation is calculated from H_0M_0 SRT data whereas the squelch is calculated from the H_0M_{90} or $H_{30}M_{180}$ as per Schleich *et al.* (2004). Both summation (5.2 dB) and squelch (4.2 and 5 dB) mean values are far superior to the 1-2 dB reported elsewhere (e.g., Schleich *et al.*, 2004). The two BCIs showing largest summation and squelch reported large differences in the character of the sound perceived from each implant, suggesting a spectral summation effect rather than any binaural unmasking.

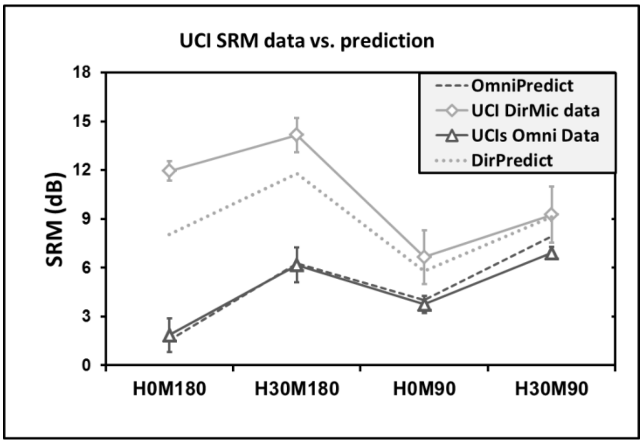


Fig. 6: SRM data vs. predictions for UCIs, omni and directional microphones.

Figure 6 illustrates the quality of match between predictions and data for the omnidirectional microphone UCIs mean SRMs. All means were indeed within 1 dB of predictions. 4.5 dB was gained from a favourable 30 deg head turn.

One UCI used a directional microphone setting, which had the effect of boosting their SRM by over 10 dB in the T_0M_{180} configuration. The relatively small discrepancy (3 dB) between data and omnidirectional predictions with a 90-deg masker separation however shows that these directional settings are not so influential for sound towards the front. New directional predictions were generated by adding the computed difference between directional and omnidirectional anechoic predictions (Cochlear HRIRs) to the omnidirectional sound-deadened room predictions. The new predictions fitted the data overall much better and, in terms of benefit of head rotation, the directional data fitted predictions within 2 dB or so. All benefits of 30-deg head-turns tested were statistically shown to be significant. *t*-tests performed between configuration pairs resulted in *p*-values of 0.03-0.04 for BCIs , $p < 0.01$ for UCIs and $p < 0.001$ for NHs across participants.

CONCLUSIONS

This study demonstrates how a modest and therefore socially acceptable head orientation away from a speaker can provide a significant benefit in understanding speech in noise. The single steady-noise masker situation studied here enables analysis of the fundamental benefit of combining optimum positioning in a room with optimum head orientation, without compromising lip-reading. With the bulk of the noise coming from the side or the rear of the listener, a head orientation of 30 deg is shown to provide a SpIN benefit between 2 and 5 dB for cochlear-implant users without disrupting lip reading. This is a welcome, potentially significant improvement in their challenging speech-in-noise listening situation. Although testing in more reverberant environments and with multiple talker interferers would more realistically mimic a social situation such as a restaurant, this simpler approach demonstrates fundamental benefits. This study also demonstrates how quickly (within minutes after guidance is provided) CI users can learn to reap the benefits a head-orientation strategy can provide. This shows how easily CIs could benefit from simple training. In that, this study has an immediate translational application.

ACKNOWLEDGMENTS

With special thanks to Rosie Lovett and UCL for access to a Chandler House lab.

REFERENCES

Bronkhorst, A., and Plomp, R. (**1988**). “The effect of head-induced interaural time and level differences on speech intelligibility in noise,” J. Acoust. Soc. Am., **83**, 1508-1516.

Culling, J.F., Jelfs, S., Talbert, A., Grange, J.A, and Backhouse, S.S. (**2012**). “The benefit of bilateral versus unilateral cochlear implantation to speech intelligibility in noise,” Ear Hearing, **33**, 673-682.

Jelfs, S., Culling, J.F., and Lavandier, M. (**2011**). “Revision and validation of a binaural model for speech intelligibility in noise,” Hear. Res., **275**, 96-104.

Lavandier, M., and Culling, J.F. (**2010**). “Prediction of binaural speech intelligibility against noise in rooms,” J. Acoust. Soc. Am., **127**, 387-399.

Litovsky, R.Y., Parkinson, A.J., Arcaroli, J., and Sammeth, C. (**2006**). “Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study,” Ear Hearing, **27**, 714-731.

Loizou, P.C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., and Roland, P. (**2009**). “Speech recognition by bilateral cochlear implant users in a cocktail-party setting,” J. Acoust. Soc. Am., **125**, 372-383.

Schleich, P., Nopp, P., and D’Haese, P. (**2004**). “Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant,” Ear Hearing, **25**, 197-204.

Van Hoesel, R., and Tyler, R. (**2003**). “Speech perception, localization, and lateralization with bilateral cochlear implants,” J. Acoust. Soc. Am., **113**, 1617-1630.

Validation of a spatial speech-in-speech test that takes signal-to-noise ratio (SNR) confounds into account

SØREN LAUGESEN*, FILIP MARCHMAN RØNNE, NIELS SØGAARD JENSEN,
AND MARIA GRUBE SORGENFREI
Eriksholm Research Centre, Oticon A/S, Rørtangvej 20, 3070 Snekkersten, Denmark

A Spatial Fixed-SNR (SFS) speech-in-speech intelligibility test is presented and the reliability and validity of the test is investigated. As part of the validation the SFS test was used to compare a linear hearing-aid setting to a setting with aggressive compression limiting. Two sub-groups of listeners were tested in a fixed-SNR paradigm at –5 and +5 dB SNR, respectively.

INTRODUCTION

Measuring speech-reception threshold (SRT) using adaptive procedures is popular, as testing yields results at the steepest, most sensitive part of the psychometric functions of individual test subjects. However, the signal-to-noise ratio (SNR) at which the SRT is achieved is not kept constant in this test paradigm. Thus, if testing involves the use of hearing-impaired (HI) test subjects, the variation in SRT measures for a single condition can easily span 10 dB. Further, if testing with normal-hearing (NH) test subjects, the SRT will often be a double-digit negative number, which compromises the ecological validity of the result (Pearsons *et al.*, 1977; Smeds *et al.*, 2012). If testing involves hearing aids (HA), extremely low SRTs mean that these devices and the signal-processing algorithms in them may be operating in conditions for which they were not intended.

Another way of testing speech intelligibility is to score %-correct words or sentences at a fixed SNR. However, as test subjects do not perform equally well at equal SNRs, it may be necessary to vary test SNR across subjects in order to obtain results in the informative 20-90% range. As above, this introduces a potential SNR confound. It would be preferable to test all subjects at the same fixed SNR and at the same time have everybody performing around the steepest part of their psychometric functions.

One way to accomplish this is to provide the experimenter with 'SRT manipulators', to control the SNR at which testing takes place for the individual listener. Using such manipulators on an individual basis could potentially reduce the spread of SRTs across a group. In an earlier study (Rønne *et al.*, 2013), three suitable manipulators were identified: changing between male and female masker speakers, changing the scoring method from word-correct to sentence-correct, and changing the spatial separation between target and maskers.

This paper presents a spatial speech-in-speech test with means of addressing ecological validity and SNR confounds. This was achieved by selecting four

*Corresponding author: slu@eriksholm.com

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

appropriate test conditions using the three SRT manipulators mentioned above, making it possible to shift the individual listener's SRT towards a common desired target SNR. Further, this paper presents the results of a perceptual validation study.

METHODS

The SFS test basics and setup

The SFS test is a speech-in-speech intelligibility test, using the Danish HINT corpus (Nielsen and Dau, 2011) as target speech. The masker speech signals are recordings of two different either male or female speakers reading from H.C. Andersen's fairytale *The Nightingale*. The masker signals are approximately 2 minutes long and were looped. Masker-speech pauses were cut down to 65 ms. Both the male target and the male and female maskers were spectrally matched to a female reference spectrum (the Dantale 2 spectrum, Wagener *et al.*, 2003).

For the trials targeting 50% words correct the Dantale 2 (Wagener *et al.*, 2003) adaptive procedure is used. For trials targeting 50% sentences correct the HINT adaptive procedure is used (Nielsen and Dau, 2011).

The test is set up in an anechoic chamber. The test subjects are seated in an adjustable chair in the centre of the room and it is ensured that the point between the subject's ears is at the same height as and distance from the surrounding loudspeakers, see Fig. 1.

The target speech is played at 70 dB SPL (C) from 0º, and the two masker signals are used in pairs and arranged symmetrically around the listener, at angles ±15º, ±30º, or ±45º. The ±60º or ±90º loudspeakers are used in conditions with target location uncertainty (see below).

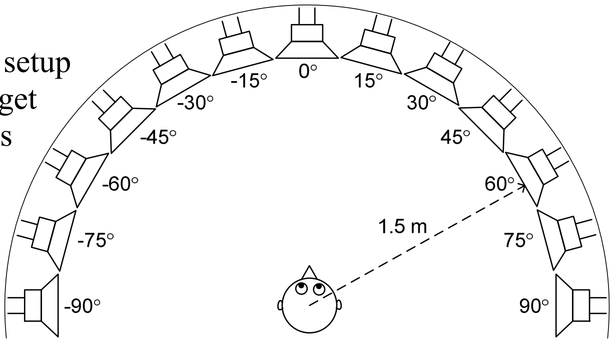


Fig. 1: The loudspeaker setup used for the experiment. Target was presented at 0º, whereas two maskers were presented from different symmetrical configurations.

Test conditions

The post-analysis outcome of the Rønne *et al.* (2013) study was a selection of SRT-manipulator settings resulting in four SFS conditions. In an adaptive-SNR test paradigm these conditions on average lead to successive 2.5-dB shifts of the SRT, as shown in Table 1. When used in a fixed-SNR paradigm, the SFS conditions will allow a group of subjects to be measured at the same target SNR and still be evaluated within the sensitive part of their individual psychometric function. This is true as long as the between-subjects spread in baseline SRT is up to about 10 dB.

SFS condition	Masker gender	Scoring	Masker positions	Expected shift of SRT
15mS	Male	Sentence	±15°	+5 dB
30mS	Male	Sentence	±30°	+2.5 dB
30mW	Male	Word	±30°	0 dB
45fW	Female	Word	±45°	-2.5 dB

Table 1: The four SFS conditions. Condition *30mW* (male maskers at ±30°, Word scoring) is chosen as the baseline.

Target location uncertainty (TGLU)

An option available in the SFS test is to include TGLU. In the SFS test this means presenting the target sentence randomly from three different loudspeaker positions. TGLU was included in the validation study, while data will be reported elsewhere.

Calibration

Calibration of the signals used in the SFS test was done with the test subject absent and a microphone positioned at the centre of the semi-circle in Fig. 1. All SNRs in this paper are referred to this reference condition. Note that the shadow and baffle effect of the head and the pinna changes the SNR at the position of the hearing aid, when the spatial configuration is changed (Rønne *et al.*, 2013), see Table 2.

SFS conditions	$\mathrm{SNR}_{\mathrm{HA}}$-$\mathrm{SNR}_{\mathrm{ref}}$ [dB]
15mS	-0.3
30mS, *30mW*	-0.9
45fW	-1.2

Table 2: Differences between the calibrated $\mathrm{SNR}_{(\mathrm{ref})}$ and the $\mathrm{SNR}_{\mathrm{HA}}$ measured at a BTE hearing aid, averaged across a pool of subjects.

VALIDATION TEST DESIGN

The purpose of the validation test was to validate that the four SFS conditions yielded the expected SNR shifts, and to examine the validity and test-retest reliability of the SFS test.

Test subjects

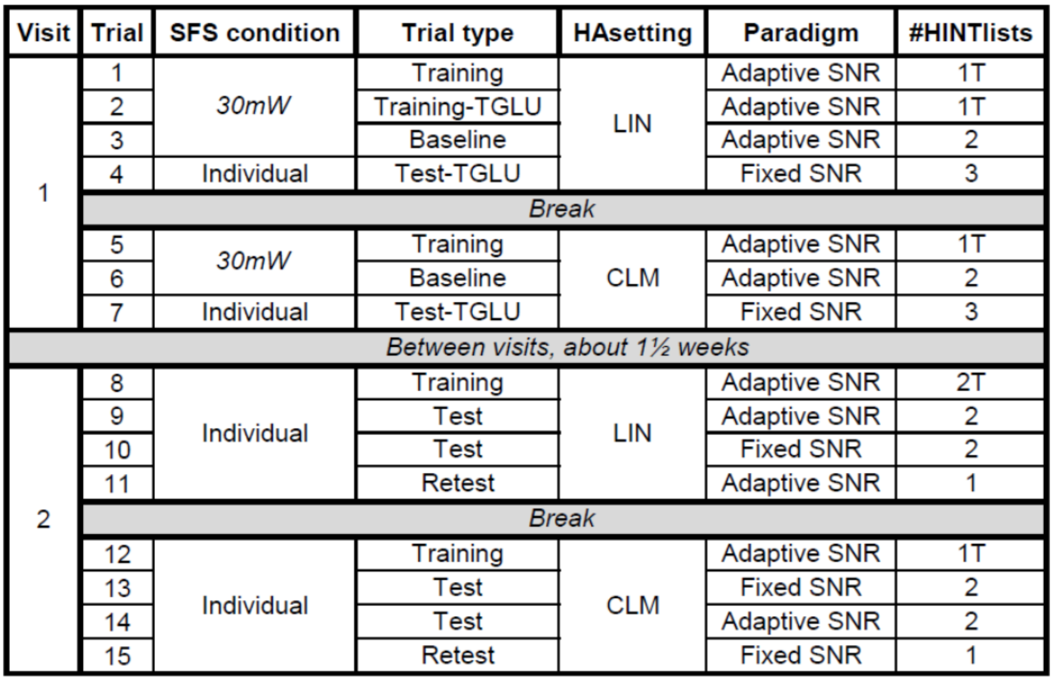
N = 26 hearing-impaired listeners with sensorineural and mixed hearing loss took part. Pure-tone-average (PTA) HTL values across 0.5, 1, 2, and 4 kHz, averaged across ears, ranged from 29 dB to 66 dB, with a mean value of 46 dB. Subjects were listening bilaterally aided through Agil Pro miniRITE hearing aids with closed ‘power domes’. Directionality and noise management were disabled.

Experimental contrast

The experimental contrast used in the validation study was the difference between hearing-aid settings with compression limiting (CLM) and linear processing (LIN). This was selected because Naylor and Johannesson (2009) found a significant change in SNR (ΔSNR) from the input to the output of an aggressive compressive hearing aid. Further, this ΔSNR was shown to depend of the input SNR, such that the ΔSNR was positive for negative input SNRs, and negative for positive input SNRs. No change in SNR was expected with a linear hearing aid. Thus, this was a clear example of an experimental contrast with an SNR confound, where different results would be expected if a test subject was tested in positive or negative SNRs. Given this contrast it was decided to include two target SNRs, one at −5dB SNR and one at +5dB SNR. Test subjects were shifted, by choosing an appropriate SFS condition, towards the target SNR that was closest to their baseline performance. The projected SNR confound should be observable in test performance as a significant interaction between target-SNR group and hearing-aid setting.

Protocol

The test protocol is sketched in Table 3. Hearing-aid setting LIN was tested against hearing-aid setting CLM. Measurements were done in either the adaptive-SNR paradigm or the fixed-SNR paradigm. Further, TGLU was included (only in fixed-



Visit	Trial	SFS condition	Trial type	HAsetting	Paradigm	#HINTlists
1	1	*30mW*	Training	LIN	Adaptive SNR	1T
	2		Training-TGLU		Adaptive SNR	1T
	3		Baseline		Adaptive SNR	2
	4	Individual	Test-TGLU		Fixed SNR	3
	Break					
	5	*30mW*	Training	CLM	Adaptive SNR	1T
	6		Baseline		Adaptive SNR	2
	7	Individual	Test-TGLU		Fixed SNR	3
Between visits, about 1½ weeks						
2	8	Individual	Training	LIN	Adaptive SNR	2T
	9		Test		Adaptive SNR	2
	10		Test		Fixed SNR	2
	11		Retest		Adaptive SNR	1
	Break					
	12	Individual	Training	CLM	Adaptive SNR	1T
	13		Test		Fixed SNR	2
	14		Test		Adaptive SNR	2
	15		Retest		Fixed SNR	1

Table 3: The test protocol. The order of hearing-aid settings, the order of test paradigms in trial pairs (9,10) and (13,14), as well as use of HINT test lists were balanced across listeners.

SNR paradigm). The test protocol included two visits. At the beginning of visit 1, two HINT training lists (40 sentences) were included, followed by the baseline *30mW* adaptive-SNR measurement. Based on the performance of the individual test subject, each subject was 'shifted' to one of the two target SNRs in this design, either +5dB SNR or −5dB SNR. One more measurement was done in the baseline setup to determine a baseline performance difference between the two hearing-aid settings.

Within-visit training effects are small (Rønne *et al.*, 2013) and are assumed to be balanced in the present test design. However, Rønne *et al.* (2013) found a 0.3-dB between-visit training effect (better performance at the second visit), which needs to be addressed here because the baseline and the TGLU measurements always were done at the first visit. Thus, between-visit training was corrected for when relevant.

RESULTS

SFS condition performance

All 26 test subjects were measured twice in the baseline condition (*30mW*, adaptive-SNR paradigm), one with hearing-aid setting LIN and one with CLM. Later in the protocol all subjects were also measured in their individually selected SFS condition (trials 3, 6, 9 and 14). In order to assess the effectiveness of the SFS test conditions, Fig. 2 shows the magnitude of the SRT shifts (two from each subject). Some subjects did in their baseline measurement perform close to one of the two target SRTs (−5 or +5 dB SNR), and were thus not shifted. The data points from these subjects are depicted at *30mW*.

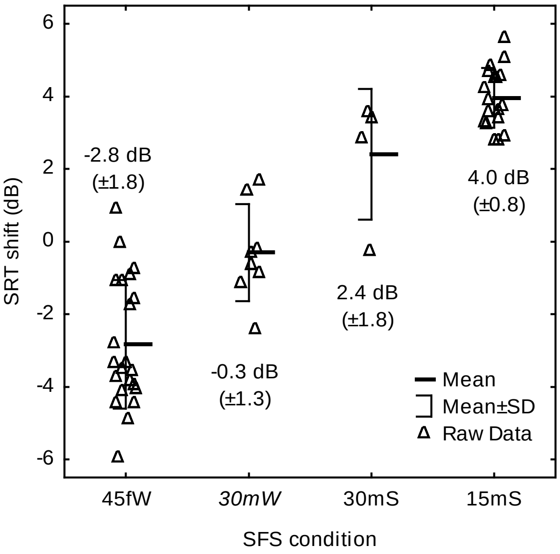


Fig. 2: SRT shifts between the adaptive-SNR baseline conditions (trials 3 and 6 in protocol) and the test subject's individually selected adaptive-SNR SFS condition (trials 9 and 14). Note the very few individual data points in the 30mS condition. Two data points were obtained from each test subject (9-3, 14-6), thus 52 data points are present in the figure.

Test-rest reliability

Test-rest data were derived from the first list (20 sentences) of trials 9 and 13, to allow direct comparison with trials 11 and 15. Note that each subject thus contributes one set of test-retest data points in the adaptive-SNR paradigm and one set in the fixed-SNR paradigm. The protocol was balanced such that half of the data points for each paradigm were measured with each hearing-aid setting.

First, the variance of the difference measure (trial pair 9-11 for the adaptive-SNR paradigm) is found as

$$V_{\Delta\text{adaptive}-\text{SRT}} = \frac{1}{N}\sum_{n=1}^{N}(SRT_{\text{test}} - SRT_{\text{retest}})^2 \; . \qquad \text{(Eq. 1)}$$

The test-retest standard deviation (SD) of a single measurement is then

$$\text{SD}_{\text{adaptive}-\text{SRT}} = \sqrt{\tfrac{1}{2}V_{\Delta\text{adaptive}-\text{SRT}}} = 0.95 \text{ dB}. \qquad \text{(Eq. 2)}$$

Similarly, for the fixed-SNR paradigm:

$$V_{\Delta\text{fixed}-\text{SNR}} = \frac{1}{N}\sum_{n=1}^{N}(\%correct_{\text{test}} - \%correct_{\text{retest}})^2 \qquad \text{(Eq. 3)}$$

$$\text{SD}_{\text{fixed}-\text{SNR}} = \sqrt{\tfrac{1}{2}V_{\Delta\text{fixed}-\text{SNR}}} = 8\ \%. \qquad \text{(Eq. 4)}$$

Validity

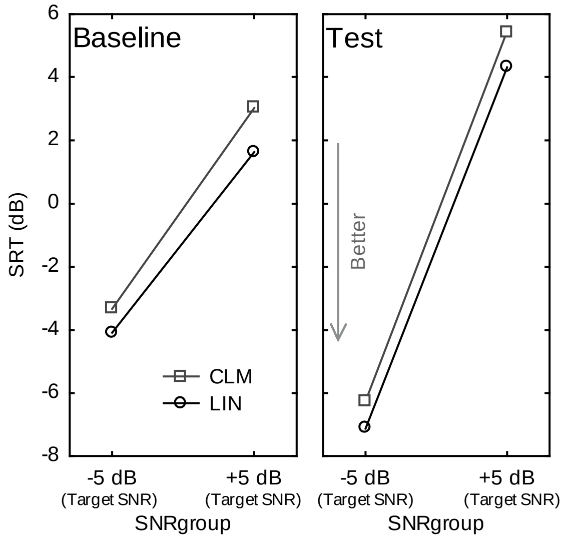
Figure 3 (left panel) shows the baseline performance with the two hearing-aid settings in the two subgroups of good performers (labelled −5 dB, typically small hearing losses) and poor performers (labelled +5 dB, typically larger hearing losses). Figure 3 (right panel) shows the performance of the same subjects in their individually selected SFS condition. The Test SRTs are forced further apart than the Baseline SRTs, and the Test SRTs are closer to the target SNRs in the +5-dB group. This indicates that the SFS conditions are working as expected. Two mixed-model ANOVAs (one for each panel) indicate that hearing-aid setting was significant for both Baseline (p = 0.0008) and Test (p = 0.004), whereas the expected interactions between SNR group and hearing-aid setting were not significant.

Figure 4 shows the average performance for the fixed-SNR paradigm. Hearing-aid setting was significant (p = 0.00009) and the interaction between hearing-aid setting and SNR group was significant (p = 0.02).

DISCUSSION

Test-retest reliability

The test-retest within-subject SD of the adaptive-SNR paradigm SFS test was determined to be 0.95 dB (Eq. 2). This is comparable to the test-retest SD of the original HINT material that was found to be 0.92 dB for hearing-impaired listeners (Nielsen and Dau, 2011). Thus, it seems that the inclusion of a spatial setup, speech maskers, and different SFS conditions, does not increase the SD of the test. For the fixed-SNR paradigm the test-retest SD was determined to be 8% (Eq. 4). For this measure no relevant literature comparison exists. However, the gradient of the



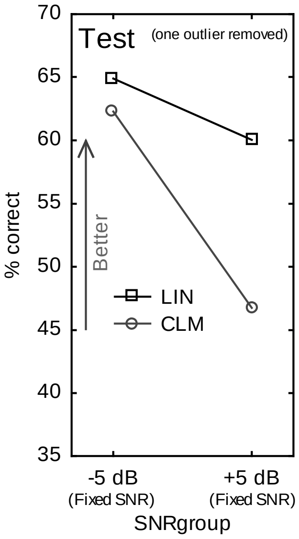


Fig. 3: Left panel shows the average performance of the test subjects in the baseline condition for each the two hearing-aid settings (trials 3 and 6). The subjects were based on their performance divided into the groups, labelled −5 or +5 dB target SNR. Right panel shows the average performance of the same subjects when measured in their individually selected SFS conditions (trials 9 and 14).

Fig. 4: Average performance across subjects tested in the fixed-SNR paradigm (trial 10 and 13). All individual %-correct scores were in the 10-84% range.

psychometric curve at the 50% correct point is 13.7%/dB for the SFS test, and it could thus be speculated that the test-retest SD of the fixed SNR paradigm should be somewhere slightly below 13.7*0.95 = 13%. That it is in fact 8% indicates that the test-retest reliability potentially is better for the fixed-SNR paradigm compared to adaptive-SNR.

The experimental contrast

According to Naylor and Johannesson (2009) hearing-aid compression affects the long-term SNR, such that the SNR at the output of the hearing aid is improved by compression in negative input SNRs and is made worse in positive SNRs. This change in long-term SNR from input to output was denoted ΔSNR. This study replicated the setup of Naylor and Johannesson (2009) and did actual measurements in a test-box to determine the magnitude of the ΔSNR for each individually-fitted hearing aid programmed to be first in LIN and then in the CLM setting. For all subjects the ΔSNR was measured to be approximately 2 dB in the expected direction, positive or negative.

The adaptive-SNR trials in this study showed a constant influence of CLM of about −1 dB independent of input SNR, whereas the fixed SNR trials showed a small but significant interaction between SNR group and hearing-aid setting. Neither of the two methods showed the expected ±2 dB ΔSNR swing and strong dependence on SNR group. Thus a major question mark has to be raised regarding the perceptual relevance of the Naylor and Johannesson (2009) output-SNR measure. Also, the results from this study contradict the perceptual correlations found between speech intelligibility performance and ΔSNR by Naylor *et al.* (2008).

It is also interesting that the two test paradigms yield different results regarding the interaction between hearing-aid setting and SNR group in Figs. 3 (right) and 4. One explanation could be that all test subjects in the fixed-SNR paradigm are tested at the same SNR, whereas subjects in the adaptive-SNR paradigm are tested at a range of SNRs around the target SNR. It can be speculated that the latter approach has made the results more variable, and thus made it harder for a contrast to be visible.

CONCLUSION

A Spatial Fixed-SNR (SFS) speech intelligibility test was designed and validated. The unique asset of the SFS test is the way individual test subjects can be evaluated in different conditions such that the SNR at which they are evaluated is the same. This study found that the SFS test conditions provide SNR shifts of the expected magnitude, that reliability is on par with the standard HINT, and that the test is able to detect relevant experimental differences with high statistical significance.

REFERENCES

Naylor, G., Rønne, F.M., and Johannesson, R.B. (**2008**). **“**Perceptual correlates of the long-term SNR change caused by fast-acting compression,” International Hearing Aid Research Conference (IHCON), Lake Tahoe, CA, USA (poster).

Naylor, G., and Johannesson, R.B. (**2009**). “Long-term signal-to-noise ratio at the input and output of amplitude-compression systems,” J. Am. Acad. Audiol., **20**, 161-171.

Nielsen, J.B., and Dau, T. (**2011**). “The Danish hearing in noise test,” Int. J. Aud., **50**, 202-208.

Pearsons, K.S., Bernett, R.L., and Fidell, S. (**1977**). “Speech levels in various noise environments,” Bolt, Beranek and Newman Inc., Canoga Park, California.

Rønne, F.M., Laugesen, S., Jensen, N.S., Hietkamp, R.K., and Pedersen, J.H. (**2013**), “Magnitude of speech-reception-threshold manipulators for a spatial speech-in-speech test that takes signal-to-noise ratio confounds and ecological validity into account,” Proc. Meet. Acoust., **19**, International Congress on Acoustics, Montréal, Canada, pp. 050069.

Smeds, K., Wolters, F., and Rung, M. (**2012**). “Realistic signal-to-noise ratios,” International Hearing Aid Research Conference (IHCON), Lake Tahoe, CA, USA (poster).

Wagener, K., Josvassen. J.L., and Ardenkjær, R. (**2003**). “Design, optimization and evaluation of a Danish sentence test in noise,” Int. J. Aud., **42**, 10-17.

Informational masking in speech intelligibility tests

Ellen Raben Pedersen[1,*], Hoda El-Samail[2], Carsten Daugaard[3], Peter Møller Juhl[1], and Ture Andersen[4, 5]

[1] *The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense M, Denmark*

[2] *Audiology and Logopedics Studies, University of Southern Denmark, Odense M, Denmark*

[3] *DELTA Technical-Audiological Laboratory, Odense C, Denmark*

[4] *Institute of Clinical Research, University of Southern Denmark, Odense M, Denmark*

[5] *Department of Audiology, Odense University Hospital, Odense C, Denmark*

It is often challenging to separate speech from a noise – especially for hearing-impaired persons. A particular difficult listening situation is when speech is obscured by speech from one or more simultaneous talkers. The purpose of this study is to investigate the effect of informational masking on the speech reception threshold (SRT) and to compare the SRT values obtained with subjective data from the SSQ questionnaire. A listening test was performed with 20 normal-hearing and 20 hearing-impaired subjects. The subjects were presented to the sentences from the Danish speech material Dantale II in four different speech-shaped interfering maskers. The maskers differ regarding fluctuation and to what extent they represent intelligible speech. The listening test shows that the three fluctuating maskers distinguish better between normal-hearing and hearing-impaired subjects than the almost stationary masker. The test-retest variation was found to be the same for the four maskers. The SRT values for the four maskers were generally found not to correlate with the hearing-impaired subjects' answers to specific questions in the SSQ questionnaire.

INTRODUCTION

Understanding speech in noise is a challenging task for people in general and especially for hearing-impaired persons – a particular difficult listening situation is when the masker is speech from one or more simultaneous talkers. Therefore speech-in-noise tests are routinely carried out in clinics in order to assess the degree of the hearing loss and the effect of treatment. However, the results of the tests are often in disagreement with the problems that the subjects report. One reason for this difference might be that the masker used in the clinical test does not represent real-life maskers well, by which the tests are dominated by energy masking and only to a

*Corresponding author: erpe@mmmi.sdu.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

limited amount involve informational masking (for a review on informational masking, see Schneider *et al.*, 2007).

The purpose of this study is to investigate how four different maskers influence the result of a speech-in-noise test – i.e., the speech reception threshold (SRT) – for both normal-hearing and hearing-impaired subjects. The maskers differ regarding fluctuation, to what extent they represent intelligible speech, and were expected to cause different amount of informational masking. Three study questions were addressed: 1) Do the different maskers influence the test sensitivity differently, i.e., are the different maskers equally good/bad at distinguishing between normal-hearing and hearing-impaired subjects?, 2) Is the test-retest variation affected by the different maskers?, and 3) Are the test results obtained with the four maskers in agreement with the subject's own experience of his/her ability to understand speech in noise?

METHODS

Target signal

As a target signal the test sentences from the Danish speech material Dantale II (Wagener *et al.*, 2003) were used. The material consists of 16 lists with ten test sentences each. The test sentences are spoken by a female speaker and have a fixed structure of five words from different word classes in the order: name, verb, numeral, adjective, and noun. As an example the first sentence in list 1 is: 'Ingrid finds seven red houses' (translation of the Danish sentence: 'Ingrid finder syv røde huse').

Masker signals

During the listening test the subjects were presented to the target signal in four different speech-shaped interfering maskers, which are expected to cause different amount of informational masking. Below is a short description of each of the four signals. The Dantale II noise is almost stationary, whereas the three other signals fluctuate comparably to natural speech. For the listening test the overall RMS level of the three fluctuating signals was adjusted to that of the Dantale II noise. The signal named 2FS was specially generated for this study and is the only of the four signals representing intelligible speech.

Dantale II noise: This signal is included in the Dantale II speech material. It is generated by superimposing the test sentences many times by which the signal becomes almost stationary (Wagener *et al.*, 2003).

ICRA-4: This signal is made by the International Collegium for Rehabilitative Audiology (the number four refer to track no. 4 on the ICRA CD). The signal is artificial and represents one female speaker (Dreschler *et al.*, 2001).

IFFM: This signal is based on the International Speech Test Signal (ISTS) but with limited pause durations (www.ehima.com). The ISTS contains fragments of recordings from female speakers talking different languages (Holube *et al.*, 2010).

2FS: This signal was generated by making a sequence of nine Dagmar sentences and a sequence of nine Asta sentences from the DAT corpus (Nielsen *et al.*, 2011) and storing them in different channels. The name 2FS refers to the signal containing '2 Female Speakers'.

Test setup

The SRT measurements were performed with the software HearVal 1.0.0.8, which is developed in LabVIEW 2010 by DELTA. Three active Genelec 1029A loudspeakers were positioned in the horizontal plane at a distance of 1.4 m from the subject at different angles. The target signal was presented frontal to the subject at an angle of incidence of 0^o, whereas the masker was presented by two loudspeakers symmetrically located at the angles of incidence of $\pm 45^{\circ}$ (in accordance with the recommendation in DS/EN ISO 8253-3:2012). A laptop (IBM ThinkPad R51 Type 1829-R6G) was used to play and control the level of the target signal, while another laptop (IBM ThinkCentre MT-M type 9210-D1G) was used to play the masker. The masker was played incoherently from the two spatially-separated loudspeakers.

Measurement procedure

In each SRT measurement the presentation levels, i.e., the signal-to-noise ratios (SNRs) at which the sentences were presented, were adjusted (to a speech understanding of 50%) according to the adaptive procedure described in Brand and Kollmeier (2002). The adjustment was done by changing the target level, whereas the level of the masker was kept constant at 65 dBC for the normal-hearing subjects and at 80 dBC for the hearing-impaired subjects. The first sentence was presented at 0 dB SNR. After the presentation of each sentence the subjects orally repeated the words that were perceived, whereupon the test operator registered whether the subject's answer was correct or incorrect to control the SNR of the next sentence. The measurement stopped when one of the two following criteria were met: 1) the subject had been presented to three entire lists, i.e., 30 sentences or 2) ten reversals of the presentation level were attained (a reversal is attained when the change in SNR alters sign). When the measurement stopped the SRT was determined as a mean of the SNRs at the four last reversals.

Subjects

The listening test was performed with 20 normal-hearing subjects (eight males and 12 females, aged 18-26 years with a mean age of 21 years) and 20 hearing-impaired subjects (seven males and 13 females, aged 18-65 years with a mean age of 45 years). The normal-hearing subjects had no otological problems and their hearing thresholds did not exceed 20 dB HL at the octave frequencies from 0.25 to 8 kHz. The hearing-impaired subjects had varying degrees of a bilateral sensorineural hearing loss. Their pure-tone averages (PTAs) for the frequencies 0.5, 1, 2, and 4 kHz were 17.5-66.9 dB HL (mean: 38.4 dB HL). The hearing-impaired subjects were hearing-aid users, but they did not use their hearing aids during the listening test.

Questionnaire

In order to investigate whether the SRT values obtained agreed with subjective data the hearing-impaired subjects were presented to part 1 of a Danish version of the SSQ questionnaire (Gatehouse and Noble, 2004). Part 1 contains 14 questions regarding hearing speech in competing contexts. The subjects were asked to respond on a scale from 0 to 10 ('not at all' to 'perfectly') and to focus on listening situations where they did not use their hearing aids when filling in the questionnaire. The subjects' answers to four of the questions (Q1, Q5, Q9, and Q11) were chosen for comparison with the SRT values obtained – one of the questions (Q5) describes a listening situation with an almost stationary background noise, whereas the three others describe listening situations with fluctuating background noises:

Q1: You are talking with one other person and there is a TV on in the same room. Without turning the TV down, can you follow what the person you're talking to says?

Q5: You are talking with one other person. There is continuous background noise, such as a fan or running water. Can you follow what the person says?

Q9: Can you have a conversation with someone when another person is speaking whose voice is different in pitch from the person you're talking to?

Q11: You are in conversation with one person in a room where there are many other people talking. Can you follow what the person you are talking to is saying?

Test course

The normal-hearing subjects participated in two test sessions, which were separated by 20-62 days (mean: 31 days). The hearing-impaired subjects participated only in one test session. During each test session the subjects were presented to four SRT measurements containing the four different speech-shaped interfering maskers. For each measurement three different lists were randomly chosen. To avoid any effect of the presentation sequences on the results, the presentation order of the maskers was counterbalanced among the subjects. Before each measurement the subjects were presented to sentences for training masked by the same masker as in the subsequent measurement. The training contained three lists before the first measurement and one list before each of the following measurements (for both test sessions). After the SRT measurements the hearing-impaired subjects filled in part 1 of the Danish SSQ questionnaire.

Statistical analyses

For the statistical analyses the computer program SPSS 11.5.1 for Windows was used (www.spss.co.in). All analyses were performed at a 0.05 significance level. The Kolmogorov-Smirnov test was used to ascertain whether data could be assumed to come from a normal distribution, and the Levene test was used to test for homogeneity of variance. To test for differences between the SRT values obtained parametric tests were used, provided that the conditions for performing those tests were satisfied. Otherwise corresponding non-parametric tests were used. All correlation analyses were consistently made with the non-parametric Spearman's rank-order correlation, even though some could be made with the parametric Pearson's product-moment correlation.

RESULTS

Normal-hearing vs. hearing-impaired

Figure 1 shows the results from the SRT measurements for both the normal-hearing and hearing-impaired subjects. From the figure it is seen that the hearing-impaired subjects obtained higher (poorer) SRT values than the normal-hearing subjects for all four maskers. The standard deviations are also higher for the hearing-impaired subjects than for the normal-hearing subjects indicating that the hearing-impaired subjects were a more inhomogeneous group. The figure shows that the SRT values obtained for the normal-hearing and hearing-impaired subjects differ more for the three fluctuating maskers than for the Dantale II noise, i.e., the three fluctuation maskers are better than the Dantale II noise to distinguish between normal-hearing and hearing-impaired subjects.

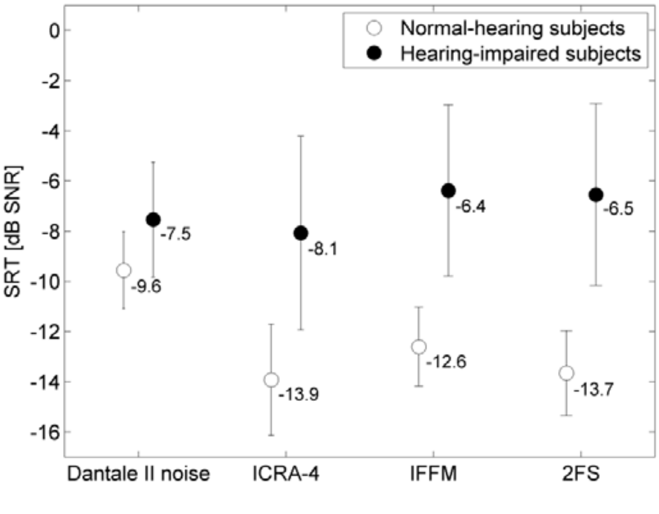


Fig. 1: Mean and one standard deviation of the SRT values for each of the four different maskers obtained with 20 normal-hearing subjects and 20 hearing-impaired subjects, respectively. The SRT values for the normal-hearing subjects are obtained at the first test session.

For the normal-hearing subjects a one-way ANOVA test showed a difference between the mean values of SRT obtained with the four maskers ($F(3,76) = 25.402$, $p = 0.000$). The post hoc Scheffe test revealed that the difference is between the mean SRT values obtained with the Dantale II noise and with the three other maskers. For the hearing-impaired subjects the non-parametric Kruskal-Wallis test showed no statistical difference between the mean SRT values obtained with the four maskers ($X^2(3) = 4.462$, $p = 0.216$).

Test-retest variation

Figure 2 shows the difference between the SRT values obtained at the two test sessions for the normal-hearing subjects. One of the normal-hearing subjects did not complete the second test session, by which the SRT values shown in the figure only are for 19 subjects. The negative differences indicate that the subjects obtained lower (better) SRT values in the second test session than in the first test session. The one-way ANOVA test showed no statistical difference between the SRT difference for the four maskers ($F(3,72) = 0.460$, $p = 0.711$), i.e., the test-retest variation was found to be independent of the type of masker used.

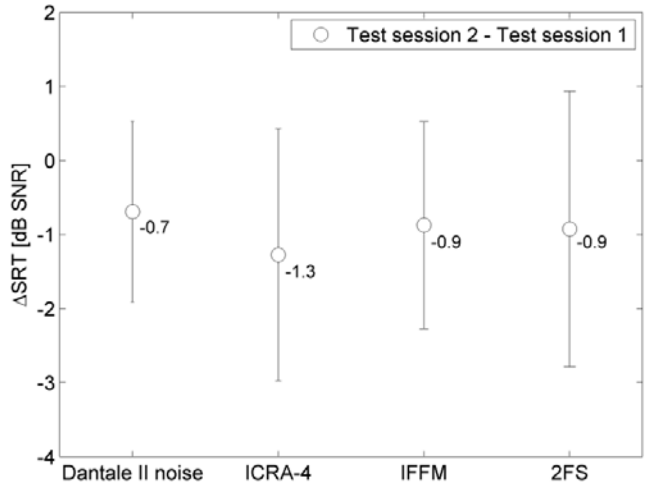


Fig. 2: Mean and one standard deviation of the SRT difference for each of the four different maskers calculated based on measurements from 19 normal-hearing subjects.

Comparison with subjective data

The mean of the answers to the four selected questions given by the hearing-impaired subjects were (one standard deviation is given in the brackets): Q1 = 5.0 (2.1), Q5 = 5.1 (2.0), Q9 = 4.0 (2.2), and Q11 = 3.4 (1.8). In order to compare the SRT values obtained with the subjects' answers to the questions correlation analyses were performed. Two of the subjects did not fill in the questionnaire. Thus the analyses only include data from 18 subjects. For the SRT values obtained with the Dantale II noise and the subjects' answers to Q5 the non-parametric Spearman's rank-order correlation showed no statistical correlation ($\rho = -0.368$, $p = 0.132$). Table 1 shows the correlations between the SRT values obtained with the three fluctuating maskers and the subjects' answers to questions Q1, Q9, and Q11. Only the correlation between the SRT values obtained with the 2FS noise and the subjects' answers to Q9 was statistically significant.

	Q1	**Q9**	**Q11**
ICRA-4	$\rho = -0.099, p = 0.695$	$\rho = -0.450, p = 0.061$	$\rho = -0.287, p = 0.248$
IFFM	$\rho = -0.005, p = 0.983$	$\rho = -0.307, p = 0.215$	$\rho = -0.103, p = 0.684$
2FS	$\rho = -0.125, p = 0.622$	$\rho = -0.536^{*}, p = 0.022$	$\rho = -0.410, p = 0.091$

Table 1: Spearman's rank-order correlation between SRT values obtained with the three fluctuating maskers and scores for specific questions in the SSQ questionnaire. Each correlation contains data from 18 hearing-impaired subjects. Significant results at a 0.05 level are marked with an asterisk (*).

DISCUSSION

For the normal-hearing subjects the SRT values were found to be lower for the three fluctuating maskers than for the almost stationary Dantale II noise, whereas no difference was found between the SRT values obtained with the four maskers for the hearing-impaired subjects. This finding may be due to the normal-hearing subjects being able to benefit from the silent intervals in the fluctuating maskers by *listening in the dips*, whereas the hearing-impaired subjects do not seem to benefit from those intervals.

The test-retest variation was found to be independent of the type of masker used for the group of normal-hearing subjects. However, the test-retest variation was expected to differ for the different maskers. This is due to the silent intervals in the fluctuating maskers, where parts of the target signal can be heard through the maskers – sometimes the part heard may contain speech sounds from which the word can be *guessed*, whereas at other times this will not be the case. Hence, a larger variation in the test results was expected, which would also result in a larger test-retest variation.

The correlation analyses show that the lowest *p*-values were obtained with the 2FS noise. This indicates that, when using a masker representing intelligible speech, the SRT value is in better agreement with the subject's own experience of his/her ability to understand speech in noise, i.e., the test seems to be more valid than when using the other maskers.

CONCLUSIONS

Analysing the results from the listening test gave the following answers to the three study questions: 1) The three fluctuating maskers distinguish better between normal-hearing and hearing-impaired subjects than the almost stationary Dantale II noise, 2) The test-retest variation was not found to be affected by the different maskers used, and 3) The SRT values for the maskers were generally found not to correlate with the hearing-impaired subjects' answers to specific questions in the SSQ questionnaire. Only the correlation between the SRT values obtained with the 2FS noise and the subjects' answers to Q9 was found to be statistically significant.

ACKNOWLEDGEMENTS

The authors would like to thank Chris Vigandt Jørgensen from DELTA Technical-Audiological Laboratory, Odense, Denmark for helping with the test setup. The authors would also like to thank all subjects who participated in the listening test.

REFERENCES

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am., **111**, 2801-2810.

Dreschler, W.A., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**). "ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. International Collegium for Rehabilitative Audiology," Audiology, **40**, 148-157.

DS/EN ISO 8253-3:2012 (**2012**). *Acoustics – Audiometric test methods – Part 3: Speech audiometry*, 2nd Ed. (Copenhagen, Denmark).

Gatehouse, S., and Noble, W. (**2004**). "The speech, spatial and qualities of hearing scale (SSQ)," Int. J. Audiol., **43**, 85-99.

Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (**2010**). "Development and analysis of an International Speech Test Signal (ISTS)," Int. J. Audiol., **49**, 891-903.

Nielsen, J.B., Neher, T., and Dau, T. (**2012**). "Towards a Danish speech material for speech-on-speech masking investigations," in Proceedings of ISAAR 2011: *Speech perception and auditory disorders*. 3rd International Symposium on Auditory and Audiological Research, Nyborg, Denmark. Edited by T. Dau, M.L. Jepsen, T. Poulsen, and J. Christensen-Dalsgaard. ISBN: 87-990013-3-0. (The Danavox Jubilee Foundation, Copenhagen), pp. 175-181.

Schneider, B.A., Li, L., and Daneman, M. (**2007**). "How competing speech interferes with speech comprehension in everyday listening situations," J. Am. Acad. Audiol., **18**, 559-572.

Wagener, K., Josvassen, J.L., and Ardenkjaer, R. (**2003**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol., **42**, 10-17.

Influence of memory effects in speech intelligibility tasks

Stefanie Keller[1,*], C hristian Wirtz[2], H anna Beike[3],
and Werner Hemmert[1]

[1] *BAI, Bio-inspired information processing, IMETUM, Technische Universität München, Germany*

[2] *MED-EL Deutschland GmbH, Germany*

[3] *Hochschule für angewandte Wissenschaften, FH München, Germany*

Testing speech reception thresholds of hearing-impaired patients is a common task in clinical routine and research. Tests consist of grammatically correct sentences containing different grammatical classes. It is expected that due to primacy and recency memory effects error rates of the first and last word are minimal. In addition, from a linguistic point of view, not only the position of a word but also its grammatical class causes different cognitive effort. This study analyses the effect of different conditions on the comprehended words belonging to different grammatical classes. So far, nine normal-hearing subjects were measured via headphones with a German speech intelligibility test with different kinds of noise and different interaural time differences. The results do not only show the expected memory effects for the noun at the first and last position of the sentences. Also significant differences for the comprehension of sentence-centered numerals were found in comparison to neighboring positions. This is impressive because in the middle, normally the attention of a listener is minimal, therefore one would expect a small recognition rate. In summary, we conclude that careful analysis of speech-reception tests also provides information on more cognitive aspects involved in speech understanding like memory capacity.

INTRODUCTION

Speech-intelligibility tasks are a common tool to measure the speech reception threshold (SRT) in noise of hearing-impaired persons. They are well-established in different Western European languages and are mostly all designed the same way: Subjects listen to a sentence in the specific language in background noise. Then they repeat all words they understood. Depending on the number of correctly understood words the signal-to-noise level is varied to determine the so-called SRT where 50% of the words were understood. For the German language, the test of choice is the *Oldenburger Satztest (Olsa)* (Wagener *et al.*, 1999).

The test consists of 40 lists composed of 30 five words sentences (Wagener *et al.*, 1999). Sentences are non-sense sentences with identical grammatical structure. As

*Corresponding author: stefanie.keller@tum.de

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Britta	kauft	fünf	alte	Ringe.
N (name)	V	Num	Adj	N
Britta	buys	five	old	rings.

Fig. 1: Example of a typical Olsa-sentence. Upper line shows the German sentence, middle line the grammatical classes, lowest line shows the English translation.

all words can be exchanged, there is no contextual information to guess them. Each position of the sentence is filled with a word belonging to a different grammatical class (cf. Fig. 1).

Each grammatical class exhibits different cognitive effort and requires therefore different complexity for processing.[1] Verbs for example carry information about person, numerus, tempus, genus verbi, and modus, nouns about numerus, genus, and case. For this study, we will assume that nouns are more simple than verbs and adjectives because of the restricted use of objects and names. Nouns coding an object are one of the earliest words in language acquisition (Dittmann, 2002), this is a hint that these are simple words.[2] Personal names have restricted use and are sometimes triggered by personal experiences (e.g., someone might know a ‘Britta’ who is a very kind person). Numerals are a special group of words; it is a collection of different kinds of numbers: cardinals, ordinals, fraction numbers, etc. Here, we will only take a look at the cardinals because that is the group used in Olsa-sentences. So it is a relatively small group of words that is closed, i.e., no new speech material joins the group. That means this position in Olsa-sentences is filled by a collection which is small and predictable and should therefore be easy to classify and remember. Hence, we derive **Hypothesis 1**: The nouns (names and objects) should be recalled best, whereas a decreasing recall should be found for numerals, adjectives, and verbs (cf. Fig. 2). This should be reflected in a stable recognition rate of each grammatical class over conditions, i.e., the values shall not differ significantly from 50%, which is the value for the determined SRT.

Syntactic structure is another point that can influence speech understanding. Carroll

[1] The following reflections are mainly German specific and may just be transferred to other languages with constraints.

[2] Because of this restricted use of the grammatical class *nouns* in Olsa-sentences complex nouns like abstract words *freedom*, *wisdom*, etc. will not be considered in the following text. That is why the assumption of the smaller complexity of used nouns can be made. Note that this is not true for all representatives of the grammatical class *noun*, for further information and detailed discussion see Leiss (2002) and Vigliocco *et al.* (2011).

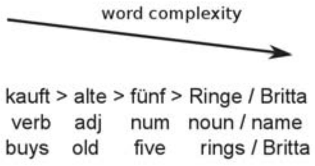

word complexity
kauft > alte > fünf > Ringe / Britta
verb adj num noun / name
buys old five rings / Britta

Fig. 2: Hypothesis 1: Increasing complexity for the grammatical classes used for the Olsa-sentences.

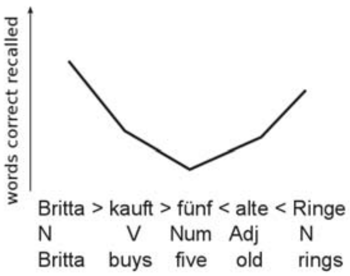


Fig. 3: Hypothesis 2: Expected primacy and recency effects for the Olsa-sentences.

and Ruigendijk (2013) found that the syntactic complexity can influence intelligibility in noise, and Uslar *et al.* (2011) showed the dependency of syntactic complex sentences on speech recognition in younger listeners. The default syntactic structure for German main clauses is subject followed by verb followed by an object, which is also known as SVO. If there is an unusual position like OVS, this causes more cognitive effort and therefore is more difficult to understand in noise (Carroll and Ruigendijk, 2013). Because the default structure is used in all Olsa-sentences, the syntactic structure will not be further analysed.

Although Carroll and Ruigendijk (2013) could not prove a connection between memory load and intelligibility, Ljung *et al.* (2013) on the other hand demonstrated that people with a higher working-memory capacity can recall words better in noise than people with a lower working-memory capacity. As postulated by Miller (1956), the working memory consists of 7 ± 2 items (Miller's law).[3] By chunking it is of course possible to remember for example a telephone number with more than nine numbers. But if you investigate the working memory, the chunks will not exceed 7 ± 2. Olsa-sentences are composed of five words, which should minimize memory effects.

Not only the grammatical class differs for cognitive load but also the position in the sentence is important. That is why we should have a look at memory effects as well. There are serial-position effects for memorising items of a list, they show a U-shaped curve: Items at the beginning and at the end tend to be better recalled (e.g., Jones and Oberauer (2013) and Oberauer *et al.* (2003)).

Investigations of these so-called primacy and recency effects have been made. Objects in the middle are not as well remembered as first and last position but the occurrance of primacy or recency effect depends on the task (Healy *et al.*, 2000). It is therefore likely that the score for the remembered words are higher at the beginning and at

[3] Working memory is a highly complex subject and cannot be fully discussed in this article.

the end of sentences. For Olsa-sentences these positions are filled by a name and an object.

According to memory load we therefore formulate **Hypothesis 2**: Recall for the first and last position of the sentences should be higher than for all other positions (Fig. 3).

METHODS

The SRT was determined with Olsa. Nine normal-hearing listeners (29.2 ± 5.2 years) were tested via earphones (Sennheiser HDA 280) in an auditory booth (IAC 350). We introduced different interaural time differences (ITDs) for speech and noise (ITDs were 0, 200, 400, and 600 μs): the speech was shifted to the side whereas the noise was presented with 0 – µs ITD on both channels of the earphones. The background noise was Fastl-Noise (Fastl, 1987) or olnoise, which is a noise that was produced by overlaying all Olsa-sentences, i.e., the spectrum of the noise is the spectrum of all sentences. Whereas olnoise has no additional temporal modulations, Fastl-noise is modulated similarly to speech. The conditions S0N0, S0N0$_{Fastl}$, S200N0, S400N0, S600N0 were tested (the numbers indicate the ITD in microseconds).

RESULTS

SRTs for the different conditions are shown in Fig. 4. The participants profit if the speech signal is presented from the side and the noise is presented frontally because the SRT values for the SΔtN0 are better (i.e., more negative) than in the S0N0 condition. If the noise is changed from olnoise to Fastl-Noise, subjects are able to listen into the dips; SRT values are almost 9 dB better.

Hypothesis 1

Increasing complexity should impact recognition scores in all measured conditions. Comparing the boxplots for each grammatical class and condition (Fig. 5) there are some apparent fluctuations: medians of numerals differ between conditions, sometimes the median values are above and sometimes below the 50% line, whereas these fluctuations could not be found for the other grammatical classes. An analysis of variance showed that none of these differences between the recognition rate for either grammatical class over conditions was significant, especially not for the fluctuations of the numerals ($F(4,40) = 1.62$, n.s.). The conditions for an analysis of variance are accomplished: Bartlett’s test and Shapiro-Wilk’s (Shapiro-Wilk’s gives very robust results) test showed no significant results so that we assume equal variances across groups (homoscedasticity) and normal distribution of samples. This ANOVA result means that the recognition rate is stable over different conditions as expected.

Furthermore, adjectives and verbs are as expected the gramatical classes with the lowest medians. They show over all conditions medians lower than 50%, whereas the recognition rates of nouns (names and objects) are always (except in one case) over 50%. The following range of means of grammatical classes from low to high can

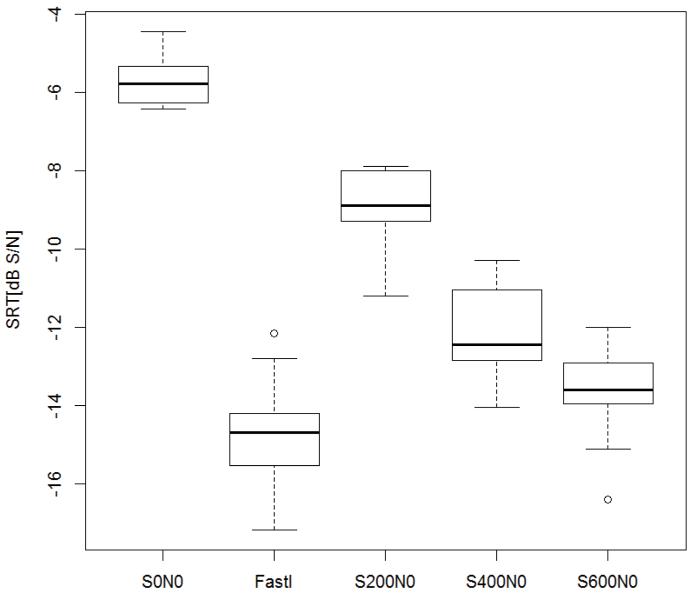


Fig. 4: Boxplots for the different spatial conditions for nine normal-hearing listeners.

be found: Adj < Verb < Num < Noun (cf. Table 1). This is almost the predicted range for Hypothesis 1, except that verbs and adjectives changed places.

Hypothesis 2

The first and last position have the best score of recognition, cf. exemplarily Fig. 6 and 7. In the mid-position an increasing value for numerals can be seen in comparison to the neighbouring positions. Therefore a U-shaped form of recall cannot be reported because the values for the middle position are not lower or not even as low as for verbs and adjectives. The values in Table 1 do not show the predicted U-shaped form but rather a W-shaped form, where the midst value is increasing a bit.

The conditions for an ANOVA were accomplished; the analysis shows the reported behaviour. There is a significant effect of position ($F(4,20) = 36.89$, $p < 0.01$). Post hoc Tukey's test showed significant differences at $p < .05$ level between the recognition rate for the first position against all positions except the last. The rate of the middle position differs significantly from neighbouring positions. This means that the middle position can be well remembered, which was not expected according to Hypothesis 2. The second and fourth position do not show a significant difference from each other.

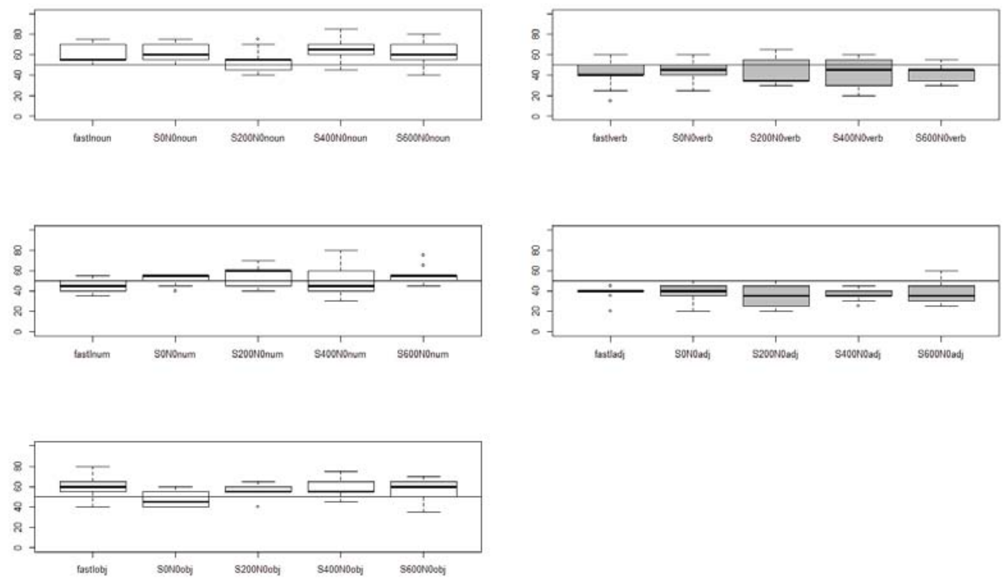


Fig. 5: Boxplots show percent correct recalled words. Each subplot shows the values for one grammatical class for the different conditions. From upper left to bottom right: nouns, verbs, adjectives, numerals, objects.

DISCUSSION

The results show an interaction between Hypotheses 1 and 2. This can be seen by the values of the middle position (numerals). Although the differences between the values for the numerals over coniditons were not significant, they show the most flucutations. Due to the grammatical class hypothesis it was not expected that numerals would show this variances over conditions in recall, and due to Hypothesis 2 it was not expected that they would show such a great recognition rate.

The results therefore show that grammatical class has an effect on the recognition rate and should be further analysed. One goal should be to clarify the explicit effect of grammatical classes by testing grammatical class and recency/primacy effects separately, e.g., by constructing lists of different grammatical classes to get rid of the syntactic structure (i.e., in this case position) or by replacing names and objects in the sentences with more difficult word classes, e.g., pronouns. Thereby the first and last position of the sentences are not filled with the least complex grammatical class.

Likewise working memory capacity should also be registered for each patient to see if his capacity is low or high. If a patient can only remember at most five things, it will be very hard to solve a task like Olsa where five items are covered with noise.

Another important fact that has to be considered is age. Larsby *et al.* (2005) and Uslar

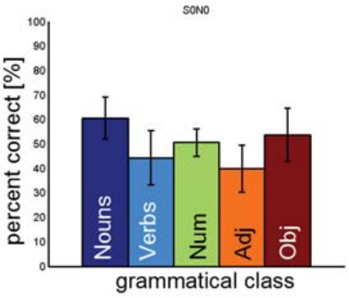


Fig. 6: Expected primacy and recency effects for the Olsa-sentences, condition S0N0.

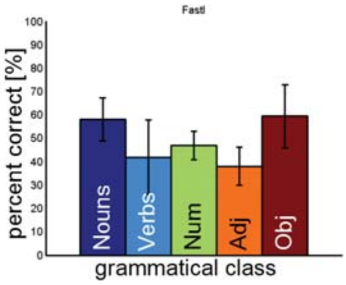


Fig. 7: Expected primacy and recency effects for the Olsa-sentences, condition S0NFastl.

Table 1: Means and standard deviations for the recognition rate for each grammatical class per condition in percent [%]. Note: SD for ALL is the standard deviation for the mean ALL.

Condition	Nouns (names)	Verbs	Num	Adj	Nouns (obj)
S0NFastl	59.44± 9.5	41.67 ± 14.79	45.55 ± 6.8	38.33 ± 7.5	60 ± 12.75
S0N0	62.22 ± 9.39	45 ±10.6	51.11 ± 5.46	39.44 ±9.17	48.3± 8.3
S200N0	54.44± 11.57	42.78 ± 13.5	56.11 ± 10.6	35 ± 10.6	56.11 ± 6.97
S400N0	65 ± 10.6	43.33 ± 6.0	50.55 ± 14.36	36.11 ± 15.7	57.77 ± 9.05
S600N0	60.55 ± 12.6	42.77± 7.95	55.55±9.17	38.89 ± 11.11	56.11 ± 11.93
ALL	**59.13**± 3.69	**43.50** ± 1.5	**52.77** ± 4.2	**37.55** ± 1.9	**55.38** ±4.11

et al. (2011) found that elderly people have more difficulties than younger people to understand speech in noise. As typically hearing-impaired people are older, they may not perform as good as teenage/young-adult groups of hearing-impaired people.

For a detailed analysis or conclusion of this interaction, it is required to test more subjects of different age, including their memory capacity.

We have shown that with a careful analysis we can evaluate working-memory effects already with a standard SRT test.

ACKNOWLEDGEMENTS

This work was supported by the Munich Bernstein Centre for Computational Neuroscience (01GQ1004B and 01GQ1004D) and MED-EL.

REFERENCES

Carroll, R., and Ruigendijk, E. (**2013**). “The effects of syntactic complexity on processing sentences in noise,” J. Psycholinguist. Res., **42**, 139-159.

Dittmann, J. (**2002**). *Der Spracherwerb des Kindes. Verlauf und Störungen.* (Beck Verlag, München), 1. Auflage.

Fastl, H. (**1987**). “Ein Störgeräusch für die Sprachaudiometrie,” Audiol. Akustik, **26**, 2-13.

Healy, A.F., Havas, D.A., and Parker, J.T. (**2000**). “Comparing serial position effects in semantic and episodic memory using reconstruction of order tasks,” J. Memo. Lang., **42**, 147-167.

Jones, T., and Oberauer, K. (**2013**). “Serial-position effects for items and relations in short-term memory,” Memory, **21**, 347-365.

Larsby, B., Hällgren, M., Lyxell, B., and Arlinger, S. (**2005**). “Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects,” Int. J. Audiol., **44**, 131-143.

Leiss, E. (**2002)** “Die Wortart Verb”, in *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen. XVII. Die Architektur des Wortschatzes I: Die Wortarten. 2. Halbband*. Edited by D.A. Cruse, (de Gruyter, Berlin, New York), pp. 605-616.

Ljung, R., Israelsson, K., and Hygge, S. (**2013**). “Speech intelligibility and recall of spoken material heard at different signal-to-noise ratios and the role played by working memory capacity.” Appl. Cognitive Psych., **27**, 198-203.

Miller, G.A. (**1956**). “The magical number seven, plus or minus two. Some limits on our capacity for processing information,” Psychol. Rev., **63**, 82-97.

Oberauer, K. (**2003**). “Understanding serial position curves in short-term recognition and recall,” J. Mem. Lang., **49**, 469-483.

Uslar, V., Ruigendijk, E., Hamann, C., Brand, T., and Kollmeier, B. (**2011**). “How does linguistic complexity influence intelligibility in a German audiometric sentence intelligibility test?” Int. J. Audiol., **50**, 621-631.

Vigliocco, G., Vinson, D.P., Druks, J., Barber, H., and Cappa, S.F. (**2011**). “Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies,” Neurosci. Biobehav. R., **35**, 407-426.

Wagener, K., Brand, T., and Kollmeier, B. (**1999**). “Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test,” Z. Audiol., **38**, 5-15.

Examination of the learning effect with the Dantale II speech material

ELLEN RABEN PEDERSEN* AND PETER MØLLER JUHL

The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense M, Denmark

This study examines the learning effect when using the Danish speech material Dantale II to determine the speech reception threshold (SRT) in noise under three different test conditions. The learning effect is shown by an improvement of the test result, i.e., by a decrease in the value of SRT at repeated measurements until a certain number of measurements has been made. A listening test was performed with 24 normal-hearing subjects. The purpose of the test was to investigate the influence of the target level on the learning effect in an open-set test format, where the subject's task is to orally repeat as much as possible of the sentence just presented. The target level was set to 50% and 80% correctly understood words, respectively. Furthermore, the purpose was to investigate whether using a closed-set test format affects the learning effect. In the closed-set test format the subject had, for each word presented, to select a response from ten alternative words. Statistical analyses of the test results did not show any significant differences in neither the within-visit learning effect nor the inter-visit learning effect for the two target levels or for the different test formats. However, the learning effect was found to be finished faster for the open-set test format with a target level of 80% than for the two other conditions.

INTRODUCTION

Over the years different speech-in-noise tests have been developed for determining the speech reception threshold (SRT). The tests have been applied both in the clinical practice and in hearing research. A commonly known speech material is the Danish Dantale II speech material (Wagener *et al.*, 2003), which consists of syntactically fixed but semantically unpredictable test sentences and an almost stationary noise signal. The speech material Dantale II is developed in analogy to the materials for the Swedish Hagerman test (Hagerman, 1982) and the German Oldenburg sentence test (Wagener *et al.*, 1999). Within the European HearCom project the material has also been developed for other languages, e.g., Polish (Ozimek *et al.*, 2010), Spanish (Hochmuth *et al.*, 2012), and French (Jansen *et al.*, 2012).

It is known that, when using a speech material as the Dantale II speech material, a learning (or training) effect is present. The learning effect is shown by an improvement of the test result, i.e., by a decrease in the value of SRT at repeated

*Corresponding author: erpe@mmmi.sdu.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7.
The Danavox Jubilee Foundation, 2014.

measurements until a certain number of measurements has been made – then the SRT values only vary with the uncertainty of the measurement. At the development of the Dantale II speech material Wagener *et al*. (2003) found a learning effect of 2.2 dB in an open-set test format. Wagener *et al*. (2003) performed eight subsequent measurements of SRT (containing 20 test sentences each) on normal-hearing subjects. The learning effect was determined as the difference between SRTs obtained at the first and eighth measurement. If two lists of 20 sentences were performed as training prior to an actual measurement, the learning effect was found to affect SRT by less than 1 dB (Wagener *et al*., 2003).

Hernvig and Olsen (2005) also investigated the learning effect using the Dantale II speech material in an open-set test format. The study distinguishes between two types of learning effects: the within-visit learning effect and the inter-visit learning effect. The within-visit learning effect corresponds to the learning effect investigated by Wagener *et al*. (2003), whereas the inter-visit learning effect is a learning effect found between SRT measurements that are substantially separated in time. Hernvig and Olsen (2005) performed six subsequent measurements of SRT (containing 30 test sentences each) on hearing-impaired subjects and found a within-visit learning effect (the difference between SRT determined at the first and the sixth measurement) of 3.2 dB. The inter-visit learning effect was found to be 1.6 dB with a median inter-visit period of 27 days (range: 14-43 days).

A within-visit learning effect in an open-set test format has also been found for the Swedish Hagerman test (Hagerman, 1984; Hagerman and Kinnefors, 1995), the German Oldenburg sentence test (Wagener *et al*., 1999) and the corresponding French test (Jansen *et al*., 2012). In a study by Brand *et al*. (2004) the within-visit learning effect has been investigated for the German material in a quasi closed-set test format and compared to that for an open-set test format. In the quasi closed-set test format the subject had, for each word presented, to select a response from ten alternative words (corresponding to the different words in the speech material) or they could answer ‘I do not know’ (each ‘I do not know’ answer was interpreted as an incorrect answer). The alternative answers for each word were listed in a matrix on a computer screen. The within-visit learning effect was found to be comparable in the quasi closed-set test format and in the open-set test format. A corresponding finding was made with the Spanish material (Hochmuth *et al*., 2012).

Even though previous studies showed that a learning effect exists and that it is needed to present a subject with training lists prior to an actual measurement, it is unknown what causes the learning effect. Some studies indicated that the observed learning effect is due to the sentences having a syntactically fixed structure and the number of different words in the material being limited (Hernvig and Olsen, 2005; Wagener and Brand, 2005). However, the studies by Brand *et al*. (2004) and Hochmuth *et al*. (2012) found no difference in the within-visit learning effect for an open-set and quasi closed-set test format. A difference would have been expected if the learning effect is caused by the composition of the speech material, since the subjects in the quasi closed-set test format were visually presented to the different words in the material.

It is interesting to study the learning effect because the number of lists required for training influences the total test time. To the authors' knowledge no previous study has investigated the influence of the target level (i.e., the level at which the sentences are presented) on the learning effect. If the learning effect is caused by the composition of the speech material, the learning effect could be expected to be influenced when the test sentences are presented at a target level higher than the normal 50% correctly understood words, i.e., when the subject hears more of the words presented. Furthermore, no study on the learning effect has to the authors' knowledge previously been performed with the Dantale II speech material in a quasi closed-set or a closed-set test format. Therefore, this study investigated whether the target level affects the learning effect (in an open-set test format) and whether the learning effect is influenced by a closed-set test format using the Dantale II speech material. In the closed-set test the subject had, for each word presented, to select a response from ten alternative words without the possibility to answer 'I do not know'. For each test condition both the within-visit learning effect and the inter-visit learning effect were determined.

METHODS

Speech material

The Danish speech material Dantale II (Wagener *et al*., 2003), which was used in this study, consists of 16 lists with ten test sentences each. The test sentences have a syntactically fixed structure of five words from different word classes in the order: name, verb, numeral, adjective, and noun. Since the test sentences are semantically unpredictable, the words cannot be predicted from the context. As an example the first sentence in list 1 is: 'Ingrid finds seven red houses' (translation of the Danish sentence: 'Ingrid finder syv røde huse'). The noise signal included in the speech material was generated by superimposing the test sentences many times by which the signal became speech-shaped without strong fluctuations.

Test versions

Three test versions were implemented: two with an open-set test format and one with a closed-set test format. In the two versions with the open-set test format the subject had to orally repeat as much heard as possible after each sentence presented. The operator then registered whether the subject's answer for each word was correct or incorrect. In the version with the closed-set test format the subject had to select a response from ten alternative words listed in a matrix for each word presented. The subject did not have the possibility to answer 'I do not know', i.e., the subject was forced to guess when a word had not been heard.

All three test versions were implemented using the adaptive procedure described in Brand and Kollmeier (2002). The presentation level, i.e., the signal-to-noise ratio (SNR) at which the sentences was presented, was adjusted from sentence to sentence depending on the number of correctly answered words given to the previous sentence, and on the advance of the test to stabilize the SNRs near the target level.

The adjusting was done by changing the level of the test sentences, whereas the level of the noise signal was kept constant at 65 dBC. The first sentence was presented at 0 dB SNR. For the two versions with the open-set test format the target level was set to 50% and 80% correctly understood words, respectively. For the version with the closed-set test format the target level was set to 50%.

Equipment

A specially-designed measurement program was developed in MATLAB 6.5 according to the three test versions and the adaptive procedure. Under the listening test a laptop with a touch screen (Acer model TravelMate C300XCi) was used. The subjects who were presented to the closed-set test format had to use the touch screen to give their answers after each sentence presented. The test sentences and the noise signal were presented to the subjects by a loudspeaker (Vifa P13WH00-08 in a 6.6-litres vented cabinet), which was connected to the laptop through a power amplifier (Bruel & Kjaer, type 2706). The subjects were seated 1.2 m in front of the loudspeaker.

Subjects

The listening test was performed with 24 normal-hearing subjects (12 males and 12 females, aged 21-39 years with a mean age of 26 years). The subjects were native speakers of Danish and had not been presented to the Dantale II speech material before the actual listening test. They had no otological problems and their hearing thresholds did not exceed 15 dB HL at the frequencies 0.5, 1, 2, and 4 kHz. The subjects participated in the study voluntarily without getting paid.

Test course

The 24 normal-hearing subjects were divided into three test groups of eight persons each, who were presented the two versions with the open-set test format and the version with the closed-set test format, respectively. For each subject eight subsequent measurements of SRT using two lists each were made to determine the within-visit learning effect. The subjects were presented to each of the 16 test lists in the speech material once. To avoid any effect of the list sequences, the presentation order of the lists was counterbalanced among the subjects. After a period of 12-16 days (mean: 14 days) one more measurement of SRT was made to determine the inter-visit learning effect. As at the first visit, the measurement of SRT included two lists. Within each of the three groups none of the subjects were presented to the same lists at the second visit.

Statistical analyses

For the statistical analyses the computer program SPSS 11.5.1 for Windows was used (www.spss.co.in). All analyses were performed at a 0.05 significance level. The Kolmogorov-Smirnov test was used to ascertain whether data for the different test conditions could be assumed to come from a normal distribution, and the Levene test was used to test for homogeneity of variance. To test for differences between the SRT values obtained, parametric tests were used, provided that the conditions for performing those tests were satisfied. Otherwise corresponding non-parametric tests were used.

RESULTS

Figure 1 shows the results from the SRT measurements for each of the three different test conditions, where the SRT values are given at the representative target level. From the figure it is seen that the mean value of SRT decreased at repeated measurements (indicating that the subjects scored ‘better’) until a certain number of measurements had been made. The curves for the three conditions have a similar shape but are vertically displaced. The highest SRT values were obtained for the subjects who were presented to the open-set test format with a target level of 80%. The higher target level causes the sentences to be presented at higher SNRs than for a target level of 50%, which results in higher SRT values. The SRT values are higher for the open-set test format than for the closed-set test format both with a target level of 50%. This can be explained by the fact that the subjects who were presented to the closed-set test format had the different words (response alternatives) listed in a matrix as a visual cue, which makes it easier to guess the correct word from the alternatives.

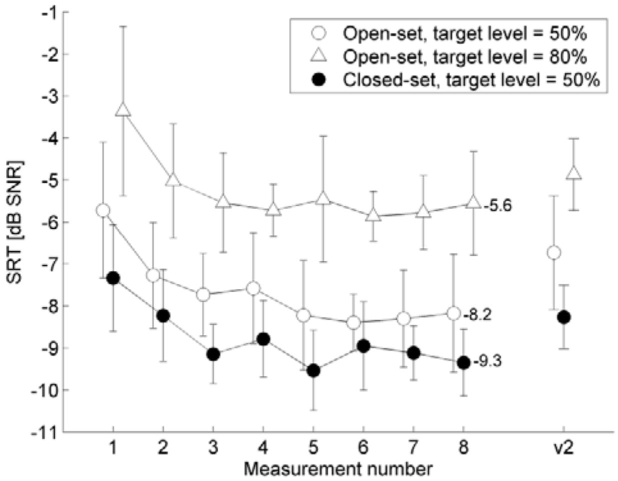


Fig. 1: Results of the SRT measurements as function of measurement number for the three test groups, which were presented to the two versions with the open-set test format and the version with the closed-set test format, respectively. For each measurement the mean SRT and one standard deviation are determined across eight normal-hearing subjects. The results marked 1 to 8 were obtained at the first visit, whereas the results marked v2 were obtained at the second visit, which took place 12-16 days after the first visit.

Figure 2 shows the within-visit learning and inter-visit learning effect. A Kruskal-Wallis test showed for the within-visit learning effect no statistical difference between the three test groups ($X^2(2) = 0.060$, $p = 0.970$). This finding for the open-set test format with a target level of 50% and the closed-set test format is in agreement with previous studies (Brand *et al.*, 2004; Hochmuth *et al.*, 2012). For all three test conditions the within-visit learning effect was comparable to the learning effect of 2.2 dB found by Wagener *et al*. (2003).

For the inter-visit learning effect shown in Fig. 2 a Kruskal-Wallis test showed no statistical difference between the three test groups ($X^2(2) = 1.005$, $p = 0.605$). For all three test conditions the inter-visit learning effect was lower than the inter-visit learning effect found by Hernvig and Olsen (2005) with hearing-impaired subjects. The within-visit learning effect in this study was also lower than in the study by Hernvig and Olsen (2005).

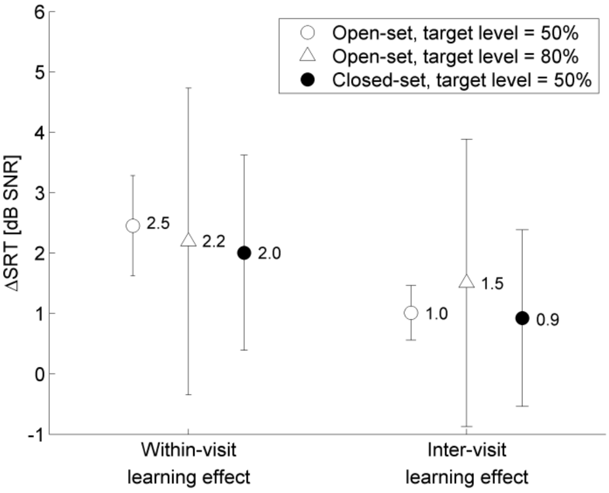


Fig. 2: Mean and one standard deviation of the within-visit and inter-visit learning effect. The within-visit is calculated as the difference between SRT obtained at the first and eighth measurement, whereas the inter-visit learning effect is calculated as the difference between SRTs obtained at the first measurement in the two visits.

To analyse when the within-visit learning effect can be assumed to be finished, paired-sampled *t*-tests (2-tailed) were performed between SRT values obtained at different measurements. For the SRT values obtained at the third and eighth measurement the tests showed no statistical difference for any of the three test conditions ($t(7) = 2.091$, $p = 0.075$; $t(7) = 0.024$, $p = 0.982$; $t(7) = 1.061$, $p = 0.324$), i.e., the learning effect can be assumed to be finished after two measurements. The difference between the second and eighth measurement were also analysed. For the open-set test format with a target level of 50% and for the closed-set test format a statistical difference was found ($t(7) = 2.927$, $p = 0.022$; $t(7) = 2.721$, $p = 0.030$), i.e., the learning effect cannot be assumed to be finished after only one measurement. For the open-set test format with a target level of 80% no statistical difference were found ($t(7) = 0.692$, $p = 0.511$), i.e., for this test condition the learning effect can be assumed to be finished after only one measurement.

DISCUSSION

The learning effect was found to be finished after only one measurement for the open-set test format with a target level of 80%. However, the learning effect for the closed-set test format was not found to be finished until after two measurements. Therefore, it is not possible to conclude whether the cause of the learning effect is dominated by the composition of the speech material or by the subjects having to adapt to the test situation and to listening for the words in the noise signal.

It could be interesting to investigate the learning effect in further details in a future study in order to obtain more insight in its causes, e.g., it could be interesting to investigate whether there is any difference between the learning effect obtained with a speech material as the Danish Dantale II speech material and a speech material containing everyday sentences (sentences without a syntactically fixed structure and with an unlimited number of words).

CONCLUSIONS

No statistical differences were found either in the within-visit or in the inter-visit learning effect for the three conditions tested. However, for the open-set test format with a target level of 80%, the learning effect was found to be finished faster than for the two other conditions.

Like previous studies this study shows the need for presenting the subjects with training lists prior an actual measurement to remove the effect of learning on the test result. Two training lists of 20 sentences seem sensible. The number of training lists might be reduced to one for an open-set test format with a target level of 80%. If the subject has been presented to the material within a short period of time training can be reduced.

ACKNOWLEDGEMENTS

The authors would like to thank Carsten Daugaard and Søren L. Jørgensen from DELTA Technical Audiological Laboratory, Odense, Denmark for providing some of the equipment and for valuable discussions. The authors also thank all the subjects who participated in the tests.

REFERENCES

Brand, T., and Kollmeier, B. (**2002**). “Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests,” J. Acoust. Soc. Am., **111**, 2801-2810.

Brand, T., Wittkop, T., Wagener, K., and Kollmeier, B. (**2004**). “Vergleich von Oldenburger Satztest und Freiburger Wörtertest als geschlossene Versionen,” (in German), Deutsche Gesellschaft für Audiologie (DGA).

Hagerman, B. (**1982**). “Sentences for testing speech intelligibility in noise,” Scand. Audiol., **11**, 79-87.

Hagerman, B. (**1984**). “Some aspects of methodology in speech audiometry – Studies of reliability, computer simulations and development of a new speech material for measuring speech reception threshold in noise,” Scand. Audiol. Suppl., **21**, 1-25.

Hagerman, B., and Kinnefors, C. (**1995**). “Efficient adaptive methods for measuring speech reception threshold in quiet and in noise,” Scand. Audiol, **24**, 71-77.

Hernvig, L.H., and Olsen, S.O. (**2005**). “Learning effect when using the Danish Hagerman sentences (Dantale II) to determine speech reception threshold,“ Int. J. Audiol., **44**, 509-512.

Hochmuth, S., Brand, T., Zokoll, M.A., Castro, F.Z., Wardenga, N., and Kollmeier, B. (**2012**). “A Spanish matrix sentence test for assessing speech reception thresholds in noise,” Int. J. Audiol., **51**, 536-544.

Jansen, S., Luts, H., Wagener, K.C., Kollmeier, B., Del Rio, M., Dauman, R., James, C., Fraysse, B., Vormes, E., Frachet, B., Wouters, J., and van Wieringen, A. (**2012**). “Comparison of three types of French speech-in-noise tests: a multi-center study,” Int. J. Audiol., **51**, 164-173.

Ozimek, E., Warzybok, A., and Kutzner, D. (**2010**). “Polish sentence matrix test for speech intelligibility measurement in noise,” Int. J. Audiol., **49**, 444-454.

Wagener, K.C., and Brand, T. (**2005**). “Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters,” Int. J. Audiol., **44**, 144-156.

Wagener, K., Brand, T., and Kollmeier, B. (**1999**). “Entwicklung und Evaluation eines Satztests für die deutsche Sprache – Teil III: Evaluation des Oldenburger Satztests,” (in German), Zeitschrift für Audiologie, **38**, 86-95.

Wagener, K., Josvassen, J.L., and Ardenkjaer, R. (**2003**). “Design, optimization and evaluation of a Danish sentence test in noise,” Int. J. Audiol., **42**, 10-17.

Prosody perception by postlingually-deafened cochlear implant recipients: a cross-language investigation

David Morris[1,*], Andrew Faulkner[2], Holger Juul[1], Lennart Magnusson[3], and Radoslava Jönsson[3]

[1] *Department of Scandinavian Studies and Linguistics-Speech Pathology and Audiology, University of Copenhagen, Denmark*

[2] *Department of Speech, Hearing and Phonetic Sciences, University College London, London, United Kingdom*

[3] *Department of Audiology, Sahlgrenska University Hospital, Gothenburg, Sweden*

Due to the inherent device limitations of cochlear implants (CI) and of auditory perception via an electrical-neural interface, the ability of CI listeners to perceive prosody is often reported as being worse than that of normal-hearing listeners. We tested the perceptual ability of postlingually-deafened adult CI listeners with stimuli where prosodic features signalled distinctive semantic contrasts. These contrasts were tested with a Danish (n=18) and a Swedish (n=21) cohort in quiet and in noise. We also tested other speech perceptual abilities that could be linked to prosody perception. These included word recognition, sentence perception in noise, and vowel identification. Results of this study show that speech-in-noise ability by CI listeners is related to abilities that underlie vowel identification, while word recognition is related to the identification of compound words and phrases. Comparison of the mean identification rates of the prosodic tasks showed that there was a disparity in the performance of Danish and Swedish CI listeners over tasks that are similar in both languages.

INTRODUCTION

Prosody is an integral part of spoken communication that imbues speech with dynamic variation. This variation is perceived as a result of variations in the acoustic cues of pitch, intensity, and rhythm during the course of an utterance. Cochlear implant (CI) listeners can have difficulty in perceiving changes in these cues. The purpose of the present investigation was to examine the prosodic ability of CI listeners where prosody provided a distinctive semantic contrast, the perception of which is necessary for accurate identification of natural utterances in both Swedish and Danish. In testing the Swedish participants we included a quiet and a noise condition in order to involve a situation relevant to everyday listening.

Noise generally has a deleterious effect on most measures of speech intelligibility by CI listeners. This is due to device factors and to limitations in the electrical-neural interface. Device factors include limited transmission of both spectral and temporal information. Electrical-neural interface limits also appear to restrict the number of effective spectral channels, so that, for example, improvements in speech

*Corresponding author: dmorris@hum.ku.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

intelligibility in noise plateau when the number of channels is increased above eight (Friesen *et al.*, 2001). This plateau may be attributable to a broad spread of in-vivo excitation caused by the distance between a stimulating electrode and the receptor site. The net effect of these factors is that spectral detail is poorly represented to CI listeners, and listening in background noise is problematic. The encoding of more rapid temporal information is also limited by neural factors beyond the limits imposed by the speech processing (Zeng, 2002).

To examine the extent to which naturally-occurring prosodic detail is available to CI listeners we compared their performance to normal-hearing (NH) listeners in tasks which required the identification of a word or phrase from two known alternatives. Together these two answer choices constitute a prosodically contrastive minimal pair, that is, the two words are segmentally identical, or close to identical, but their prosodic characteristics differ. The prosodic features that were investigated were word stress, vowel length, stød, and compounds and phrases.

Vowel duration and word stress are critical features in many Scandinavian languages as they can denote semantic distinctions between words. For instance, in Swedish, ‘väg’ /vɛːg/, with a long vowel means ‘road’ and ‘vägg’ /vɛg/, with a short vowel, means ‘wall’. An example of word stress arises in the English word ‘convict’ which can be either a noun or a verb according to the position of the stress: for example, ‘Because of crimes committed in prison, the judge will ˳conˈvict (verb) the ˈcon˳vict (noun)’. Stød is a prosodic feature that is peculiar to the Danish language. It is a syllabic feature characterized by irregular laryngeal vibrations that affect a long vowel or consonant (Grønnum, 2001). In the Swedish study, we also tested the ability of CI and NH listeners to identify compound words as opposed to phrases. This perceptual ability is critical to parsing the speech stream, because there are semantic distinctions between compound words and phrases. An example of a compound word and a phrase pair is the word ‘blackbird’ and the phrase ‘black bird.’ The distinction between the two is exemplified in the sentence, ‘the raven is a black bird, but not a blackbird.’

Our interest in examining naturally uttered minimal pairs was also to consider performance on these tasks with testing that is carried out as a part of clinical routine. Cullington and Zeng (2011) reported that neither performance on the HINT sentence test in quiet nor in babble noise correlated with prosodic scores from tasks where adult CI listeners were required to identify affective emotion. In contrast, Rogers *et al.* (2006) reported that there was a link between performance on prosodic discrimination and sentence perception in quiet. They found that the ability of adult CI listeners to discriminate changes in prosodic stress signalled by concurrent F0 and intensity modification correlated with speech perception as measured with the CNC words. To further investigate the speech perceptual ability of CI listeners as measured with standard clinical tests in relation to their ability to perceive prosody, we used a series of minimal-pair identification tasks where prosodic features of words provided a distinctive semantic contrast.

METHOD

Participants

Swedish CI participants (n=21) were recruited from the Sahlgrenska University Hospital register. These were 8 males and 13 females with a mean age of 64.3 years (range 40-82). All participants had been unilaterally implanted within the last nine years and, with the exception of two participants, all had word recognition scores in quiet above 50% at their most recent post-implant testing. The processing strategies used by the Swedish listeners included ACE (18), FSP (2), and HDCIS (1).

The Danish CI participants (n=18) were recruited by advertisements displayed online and in social media. There were no selection criteria based on prior knowledge of hearing ability. These participants had a mean age of 53.3 years (range 41-70). Eleven of the CI participants had been bilaterally implanted. However, all were tested with only one implant. The Danish participants used ACE (12), HiRes fidelity 120 (3), SPEAK (1), MP3000 (1), and CIS (1).

The Swedish CI participants had a mean of 4.8 yrs of experience listening via their CI while the Danish CI participants had 4.7 yrs. No participant suffered from a cognitive handicap or reading impairment and all were native speakers of their respective language. All participants were reliant on oral communication in their everyday lives. In instances where a participant wore a contralateral hearing aid, this was removed prior to testing. In cases where a participant used bilateral implants, they were instructed to select the side from which they believed they derived the most benefit. An earplug was used to block the contralateral ear of participants that reported residual hearing in that ear.

The Swedish control group (n=10) consisted of NH adults with a mean age of 51 years (range 36-70). The Danish control group consisted of NH (n=16) participants with a mean age of 30.4 years (range 23-61).

Minimal-pair testing

The minimal-pair tasks were chosen on the basis of their phonetic characteristics by the first author. A panel of native speakers reviewed the Swedish items and two expert phoneticians reviewed the Danish stimuli. The words in each pair were spelt differently so that participants could distinguish between the orthographically presented response alternatives. Table 1 provides an overview of the test stimuli in both languages.

The vowel identification stimuli from the Danish study consisted of both short and long vowels, which were tested separately. Short- and long-vowel stimuli were presented in factorial combination such that each stimulus item was presented with all other items in that series as response alternatives. For example, the word ‘mit’ was presented as a response alternative for the stimuli ‘mit,’ ‘midt,’ ‘mæt’, and ‘mat’.

Language	*Vowel length*	*Word stress*	*Compound word/phrases*	*Stød*	*Vowel quality - short*	*Vowel quality - long*
Swedish	väg [vɛːg] vägg [vɛg]* (34)	korset [ˈkɔrɕɛt] korsett [kɔrˈɕɛt]* (18)	rökfritt rök fritt* (82)	-	-	-
Danish	læser [lɛːsʌ] læsser [lɛsə] (44)	August [ˈɑwgɔsd] august [ɑwˈgɔsd] (14)	-	Brugsen [bʁuˀsən] brusen [bʁuːsən] (22)	mit [mɪd] midt [med] mæt [mɛd] mat [mad] (12)	mile [miːlə] mele [meːlə] mæle [mɛːlə] male [mæːlə] (12)

Table 1: Examples of the minimal-pair stimuli used in these experiments from both languages. Values in parentheses are the number of stimulus items per test. * indicates contrasts that were tested in quiet and in a noise condition where ICRA unmodulated speech-shaped noise was combined with the stimuli at an SNR of 10 dB.

Two male adult non-professional speakers recorded the stimuli. Both Swedish and Danish speakers could be considered as having representative dialects with which all participants would be very familiar. Recordings of all stimuli were made in a sound-treated environment with a quality microphone. The individual members of each minimal pair were edited so that they were preceded by the carrier phrase 'Ordet är ___' (Swedish) or 'Ordet er ___' (Danish) [The word is ___]. The root mean square (RMS) level of each stimulus item and the carrier phrase were adjusted to a uniform value. To create the noise conditions used in the Swedish study, an unmodulated speech-spectrum-shaped random noise from the International Collegium of Rehabilitative Audiology (ICRA) collection was used. This noise was added at an SNR of 10 dB. This SNR was considered to be challenging but not impossible for CI listeners based on the mean CI perceptual results for the steady noise reported in Fu and Nogaki (2005) and Nelson *et al.* (2003).

Procedure

Stimuli were presented via a single loudspeaker placed at a distance of one meter from the listener. CI participants used their clinically-assigned speech processor in the preferred setting. Participants were instructed to identify the interval that was different and respond on a keyboard. Repetitions of stimuli were not permitted. Although no training round was provided, it was noted that no participant was consistently hesitant in identifying answer choices. The presentation level of the minimal-pair stimuli was 70 or 75 dB(A).

Speech in noise and word identification testing

During the testing session the speech reception threshold (SRT) in noise of the Danish CI listeners was measured with the Danish HINT (Nielsen and Dau, 2011). The speech material in this test consists of five-word sentences read by male speakers presented in a speech-shaped noise with minor amplitude modulation. Participants were required to repeat the sentence that was presented and scoring was by words correct.

The Swedish PB word identification test was measured during a routine follow-up appointment with the Swedish CI listeners. For a description of this test, see Magnusson (1995). One list containing 50 words was used and was presented without the addition of noise, from a loudspeaker positioned one meter in front of the subject in a sound treated room.

RESULTS

Minimal pair tasks

The individual and mean results for the minimal-pair identification tasks performed by the Swedish group can be seen in Table 2 (upper panel). The lower panel shows the mean results for the minimal-pair identification tasks performed by the Danish group. It can be noted that in both languages the NH participants showed ceiling levels of performance on all tasks, and the addition of noise in the Swedish study did not markedly affect this.

	vowel length quiet	*vowel length noise*	*word stress quiet*	*word stress noise*	*compounds/ phrases quiet*	*compounds/ phrases noise*
CI	0.95	0.93	0.93	0.87	0.82	0.77
NH	1	1	0.99	1	0.97	0.95

	vowel length	*word stress*	*stød*	*vowel quality - long*	*vowel quality - short*
CI	0.83	0.74	0.77	0.86	0.81
NH	1	0.97	0.96	0.99	0.98

Table 2: Upper panel shows mean proportions correct on the Swedish minimal-pair tasks and lower panel shows the Danish results.

Other measures of speech perceptual ability

The mean results from the Danish listeners in the vowel identification tasks are provided in Table 2 (lower panel). In both the long and short vowel sets, the open and closed vowels at the extremities of the production plane from which the stimuli were drawn were often successfully identified. The mean proportions of correct identifications for these stimuli were 0.93 for /iː/, 0.87 for /i/, 0.96 for /æː/, and 0.97 for /æ/. In contrast, mean identification rates for vowels from the middle of the frontal plane were poorer. For /eː/ they were 0.7, for /ɛː/ 0.84, and for /e/ 0.66.

Group mean results from the Swedish CI listeners on the PB word test were 73.3% (S.D. = 15). The mean Danish HINT results from the CI group were 17.8 dB SNR (S.D. = 9).

Correlations between minimal-pair results and standard measures

In comparing the results from the Swedish CI listeners to their most recent performance on the PB word test a correlation was observed between the compound word and phrase task in quiet ($r = 0.58$, $p < 0.01$) and in noise ($r = 0.64$, $p < 0.01$). Correlation coefficients for the word stress results in noise were high but were not significant ($r = 0.39$, $p = 0.09$). Neither the word stress in noise nor the vowel-length results in quiet nor in the noise condition were significantly correlated with the PB word-recognition scores. Significant negative correlations were found between the SRT from the Danish HINT and the individual mean scores on the tests of vowel length ($r = -0.78$, $p = 0.001$) and long vowel identification ($r = -0.73$, $p = 0.001$). The negative correlations indicate that participants with lower SRTs performed better on the identification tests. The word-stress identification ($r = -0.59$, $p = 0.01$) and short-vowel identification ($r = -0.51$, $p = 0.03$) were also found to correlate with the Danish HINT results, but these were not significant after Bonferroni-adjusted corrections.

DISCUSSION

This study shows that CI participants could not identify naturally-uttered prosodic cues that distinguish words as well as NH participants in both of the Scandinavian languages that were tested. The performance of Swedish CI participants on all minimal-pair tasks was significantly negatively affected by the introduction of noise. This noise condition had little effect on the performance of the NH participants, and the scores from this group were close to ceiling. The results from the identification tasks from the Danish and the Swedish listeners are of interest as they can provide a cross-language comparison of prosodic features that are similar. Although minimal-pair test items in both languages were not cognate words, the Swedish CI listeners generally performed better than the Danish CI listeners.

It is possible that the selection of the Swedish CI participants according to results on the PB words test contributed to the performance differences that were noted. This meant that only Swedish CI listeners with word-recognition scores above approximately 50% were admitted to the study, whereas the admittance of Danish participants to the study was not based on any speech-performance measures. Nonetheless, due to the magnitude of the performance differences between the CI listeners from both languages (0.12 for vowel length and 0.19 for word-stress identification), other factors may have been involved. One such factor that may explain the poorer performance observed in the Danish cohort is the lenition process to which the Danish language has been subject. The consonant reduction and schwa assimilation and deletion (Bleses *et al.*, 2008) that has been a consequence of this may have made the minimal-pair identification tasks harder for Danish CI listeners as segmental reductions would diminish the distinctive cues on which identification was based. A factor that may have contributed to the better performance observed in the Swedish cohort on the word-stress identification task is that the Swedish language exhibits prominent stress patterns that can be transcribed by listeners even when the language is hummed (Svensson, 1974; Bruce, 1998).

Another reason that may have benefited the Swedish CI listeners' performance on both vowel-length identification and word stress is that intrasyllabic durational cues in Swedish are relatively symmetrical. This relationship is what Lehiste (1970) termed the 'mutual complementation of vocalic and consonantal quantity', which means that a short vowel is followed by a long consonant and a long vowel by a short consonant. These relationships deliver essentially a doubling of durational cues to the listener as the durations of both the vowel and the postvocalic consonant signal the length of the vowel. It is also possible that the higher identification score observed for the Swedish CI listeners on the vowel-duration test were promoted by Swedish orthography, in which the representation of the postvocalic consonant clearly marks vowel durational oppositions with a double letter for the short vowel and a single letter for the long vowel, for instance, 'söt' is a long vowel and 'sött' is a short vowel. However, neither of these explanations accounts for the difference in word-stress identification scores observed in the Swedish and Danish CI participants.

The correlation that was observed in the Swedish cohort between PB word results and compound word and phrase identification scores indicates that there is a link between the ability to identify monosyllables and the accurate placement of word boundaries. This correlation was found to be stronger in the noise condition than in quiet, suggesting that those listeners that are adept at placing word boundaries when the stimuli are marred are also better at word recognition in quiet. It has been suggested that sensitivity to the stress patterns of a language assists in the division of the speech stream during language acquisition (Jusczyk *et al.*, 1999). One of the perceptual abilities that are believed to underlie the placement of word boundaries is differentiation of stress patterns that are contained within words (word stress) and those which are contained within phrases (phrasal stress). The correlation that we found, albeit weak, between PB word scores and word-stress identification in noise supports this notion. It would be of interest to investigate how postlingually-deafened adult CI listeners use their knowledge of other language-specific characteristics, for instance, phonotactic properties and distributional characteristics, to perform speech segmentation.

This study highlights the limitations of the everyday listening abilities of CI listeners and the problems that they face in perceiving prosody as it occurs in natural utterances. The correlation results suggest that there is a link between sentence perception in noise and abilities that underlie vowel identification by CI listeners. Also, word recognition in quiet appears to be linked to the ability to assign word boundaries accurately. Furthermore, these findings indicate that results from some common speech audiometric routines can yield information about the prosodic perceptual abilities of CI recipients.

REFERENCES

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T.O., and Basboll, H. (**2008**). "Early vocabulary development in Danish and other languages: a CDI-based comparison," J. Child. Lang., **35**, 619-650.

Bruce, G. (**1998**). "Allmän och svensk prosodi" [General and Swedish prosody], Vol. 16. Lund: Reprocentralen.

Cullington, H., and Zeng, F.-G. (**2011**). "Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination and talker identification," Ear Hearing, **32**, 16-30.

Friesen, L.M., Shannon, R.V., Baskent D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am., **110**, 1150-1163.

Fu Q.J., and Nogaki G. (**2005**). "Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol., **6**, 19-27.

Grønnum, N. (**2001**). "Fonetik og fonologi – almen og Dansk" [Phonetics and phonology – general and Danish], Akademisk forlag.

Jusczyk, P.W., Houston, D.M., and Newsome, M. (**1999**). "The beginnings of word segmentation in English-learning infants," Cogn. Psychol., **39**, 159-207.

Lehiste, I. (**1970**). *Suprasegmentals* (Cambridge, Mass.: M.I.T. Press).

Magnusson, L. (**1995**). "Reliable clinical determination of speech recognition scores using swedish PB words in speech-weighted noise," Scand. Audiol., **24**, 217-223.

Nelson, P.B., Jin, S.H., Carney, A.E., and Nelson, D.A. (**2003**). "Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am., **113**, 961-968.

Nielsen, J.B., and Dau, T. (**2011**). "The Danish hearing in noise test," Int. J. Audiol., **50**, 202-208.

Rogers, C.F., Healy, E.W., and Montgomery, A.A. (**2006**). "Sensitivity to isolated and concurrent intensity and fundamental frequency increments by cochlear implant users under natural listening conditions," J. Acoust. Soc. Am., **119**, 2276-2287.

Svensson, S.G. (**1974**). "Prosody and grammar in speech perception," Monographs from the Institute of Linguistics, **2**.

Zeng, F.G. (**2002**). "Temporal pitch in electric hearing," Hear. Res., **174**, 101-106.

Facial configuration and audiovisual integration of speech: a mismatch negativity study

KASPER ESKELUND[1,2,*], LAURA FRØLICH[1], AND TOBIAS S. ANDERSEN[1,2]

[1] *Section for Cognitive Systems, Department of Applied Mathematics and Computer Science, Technical University of Denmark*

[2] *CHeSS, Oticon Centre for Hearing and Speech Sciences, Technical University of Denmark*

Visual speech plays a central role in general speech perception. Through audiovisual integration, visual speech may facilitate auditory detection and identification for people with normal hearing in noisy conditions. Further, a visual syllable may alter the auditory phonetic percept, as can be seen in the McGurk illusion. In this study, we investigate the role of the configuration of facial features in perception of audiovisual speech. Face perception is known to be highly sensitive to specific arrangements of facial features. By nature, visual speech perception – and thus bimodal integration of audio-visual speech – relies on information from the talking face. However, visual speech encoding and face perception are known to be functionally separate. Previous behavioral findings have shown that for some speech tokens, audiovisual speech perception is altered when the facial configuration is manipulated, even though the constituent features are unchanged. This suggests a functional dependency between the encoding of audiovisual speech and face perception. Here, we investigate the effect by means of electrophysiology in a mismatch-negativity paradigm. Specifically, we present stimuli that support face perception and stimuli that do not, but only find mismatch negativity indicating audiovisual integration with the former.

INTRODUCTION

The integration of acoustic and visual speech signals is known to be beneficial for speech reception in many ways. Acoustic speech is detected at lower intensities if accompanied by a corresponding talking face (Grant and Seitz, 2000; Eskelund *et al.*, 2010). Visual speech can also facilitate speech comprehension (Sumby and Pollack, 1954). When visual and auditory speech are phonetically incongruent, an illusory alteration of the phoneme perceived in the voice can occur. This is known as the McGurk illusion (McGurk and MacDonald, 1976).

Perception of natural audiovisual speech evidently relies on visual perception of the talking face. Here, the visual signal emanating from the lip area plays a central role. Silent, visual speech is, e.g., known to modulate activity in auditory cortex (Calvert and Campbell, 2003). When considering visual perception of the talker in general, face perception directed towards the configuration of facial features may also be involved.

[*] Corresponding author: kaes@dtu.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Interestingly, the functions of visual speech and face perception have been suggested to be separate cognitive modules (Bruce and Young, 1986). But does facial configurational information influence audiovisual perception of speech?

The importance of configurational information for face perception is demonstrated by the so-called Thatcher illusion (Thompson, 1980). This striking illusion is based on four different manipulations of a face stimulus: a) a normal face with upright facial context and upright mouth (UF-UM), b) facial context kept upright, but mouth area inverted vertically (UF-IM), c) facial context inverted vertically but mouth area kept upright (IF-UM), d) facial context and mouth area both inverted vertically (IF-IM). When presenting these stimuli, Thompson (1980) observed that they were all perceived as normal faces, except for stimulus UF-IM, which was perceived as strikingly grotesque. Although the relation between directions of facial context and mouth area in stimuli UF-IM and IF-UM are identical, holistic or configurational mismatch is only perceived in the upright facial context. Thus, facial configuration is only perceived for stimuli with upright facial context (UF).

To investigate the influence of facial configuration on audiovisual speech perception, Rosenblum and colleagues (2000) used video stimuli based on the Thatcher illusion. The four visual stimulus modifications were combined with audio, forming congruent and incongruent (McGurk-type) audiovisual speech tokens, which according to direction of facial context supported or did not support perception of facial configuration. For the incongruent audiovisual syllable consisting of an auditory /ba/ and a visual /va/, Rosenblum *et al.* reported 90% visually-driven (McGurk) responses for UF-UM stimuli, while this tendency was reduced to 45% for UF-IM stimuli. Thus, audiovisual integration was reduced when perception of facial configuration was obstructed. This finding suggests a role for face perception in audiovisual speech perception.

The hypothesis of some degree of dependency of audiovisual integration in speech perception on holistic properties of the talking face is intriguing. In the present study, we investigate if the behavioral findings are mirrored in a neural differential response.

Specifically, we attempt testing the influence of configurational face information by electrophysiological means, using the mismatch negativity (MMN) paradigm developed by Näätänen *et al.* (1978). MMN is a component in the auditory event-related potential (ERP), generated by presenting a sequence of identical standard auditory stimuli at a constant inter-trial interval. At random places in the sequence, usually in 9-15% of stimulus presentations, the stimulus is altered (deviant trials). The alteration must be noticeable and can be in, e.g., intensity, pitch, modulation frequency, spatial location, or, in the case of speech stimuli, phoneme. When averaging ERPs due to standard and deviant stimuli, a negative deflection of the deviant ERP is observed. This has been hypothesized to be due to a memory process comparing each incoming stimulus with the established trace of standard stimuli (Näätänen, 2003). Whenever a deviant stimulus occurs, a differential neural response is evoked.

Sams and colleagues (1991) showed that the McGurk illusion can elicit MMN without any acoustic difference between standard and deviant stimuli. In this McGurk MMN paradigm standard trials are congruent combinations of, e.g., an audiovisual /ba/. Phonetic deviance is then induced by McGurk-type audiovisual integration with incongruent audiovisual stimuli, e.g., /ba/ + /va/. Thus, only the visual phoneme is altered in deviant trials.

We chose phonemes /ba/ and /va/ as in Rosenblum's study (2000) and generated new stimuli for use with native Danish-speaking subjects. To keep the duration of the MMN paradigm within practical limits for EEG recordings, only two visual stimulus types were used, i.e., UF-UM and UF-IM, which yielded normal audiovisual integration and reduced audiovisual integration responses in Rosenblum's study, respectively. For the UF-UM stimuli, we would expect normal bimodal integration, resulting in a McGurk-type percept with deviant stimuli, and thus an MMN signature in the ERP. UF-IM stimuli, on the other hand, are expected not to support audiovisual integration due to their disruption of normal face perception. Thus, deviant stimuli should not induce an MMN response with UF-IM stimuli.

To ensure that audiovisual integration was present in all subjects, a behavioral task was devised after the EEG recordings. In the behavioral task, subjects were asked to identify the same stimuli as presented in the EEG experiment.

METHODS

Subjects

24 engineering students and university faculty members participated, 11 female. Mean age 29 years, age range 21-59. Five subjects were excluded due to electrode failure or movement artifacts.

Stimuli

Stimulus material was generated from a video recording of syllables /ba/ and /va/. Each video was recorded at 30 fps and lasted 31 frames. Sound was recorded at 44.1-Hz sampling rate and 16-bit depth. The single auditory /ba/ was combined with four different visual stimuli: a visual /ba/ with upright face and upright lips and a visual /va/ with upright face and vertically-inverted lips. This yielded congruent and incongruent UF-UM syllables, and congruent and incongruent UF-IM syllables.

Stimuli were presented on a 19" CRT screen and with Etymotic Research ER-2 ear probes at an intensity of 60 dB SPL. Subjects were seated in a comfortable armchair in a dimly lit, shielded EEG booth at a distance of 1.2 meters from the visual display.

Behavioral task

The behavioral task consisted of a random presentation of 25 trials of each of the four audiovisual stimuli. After each trial, subjects were prompted to identify what they just heard in response categories 'ba', 'da', 'fa', or 'va'.

EEG recordings

EEG was recorded on a BioSemi ActiveTwo 64-channel system with six EOG and two mastoid electrodes. The data were sampled at 512 Hz.

The four stimuli were presented in the following sequence: Two conditions were constructed, consisting of UF-UM and UF-IM audiovisual stimuli, respectively. In each of these conditions, a congruent /ba/ + /ba/ combination was used as standard, while a /ba/ + /va/ was used as deviant stimulus. Each grand condition was presented in two blocks, consisting of a total of 550 trials each. In each block, 15% of trials were deviant stimuli, which were distributed randomly in the sequence, with the condition that at least 2 and maximally 9 standards followed each deviant. 30 standard stimuli preceded each block as a training sequence so that the memory trace for the standard stimulus could be established. To counter movement artifacts, the stimulus sequence was paused every two minutes to allow for a 20-second break where subjects were instructed to relax. In total, 1100 stimuli were presented in each condition, of which 165 were deviants. The duration of each EEG recording was approx. 1 hour and 30 minutes, including breaks between blocks.

RESULTS

EEG recordings

All analyses were performed with the EEGLAB toolbox developed for MATLAB (Delorme and Makeig, 2004). Continuous data from the EEG recordings were bandpass-filtered between 1 and 30 Hz and referenced to averaged mastoids. Noisy electrodes were detected by a measure of kurtosis, and if any were found, their original channel data were replaced with data interpolated from surrounding electrodes. Data was segmented to epochs from 100 ms before to 600 ms after auditory onset and baselined to the 100-ms period preceding auditory onset. As a means of artifact rejection, an independent component analysis was used to reveal activity distributions and time-series attributable to non-neural sources such as eye-blinks, muscular artifacts, loose electrodes, etc. After decomposition, artifactual components were selected and removed upon visual inspection of spatial distributions and time-series. Residual artifacts were removed by applying a simple threshold of −100/+100 µV on all electrodes.

Pre-selection of subjects

The MMN paradigm of the experiment relies on multiple perceptual and neuro-physiological effects. These are well-known effects, but do not occur in all members of a given population. The prevalence of acoustic MMN is high, but not universal. This is also the case for the McGurk effect, which is the auditory illusion that drives the audiovisual MMN. In the present experiment, we look for changes in audiovisual MMN when the facial configuration is altered. To be able to securely observe this, we pre-selected subjects that displayed an audiovisual MMN driven by the McGurk effect with a normal face (the UF-UM condition). Eight subjects were pre-selected on the criterion of an audiovisual MMN with UF-UM stimuli of > 1 µV 200-400 ms post-stimulus.

ERPs from the vertex electrode (Cz) are shown in Fig. 1. For the UF-UM condition presented in Fig. 1, the standard and deviant ERPs follow the same pattern until approx. 200 ms post stimulus, where a negative deflection of the deviant ERP starts.

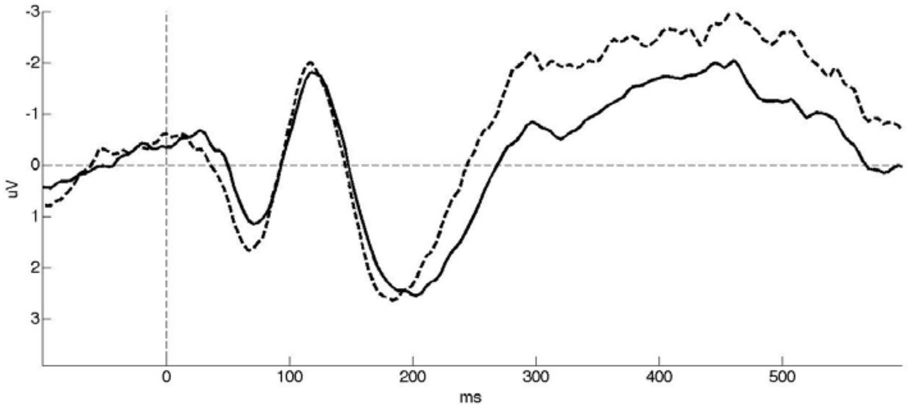


Fig. 1: Average ERPs recorded from UF-UM stimuli at electrode Cz. Auditory onset at 0 ms. Full line represents ERPs due to standard stimuli. Dashed line represents ERPs due to deviant stimuli.

Interestingly, ERPs from the UF-IM condition displayed in Fig. 2 do not show the same tendency. Here, deviant ERPs show a general, but less articulate positive shift, which starts at the beginning of the auditory stimulus.

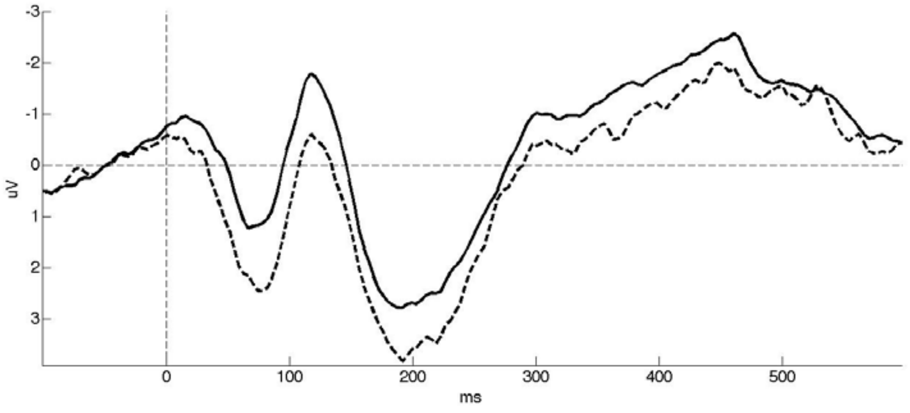


Fig. 2: Average ERPs recorded from UF-IM stimuli at electrode Cz. Auditory onset at 0 ms. Full line represents ERPs due to standard stimuli. Dashed line represents ERPs due to deviant stimuli.

In the UF-UM condition, a mismatch negativity pattern is easily seen in the difference between deviant and standard ERPs. As is evident in Fig. 3, the UF-UM condition generates an MMN response beginning at approx. 200 ms and culminating with an amplitude of −1.43 µV at 280 ms. To detect reliable differences in the MMN from zero, we submitted the ERPs producing the difference wave to a repeated measures, two-tailed permutation test based on the *tmax* statistic (Blair and Karniski, 1993), using a family-wise alpha of 0.05. All time-points between 200 and 600 ms were included in the test. 2500 random within-subject permutations of the data were used to estimate the distribution of the null hypothesis (i.e., no difference between ERPs, or difference wave at zero). Based on this estimate, a critical *t*-score of +/−4.31 was derived, i.e., any differences between the ERPs that exceeded this *t*-score were deemed statistically significant. This was the case for portions from 240 to 360 ms and 460 to 530 ms. The maximal *t*-score was −10.8 at 290 ms.

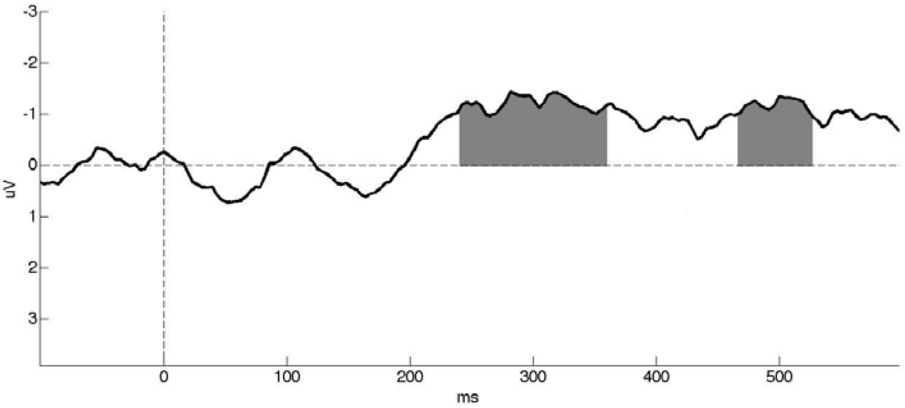


Fig. 3: Difference wave representing the difference between deviants and standards in the UF-UM condition at electrode Cz. Auditory onset at 0 ms. Shaded area marks statistically significant portions of the difference wave (exceeding the critical *t*-score of +/−4.31).

As can be seen in Fig. 4, the UF-IM condition generated a differential response (deviant ERP minus standard ERP) with less amplitude and reverse polarity. In this case, a permutation test identical to the one used for UF-UM data above revealed no portions of the UF-IM standard and deviant ERPs (see Fig. 2) to differ significantly (critical *t*-score +/−3.54, maximal *t*-score in the window 200 to 600 ms was +1.38 at 450 ms).

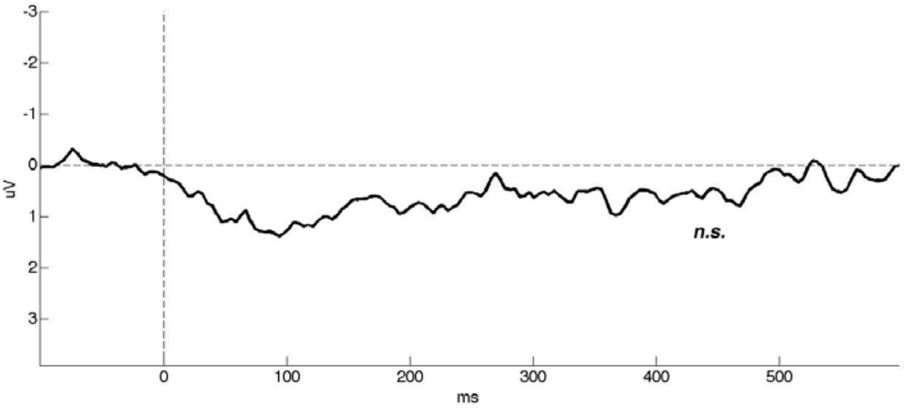


Fig. 4: Difference waves representing the difference between standards and deviants in the UF-IM condition at electrode Cz. Auditory onset at 0 ms.

Behavioral task

Observers' responses in the behavioral task were re-categorized as correct ('ba') and incorrect (all other responses). Here, we consider the mean percentage incorrect identifications as a measure of the strength of the McGurk illusion (listed in Table 1).

As can be seen in Table 1, incongruent UF-UM stimuli produced clear audiovisual integration responses, whereas incongruent UF-IM stimuli produced a less clear result, suggesting reduced bimodal integration. Responses were arcsine-transformed to correct for the heterogeneity of variances and analyzed using a two-way (syllable × mouth direction) repeated-measures ANOVA. Arcsine-transformation did not change the outcome of any of the hypothesis tests. Factor 'syllable' had two levels (congruent and incongruent). Factor 'mouth direction' had two levels (upright mouth and inverted mouth). *p*-values were Greenhouse-Geisser-corrected when appropriate.

	UF-UM	UF-IM
Congruent /ba/ +/ba/	1.5 (0.7)	4.5 (1.5)
Incongruent /ba/ + /va/	93.0 (2.1)	27.0 (6.4)

Table 1: Percentage incorrect identifications of the acoustic phoneme /ba/ in the behavioral task after EEG recordings. First value is mean proportion incorrect identifications, numbers in brackets represent standard error of mean.

The results showed that the interaction between syllable and mouth direction was significant ($F(3,21) = 120.1$, $p < 0.001$), indicating an effect of mouth direction on syllable identification. We further performed repeated measures ANOVAs to compare identification performance pairwise between syllables and between mouth directions. Performance differences between congruent and incongruent syllables were significant for UF-UM ($F(1,7) = 238.1$, $p < 0.001$) and UF-IM stimuli ($F(1,7) = 14.5$, $p < 0.01$). The difference in congruent syllable identification between UF-UM and UF-IM stimuli was not significant ($F(1,7) = 2.3$, $p > 0.1$). Finally, the difference in incongruent syllable identification between UF-UM and UF-IM stimuli, i.e., the difference in audiovisual integration responses between the two facial configurations, was significant ($F(1,7) = 69.0$, $p < 0.001$).

DISCUSSION

Results from the behavioral task match the findings of Rosenblum *et al.* (2000). In the present results, the difference in audiovisual integration responses was even slightly more articulate, with 93% in the UF-UM condition vs. 27% in the UF-IM condition.

MMN results mirrored the behavioral findings. Here, the large MMN generated by visual phonetic deviance with UF-UM stimuli effectively vanished with UF-IM versions of the same stimuli. The minor, positive deflection observed was not found to reliably differ from zero, and it is hypothesized to be due to random fluctuations. Thus, we conclude that facial configuration had a significant impact on MMN generated by audiovisual integration.

It is worth noting, that subjects were pre-selected for analysis on the basis of their MMN in the UF-UM condition. However, the object of the present study was the change in audiovisual integration between UF-UM and UF-IM conditions and not audiovisual MMN in isolation. Because the audiovisual MMN per se is not universally present in subjects, a pre-selection was necessary. The pre-selection in the present study, however, does not differ much from selection rates in other audiovisual MMN studies (cf. Colin, 2002).

Our behavioral and neurophysiological findings support the findings of Rosenblum and colleagues (2000) in suggesting that facial configuration information influences audiovisual integration in speech perception.

REFERENCES

Blair, R.C., and Karniski, W. (**1993**). "An alternative method for significance testing of waveform difference potentials," Psychophysiology, **30**, 518-524.

Bruce, V., and Young, A. (**1986**). "Understanding face recognition," Br. J. Psychol., **77**, 305-327.

Calvert, G.A., and Campbell, R. (**2003**). "Reading speech from still and moving faces: The neural substrates of visible speech," J. Cogn. Neurosci., **15**, 57-70.

Colin, C. (**2002**). "Mismatch negativity evoked by the McGurk–MacDonald effect: a phonetic representation within short-term memory," Clin. Neurophysiol., **113**, 495-506.

Delorme, A., and Makeig, S. (**2004**). "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," J. Neurosci. Meth., **134**, 9-21.

Eskelund, K., Tuomainen, J., and Andersen, T.S. (**2010**). "Multistage audiovisual integration of speech: dissociating identification and detection," Exp. Brain Res., **208**, 447-457.

Grant, K.W., and Seitz, P.-F. (**2000**). "The use of visible speech cues for improving auditory detection of spoken sentences," J. Acoust. Soc. Am., **108**, 1197-1208.

McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature, **264**, 746-748.

Näätänen, R., Gaillard, A.W.K., and Mäntysalo, S. (**1978**). "Early selective-attention effect on evoked potential reinterpreted," Acta Psychol., **42**, 313-329.

Näätänen, R. (**2003**). "Mismatch negativity: clinical research and possible applications," Int. J. Psychophysiol., **48**, 179-188.

Rosenblum, L.D., Yakel, D.A., and Green, K.P. (**2000**). "Face and mouth inversion effects on visual and audiovisual speech perception," J. Exp. Psychol. Hum. Percept. Perform., **26**, 806-819.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O.V., Lu, S.-T., and Simola, J. (**1991**). "Seeing speech: visual information from lip movements modifies activity in the human auditory cortex," Neurosci. Lett., 127, 141-145.

Sumby, W.H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am., **26**, 212-215.

Thompson, P. (**1980**). "Margaret Thatcher: a new illusion," Perception, **9**, 483-484.

Context-dependent quality parameters and perception of auditory illusions

STEPHAN WERNER[1,*], FLORIAN KLEIN[1], AND TAMÁS HARCZOS[2]

[1] *Technische Universität Ilmenau, Institute for Media Technology, Ilmenau, Germany*

[2] *Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany*

This contribution introduces context-dependent quality elements, which have significant influence on perception of an auditory illusion. Binaural synthesis of an acoustic scene via a personalized headphone system is used. The investigated elements are divergent between synthesized scene and listening room, visibility of the scene, and personalization of the system. Two rooms with different acoustic parameters are used as recording and listening room. The test persons listen either to the same room as the listening room or to the other room. The plausibility of the perceived auditory scene is described by the probands with the help of the parameter perceived externality of the auditory event. Because it is unknown if the relevant quality elements are acoustically or visually based, two groups of test persons are used. The first group has no visual cues (dark room), while the second group sees the synthesized source positions and listening room. We have found significant differences in perceived externality depending on the synthesized and listening room, on the two groups, and on personalization of the system.

MOTIVATION

The development of audio systems is motivated by the purpose to create perfect auditory illusions with a high degree of immersion and plausibility (Heeter, 1992; Lindau and Weinzierl, 2011). A lot of work is done to increase the technical quality of such systems. Systems which use the principles of binaural synthesis are one possibility to achieve auditory illusion. Binaural synthesis takes the underlying perceptual processes conditioned by the direct synthesis of the corresponding sound pressure at the ear drums of a listener into account. The technical parameters are therefore well understood and controllable (see, e.g., Hess, 2006; Silzle, 2007). Sound sources in rooms can be described by binaural room impulse responses (BRIRs). The BRIRs can be derived from acoustic room simulations or from measurements of real sound sources in real rooms. A personalization of the binaural system is achievable by using individual BRIRs and individual headphone equalization for example. In addition to the technical realization of the correct binaural synthesis and signals, many psychoacoustic effects in perception of auditory scenes and their interconnections are not completely understood until now.

*Corresponding author: stephan.werner@tu-ilmenau.de

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Such effects cover for example multimodal interactions between acoustical and visual stimuli like the McGurk-effect (McGurk and MacDonald, 1976) or the ventriloquism-effect (Bertelson and Radeau, 1981; Seeber and Fastl, 2004; Werner *et al.*, 2012). Other perceptual effects depending on the congruence or divergence between the synthesized scene (including room) and the listening situation also seem to have a not neglectable influence on perception (Werner and Siegel, 2011). The quality of experience of an audio reproduction system depends on technical quality elements of the system but also on context-dependent quality parameters. To contribute to the improvement of binaural synthesis this paper focuses on investigations on acoustic divergence between listening room and synthesized room, visibility of the listening room and simulated source positions, and on personalization of the binaural synthesis system. The quality of experience is measured with listening experiments. The ratings of perceived externalization of the auditory event are shown. However, this quality feature is only one possible feature that has an influence on a plausible perception of an auditory illusion (Raake and Blauert, 2013).

BINAURAL SYNTHESIS VIA HEADPHONES

For generating test stimuli, binaural recordings of individual and ‘mean’ (manikin KEMAR) BRIRs for the used rooms and sound source positions and the auralization via headphones were prepared. The binaural system was customized for each participant to avoid within-cone and out-of-cone of confusion errors (Møller *et al.*, 1996) and to increase the simulation’s similarity compared with the real loudspeakers (Begault and Wenzel, 2001). A listening lab and seminar rooms with defined room acoustics and an adequate source-receiver distance were chosen to include reverberation. Reverberation encourages the perception of externalization of an auditory illusion and the impression of distance (Laws, 1973). The headphones were equalized using individual headphone transfer functions (HPTFs) if individual BRIRs were used. HPTFs from the head-and-torso simulator (KEMAR) were used if ‘mean’ BRIRs were used. In-ear microphones were used to measure individual BRIRs and HPTFs at the entrance of the blocked ear canal of each subject. The microphones are not removed between the BRIR and HPTF measurements. The measurements of the HPTFs were averaged over five recordings, repositioning the headphones for each recording. The inverse of a HPTF was calculated by a least-square method with minimum phase inversion (Schärer and Lindau, 2009). The measurements of the BRIRs were averaged over three recordings. Stax Lambda Pro headphones were used for playback.

OBJECTS OF INVESTIGATION

The listening experiments were focused on the evaluation of context-dependent quality parameters and their influence on the perception of externality of the auditory event. Two listening tests were conducted. Both tests investigated the combinations of listening room and synthesized room. Additional context-dependent quality parameters like visibility of the listening room and personalization of the binaural synthesis were investigated in the first test. The second test was focused on perceived externalization depending on different distances of the synthesized sound source. Binaural recordings of non-individual BRIRs (KEMAR head-and-torso simulator) were prepared for the used rooms and sound source positions to generate the test stimuli in the second test.

Acoustic divergence between rooms

A listening lab (Rec. ITU-R BS.1116, V = 179 m^3, RT60distance (2m) = 0.16 s), a depleted seminar room (V = 182 m^3, RT60reference distance (2m) = 1.4 s), and another seminar room (V = 182 m^3, RT60distance (2m) = 0.9 s) with different room acoustic characteristics were used for the listening tests and the measurement of the BRIRs at a distance of 2.2 m. The tests were conducted in the same listening lab (HL) and the same seminar rooms (SR) to evaluate the influence of the listening situation. The left part of Fig. 1 shows the combinations of listening room and synthesized room used in the tests.

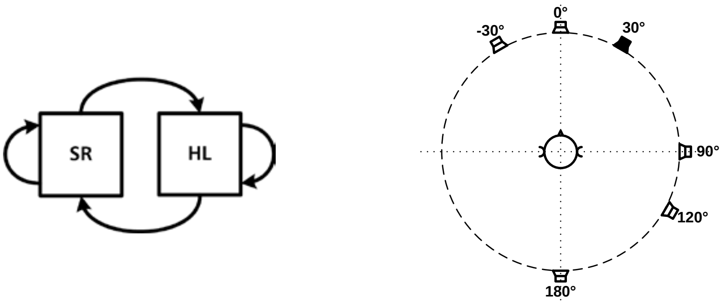


Fig. 1: Left: Combinations of listening room and synthesized room used in the listening tests; SR = seminar room, HL = listening lab; Right: Positions of the binaural synthesized sound sources for playback via headphones; distance of the sources to the listener (midpoint of the figure) approx. 2.2 m; the filled position (30°) was used in test two.

Visibility of the listening room

The test persons were randomly divided into two groups depending on the presence of visual cues within the tests. For the first group the illumination of the listening rooms was minimized (nearly complete darkness) and a sound-transparent black curtain with a distance of 2.2 m was arranged around the test persons. The test persons should have no visual impression or visual cues of the listening room. The test persons in the second group were placed in the illuminated listening rooms and dummy loudspeakers were placed at each hour position on a clock-like circle to provide additional visual cues. This situation was also used in the second test.

Sound source positions

Five sound source directions were checked for test one and one direction was used in test two. A Genelec 1030A loudspeaker was used to measure the BRIRs for each position. The right part of Fig. 1 shows the different positions. The distance from the loudspeaker to the listening point was approx. 2.2 m for test one and two. The height of the source position was approx. 1.3 m (ear position of a sitting person). The BRIRs for each position and for each test person were recorded in the two rooms. The recording position was the same as the listening position in the test.

Personalization of the binaural synthesis system

The individual BRIRs of the test persons from the two rooms and source directions and the individual headphone transfer function were recorded in a preceding session. Furthermore, the BRIRs and HPTFs of a KEMAR head-and-torso simulator (45BA) were recorded. Both the individual and 'mean' BRIRs were used to create the binaural test stimuli for test one. For test two only the 'mean' BRIRs were used.

LISTENING TESTS

Test one: The listening test was conducted in the listening lab and the seminar room separately in two sessions at different days. In every session every test person listened to individually synthesized and dummy-head synthesized source positions of both recording rooms. The stimuli were presented two times in a random order. The perceived incidence angle could be rated by choosing the respective direction on a top-down view. Externalization could be rated by choosing the midpoint, inner circle, or outer circle. The attribute externalization was oriented to definitions given by Hartmann and Wittenberg (1996). The following definitions were used in the test: a) midpoint: "The sound event is entirely in my head or it is very diffuse."; b) inner circle: "The sound event is external but it is next to my ears or head."; c) outer circle: "The sound event is external and good locatable." Note that the definitions were given in German.

Test two: The test persons rated the externalization in the listening test. The test persons indicated the externality of the auditory event by pressing one of three buttons on a graphical user interface. The same scale as in test one was used. The synthesized BRIRs of several distances from the listening lab and the seminar room were used as stimuli. A more detailed description about the BRIR synthesis and the test design can be found in Werner and Sass (2013).

Twenty-one test persons participated in the first and 16 test persons in the second listening test. The test persons were well experienced with listening tests and were trained before each test. For the first test the training consisted of an oral and written introduction and a definition of the used attributes localization and externalization. Each subject had to listen to all different test items. The test persons could compare each item with the others and could listen to each item several times. The test persons had to rate each test item on the same rating sheet as in the main test session. For the second test a presentation of non-binaural stereo panned signals, a playback via the reference loudspeaker, and a binaural synthesis of the reference loudspeaker were used as training. The test persons should build up an own internal reference and had to define differences between the items for the attributes localization and externalization.

RESULTS

The ratings of the test persons for externalization were counted as frequencies. The frequencies showed no significant dependency from the used sound signal. Both signals were put together for analysis. An externalization index was calculated as ratio between the ratings of extern (outer circle on the rating sheet) and all ratings within the test. An index of 0 indicates in-head localization, while an index of 1 indicates out of the head localization of the auditory event.

Figure 2 shows the rating of perceived externalization depending on the presence of visual cues, personalization method, and combinations of listening room and synthesized room. The midpoints of the polar plots represent an externalization index of 0 while the outer circle represents an index of 1 (linear scale in between). Wilcoxon signed rank tests at the 5% confidence level were conducted for statistical testing. The upper row of Fig. 2 shows the externalization indexes for the reverberant seminar room as listening room, while the lower row shows the ratings for the less reverberant listening lab as listening room.

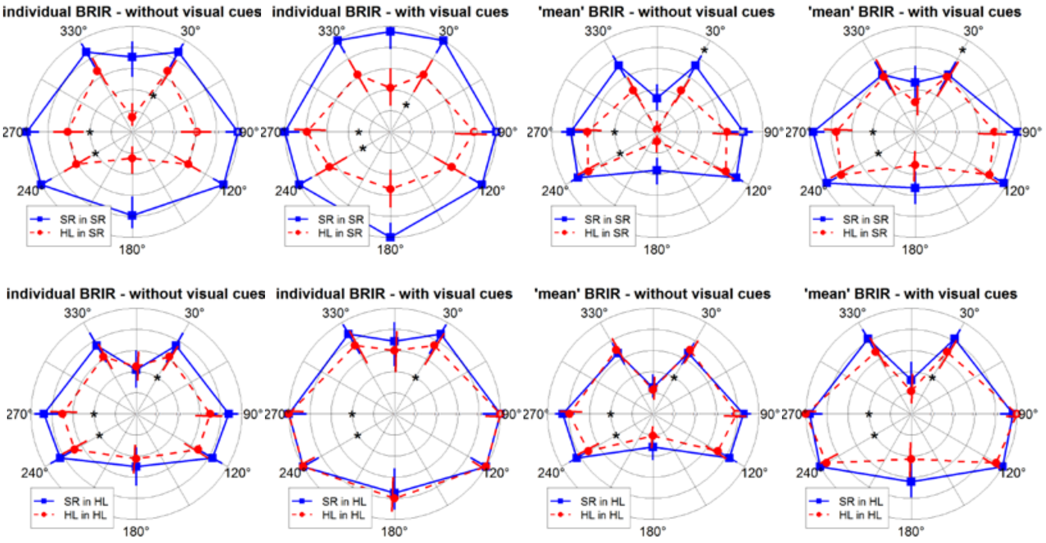


Fig. 2: Ratings for perceived externalization as externalization indexes depending on the combinations of listening room and synthesized room, personalization of the binaural synthesis, and presence of visual cues with 95% confidence intervals from test 1; SR = seminar room, HL = listening lab, * are the mirrored ratings at the 0° to 180° axis.

A general lower externalization index is achieved for binaural synthesis using ‘mean’ BRIRs compared to individual BRIRs. Very low indexes are visible especially for the direct front and back directions. The usage of an individualized synthesis increases the perceived externalization of the auditory event significantly for the direct front and back directions. Furthermore, a higher index is visible for congruence between the listening room and synthesized room (SR in SR) related to divergence between the rooms (HL in SR). This effect is mostly significant if individual BRIRs are used. The ratings show no significant differences if ‘mean’ BRIRs are used for the synthesis. However, the magnitude of the indexes is decreased compared to individual BRIRs at congruence between listening and synthesized room (SR in SR). The room effect is maybe covered by the effect caused by the personalization of the binaural synthesis. Further research is needed to determine the interconnection between these two context-dependent quality elements. The effect caused by room divergences seems to be independent of the visibility of the listening room. However, the visibility of the room increases the indexes especially for the front and back directions. The lower row of Fig. 2 shows the ratings of test one for the less reverberant listening lab as listening room. Significant differences depending on divergence or congruence between the listening and synthesized room are visible in contrast to the seminar room as listening room for the direct front and back directions. The visibility of the room also increases the externalization indexes for all conditions. The room effect seems to be much more present for synthesis of a less reverberant scene in a more reverberant room.

Figure 3 shows the rating as externalization index for different combinations of listening room and synthesized room and additionally for different distances of the auditory event. A similar effect of dependencies of the rooms is visible as in test one. Clearly higher ratings are reached if the synthesized room is the same as the listening room especially for the more reverberant seminar room (SR in SR compared to HL in SR). The source distance of one meter is rated with the lowest externalization indexes while the more far away distances are rated with higher values. Saturation is visible for the synthesis of the seminar room but not for the less reverberant listening lab. An increase of the externalization index is visible for synthesis of the listening lab in the listening lab (HL in HL) compared to the synthesis of the listening lab in the seminar room (HL in SR) for the distance of 5 m. The ratings of test two are consistent with the ratings of test one for the 2.2-m distance, 30° direction, and ‘mean’ personalization of the binaural synthesis.

CONCLUSIONS

The ratings from two listening tests to evaluate the perceived externalization of an auditory event using a binaural auralization via headphones were reported. Five source positions, four combinations of listening room and synthesized room, and two personalization methods were investigated. A dependency of the perceived externalization of an auditory event from the used personalization method was shown. Higher externalization indexes are reached especially for the direct front and back direction and for the frontal lateral direction. This is in contrast to own former investigations (Werner and Siegel, 2011). It would be insightful to investigate the correlation between externalization and errors in perception of direction.

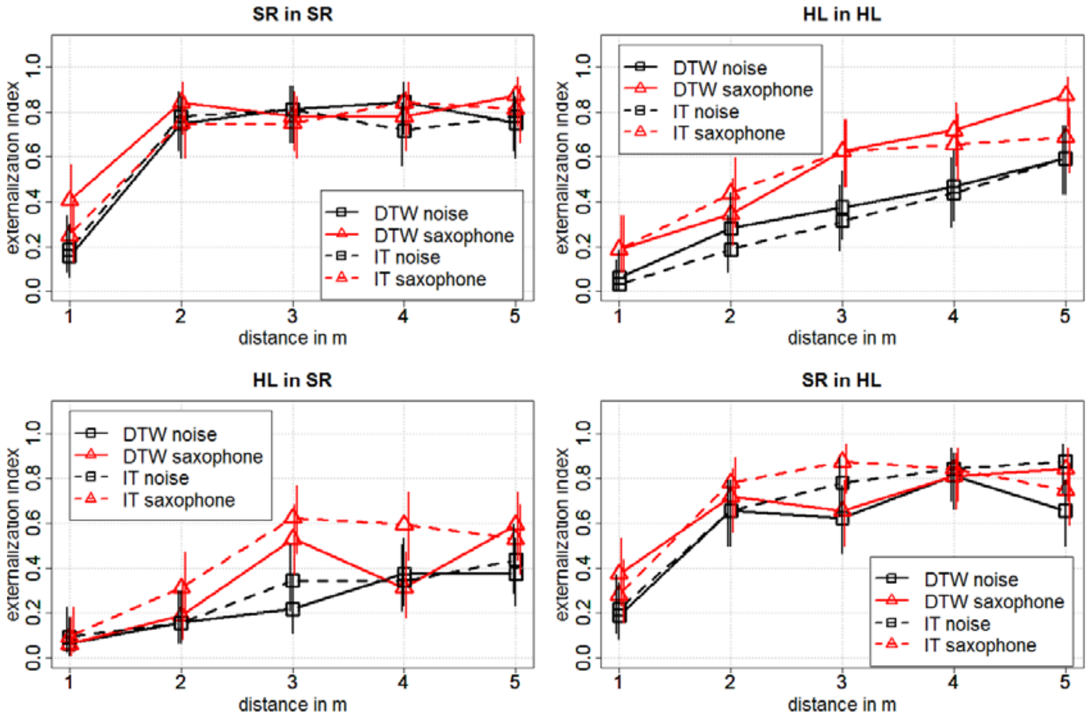


Fig. 3: Perceived externalization depending on distance of synthesized sound sources from test 2; BRIRs of different distances using interpolation methods (Werner and Sass, 2013) with 95% confidence intervals; azimuth of source direction = +30°; IT = interpolation in time domain; DTW = interpolation + dynamic time warping; 1 m = measured start-BRIR; 5 m = measured target BRIR; HL = listening lab, SR = seminar room.

Furthermore, low externalization indexes were found for synthesis of the less reverberant room in the more reverberant room. The highest externalization indexes were found for playback of test signals from the reverberant room in the same room. The personalization method maybe covers the room effect. The interconnection between personalization and room divergences is not well-known until now. The presence of visual cues has a supporting effect on the perceived externalization independent of the personalization method and combination of listening and synthesized room. The effect of perceived externalization depending on room divergences seems to be an acoustically based context-dependent quality element. Further investigations in evaluation of detailed quality elements based on a variety of plausibility features are meaningful.

REFERENCES

Begault, D.R., and Wenzel, E.M. (**2001**). "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," J. Audio Eng. Soc., **49**, 904-916.

Bertelson, P., and Radeau, M. (**1981**). "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance," Percept. Psychophys., **29**, 578-584.

Hartmann, W.M., and Wittenberg, A. (**1996**). "On the externalization of sound images," J. Acoust. Soc. Am., **99**, 3678-3688.

Heeter, C. (**1992**). "Being there: The subjective experience of presence," in *Presence: Teleoperators and Virtual Environments* (MIT Press).

Hess, W. (**2006**). *Time-Variant Binaural-Activity Characteristics as Indicator of Auditory Spatial Attributes*, PhD Thesis, Ruhr-Universität Bochum, Bochum, Germany.

Laws, P. (**1973**). "Entfernungshören und das Problem der Im-Kopf-Lokalisiertheit von Hörereignissen [Auditory distance perception and the problem of 'in-head localization' of sound images]," Acustica, **29**, 243-259 (NASA Technical Translation TT-20833).

Lindau, A., and Weinzierl, S. (**2011**). "Assessing the plausibility of virtual acoustic environments," Forum Acusticum, European Acoustic Association, Aalborg, Denmark, pp. 1187-1192.

McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," *Nature*, **264**, 746-748.

Møller, H., Sørensen, M.F., Jensen, C.B., and Hammershøi, D. (**1996**). "Binaural technique: Do we need individual recordings?" J. Audio Eng. Soc, **44**, 451-469.

Raake, A., and Blauert, J. (**2013**). "Comprehensive modeling of the formation process of sound quality," 5th Int. Workshop on Quality of Multimedia Experience (QoMEX), Klagenfurt, Austria, pp.76-81.

Schärer, Z., and Lindau, A. (**2009**). "Evaluation of equalisation methods for binaural signals," Proc. of the 126th AES Conv., preprint 7721.

Seeber, B., and Fastl, H. (**2004**). "On auditory-visual interaction in real and virtual environments," Proc. ICA 2004, 18th Int. Congress on Acoustics, Kyoto, Japan, volume III, Int. Commission on Acoustics, pp. 2293–2296.

Silzle, A. (**2007**). *Generation of Quality Taxonomies for Auditory Virtual Environments by Means of Systematic Expert Survey*, PhD Thesis, Ruhr-Universität Bochum, Bochum, Germany.

Werner, S., and Siegel, A. (**2011**). "Effects of binaural auralization via headphones on the perception of acoustic scenes," Proc. of the 3rd International Symposium on Auditory and Audiological Research, ISAAR, Denmark, pp.215-222.

Werner, S., Liebetrau, J., and Sporer, T. (**2012**). "Audio-visual discrepancy and the influence on vertical sound source localization," Quality of Multimedia Experience (QoMEX), Fourth International Workshop, Australia, pp.133-139.

Werner, S., and Sass, R. (**2013**). "Synthesis of binaural room impulse responses," AIA-DAGA Merano 2013, pp. 572-575.

Attitudes, rewards, and listening-habits in Danish youth

MORIN REINESS[1], CARSTEN DAUGAARD[2,*], AND PER NIELSEN[3]

[1] *University of Copenhagen, Department of Scandinavian Studies and Linguistics, Njalsgade 120, DK-2300 Copenhagen S, Denmark*

[2] *DELTA, Technical-Audiological Laboratory, Edisonsvej 24, DK-5000 Odense C, Denmark*

[3] *Københavns Kommune, CSV, Frankrigsgade 4, DK-2300 Copenhagen S, Denmark*

This study surveyed more than 1,800 Danish teenagers' habits and attitudes towards MP3 listening. The questionnaire registered self-reported sound exposure, listening behavior, perceived rewards of listening and the effect and media preferred for prophylactic information. A 'risk group' of approx. 10% of respondents was defined, which in terms of relative size corresponds well to other recent studies. In general, the risk group indicated more reasons for listening to loud music. However, the three most popular reasons, independent of risk categorization, were: "I can better feel/enjoy music when it is loud", "I can lose myself in loud music", and "I get energy from listening to loud music". More than 40% of the risk group indicated "I relax better with loud music" and "I get a pleasant bodily effect with loud music". Not surprisingly, the pattern of use revealed that the risk group use their MP3-player in more situations, and for notably longer periods of time, such as reading, sleeping, and by the computer. The respondents indicated that information on potential hearing risks from MP3-usage is preferably received via television and commercials or from nurses and doctors. The most effective examples seen in the survey were actual case stories, medical argumentations, or the experience of hearing-loss symptoms.

BACKGROUND

The MP3 player is often criticized for exposing younger generations to music at excessive sound levels, increasing the risk of noise-induced hearing loss later in life. Unlike previous sound systems, the MP3 players are easy to carry and capable of delivering uninterrupted music for prolonged periods at high listening levels, which notably increases the possible exposure. This observation is based on the general assumption that sound energy (product of time and level) is the cause of noise-induced hearing loss. To prevent such hearing damage in the younger generations the European Union has issued a regulation which prohibits MP3 players from delivering more than 100 dB SPL (SCENHIR, 2008). However, it seems that the information campaigns on the hazardous effects of MP3-listening had a small impact on the MP3-users' behaviour. The hypothesis behind this study is that the limited effect might stem from the rewards experienced when listening to (loud) music,

[*] Corresponding author: cd@delta.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

exceeding the perceived risk from the sound exposure. To develop effective prevention strategies, this study was designed to investigate the listening behaviour among teenagers and the experienced effects from listening to (loud) music, as well as their self-reported music exposure.

DEVELOPING THE QUESTIONNAIRE

A 25-item web-based questionnaire was developed with inspiration from the thoughts of Barry Blesser that the instant rewards from loud music might outweigh the long-time negative effects to hearing (Blesser and Salter, 2008). As there were no existing questionnaires with this exact focus, a new one was developed. It was assumed that assessing the listening habits and the rewards from listening could not be extracted from an 'open' questionnaire. Therefore, a predetermined or 'closed' set of answers was developed. The respondents could respond to the questions by selecting the appropriate option or selecting their degree of agreement with a statement, allowing their attitude, behavior, and habits to MP3-player usage to be determined. The closed sets of answers were adjusted through pretests, which aimed to locate the largest set of realistic situations and rewards. The questionnaire was developed to be administered online through Enalyzer.com. The questionnaire items were divided into six main categories: 'Demographic information', 'Listening habits', 'Rewards', 'Symptoms of hearing loss', 'Knowledge and attitudes towards MP3 loud music listening', and 'The effect of media and prophylactic information".

Statistical analyses

The data were extracted in pivot tables in Excel, and the risk and non-risk groups were tested for differences compared to the population using the Chi-square statistics. A probability level of $p < 0.05$ was used to determine the statistical significance of the results.

REWARDS FROM LOUD MUSIC

Since ancient times, music has been used as a tool to change the emotional state of the listener; raising the spirit, calming or soothing the mind, as well as experiencing physical effects such as relaxation or as an energy booster (Blesser, 2007). Loud music drowns out fainter sounds, which enables some kind of territorial dominance. Since non-amplified music requires much effort to be played out loud, the loud played music represents more strength and power. If the music is kept in headphones, it offers the listener the possibility of getting lost in the music and mask unpleasant sounds from the outside surroundings. Louder music is more effective in the masking. Also, MP3 music can provide the listener with physical pleasure, improve his/her mood and concentration level Vogel *et. al.* (2011). Salimpoor *et al.* (2011) showed that music can trigger dopamine and thus activate the reward center of the brain. This provides a neurologic explanation as to why music can act as a mood agent. Music is simply a stimulant, like sugar, caffeine, exercise, sex, etc. Florentine *et al.* (1998) questioned 90 young people in their musical behavior, based upon the 'Michigan Alcoholism Screening Test', and concluded that 9% of the questioned scored high enough to be qualified as ‘music alcoholics’. They indicated a behavior corresponding to alcoholics, as they felt the necessity of loud music, as a way of releasing tension. They ignored negative consequences (i.e., tinnitus) and they even showed withdrawal symptoms when deprived of loud music.

DEFINITION OF RISK GROUP

To investigate the behavior of teenagers’ risky MP3-player listening, compared to a control group not at risk of hearing damage, a so-called ‘risk group’, based on self-reported data, was defined. Portnuff *et al.* (2011) reports the average maximum outputs in dB(A) of three different types of headphones across all MP3 players and all music signals. In this way, self-reported volume settings can be converted to listening levels, and combined with listening time an exposure value can be calculated and compared with 50% daily noise dose limits from MP3 alone (cf. Table 1 and Fig. 1), given that a person may be exposed to other intense noise during the day. Based on this a risk group of about 10% of the respondents was constructed. The relative size of the risk group corresponds well to earlier studies (Degn, 2009; SCENHIR, 2008).

Maximum listening time per day depending on headphone type and volume control setting			
% of VC	Earbud	Isolator	Supra-aural
10-50%	No limit	No limit	No limit
60%	No limit	14 h	No limit
70%	6 h	3.4 h	19 h
80%	90 min	50 min	4.6 h
90%	22 min	12 min	66 min
100%	5 min	3 min	16 min

Table 1: The table shows an average listening time to 50% noise dose (8 hours 85 dB LAeq) using the criteria for a noise risk from the National Institute for Occupational Safety and Health (NIOSH) (Portnuff *et al.*, 2011:669).

MAIN RESULTS

The main results are presented with five headlines ‘Demographic data’, ‘Listening habits’, ‘Rewards’, ‘Knowledge and attitudes’, and ‘The effect of media and prophylactic information’, and correspond roughly to the six dimensions defined in the construction of the questionnaire.

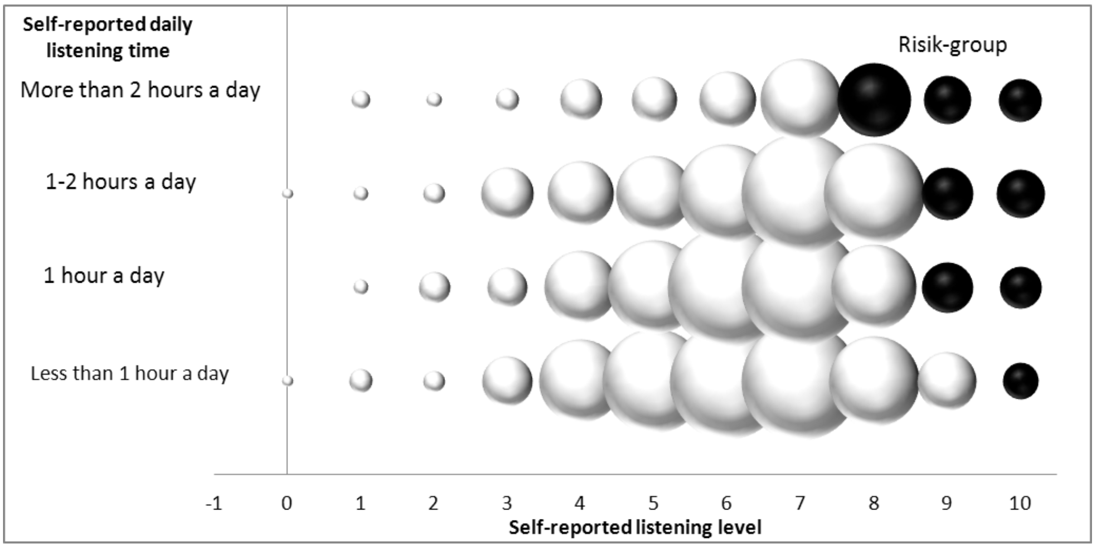


Fig. 1: The figure shows the distribution of self-reported listening time and MP3-player volume setting. The black balls indicate the people in the risk-group defined as too loud and/or to long listening sessions.

Demographic data

Although 1,828 completed the questionnaire, the distribution of education and gender of the received answers were not representative for the population. The answers showed that women attending the Danish secondary school, ‘Gymnasium (Stx)’ were more prevalent than in the population. Normally around 30% of the whole population attends gymnasium, and among the respondents of this questionnaire there were more than 80%. Furthermore, 66% of the questionnaires were answered by girls, and 34% by boys. Via the homepages of a number of gymnasiums, it was possible to distribute the questionnaires, which explains the higher prevalence from this education, and possibly also the gender effect, since there are probably more females attending gymnasium. Furthermore, a geographical bias in the respondents was uncovered. These factors must be kept in mind when interpreting the results from the questionnaire. It is easy to imagine a profile of a young male who has a practical education (perhaps operating a noisy machinery) having more loud listening habits, compared to a young woman attending classes most of the day. Nevertheless, the risk group also included a significant share of secondary school attendants.

Listening habits

One main question to be asked was of course: Do you listen to music at a volume level that might affect your hearing? The pie-charts in Fig. 2 indicate the distribution of answers in the risk and the non-risk group respectively. Clearly the risk-group is aware of its own risky behavior.

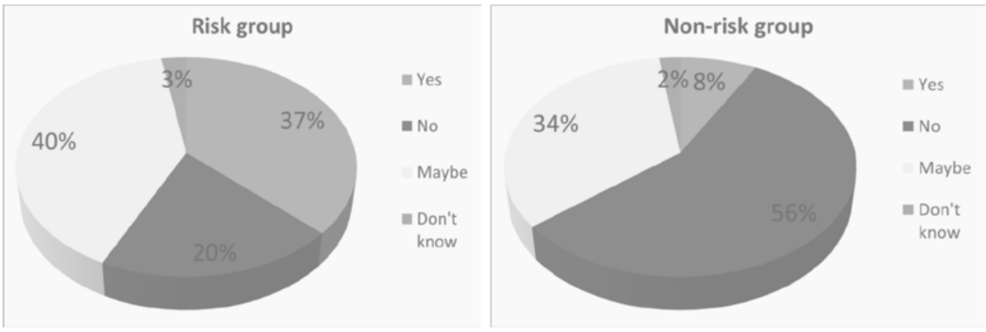


Fig. 2: Distribution of the answers to the question 'Do you listen to music at a volume level that might affect your hearing?' in the risk and in the non-risk group.

A supplementary question on the users' own evaluation of the volume setting on their MP3 player indicated that 31% in the non-risk group considered their typical volume setting to be 'high' or 'very high', whereas 81% in the risk group were of the same opinion. The most commonly reported reasons for not listening at louder volume levels were to protect the hearing and not to bother others with their music in the non-risk group. The most reported reason in the risk group was also not to bother others with their music, but the second most common reason was that their MP3 players were unable to play louder, as shown in Fig. 3. Another question reveals that more than 1/3 in the non-risk group were often, or constantly, having trouble hearing their surroundings while listening to MP3 music, whereas in the risk group the corresponding number was 2/3.

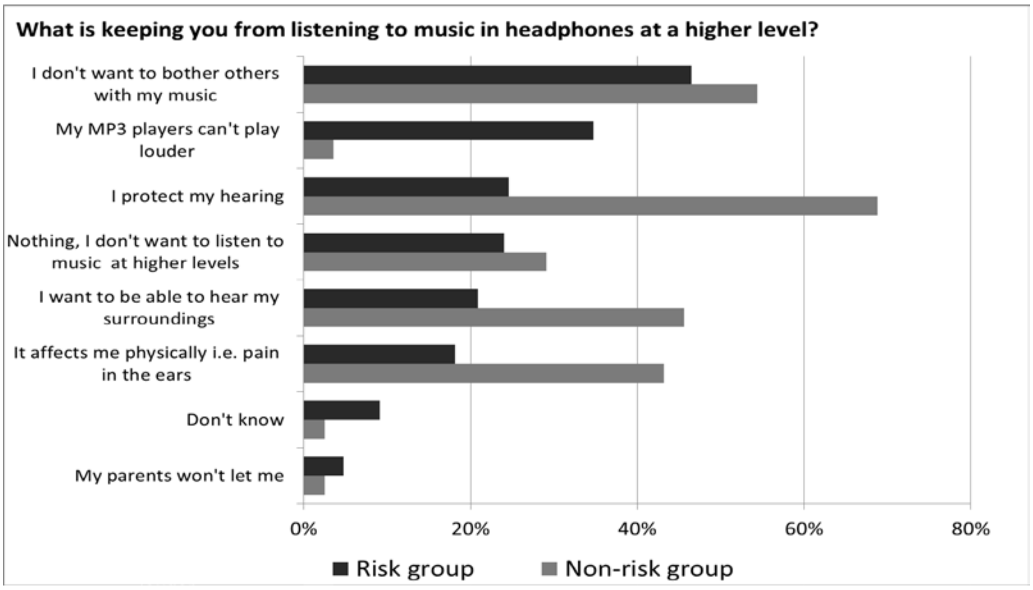


Fig. 3: Factors prohibiting music listening at louder levels, and how often they were chosen as answers respectively in the risk and the non-risk group.

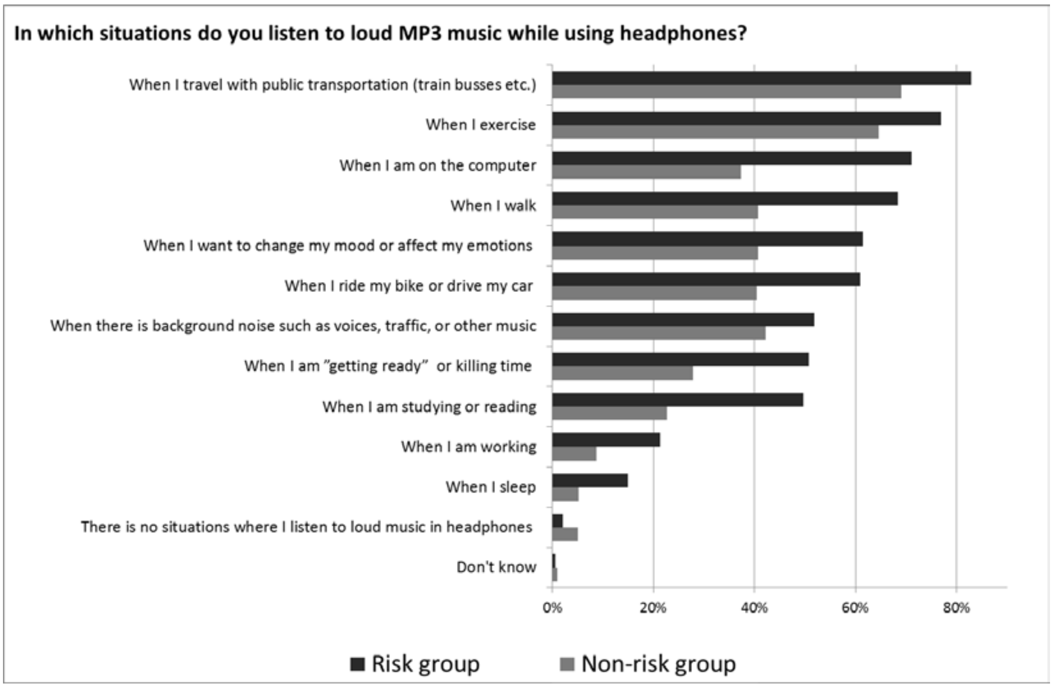


Fig 4: Situations of MP3-listening sorted after percentage of answers relatively to population.

Rewards from listening at high levels

The highest percentage (34%) in the non-risk group indicated that they enjoy listening to loud music on their MP3 player, while the majority of the risk group (62%) indicated that they *really* enjoy this. Figure 4 shows the settings for music listening. The most popular in both groups was during transportation and physical exercise. The risk-group uses the MP3 player far more in situations like 'walking', 'at the computer (71% vs. 37%)', 'studying/reading', and even 'sleeping'. An average person from the risk group indicated more reasons for listening to loud music than the non-risk group. This indicates that the risk group persons are rewarded far more for their listening than the non-risk group.

More than half of the risk group indicated that they get in better mood with loud music and that they feel a pleasant effect in the body. A large percentage compared to the non-risk group also indicated that they were able to relax better while listening to loud music. All the possible reasons and their frequency of selection can be seen in Fig. 5.

Knowledge and attitudes

More than 3/4 of the respondents were knowledgeable as to how to protect their hearing, if the music was played at a moderate level. Knowledge on protection methods were considerably smaller in the risk group compared to the non-risk group. In the non-risk group, 42% reported concern about the potential damaging effects of loud music, compared to 28 % in the risk group.

The effect of media and prophylactic information

A series of questions showed the effect of prophylactic information and the way it is

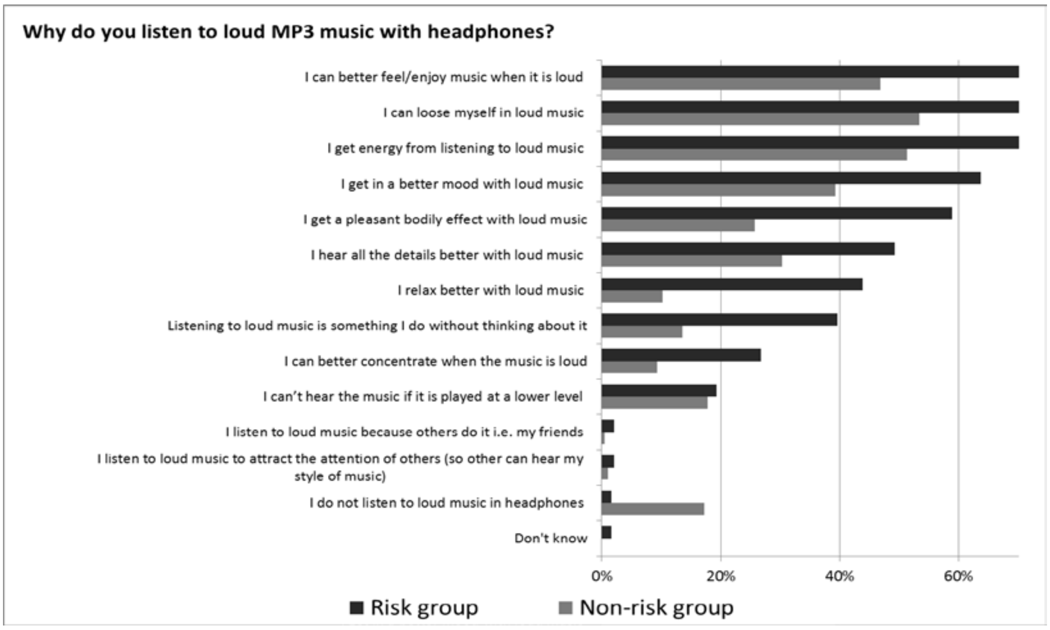


Fig. 5: Reported reasons for listening to loud music respectively for the risk and the non-risk group.

preferably received. Both groups indicated that information on potential hearing risks from MP3 listening are preferably brought to them by television and commercials or by nurses and doctors. Information received from newspapers, friends and family were reported to have an effect of less than 30% for both groups. Finally, the three most effective ways of prophylaxis, according to the respondents, were medical argumentations like doctors and nurses advice on volume levels, if they themselves experienced symptoms of hearing loss, or case stories like being exposed to examples of other young people with damaged hearing (cf. Fig. 6).

SUMMARY

Data showed that a considerably large number of teenagers have a behavior and an attitude towards loud music listening from MP3 players, which pose a threat of hearing loss later in life. Interestingly, the results clearly indicate that these teenagers are fully aware of their behavior, and actually do it for a kick, despite the risk of hearing loss. It seems that the teenagers who experience a greater emotional reward from listening to loud portable music use the players in several situations and in longer periods of time, which then increases the exposure time.

PERSPECTIVE

Most experts agree on the clear benefits of protecting the ears from music played too loud for too long. However, most young people seem not to act on the warning in due time. The results of this survey seem to support this observation, whilst at least some of the explanation of this phenomenon is that the positive effects from listening to loud music are felt to outweigh the long term negative effects for the individual. In this respect, listening to loud music might not be so different from other unhealthy lifestyle issues such as obesity, alcohol, stress, smoking, etc., and consequently campaigns directed at lowering exposure to loud music should take this factor of pleasure/addiction into account when designed.

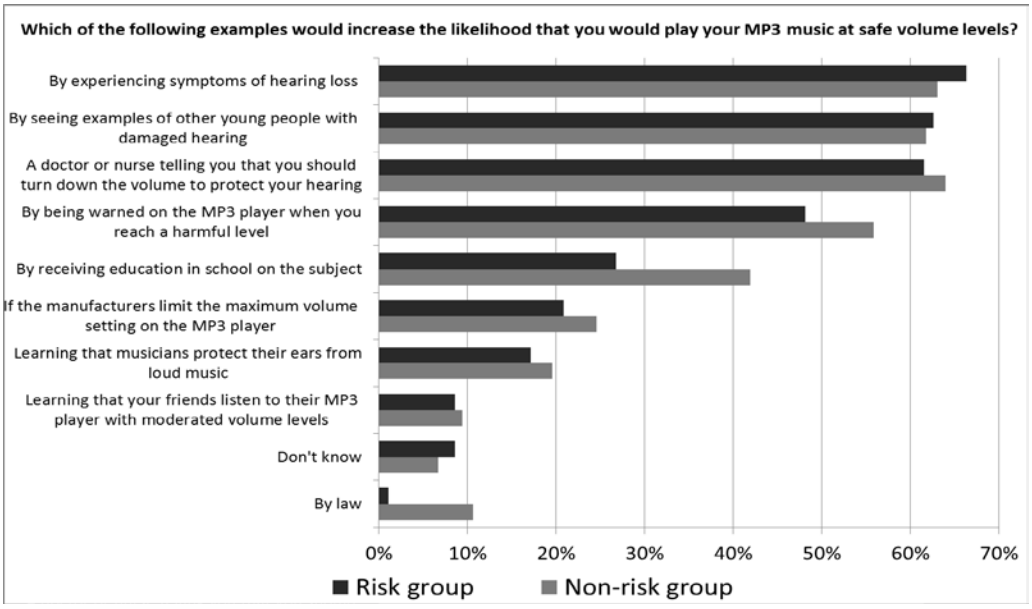


Fig. 6: The most effective ways of prophylaxis.

REFERENCES

Blesser, B. (**2007**). "The seductive (yet destructive) appeal of loud music," eContact! (http://cec.sonus.ca/econtact/9_4/blesser.html).

Blesser, B., and Salter, L.R. (**2008**). "The unexamined rewards for excessive loudness," Communications: 9th International Congress on Noise as a Public Health Problem (http://blesser.net/downloads/ICBEN%202008%20Final.pdf).

Degn, C. (**2009**). *Unges brug af MP3 afspillere*. Bachelor Thesis, University of Southern Denmark.

Florentine, M., Hunter, W., Robinson, M., Ballou, M., and Buus, S. (**1998**). "On the behavioral characteristics of loud music listening," Ear Hearing, **19**, 420-428.

Portnuff, C.D.F., Fligor, B.J., and Arehart, K.H. (**2011**). "Teenage use of portable listening devices: a hazard to hearing?" J. Am. Acad. Audiol., **22**, 663-677.

Salimpoor, V.N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R.J. (**2011**). "Anatomically distinct dopamine release during anticipation and experience of peak emotion to music," Nat. Neurosci., **14**, 257-262.

SCENHIR (**2008**). "Potential health risks of exposure to noise from personal music players and mobile phones including a music player function," Scientific committee on emerging and newly identified health risks, European Commission, Brussels (http://ec.europa.eu/consumers/safety/projects/#mp3).

Vogel, I., Brug, J., Van der Ploeg, C.P.B., and Raat, H. (**2011**). "Adolescents risky MP3-player listening and its psychosocial correlates," Health Educ. Res., **26**, 254-264.

Aspects of music with cochlear implants – Music listening habits and appreciation in Danish cochlear-implant users

BJØRN PETERSEN[1,2], MADS HANSEN[1,3], STINE DERDAU SØRENSEN[4,*],
THERESE OVESEN[5], AND PETER VUUST[1,2]

[1] *Center of Functionally Integrative Neuroscience, Aarhus University Hospital, DK-8000 Aarhus, Denmark*

[2] *Royal Academy of Music, DK-8000 Aarhus, Denmark*

[3] *Department of Psychology and Behavioural Sciences, Aarhus University, DK-8000 Aarhus, Denmark*

[4] *Department of Aesthetics and Communication, Aarhus University, DK-8000 Aarhus, Denmark*

[5] *ENT department, Aarhus University Hospital, DK-8000 Aarhus, Denmark*

Cochlear-implant users differ significantly from their normal-hearing peers when it comes to perception of music. Several studies have shown that structural features – such as rhythm, timbre, and pitch – are transmitted less accurately through an implant. However, we cannot predict personal enjoyment of music solely as a function of accuracy of perception. But can music be pleasant with a cochlear implant at all? Our aim here was to gather information of both music enjoyment and listening habits before the onset of hearing loss and post-operation from a large, representative sample of Danish recipients. A hundred and sixty three adult cochlear-implant users (101 females, 62 males) completed a survey containing questions about musical background, listening habits, and music enjoyment. The results indicate a wide range of success with music, but in general, the results show that the CI users enjoy music less post-implantation than prior to their hearing loss. Nevertheless, a large majority of CI listeners either prefer music over not hearing music at all or find music as pleasant as they recall it before their hearing loss, or more so.

BACKGROUND

A cochlear implant (CI) is a neural prosthesis that restores hearing sensation in deaf individuals. The clinical impact of the evolution of CIs has been nothing less than extraordinary, and over 250,000 individuals worldwide use the device (Peters *et al.* 2010). While the majority of adult CI users achieve good speech perception in quiet, auditory processing in general and music perception in particular are hampered. This is supported by several studies showing that discrimination of pitch, melody, timbre, and emotional prosody is significantly poorer in CI-users than in normally-hearing controls (Gfeller *et al.*, 2007; Cooper *et al.*, 2008; Petersen *et al.*, 2012).

*Corresponding author: stinederdau@gmail.com

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Nevertheless, some users seem to overcome the technical limitations of the implant and enjoy music immensely (Gfeller *et al.*, 2000). Because music is an important part of our everyday life with great emotional and social aspects, it is reasonable to evaluate the extent of music listening in CI users and identify possible factors that impacts music appreciation. With this study, we aimed to gather information about music listening habits and music appreciation before the onset of hearing loss and after receiving an implant from a large, representative sample of Danish CI users. Furthermore, we aimed to correlate this information with self-reported measures of quality of life (QOL).

PARTICIPANTS

All adult CI recipients (≥ 18) implanted at the ENT department, Aarhus University Hospital, between January 1st 2000 and December 31st 2010, were invited to take part in the study. Of the 250 patients, 163 responded (101 female; $M_{\text{age}} = 56.4$ y; *SD* = 15.7; age range: 18 to 86 y; 65% response rate). A hundred and seventeen respondents filled out the questionnaire online, while 46 requested the printed version. The implant experience ranged from 0.4 years to 11.2 years (*M* = 4.3 y, *SD* = 2.65). One hundred and thirty seven (84%) participants used an implant from Cochlear® and 26 (16%) participants used an implant from Advanced Bionics®. The demographic data of the respondents are listed in Table 1.

Respondents (M/F)	**Mean age (years)**	**Duration of profound deafness**	**Mean CI experience**
163 (62/101)	56.44 (±15.7; 18-86)	34.5 (± 18.2; 75.3-1.1)	4.3 (± 2.6; 0.4-11.2)
Unilateral users (R/L)	**Bilateral users**	**Users of hearing aid on non-implanted ear**	**Able to speak on the phone**
147 (108/39)	16	73	106

Table 1: Demographic data for the 163 respondents in the study.

METHOD

The questionnaire used in the study was a modified, Danish version of the IOWA Musical Background Questionnaire (Gfeller *et al.*, 2000). The 21 questions in the survey included multiple-choice, Likert rating scales, visual analog scales, and open-ended questions concerning musical background, listening habits, the quality of musical sound heard through the implant, and music enjoyment prior to hearing loss and after cochlear implantation. In addition, respondents were required to fill out two questionnaires concerning their quality of life (QOL) post-implantation: the

Short Form 36 (SF 36, Ware and Sherbourne, 1992) and the Glasgow Benefit Inventory (GBI, Robinson *et al.*, 1996). Here, the QOL data were used for correlational analyses.

RESULTS

Musical background

23.9% of the participants had received singing and/or instrument lessons (in primary school: *M* = 3.6 y; in high school: *M* = 1.5 y). 12.9% had been a member of a band, choir, or an orchestra. Table 2 sums up the respondents' self-assessed knowledge and experience with music. In total, 77% were involved in music to a lesser or larger extent. This is in agreement with Gfeller *et al.* (2000) and considered representative of the general population.

Category	**Percentage**
No formal training and only limited knowledge about music	23 %
No formal training or knowledge about music, but informal listening experience	56 %
Autodidact musician	3 %
Some musical training and have basic knowledge of musical terms	12 %
Several years of musical training, knowledge about music, and involvement in music groups	4 %

Table 2: Self-assessment of musical experience.

Music listening habits

The participants indicated on a four-point Likert-scale to what degree they would consider themselves as a person who often chose to listen to music (i) before the hearing loss and (ii) after receiving their implant (from 1 point = strongly disagree to 4 point = strongly agree). Furthermore, they indicated how often they chose to listen to music before their hearing loss and after getting accustomed to their implant, respectively (from 1 point = 0-2 hours per week to 4 points = 9 hours or more per week). Summed and averaged, the scores were used as mean composite scores for pre- and post-music listening habits. The mean composite score for music listening habits prior to hearing loss was 4.96 (*SD* = 1.86). The mean composite score for listening habits post-implantation was lower, at 4.23 (*SD* = 1.76). A paired *t*-test showed that the difference was significant (*t* = 3.6, *p* = 0.000).

Quality of musical sound

Figure 1 shows the mean values for the seven adjective descriptors of music through the implant. The average quality rating across all descriptors was 56.1, indicating a positive trend.

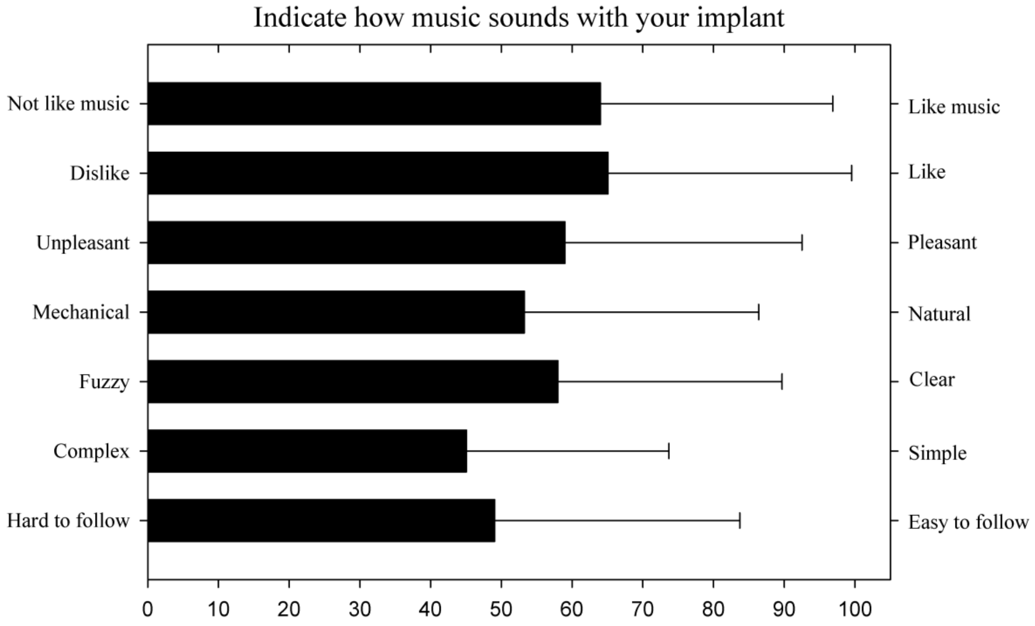


Fig. 1: Mean scores for adjective descriptors of music through the implant.

Music enjoyment

Figure 2 shows the respondents' evaluation of how their music enjoyment has changed after receiving their implant. The two rightmost categories (37%; 44%) indicate a range of music enjoyment. The left category (19%) indicates no music enjoyment.

Correlations

The ability to talk on the phone showed a weak positive correlation with both music listening habits (*r* = 0.233, *p* = 0.003), quality of musical sound (*r* = 0.361, *p* = 0.000), and enjoyment (*r* = 0.138, *p* = 0.013). Furthermore, age was negatively correlated with music listening habits (*r* = −0.264, *p* = 0.000), quality of musical sound (*r* = −0.245, *p* = 0.001), and enjoyment (*r* = −0.389, *p* = 0.000). No other demographic factors showed any significant correlation with any measures of music listening. The composite scores of the GBI questionnaire showed a significant correlation with music listening habits (*r* = 0.329, *p* = 0.000), quality of musical sound ($r = 0.408$, $p = 0.000$), and enjoyment ($r = 0.326$, $p = 0.000$). Furthermore, the social functioning subscale of the SF 36 questionnaire data showed correlations of similar strength with the three music listening measurements.

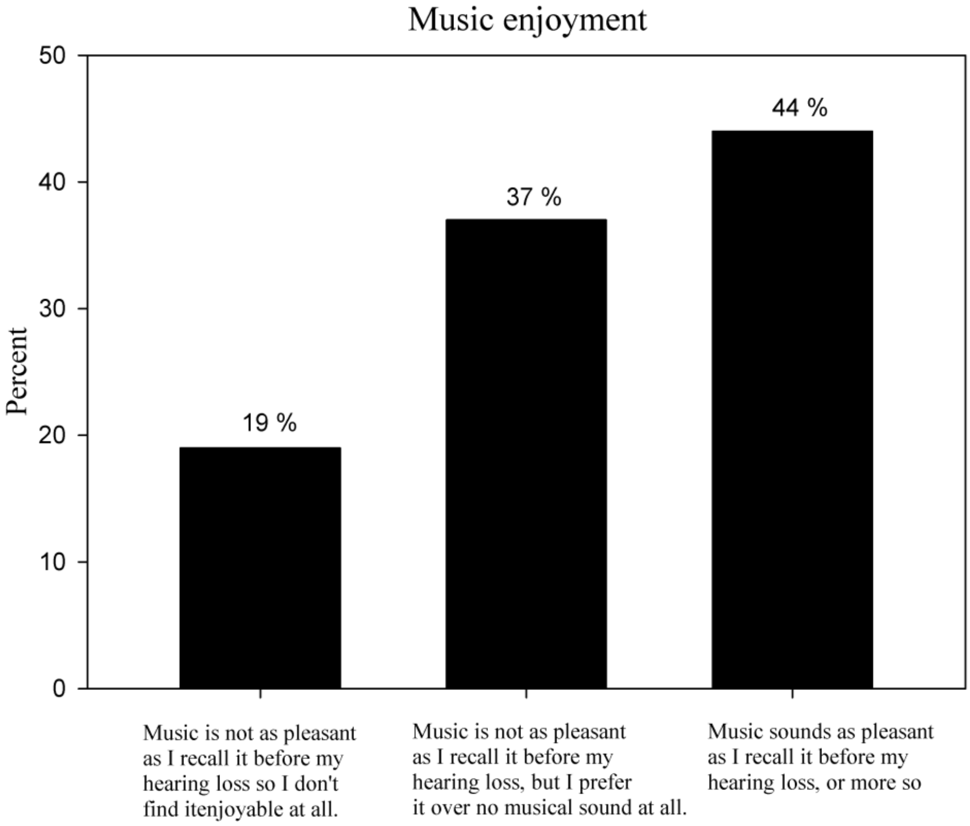


Fig. 2: Music enjoyment after implantation.

DISCUSSION

In line with findings by Gfeller *et al.* (2000), this study shows that in general adult CI users enjoy music less post-implantation than prior to hearing loss. In addition, the findings show a wide range of success with music. Interestingly, a large majority of CI listeners seem to listen to and enjoy music ranging from modest satisfaction to great enthusiasm, despite the technical disadvantages of the CI's music presentation. Furthermore, on average, the respondents describe their appreciation of different aspects of music slightly more positively than those in the Gfeller *et al.* (2000) study. This difference may suggest a benefit from the technical improvements achieved in the last decade. Interestingly, our findings indicate that solely the ability to talk on the phone is associated with success in all aspects of music listening. Previous studies found that both use of contralateral hearing aid and duration of deafness were predictive for music perception with a CI (Looi *et al.*, 2008). However, no such correlations were found in the present study. In accordance with Lassaletta *et al.* (2007) our findings suggest an association between QOL and success in music listening. Although the causes for this association may be manifold, this suggests that music exposure or training could be beneficial not only for CI users' perception of music, but also for their QOL.

ACKNOWLEDGMENTS

This study was supported by the Augustinus Foundation, Ejner Danielsens Foundation, and Danaflex A/S.

REFERENCES

Cooper, W.B., Tobey, E., and Loizou, P.C. (**2008**). "Music perception by cochlear implant and normal hearing listeners as measured by the Montreal Battery for Evaluation of Amusia," Ear Hearing, **29**, 618-626.

Gfeller, K, Christ, A., Knutson, J.F., Witt, S., Murray, K.T., and Tyler, R.S. (**2000**). "Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients," J. Am. Acad. Audiol., **11**, 390-406.

Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., and Olszewski, C. (**2007**). "Accuracy of cochlear omplant recipients on pitch perception, melody recognition, and speech reception in noise," Ear Hearing, **28**, 412-423.

Lassaletta, L., Castro, A., Bastarrica, M., Pérez-Mora, R., Madero, R., De Sarriá, J., and Gavilán, J. (**2007**). "Does music perception have an impact on quality of life following cochlear implantation?" Acta Oto-Laryngologica, **127**, 682-686.

Looi, V., McDermott, H., McKay, C., and Hickson, L. (**2008**). "Music perception of cochlear implant users compared with that of hearing aid users," Ear Hearing, **29**, 421-434.

Peters, B.R., Wyss, J., and Manrique, M. (**2010**). "Worldwide trends in bilateral cochlear implantation," Laryngoscope Suppl., **120**, 17-44.

Petersen, B., Mortensen, M.V., Hansen, M., and Vuust, P. (**2012**). "Singing in the key of life: A study on effects of musical ear training after cochlear implantation" Psychomusicology: Music, Mind and Brain, **22**, 134-151.

Robinson, K., Gatehouse, S., and Browning, G.G. (**1996**). "Measuring patient benefit from otorhinolaryngological surgery and therapy," Ann. Otol. Rhino. Laryn., **105**, 415-422.

Ware, J.E., Jr, and Sherbourne, C.D. (**1992**). "The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection," Med. Care, **30**, 473-483.

Addendum to the proceedings of ISAAR 2011:

Speech Perception and Auditory Disorders

Jont B. Allen and Woojae Han

“Sources of decoding errors of the perceptual cues, in normal and hearing impaired ears”

pp. 495-510

www.isaar.eu

The Danavox Jubilee Foundation, 2012

Sources of decoding errors of the perceptual cues, in normal and hearing impaired ears[a]

JONT B. ALLEN[1,*] AND WOOJAE HAN[1]

[1] *University of Illinois, Urbana-Champaign, IL, USA*

[2] *Hallym University, Korea*

After many decades of work it is not understood how the average normal-hearing (NH) ears, or significantly hearing-impaired (HI) ears, decode consonants. We wish to discover the strategy HI persons use to recognize consonants in a consonant-vowel (CV) context. To understand how NH ears decode consonants, we have repeated the classic consonant perception experiments of Fletcher, French and Steinberg, G.A. Miller, Furui, and others. This has given us access to the raw data (e.g., to allow for ANOVA testing) and the ability to verify many widely held (typically *wrong*) assumptions. The first lesson of this research is the *sin* of averaging: While audiology is built on average measures, most of the interesting information is lost in these averages. It has been shown, for example, that averaging across consonants is a grievous error, as is averaging across talkers for a given consonant. It will be shown how an average entropy measure (a measure of dispersion in probability) has higher utility than the average error.

INTRODUCTION

A fundamental problem in auditory science is the perceptual basis of speech, that is, phoneme decoding. How the ear decodes basic speech sounds is important for both hearing-aid and cochlear-implant signal processing, both in quiet and in noise. The object of our studies are three-fold (We are at the out-set of objective 3, objectives 1 and 2 being mostly complete):

1. We have isolated the acoustic cues in >100 consonant-vowel (CV) utterances.
2. We have measured the full-rank confusions in ≈ 50 hearing-impaired (HI) ears.
3. We are attempting to relate the measured HI confusions to the NH cues.

Objective 1): An *acoustic cue* is defined as the time-frequency features of the acoustic signal which are decoded by the auditory system for representing the consonant-vowel (CV) combination (Cole and Scott, 1974). The acoustic cues used by the average normal-hearing (ANH) ear are made up of at least four different cues (Li and Allen, 2009): a) onset bursts, b) low-frequency “edges,” c) durations, and d) F0 modulation.

[a] Page numbers starting with A refer to the ISAAR 2011 proceedings numbering.

*Corresponding author: jontalle@illinois.edu

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. C. Dalsgaard, and T. Poulsen. ISBN: 978-87-990013-3-0. The Danavox Jubilee Foundation, 2012.

The timing of the onset burst is relative to the onset of voicing of the vowel. A low-frequency *edge* is defined as the lowest frequency of the fricative region (Li and Allen, 2011).

Objective 2): We shall see that individual differences are the rule in HI confusions. No two ears are the same.

Objective 3): Our underlying hypothesis is that the consonant loss experienced by the HI ear is due to degradations in the cochlea, that cause a specific loss of detectability of specific classes (e.g., onset-burst, F0 detection) of consonant cues. Based on the HI data obtained, the most likely character of the consonant loss is *cochlear dead regions*, e.g., regions where the synapse is poorly connected to the auditory nerve (Allen *et al.*, 2009).

We hypothesize that when one or more of these cues is diminished in the ANH ear, certain consonants are confused with others in a predictable way. This hypothesis seems in agreement with our present findings, however the precise relationships are yet to be determined. It is significant that a) there are large individual differences, that appear to be b) uncorrelated to the audiograms, and that c) the HI ears are consistent in their judgments.

Questions being addressed in our publications include the following (many papers are still under review, as identified in Table 1):

1. What is the the phone error rate in NH and HI ears? (Phatak and Allen, 2007)
2. What is the source of this error (which consonants and confusions vs. SNRs)? (Singh and Allen, 2012)
3. What are the invariant acoustic cues used by NH and HI ears to identify consonants? (Li *et al.*, 2010)
4. Is audibility of an acoustic cue sufficient (it is necessary), and how may this be measured? (Li and Allen, 2011)
5. How does the HI ear differ from the NH ear in detecting invariant acoustic cues? (Han, 2011)
6. When does enhancing the SNR of a missed cue improve the robustness to noise of a consonant (Kapoor and Allen, 2012)?
7. What is the impact of NAL-R amplification on consonant perception (Phatak *et al.*, 2009; Han, 2011)?
8. How can we clinically quantify and diagnose the HI ear using speech (Han, 2011)?

Additional questions for future research include:

1. How do invariant acoustic cues depend on the following vowel?
2. Can we fit a hearing aid using consonant confusion profiles?

HISTORICAL STUDIES

The first speech studies were done in by Lord Rayleigh (1908) following the telephone's commercialization. Within a few years, Western-Electric's George Campbell (1910) developed the electrical wave filter to high and lowpass speech signals, as well as probabilistic models of speech perception such as the *confusion matrix method* of analysis. With these tools established, Harvey Fletcher (1921) extended these with related studies. He soon discovered that by breaking the speech into bands having equal scores, he could formulate a rule relating the errors in each band to the wide-band error. This method became known as the *articulation index* (AI). Even today it is not clear why the AI is well correlated to the average speech score (Singh and Allen, 2012). Today we know that Fletcher's 1921 AI formulation is similar to Claude Shannon's theory of information (1948) (Allen, 2004).

Contemporary studies

In 1970-80 a number of papers explored the role of the transitional and burst cues in consonant-vowel context. In a review of the literature, Cole and Scott (1974) argued that the burst must play at least a partial role in perception, along with transition and speech energy envelope cues. Explicitly responding to Cole and Scott (1974), Dorman *et al.* (1977) executed an extensive experiment, using natural speech consisting of nine vowels, preceded by /b,d,g/. The experimental procedure consisted of truncating the consonant burst and the devoiced transition (following the burst), of a CVC, and then splicing these onto a second VC sound, presumably having no transition component (since it had no initial consonant). Their results were presented as a complex set of interactions between the initial consonant (burst and devoiced cue) and the following vowel (i.e., coarticulations).

The same year Blumstein *et al.* (1977) published a related /b,d,g/ study, using synthetic speech, that also presented a look at the burst and a host of transition cues. They explored the possibility that the acoustic cues were *integrated* (acted as a whole). This study was looking to distinguish the *necessary* from the *sufficient* cues, and first introduced the concept of *conflicting cues*, in an attempt to pit one type (burst cues) against the other (transition cues).

While these three key publications highlighted the relative importance of the two main types of acoustic cue, burst and transition, they left unresolved their identity, or even their relative roles. In these three studies, no such masking noise was used, ruling out any form of information analysis. Masking is the classical key element basic to an information theoretic analysis of any communication channel (Fletcher, 1922; Shannon, 1948; Allen, 1994, 1996). As discussed by Allen (2005), based on the earlier work of Fletcher and Galt (1950), Miller and Nicely (1955) and inspired by

Shannon's source-channel model of communication, we repeated many of the classic experiments (Phatak and Allen, 2007; Phatak *et al.*, 2008; Li and Allen, 2009). A table summarizing the speech experiments done at UIUC between 2003-2011 is shown in Table 1.[1]

Year	Experiment	Student &Allen	Details	Publications
2004	MN04(MN64)	Phatak	MN14	Phatak and Allen (2007)
2005	MN16R	Phatak, Lovitt	MN55R	Phatak *et al.* (2008)
2005	HIMCL05	Yoon, Phatak	10 HI ears	Phatak *et al.* (2009)
2006	HINALR05	Yoon *et al.*	10 HI ears	Yoon *et al.* (2011)
2006	Verification	Regnier	/ta/	Régnier and Allen (2008)
2006	CV06-s/w	Phatak/Regnier	8C+9V SWN/WN	–
2007	CV06	Pan	CV06	–
2007	HL07	Li	Hi/Lo pass	Li and Allen (2009)
2008	TR08	Li	Furui86	ASSP
2009	3DDS	Li	plosives	Alen and Li (2009); Li *et al.* (2010); Li and Allen (2011)
2009	Verification	Kapoor/Cvengros	burst mods	Kapoor and Allen (2012)
2009	MN64 NZ-Error	Singh	PA07	Submitted JASA
2010	HI-MCL10 1,2,3	Han	46 HI ears @MCL	Submitted EH
2011	3DDS	Li	Fricatives	Submitted JASA
2011	HI-NAL11 4	Han	17 HI ears w NALR	Thesis Ch. 3

Table 1: Table of HSR experiments performed at UIUC from 2004-2011

Methods

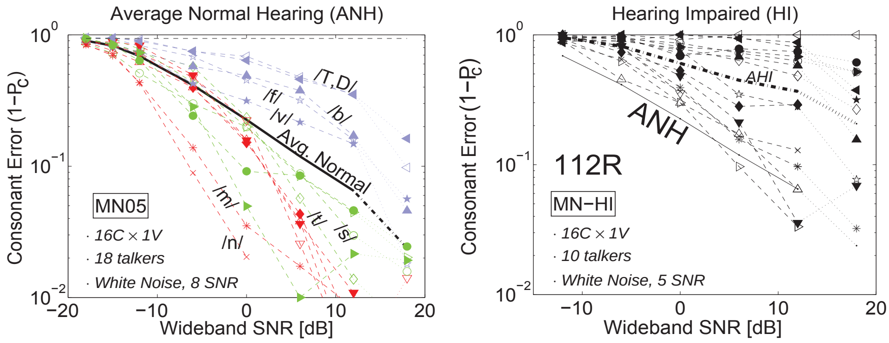
Isolated CVs were taken from real speech, with up to 20 talkers. Noise was added to the speech with a range of between 4-8 SNRs, from -26 to quiet (Q). The speech was high- and lowpass-filtered with up to 10 high/lowpass cutoff frequencies. Both white and speech-weighted additive noise was used. The listener corpus consisted of more than 200 NH subjects, 45 HI ears, up to 18 consonants and 8 vowels, and always maintaining a high source entropy (e.g., 4 bits) to eliminate guessing. To assure the estimates of the error are reliable, a minimum of 20 trials per consonant and SNR are required.

In Fig. 1 the average probability of the error $P_e(SNR)$ is shown (for speech-weighted noise the SNR *is* the articulation index). In Fig. 2 *confusion patterns* (CPs) are displayed vs. SNR.

RESULTS

From Fig.1 we see the ANH score $P_e(SNR)$ (black line), along with the score for each heard consonant h given spoken consonant s [i.e., $P_{h|s}(SNR)$], as a function of the SNR. What is most obvious is the large variation in scores: the SNR corresponding

[1] `http://hear.beckman.illinois.edu/wiki/Main/Publications`



(a) Consonant errors ANH (b) Error for HI ear 112R.

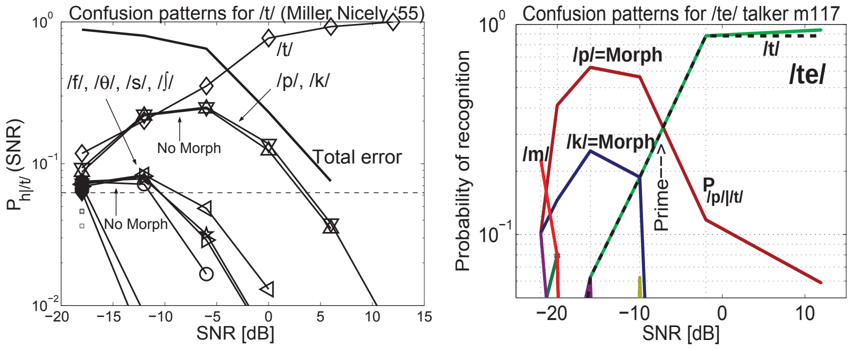
Fig. 1: Due to the large variation across consonants, the average error [e.g., $P_e(SNR) \equiv 1 - P_c(SNR)$, black line] fails to characterize speech loss. This *sin of averaging* results from (a) averaging across the natural variance across consonants (left: NH listeners), (b) across consonants for individual HI listeners (right). HI data for 112R from Phatak *et al.* (2009).

to the 50% point ranges from −12 dB [/m, n/ to +8 dB /θ, ð/ (shown as /T/ and /D/ in the figure)]. Such a large range of scores is not well captured by an average. The same is true for HI ear 112R shown in Fig. 1(b): The average score (black dashed curve) does not meaningfully represent the consonant scores. Although not shown, every consonant in our database has a wide range of scores, varying from zero error on most cases, to chance, over a wide range of SNRs (Singh and Allen, 2012).

CPs allow one to determine the precise nature of the confusions of each sound as a function of the SNR. The confusion set, and their dependence on SNR, are not predictable without running masking experiments. These confusions, and their masked dependence, are important because they reveal the mix of underlying perceptual cues. From the CP it is easy to identify a sound that *primes*, meaning that it can be heard as one of several sounds, by changing one's mental bias. In this case the confusion patterns show subject responses that are equal (the curves cross each other), similar to the CP of Fig. 2(b) at −8 dB, where one naturally primes /p/, /t/ and to a lesser extent /k/ (at −15 dB).

Identifying perceptual cues

Li *et al.* (2010) first described the 3DDS method, used to identify speech cues for a variety of real speech sounds. This method uses extensive psychophysical experimental on-CV speech-by-noise masking at a variety of SNRs, along with time-truncation and high- and lowpass filtering. These experiments made it possible, for the first time, to reliably locate the subset of perceptually relevant cues in time and frequency, while the noise-masking data characterizes the feature's masked threshold (i.e., its strength). In Fig. 3 the speech was displayed by an *AIgram* (Régnier and



(a) Average over all /t/s.

(b) Talker m117 /te/ $P_{h|/ta/}(SNR)$

Fig. 2: The *sin of averaging* extends down to the utterance level. On the left (a) we see CPs for the average score across /ta/ from Miller and Nicely (1955), while on the right (b) we see the CPs for a specific /te/. As in Fig. 1, one must conclude that averaging across utterances removes critical information from the ANH scores. As we shall see, this sin is much worse for HI ears, at the utterance level. *Priming* is reporting the sound one is thinking of, typically from a small group of sounds (Li and Allen, 2011).

Allen, 2008). The AIgram resolves acoustic features than are not easily visualized in the traditional spectrogram due to its fixed frequency resolution. First the AIgram is normalized to the noise floor. This is similar to the cochlea which dynamically adapts to the noise floor due to outer-hair-cell (OHC) nonlinear (NL) processing (Allen, 2003; Allen *et al.*, 2009). Second, unlike a fixed-bandwidth spectrogram, the AIgram uses a cochlear filter bank, with bandwidths given by Fletcher critical bands (ERBs) (Allen, 1996). Finally the intensity scale in the plot is proportional to the signal-to-noise ratio, in dB, in each critical band, as in AI-band densities $AI_k(SNR)$ for the kth band (Li *et al.*, 2010; Li and Allen, 2011). At the present time the AIgram is linear as it contains no on-frequency neural masking, nor forward and upward spread of neural masking. As a result the AIgram shows details in the speech that are not actually audible. Much work remains to be done on time-domain NL cochlear models of speech.

A summary of the audible sound cues at the threshold of masking are shown in the AIgram, as exampled in the lower-left panel for each of the six consonants in Fig. 3.

Plosives

In Fig. 3 there are six sets of 4 panels, as described in the caption. Each of the six sets corresponds to a specific consonant, labeled by a character string that defines the gender (m,f), subject ID, consonant, and SNR for the display. For example, in the upper-left 4 panels we see the analysis of /ta/ for female talker 105 (`f105ta0dB`) at 0 dB. Along the top are unvoiced plosives /t/, /k/, and /p/ while along the bottom are voiced plosives /d/, /g/, and /b/. Data from the same talker were not always available

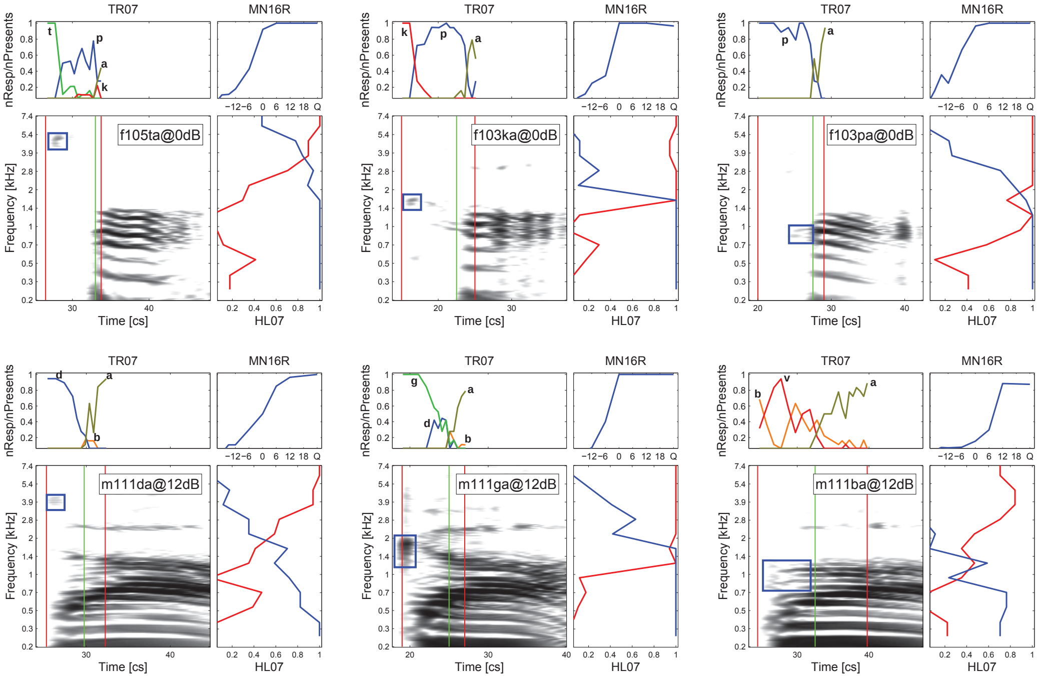


Fig. 3: Identification of cues by time, frequency, and intensity bisection using the 3-dimensional deep search (3DDS) methods, as shown here. Along the top we have unvoiced consonants /t/, /k/, and /p/, while along the bottom, the corresponding voiced consonants /d/, /g/, and /b/. Each of the six sounds consists of 4 sub-panels. For example, for /t/, upper-left, shows four panels consisting of the time-truncation confusions (upper-left), the score vs. SNR (upper-right), the AIgram (lower-left), and the score as a function of low and highpass filtering (lower-right). This last panel is rotate by 90 degrees with the score along the abscissa and the frequency along the ordinate, to line-up with the AIgram frequency axis.

in the LDC database (Fousek *et al.*, 1974), so different talkers are sometimes used for this analysis.

Three different modifications have been made to the speech: The *first* was the reported experiments (MN64, MN16R) (Phatak and Allen, 2007; Phatak *et al.*, 2008) where each CV sound was subjected to a variable signal-to-noise ratio, from −12 dB SNR to quiet, and the average score was measured by 23 NH listeners.

Next each sound was time-truncated from the onset in 10-ms steps (Exps. TR07, TR08) (Furui, 1986), and played back in random order to 14 listeners. Noise was added to the truncated sound at 12 dB SNR to remove any low-level artifacts. The results of this *truncation experiment* are presented in the top upper-left panel (labeled as TR07). Each curve is the probability $P_{h|s}(t_k)$, where h is the *heard* (reported) sound as a function of the *spoken* sound s at a truncation time t_k, labeled with the identified consonant.

Finally each CV sample was high- and lowpass filtered to a variable cutoff frequency (Li and Allen, 2009, Exps. HL05 & HL07), as indicated on the frequency axis. These HL07 data are rotated by 90 degrees so that the frequency axis lines up with that of the AIgram on the far left.

One may learn to identify perceptual cues from the 3DDS display (Li *et al.*, 2010). For example, the feature that labels the sound is indicated by the blue rectangle in the AIgram (lower-left panel) of each of the six sounds. When this burst is time-truncated (the TR07 experiment), the /t/ morphs to /p/. The term *morphs* means that one sound can be primed, i.e., is heard as several different sounds. When masking noise is added to the sound, such that it masks the boxed region, the percept of /t/ is lost. When the high- and lowpass filters remove the frequency of the /t/ burst, again the consonant is lost. Thus the three experiments are in agreement, and collectively they uniquely identify the location of the acoustic cue responsible for /t/. This generalizes to the other plosive consonants shown (i.e., voiced /k/, /p/, and unvoiced /d/, /g/, /b/), fricatives, as well as consonants followed by other vowels (not shown).

Looking at specific examples in the individual 3DDS plots is helpful. From the top-left 4 panels we see that /t/ is defined by a 4-5.4 kHz burst of energy, ≈10 cs (100 ms) before the vowel, whereas /k/ is defined as a 1.4-2 kHz burst, also ≈10 cs before the vowel. The consonant /p/ shows up as a burst of energy between 0.7-1 kHz, sticking out in front of the vowel, but connected. The three voiced sounds /d/, /g/, and /b/ have similar frequencies but onset with the vowel. The case of /b/ is not obvious, and the low score seem to reflect this weak burst. Many of the sounds in our consonant database (≈100 consonants) were analyzed using this 3DDS method, and gave similar results.

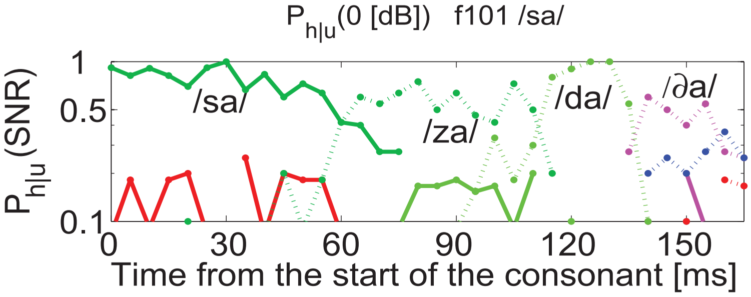


Fig. 4: Frication sound female 101 saying /sa/ (Exp. TR07). As the sound is truncated from the onset, the /s/ is heard as /z/, then /d/ and finally /ð/. Each time the conversion happens at about a factor of two in frication duration.

Fricative sounds

Not surprisingly, the perceptual cues associated with fricative sounds are quite different from the plosives. Timing and bandwidth remain important variables. For the fricative sounds, a swath of bandwidth of fixed duration and intensity is used to indicate the sound.

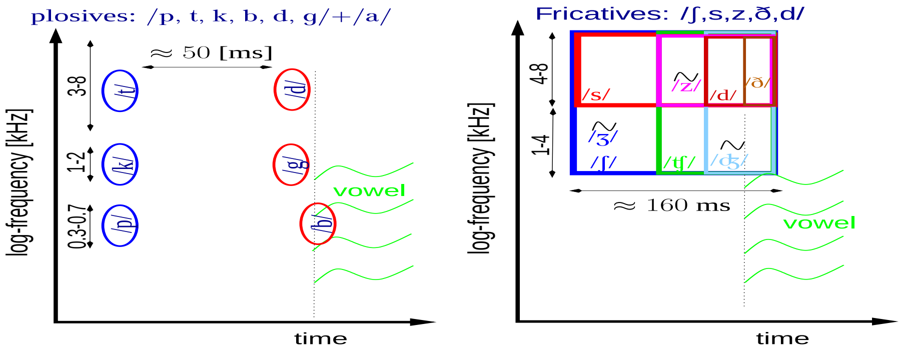


Fig. 5: Time-frequency allocation of the plosives and the fricatives. Mapping these regions into perceptual cues requires extensive perceptual experiments. Once the sounds have been evaluated, it is possible to prove how the key noise-robust perceptual cues map to acoustic features. The three consonants with the tilde over them (/z,ʒ,ʤ/), indicating that they are modulated at the pitch frequency, are voiced.

Using a time-truncation experiment similar to Furui (1986), as reported in Régnier and Allen (2008), we see the importance of duration to these consonants. In Fig. 4, a /sa/, spoken by female talker 101 and presented at 0 dB, was truncated in 10-ms steps. After about 60 ms of truncation from the onset of the sound, our pool of subjects reported /za/ instead of /sa/. After 30 additional ms of truncation, /d/ was heard. Finally at the shortest duration /ða/ was reported. A related experimental result found ʃa → ʧ→ ʤ→ d. At the end of this chain is the plosive. Thus the fricatives and the voiced-plosives seem to form a natural continuum, in the limit of very-short duration sounds.

The 3DDS results for the plosive consonants are summarized in the left half of Fig. 5, and for the fricatives in the right half of the figure. A small subset of acoustic cues define perceptual cues. Figure 5 is a modified version of the graphic by Alen and Li (2009), detailing the various *acoustic cues* for CV sounds, specifically with the vowel /a/, that were established to be perceptual cues, using a method denoted the *three-dimensional deep search* (3DDS) (Li *et al.*, 2010). Briefly summarized, the CV sounds /ta, da/ are defined by a burst at high frequencies, /ka, ga/ are defined by a similar burst in the mid frequencies, and /ba, pa/ were traced back to a wide-band burst. As noise is added, the wide-band burst frequently degenerates into a low-frequency burst, resulting in many low-level confusions. The recognition of burst-consonants depends on the delay between the burst and the sonorant onset, defined as the voice onset time

(VOT). Consonants /t, k, p/ are voiceless sounds, occurring about 50 ms before the onset of F0 voicing while /d, g/ have a VOT <20 ms. Plosive /b/ may have a negative VOT.

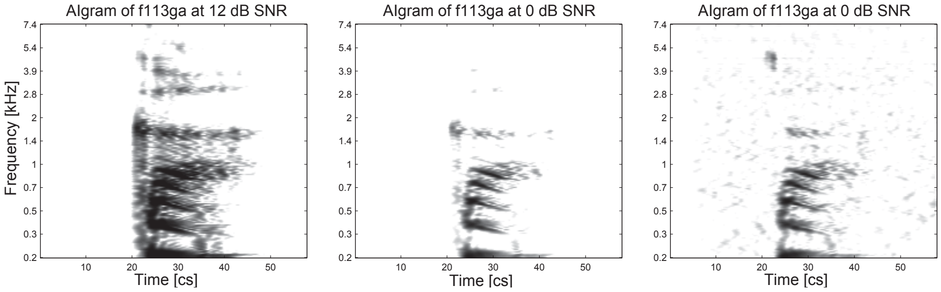


Fig. 6: On the left we see an AIgram of the original sound f113ga at 12 dB SNR, and in the middle, at 0 dB. The sound is identified 100% of the time, at and above 0 dB, 90% at −6 dB, and 30% at −12 dB. On the right is an AIgram of the sound after modification by the STFT method, where the mid-frequency burst at [20 cs, 1.5 kHz] was removed, along with remnants of the pre-vocalic burst, and 12 dB of gain was applied at 20 cs between 3.9-5.4 kHz, amplifying the low-level burst of energy, unmasked at 12 dB (left panel), as seen in the right panel. These two modifications resulted in the sound being reported as /da/.

Verification methods

To further verify all these results we have developed a method to modify the speech sounds using *short-time Fourier transform* (STFT) methods (Allen, 1977; Allen and Rabiner, 1977), to attenuate and amplify these bursts of energy. These studies have confirmed that the narrow-band bursts of energy shown in Fig. 3 are both necessary and sufficient to robustly label the plosive consonants (Li and Allen, 2011). Above the feature's masked threshold, the score is independent of SNR (Régnier and Allen, 2008; Singh and Allen, 2012).

Verification methods using STFT modifications are exampled in Fig. 6. On the left is the unmodified sound at 12 dB SNR, and in the middle again the unmodified sound at 0 dB SNR. For the right panel the /g/ perceptual cue at 1.4-2 kHz has been removed and the /d/ perceptual cue between 4-5.5 kHz has been enhanced. Following the two modifications, noise was added at 0 dB. The two modifications resulted in the morph /ga/ → /da/.

Summary

Based on such 3DDS results along with the verification experiments on the ≈100 CV in our database, we are confident that these bursts of energy label the identity of these consonants.

CONFUSIONS IN HEARING IMPAIRED EARS

As a direct extension of earlier studies (Phatak *et al.*, 2009; Yoon *et al.*, 2011), four experiments were performed (Han, 2011), two of which will be reported on here. In experiment I (Exp-I), full-rank confusion matrices for the 16 Miller-Nicely CV sounds were determined, at 6 signal-to-noise ratios (SNRs) [Q, 12, 6, 0, -6, -12], for 46 HI ears (25 subjects). In experiment II (Exp-II) a subset of 17 ears were remeasured, but with the total number of trials per SNR per consonant raised from 2-8 (Exp-I), to as high as 20 (Exp-II), to statistically verify the reliability of the subjects' responses in doing the task.

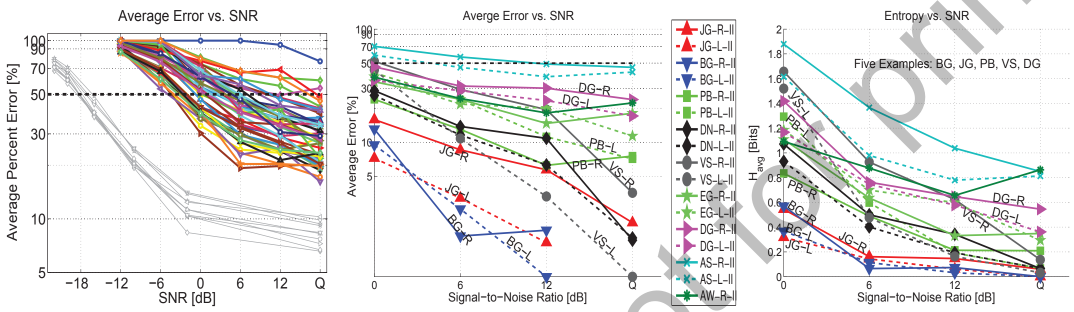


Fig. 7: Left: Average consonant error for 46 HI ears of Exp. I (Solid colored lines) and 10 NH ears (gray lines). **Middle:** Average consonant errors for the 17 HI ears of Exp. II (solid colored lines), as function of signal-to-noise ratio (SNR) using speech-weighted noise. **Right:** Average entropy for Exp. II.

The average error as a function of SNR for the 46 ears from Exp-I is shown on the left most panel of Fig. 7. The intersection of the thick horizontal dashed line at the 50% error point and the plotted average error line for each ear, marks the *consonant recognition threshold* (CRT) in dB. The data for 10 NH ears are superimposed as solid gray lines for comparison. NH ears have a similar and uniform CRT of −18 to −16 dB (a 2-dB range), while the CRT of HI ears are spread out between −5 to +28 dB (a 33-dB range). Three out of 46 ears had greater than 50% error in quiet (i.e., no definable CRT).

The data for the 17 ears (Exp. II) are mostly from the <0 dB CRT region, thus the mean error is much smaller (1% or so) compared to Exp. I, where the mean error is 15%. The minimum error for Exp. II is much lower because two high-error consonants [θ,ð Fig. 1(a)] were removed.

As discussed earlier the average score is a crude metric due to its high variance (i) across consonants, (ii) across utterances for each consonant, and (iii) across HI subjects, across both consonants and utterances. Entropy (Fig. 7, right) gives a direct measure of consistency and is insensitive to mislabeling errors (e.g., consistently across a voicing error, as in reporting /d/ given /t/). Given the observed increased mislabeling of sounds in HI ears, a high-consistency measure (i.e., entropy) seems like a better measure.

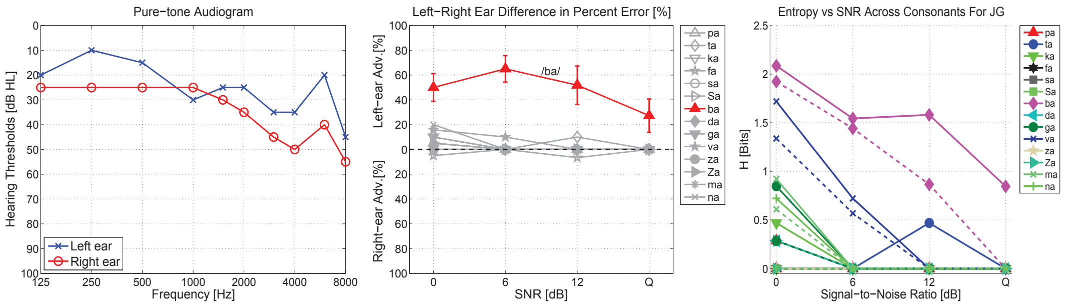


Fig. 8: Subject JG (HI36) has similar audiograms in the two ears, but a dramatic difference in the scores for /b/, of more than 50% difference between the two *consonant loss profiles* (ΔCLPs). On the right is the entropy for each consonant vs. SNR (dashed=left ear, solid=right ear).

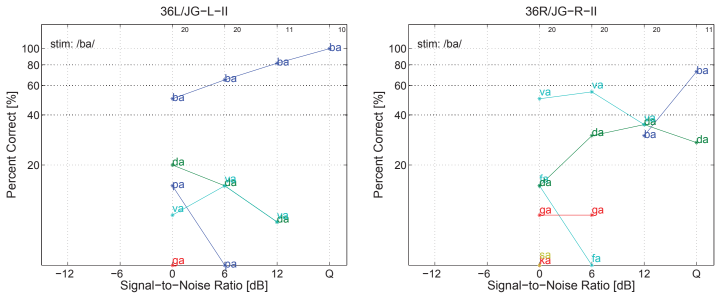


Fig. 9: Here we show the /ba/ confusion patterns $P_{h|s}(SNR)$ for subject JG (HI36). On the right we see that /b/ is confused with /v/ and /d/, even in quiet, while on the left the error is zero in quiet.

Comparison between the audiogram and confusion patterns

The observation that HI ears can exhibit large individual differences in their average consonant loss given similar *pure tone average* (PTA) (Phatak *et al.*, 2009), is further supported by the data of Fig. 8. Subject JG (HI36) has (left panel) 10-20 dB better thresholds in the left ear (blue-x) and (right panel) has a large left-ear advantage for /ba/. In the middle panel is ΔCLP(SNR), defined as the difference in consonant scores between the ears, as a function of SNR. The left-ear advantage for /ba/ peaks at 6 dB SNR at 60%. Other than /b/, subject JG heard most consonants similarly in both ears (less than 20% difference), and with no difference in /pa/, whose burst spectrum has energy in the same frequency range of .3–2 kHz with /ba/. The results for HI36 in Exp. I, collapsed over SNR, showed little difference in consonant loss between left and right ears. However in Exp. II a left-ear advantage in the /ba/ syllable was clearly indicated. This illustrates the utility of the 20 trials/condition for Exp. II, which allowed us to determine the loss as a function of SNR.

Subject HI30/DG (Fig. 10) has a 30-dB right-ear advantage for /za/, and has a distinct left-ear advantage for syllables /va, sa, fa/ and a 60% left ear advantage for /va/, at 12 dB.

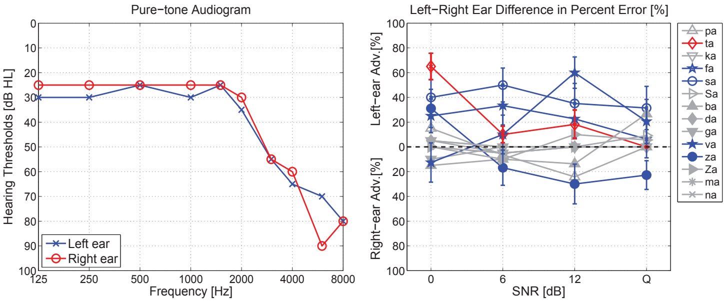


Fig. 10: Subject 30/DG L/R PTA (left) along with the difference in the confusions vs SNR (ΔCPL) on the right. While the two HLs are virtually identical, the scores are highly biased toward the left ear (left-ear advantage). These data are from Exp. II where the number of trials was up to 20 per consonant per SNR.

Summary

This article has reviewed some of what we have recently learned about speech perception of consonants, and how this knowledge might impact our understanding of NL cochlear speech processing. The application of NL OHC processing in speech is still an under-developed application area (Allen, 2008; Alen and Li, 2009) Many new ideas and methods for testing and analysis have been suggested and evaluated. The jury is out.

It is now widely accepted that outer hair cells (OHCs) provide dynamic range and are responsible for much of the NL cochlear speech signal processing, thus the common element that link all the NL data (Allen *et al.*, 2009). OHC dynamics must be understood before any model can hope to succeed in predicting basilar-membrane, hair-cell, neural tuning, and NL compression. Understanding the outer hair cell's two-way mechanical transduction is viewed as the key to solving the problem of the cochlea's dynamic range and dynamic response (Allen, 2003).

However, the perception of speech by the HI ear does not seem to be consistent with the above commonly held view. For example the large individual differences seem inconsistent with the OHC as the tying link, and seem more likely related to synaptic dead regions. Continued analysis of these confusions will hopefully provide further key insights into this important question. The detailed study of how a complex system fails can give deep insights into how the normal system works. The speech HI perception results provided here may provide further insight into normal speech perception.

The key open problem here is “How does the auditory system (e.g., the NL cochlea and the auditory cortex) processes human speech?” There are many applications of these results including speech coding, speech recognition in noise, hearing aids, cochlear implants, as well as language acquisition and reading disorders in children. If we can solve the *robust phone decoding problem,* we will fundamentally change the effectiveness of human-machine interactions. For example, the ultimate hearing aid is the hearing aid with built-in robust speech feature detection and phone recognition. While we have no idea when speech-aware hearing aids will come to be, and the time is undoubtedly many years off, when it happens it will be a technological revolution of some magnitude.

ACKNOWLEDGMENTS

We gratefully acknowledge Phonak for their generous support of this research. Much of the work was supported by an NIH grant #RDC009277A.

REFERENCES

Allen, J.B. (**1977**). “Short time spectral analysis, synthesis, and modification by discrete Fourier transform,” IEEE T. Acoust. Speech, **25**, 235-238.

Allen, J.B., and Rabiner, L.R. (**1977**). “A unified approach to short-time Fourier analysis and synthesis,” Proc. IEEE, **65**, 1558-1564.

Allen, J.B. (**1994**). “How do humans process and recognize speech?” IEEE T. Speech Audio P., **2**, 567-577.

Allen, J.B. (**1996**). “Harvey Fletcher’s role in the creation of communication acoustics,” J. Acoust. Soc. Am., **99**, 1825-1839.

Allen, J.B. (**2003**). “Amplitude compression in hearing aids,” in *MIT Encyclopedia of Communication Disorders*. Edited by R. Kent (MIT Press, MIT, Boston, MA), Chapter IV, pp. 413-423.

Allen, J.B. (**2004**). “The articulation index is a Shannon channel capacity,” in *Auditory signal processing: physiology, psychoacoustics, and models*. Edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer Verlag, New York, NY), Chapter Speech, pp. 314-320.

Allen, J.B. (**2005**). *Articulation and Intelligibility* (Morgan and Claypool, 3401 Buckskin Trail, LaPorte, CO 80535), ISBN: 1598290088.

Allen, J.B. (**2008**). “Nonlinear cochlear signal processing and masking in speech perception,” in *Springer Handbook on speech processing and speech communication*. Edited by J. Benesty and M. Sondhi (Springer, Heidelberg Germany), Chapter 3, pp. 1-36.

Allen, J.B., and Li, F. (**2009**). “Speech perception and cochlear signal processing,” IEEE Signal Proc. Mag., **26**, 73-77.

Allen, J.B., Régnier, M., Phatak, S., and Li, F. (**2009**). "Nonlinear cochlear signal processing and phoneme perception", in *Proceedings of the 10th Mechanics of Hearing Workshop*. Edited by N.P. Cooper and D.T. Kemp (World Scientific Publishing Co., Singapore), pp. 93-105.

Blumstein, S.E., Stevens, K.N., and Nigro, G.N. (**1977**). "Property detectors for bursts and transitions in speech perceptions," J. Acoust. Soc. Am., **61**, 1301-1313.

Campbell, G.A. (**1910**). "Telephonic intelligibility," Phil. Mag., **19**, 152-159.

Cole, R., and Scott, B. (**1974**). "Toward a theory of speech perception," Psychol. Rev., **81**, 348-374.

Dorman, M., Studdert-Kennedy, M., and Raphael, L. (**1977**). "Stop-consonant recognition: Release bursts and formant transitions as functionlly equivialent, contextdependent cues,", Percept. Psychophys., **22**, 109-122.

Fletcher, H. (**1921**). "An empirical theory of telephone quality," AT&T Internal Memorandum, **101**.

Fletcher, H. (**1922**). "The nature of speech and its interpretation," J. Franklin Inst., **193**, 729-747.

Fletcher, H., and Galt, R. (**1950**). "Perception of speech and its relation to telephony," J. Acoust. Soc. Am., **22**, 89-151.

Fousek, P., Svojanovsky, P., Grezl, F., and Hermansky, H. (**2004**). "New nonsense syllables database – analyses and preliminary ASR experiments," Proceedings of International Conference on Spoken-Language Processing (ICSLP).

Furui, S. (**1986**). "On the role of spectral transition for speech perception," J. Acoust. Soc. Am., **80**, 1016-1025.

Han, W. (**2011**). *Methods for robust characterization of consonant perception in hearing-impaired listeners*, Ph.D. thesis, University of Illinois at Urbana-Champaign.

Kapoor, A., and Allen, J.B. (**2012**). "Perceptual effects of plosive feature modification," J. Acoust. Soc. Am., **131**, 478-491.

Li, F., and Allen, J.B. (**2009**). "Additivity law of frequency integration for consonant identification in white noise," J. Acoust. Soc. Am., **126**, 347-353.

Li, F., Menon, A., and Allen, J.B. (**2010**). "A psychoacoustic method to find the perceptual cues of stop consonants in natural speech," J. Acoust. Soc. Am., **127**, 2599-2610.

Li, F., and Allen, J.B. (**2011**). "Manipulation of consonants in natural speech," IEEE T. Audio Speech, **19**, 496-504.

Miller, G.A., and Nicely, P.E. (**1955**). "An analysis of perceptual confsions among some english consonants," J. Acoust. Soc. Am., **27**, 338-352.

Phatak, S., and Allen, J.B. (**2007**). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am., **121**, 2312-2326.

Phatak, S., Lovitt, A., and Allen, J.B. (**2008**). "Consonant confusions in white noise," J. Acoust. Soc. Am., **124**, 1220-1233.

Phatak, S.A., Yoon, Y., Gooler, D.M., and Allen, J.B. (**2009**). "Consonant loss profiles in hearing impaired listeners," J. Acoust. Soc. Am., **126**, 2683-2694.

Rayleigh, L. (**1908**). “Acoustical notes – viii”, Philos. Mag., **16**, 235-246.

Regnier, M.S., and Allen, J.B. (**2008**). “A method to identify noise-robust perceptual features: application for consonant /t/,” J. Acoust. Soc. Am., **123**, 2801-2814.

Shannon, C.E. (**1948**). “The mathematical theory of communication,” AT&T Tech. J., **27**, 379-423 (parts I, II), 623-656 (part III).

Singh, R., and Allen, J.B. (**2012**). “The influence of stop consonants’ perceptual features on the Articulation Index model,” J. Acoust. Soc. Am., **131**, 3051-3068.

Yoon, Y., Allen, J., and Gooler, D. (**2012**). “Relationship between consonant recognition in noise and hearing threshold,” J. Speech Lang. Hear. Res., **55**, 460-473.

List of authors