Adaptive Processes in Hearing

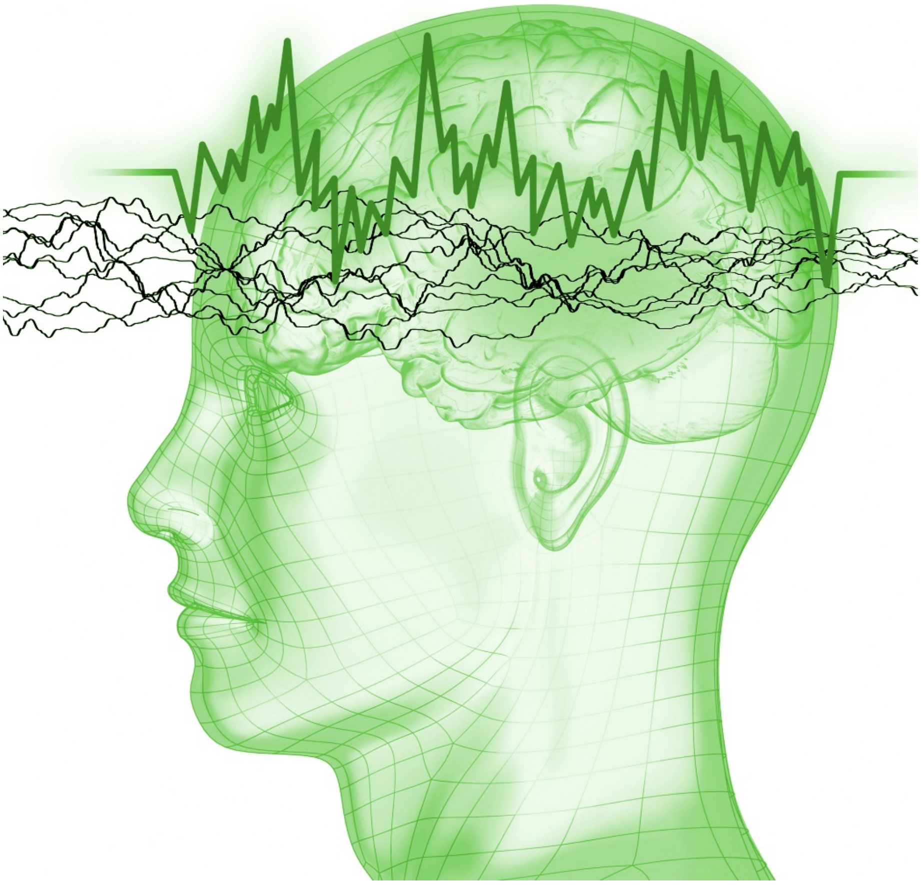


Illustration by Wet Designer Dog (www.wetdesignerdog.dk)

www.isaar.eu

ISAAR is organized by the Danavox Jubilee Foundation
and supported by

ISBN 978-87-990013-6-1 EAN 9788799001347

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

Layout: Graphic Mind
Printed by Centertryk A/S 2018

Preface

The 6th International Symposium on Audiological and Auditory Research (ISAAR) was held at Hotel Nyborg Strand in Nyborg, Denmark, from August 23 to 25, 2017. Two-hundred colleagues from all over the world participated; 30 talks and 56 posters were presented. Many of these contributions can be found as written articles in the present proceedings book.

The focus of this ISAAR was on adaptive processes in hearing and also included contributions covering a wide range of other topics within auditory and audiological research. Different perspectives were presented and discussed, including current physiological concepts, perceptual measures and models, as well as implications for new technical applications.

The goal of the symposium was to gain insight from current research in different areas and disciplines within hearing science and to relate the findings across these disciplines. The programme was comprised of the following sections: adaptive behavior in complex listening environments; neural mechanisms and modeling of adaptive auditory processes; “maladaptive” processes in hearing; electrophysiological correlates of auditory adaptation; and adaptive and learning processes with hearing devices. The various presentations reviewed current knowledge in the respective areas and shared new developments, hot topics, and future challenges. In addition to the presentation of the scientific topics, one of the major aims of ISAAR is to promote networking and dialogue between researchers from the various institutions and research centres. ISAAR enables young scientists to approach more experienced researchers and vice-versa and supports links across disciplines. At the symposium, there was a very lively discussion between the researchers spanning a large variety of academic backgrounds.

The organising committee would like to thank GN Hearing for the financial support that made this symposium possible. A special thank goes to Nikolai Bisgaard for his help and support in various matters during the planning and implementation of the symposium. Thank you also to GN Hearing for preparing and providing the symposium material. Last, but not least, the committee thanks all of the authors for their excellent presentations and all of the participants for the lively discussions.

On behalf of the organizing committee,

Torsten Dau

Organizing committee, ISAAR 2017

Scientific

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

Sébastien Santurette, Technical University of Denmark, Kgs. Lyngby
& Rigshospitalet, Copenhagen, Denmark

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

Lisbeth Tranebjærg, Rigshospitalet & University of Copenhagen, Copenhagen, Denmark

Ture Andersen, Odense University Hospital, Odense, Denmark

Torben Poulsen, Technical University of Denmark, Kgs. Lyngby, 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
& Rigshospitalet, Copenhagen, Denmark

About ISAAR

The “International Symposium on Auditory and Audiological Research” is formerly known as the “Danavox Symposium”. The 2017 edition was the 27th symposium in the series and the 6th 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 Hearing (formerly GN Danavox and later GN ReSound) and are managed by a board consisting of hearing science specialists who are entirely independent of GN Hearing. 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.

Proceedings from previous symposia are openly accessible in electronic form at the ISAAR proceedings website: http://proceedings.isaar.eu

Contents

I: Adaptive behavior in complex listening environments

II: Neural mechanisms, modeling, and physiological correlates of adaptation

III: Adaptive and learning processes with hearing devices

IV: Assessment of specific auditory functions and hearing ability

V: Speech perception: Behavioral measures and modelling

VI: Advances in hearing-instrument features and related effects

Short-term auditory learning in older and younger adults

HANIN KARAWANI[1,2], LIMOR LAVIE[1], AND KAREN BANAI[1,*]

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

[2] *Department of Hearing and Speech Sciences, University of Maryland, College Park, MD, USA*

Why speech perception in noise declines with aging remains under substantial debate. One hypothesis is that older adults adapt to perceptually-difficult listening conditions to a lesser extent than younger adults, and this, in turn, contributes to their difficulties. To test this hypothesis, we are conducting an ongoing study on the association between speech perception and perceptual learning. Here we compared the rapid learning of speech in noise between normal-hearing older and younger adults. All participants completed 40 minutes of training during which they listened to auditory passages embedded in adaptively-changing babble noise and answered content questions. To assess learning and transfer, participants were tested on the trained task and on two untrained tasks (pseudoword discrimination and sentence verification) before and after training. Both groups showed improvements over the course of the training session. Pre- to post-test improvements were observed on the trained task but not on either of the untrained ones. Consistent with the idea that poor rapid learning might limit perception in older adults, strong correlations were found between the amount of improvement during training and baseline performance of the untrained tasks.

INTRODUCTION

Aging negatively influences speech perception in noise even in individuals who maintain audiologically-normal hearing (e.g., Dubno *et al.*, 1984; Pichora-Fuller *et al.*, 1995). However, life-long experiences (e.g., playing a musical instrument) can partially offset this deleterious effect (Parbery-Clark *et al.*, 2011). Whether relatively short training protocols can yield similar effects is debated because the effects of such protocols in clinical populations are often disappointing (Henshaw and Ferguson, 2013). Improvements on perceptual tasks are variable across studies, and generalization effects, when shown, are not robust (e.g., Anderson *et al.*, 2013; Ferguson *et al.*, 2014; Karawani *et al.*, 2015). This could arise from age-related declines in perceptual learning, similar to the well documented deterioration in speech perception in noise. Because it is thought that one possible role of perceptual learning in 'real life' is to allow adaptation to challenging listening situations through rapid learning (Samuel and Kraljic, 2009), we are interested in the effects of age on this learning. The majority of previous studies

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

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

investigated the effects of multi-session training protocols. Therefore, the effects of age on rapid learning and on the relationships between rapid learning and the recognition of perceptually-difficult speech are not well understood.

Thus, the goal of this study was twofold: (1) To directly compare learning between older and younger adults and to determine whether there are age-related declines; (2) To assess the pattern of correlations between perceptual learning in one task and performance in other speech in noise (SIN) tasks. Specifically, we asked whether poor perceptual learning in one trained task is associated with poor baseline performance in two other untrained tasks.

MATERIALS AND METHODS

Twenty-two older adults (17 females) aged 60–81 years (mean age: 70 years, SD: 5) and twenty-eight younger adults (18 females) aged 20–30 years (mean age: 25 years, SD: 3) volunteered to participate in the study. All participants were native Hebrew speakers, with no history of neurological disorders and with normal hearing. Hearing was defined as normal according to the World Health Organization criteria (4-frequency pure-tone average thresholds ≤ 25 dB HL). On the first session all participants underwent a series of three SIN tests (pre-test) and then immediately completed a 40-minute training session. They were tested again on the same series of tests on the next day (post-test). All tests and training stimuli were embedded in four-talker babble noise and presented via headphones. Noise and all stimuli were normalized to 70 dB SPL. This design replicates that of Karawani *et al.* (2015), except that brief training was administered here. Given the differences in SIN performances between younger and older adults (Dubno *et al.*, 1984; Lavie *et al.*, 2014), the initial signal-to-noise (SNR) ratios of each task differed between groups (see below), such that the older group started with a more favourable SNR than the younger group.

Pre- and Post-tests included SIN tests on the trained task (A. Passages test) to assess the learning effect, and on two other untrained tasks SIN (B. Pseudoword discrimination test and C. Sentence verification test) to assess generalization. A. Participants listened to thematic passages (e.g., about energy conservation) taken from popular science articles (specific details can be found in Karawani *et al.*, 2015) and embedded in noise, and were asked to answer visually-presented multiple choice questions related to the content of the passage. Passages were 6-9 minutes long and a question was presented every 2-3 sentences. The initial SNR value of the test was +10 dB for older participants and 0 dB for younger participants. Mean SNR thresholds (in dB) were calculated for each participant. B. Pseudoword discrimination: Participants performed a same/different discrimination task in which 60 pairs of two-syllable pseudowords embedded in noise were presented aurally by a native female speaker, with equal numbers of "same" and "different" trials (e.g., "same": /damul/-/damul/, "different": /malud/-/maluk/), with equal number of pairs from each phonetic contrast and vowel template (for details see Karawani *et al.*, 2015). Discrimination thresholds (in dB) were calculated for each listener from the staircase data. C. The sentence verification test required listeners to make plausibility judgments on 60 simple sentences (e.g., "The young child climbed the high tree.") embedded in noise. After hearing a sentence, listeners had to determine whether the sentence was semantically plausible (“true”) or not (“false”). Mean SNR thresholds were calculated for each participant from the staircase data. Both pseudoword discrimination and sentence verification tests were administered with a starting SNR value of +5 dB for older adults (similar to Karawani *et al.*, 2015) and 0 dB for younger participants. SNR levels then were adapted by steps of 1.5 dB based on their responses with a 2-down/1-up adaptive staircase procedure. Across tasks, visual feedback was provided for both correct and incorrect responses. 4-talker babble noise was used for all tasks. Participants completed all tasks by making their decisions through a computer interface which recorded their responses and calculated the thresholds.

SIN training included seven blocks of training on passages embedded in 4-talker babble noise (similar to the passages task used in the pre- and post-tests, but with passages on different topics). An adaptive 2-down/1-up staircase procedure was used to adjust the level of difficulty to the performance of each listener based on their individual performance. The adaptive parameter was the SNR, where the noise level changed by 1.5 dB. Mean SNR thresholds of each block was calculated for each participant. The intensity level of the signal at the initial presentation of the first block was 10 dB greater than that of the noise (+10 dB SNR) for older participants. For younger participants the starting SNR was 0 dB. Improvement with training is reflected by a reduction in the threshold, suggesting that as training progressed listeners could maintain a good level of accuracy even with a more “difficult” (lower quality) stimulus. For each listener, the starting SNR for each block of training was based on the SNR at the end of the previous block.

RESULTS

Learning following brief training

Training effects across the seven training blocks were analysed for each group separately (Fig. 1A). To enable comparisons between groups with different starting SNRs, “normalized” scores were used. For each participant SNRs were adjusted such that block 1 values were fixed to 0. Then, for each subsequent block, SNR was presented as the difference (in dB) from block 1. To determine whether participants improved during training, linear curve estimation was performed on the group data across blocks (Fig. 1B). These analyses revealed a good fit of the linear curves to the data with significant R-squared values suggesting that a linear improvement across blocks accounts for a significant amount of the variance in performance [younger: $R^2 = 0.578$, $F(1,5) = 6.84$, $p = 0.04$; older: $R^2 = 0.934$, $F(1,5) = 70.92$, $p < 0.0001$]. To compare the amount of training-induced changes between groups, the linear slopes of the individual learning curves were calculated for each participant. Mean slopes were significantly negative in both younger and older groups. Although visual inspection of the learning curves show steeper slopes in the older than in the younger group, this was not statistically significant [older: $a = -1.54$; 95% CI: −2.23, −0.674; younger: $a = -0.79$; 95% CI: −1.26, −0.32; $t(48) = 1.56, p = 0.124$]. The younger group show some insignificant deterioration towards the end of training. We are not sure whether this deterioration might be due to lack of concentration, boredom or poor motivation of the young adults.

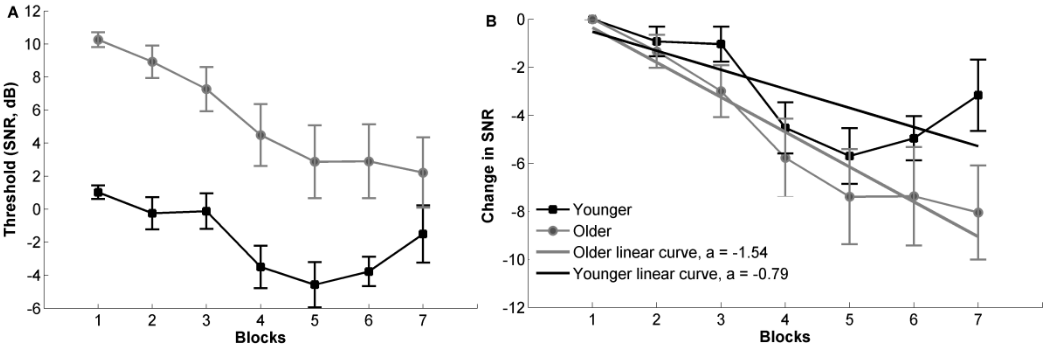


Fig. 1. A. Learning curves. Thresholds as a function of the trained block for younger (black squares) and older (grey circles) trainees are shown. **B. Adjusted learning curves.** Regression lines and slopes of the learning curves for younger (black linear lines) and older (grey linear lines) are also shown. Error bars reflect standard errors of the mean.

Pre-to-post training changes

Paired samples *t*-tests were conducted on each test to determine whether training-induced learning occurred on trained and untrained tasks (Table 1). Since the initial starting values differed between groups (see Materials and Methods), each group was analysed separately. Pre- to post-test changes (reflecting training effects) were observed only for the passages test with significant effects in both the younger and the older groups [younger: $t(27) = 4.16, p < 0.001$; older: $t(21) = 2.131, p = 0.04$]. On the other hand, no significant changes between pre- and post-sessions were shown for either the pseudoword discrimination or the sentence verification tests [pseudoword discrimination: younger: $t(27) = 0.11, p = 0.92$; older: $t(21) = 0.78, p = 0.44$; sentence verification: younger: $t(27) = 0.74, p = 0.47$; older: $t(21) = 0.68, p = 0.50$]. In order to compare the amount of change between groups in the passages tests, independent *t*-test analysis was conducted on the difference between the pre- and post-test values (calculated as the post threshold minus the pre threshold for each participant). No significant difference was observed between groups [$t(48) = 0.42, p = 0.68$; mean difference younger = −2.17, SD = 2.76; mean difference older = −1.78, SD = 3.91].

Correlation effects

The correlations (with *r* and *p* values) between the rapid learning and the three pre-test measures are shown in Fig. 2. Rapid learning over the course of training was calculated as the difference between the last and the first training blocks. The results show that older participants who improved less over the course of training also had poorer starting performance on the trained task ($r = 0.49$) and on the two untrained SIN tasks – pseudoword discrimination ($r = 0.57$) and sentence verification ($r = 0.66$). This is consistent with the idea that declines in rapid learning might limit perception. The correlations were not significant in the younger group even when re-calculated after the exclusion of the participants that improved the least during training (the rightmost data point on each panel of Fig. 2).

	Passages		Pseudoword Discrimination		Sentence Verification	
	pre	post	pre	post	pre	post
Younger	0.25 (0.47)	-0.92 (0.36)	-1.48 (0.48)	-0.55 (0.52)	-2.11 (0.52)	-2.53 (0.50)
Older	8.83 (0.68)	7.06 (0.58)	4.37 (0.98)	3.50 (0.99)	2.80 (1.27)	3.43 (1.54)

Table 1. Mean performance (with standard error of the mean, SEM) in younger and older participants, in the pre- and post-test for the Passages test, Pseudoword Discrimination and Sentence Verification tests.

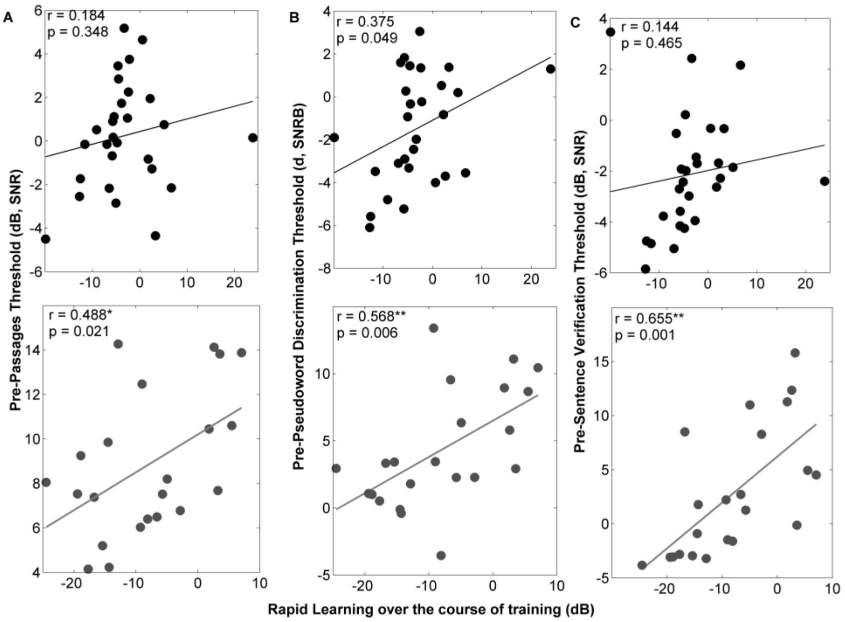


Fig. 2. Pre-test performance as a function of learning during training. A. Passage test, B. Pre-pseudoword discrimination test, and C. Sentence verification test, for younger (black dots, top row) and older (grey dots, bottom row) participants. Pearson correlation coefficient values (*r*) and *p* values are shown for each graph; * $p < 0.05$, ** $p < 0.01$.

If the amount of rapid perceptual learning explains how well an individual should do under difficult perceptual conditions, then rapid learning on the trained task should account for unique variance in the performance of the other tasks, even after we take into account the potential correlations between the different pre-test assessments of SIN. To test this idea, we used regression models to predict baseline performance on each of the untrained tasks using the amount of learning over the training session and baseline performance on the passages test as predictors. Table 2 shows that in older adults, initial performance on the passages test and rapid learning account for 34% of the variance in pre-test pseudo-words discrimination. Out of these, 30% were attributed to rapid learning. The same predictors also account for 48% of the variance in initial sentence verification, and 47% can be attributed to learning.

		R^2	R^2 change	F_{change}	df1,df2	p
Pseudoword discrimination	**Younger**	0.15	0.12	3.62	1,25	0.069
	Older	0.34	0.30	8.66	1,19	0.008
Sentence verification	**Younger**	0.02	0.02	0.46	1,25	0.502
	Older	0.48	0.47	17.19	1,19	0.001

Table 2. Regression models: Speech perception in noise predicted by rapid learning. R^2 (for full model), R^2 change (following the addition of rapid learning), F-values with degrees of freedom and p-values are presented across pre-test measures for younger adults and older adults groups.

DISCUSSION

The present study compared the effects of a short-term SIN training on speech perception between normal-hearing younger and older adults. The effects of age and the relationships between rapid learning in one task and performance in other SIN tasks were assessed. The major outcomes of the current study were: (i) Robust training-induced learning effects were found in both younger and older adults. (ii) Learning patterns were similar between younger and older adults. Although this could stem from the deterioration in performance of the younger group towards the end of training we do not think this is the case because the slopes of the learning curves appear similar even when based on fewer training blocks. Furthermore, deterioration of performance towards the end of training is not unique to the current study (see Karawani *et al.*, 2015 for another example) and is typically thought to reflect boredom or the expectation to finish training. (iii) Performance improvements were specific to the trained task with no transfer of learning to either of the untrained tasks (pseudowords and sentences in noise). (iv) Finally, the amount of improvement during training was significantly correlated to the starting performance of the untrained tasks in older adults even when the correlations between different measures of SIN were accounted for. Together, these findings suggest that rapid learning remains robust in normal-hearing older adults. Consistent with the outcomes of longer training protocols (e.g., Karawani *et al.*, 2015), generalization was limited.

Although correlation does not suggest causation, the current findings (Table 2) raise the intriguing possibility that perceptual difficulties could arise as a result of less than optimal rapid learning mechanisms. This is consistent with the view that perceptual learning serves to allow for rapid adaptation to changing acoustic circumstances (Samuel & Kraljic, 2009). Since the link between baseline SIN measures and rapid learning was robust in older adults, we suggest that the relationships between rapid learning and full training programs should be assessed because according to this idea, training will only be useful if it contributes to rapid learning in changing acoustic environments. The rich literature available on aging suggests that many behavioural and neural processes change with aging. Age-related declines have been documented in hearing, vision (e.g., Baltes and Lindenberger, 1997) and cognitive processing (e.g., Birren, 1970) such as working memory (e.g., Lyons-Warren *et al.*, 2004), attention (e.g., Kramer and Madden, 2008), executive function (Zelazo *et al.*, 2004), reasoning abilities (e.g., Salthouse, 2005), processing speed (Salthouse, 1996) and other factors. While the comprehension of the meaning of words is typically well-preserved in older age, older adults generally have difficulties understanding spoken language that is distorted (Wingfield and Grossman, 2006), especially by background noise (Schneider *et al.*, 2002). These factors are all important to new learning (Park and Reuter-Lorenz, 2009). It was shown that the ability to learn new outcome contingencies declines over the course of healthy aging (Burke and Barnes, 2006), and that explicit and implicit learning declines in the course of normal aging (Howard Jr and Howard, 2013). However, while younger and older listeners show the same amount of learning in the initial adaptation phase, older listeners' performance plateaus earlier in adapting to unfamiliar speech (Peelle and Wingfield, 2005). Older adults show less transfer of learning to similar conditions (Peelle and Wingfield, 2005) and exhibit slower consolidation of learning (Sabin *et al.*, 2013).

In conclusion, against the declines in learning described above, this study shows that when SNRs are selectively chosen to account for age-related differences in SIN perception, the rapid learning that follows short-term SIN training is still robust in older adults. It is interesting that older participants who improved less over the course of training also had poorer starting performance on the trained task as well as poorer performance on untrained SIN tasks. Future work should thus attempt to decipher the reciprocal relations between perception and learning. If good perception is pre-requisite for robust learning, training is likely to fail those listeners who need it most. On the other hand, if rapid learning contributes to the perception of perceptually-difficult speech by making individuals with better rapid learning skills more adept at adjusting to ever-changing acoustic environments, we need to consider the effects of available longer-term training programs on this rapid learning.

ACKNOWLEDGEMENTS

The authors declare no conflict of interest. This study was supported by the National Institute of Psychobiology in Israel.

REFERENCES

Anderson, S., White-Schwoch, T., Choi, H.J., and Kraus, N. (**2013**). "Training changes processing of speech cues in older adults with hearing loss," Front. Sys. Neurosci., **7**, 97. doi: 10.3389/fnsys.2013.00097

Baltes, P.B., 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. doi: 10.1037/0882-7974.12.1.12

Birren, J.E. (**1970**). "Toward an experimental psychology of aging," Am. Psychol., **25**, 124.

Burke, S.N., and Barnes, C.A. (**2006**). "Neural plasticity in the ageing brain," Nat. Rev. Neurosci., **7**, 30.

Dubno, J.R., Dirks, D.D., and Morgan, D.E. (**1984**). "Effects of age and mild hearing loss on speech recognition in noise," J. Acoust. Soc. Am., **76**, 87-96. doi: 10.1121/1.391011

Ferguson, M.A., Henshaw, H., Clark, D.P.A., and Moore, D.R. (**2014**). "Benefits of Phoneme Discrimination Training in a Randomized Controlled Trial of 50- to 74-Year-Olds With Mild Hearing Loss," Ear Hearing, **35**, e110-e121.

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. doi: 10.1371/journal.pone.0062836

Howard Jr, J.H., and Howard, D.V. (**2013**). “Aging mind and brain: is implicit learning spared in healthy aging?” Front. Psychol., **4**, 817.

Karawani, H., Bitan, T., Attias, J., and Banai, K. (**2015**). “Auditory perceptual learning in adults with and without age-related hearing loss,” Front. Psychol., **6**, 2066.

Kramer, A., and Madden, D. (**2008**). *The Handbook of Aging and Cognition*. New York, NY: Psychology Press.

Lavie, L., Banai, K., Attias, J., and Karni, A. (**2014**). “How difficult is difficult? Speech perception in noise in the elderly hearing impaired,” J. Basic Clin. Physiol. Pharmacol., **25**, 313-316. doi: 10.1515/jbcpp-2014-0025

Lyons-Warren, A., Lillie, R., and Hershey, T. (**2004**). “Short-and long-term spatial delayed response performance across the lifespan,” Dev. Neuropsychol., **26**, 661-678. doi: 10.1207/s15326942dn2603_1

Parbery-Clark, A., Strait, D.L., Anderson, S., Hittner, E., and Kraus, N. (**2011**). “Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise,” PloS One, **6**, e18082. doi: 10.1371/journal.pone.0018082

Park, D.C., and Reuter-Lorenz, P. (**2009**). “The adaptive brain: aging and neurocognitive scaffolding,” Annual Review of Psychology, **60**, 173-196. doi: 10.1146/annurev.psych.59.103006.093656

Peelle, J.E., and Wingfield, A. (**2005**). “Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech,” Journal of Exp. Psychol. Human, **31**, 1315. doi: 10.1037/0096-1523.31.6.1315

Pichora-Fuller, M.K., Schneider, B.A., and Daneman, M. (**1995**). “How young and old adults listen to and remember speech in noise,” J. Acoust. Soc. Am., **97**, 593-608. doi: 10.1121/1.412282

Sabin, A.T., Clark, C.A., Eddins, D.A., and Wright, B.A. (**2013**). “Different patterns of perceptual learning on spectral modulation detection between older hearing-impaired and younger normal-hearing adults,” JARO, **14**, 283-294. doi: 10.1007/s10162-012-0363-y

Salthouse, T.A. (**1996**). “The processing-speed theory of adult age differences in cognition,” Psychol. Rev., **103**, 403. doi: 10.1037/0033-295X.103.3.403

Salthouse, T.A. (**2005**). “Effects of aging on reasoning,” in *The Cambridge Handbook of Thinking and Reasoning*. Eds. Holyoak, K.J., and Morrison, R.G. (Cambridge University Press), pp. 589-605.

Samuel, A.G., and Kraljic, T. **(2009**). “Perceptual learning for speech,” Atten. Percept. Psycho., **71**, 1207-1218. doi: 10.3758/APP.71.6.1207

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. doi: 10.1037/h0087392

Wingfield, A., and Grossman, M. (**2006**). “Language and the aging brain: Patterns of neural compensation revealed by functional brain imaging,” J. Neurophysiol., **96**, 2830-2839.

Zelazo, P.D., Craik, F.I., and Booth, L. (**2004**). “Executive function across the life span,” Acta Psychol., **115**, 167-183. doi: 10.1016/j.actpsy.2003.12.005

"Turn an ear to hear": How hearing-impaired listeners can exploit head orientation to enhance their speech intelligibility in noisy social settings

JACQUES A. GRANGE[1,*], JOHN F. CULLING[1], BARRY BARDSLEY[1],
LAURA I. MACKINNEY[1], SARAH E. HUGHES[2], AND STEVEN S. BACKHOUSE[2]

[1] *School of Psychology, Cardiff University, Cardiff, UK*

[2] *South Wales Cochlear Implant Programme, Princess of Wales Hospital, Bridgend, UK*

Head orientation enhances the spatial release from masking. Here, with their head free, listeners attended to speech at a gradually diminishing signal-to-noise ratio (SNR) and with the noise source azimuthally separated from the speech source by 180 or 90°. Young normal-hearing listeners spontaneously turned an ear towards the speech source to improve speech intelligibility in 64% of audio-only trials, but a visible talker's face and/or cochlear implant use significantly reduced this head-turn behaviour. Instructed to explore the potential benefit of head turns, all listener groups made more head movements and followed the speech to lower SNRs. Unilateral CI users improved the most. In a virtual restaurant simulation with 9 interfering noises/voices, hearing-impaired listeners and simulated bilateral CI users typically obtained a 1-3 dB head-orientation benefit from a 30° head turn away from the talker. In this diffuse interference, the effect is due to improved target level rather than reduced noise at the better ear. Surveys of UK CI users, CI clinicians and internet-based communication advice, showed that most advice was to face the talker head on. CI users would benefit from guidelines that recommend looking sidelong to present their better hearing implanted ear towards the talker.

INTRODUCTION

Spatial release from masking (SRM) improves intelligibility through spatial separation of target speech and interfering sources. Typically, listeners face the speech head on. It was assumed by researchers and professionals that facing the speech was a more natural attitude (Bronkhorst and Plomp, 1990). However, Kock (1950) found a large benefit of orienting the head away, a benefit also predicted by the Jelfs *et al.* (2011) model of SRM. Grange and Culling (2016a) demonstrated that young normal-hearing (NH) listeners could obtain as much as 8 dB improvement in speech reception threshold (SRT) in a sound-treated room. Most of this head-orientation benefit (HOB) was obtained with a modest 30° head turn. Grange and Culling (2016b) showed that a significant HOB was also obtained by CI users alike, with a 30° turn and 180° or

[*] Corresponding author: grangeja@cardiff.ac.uk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.

ISBN: 978-87-990013-6-1.

90° source separation. Unilateral CI users obtained the same HOB (up to 4.5 dB) as age-matched NH controls. Bilateral CI users obtained less but still significant HOB (up to 2 dB). Testing listeners in audio-visual modality (AV) in addition to audio-only (A), Grange and Culling (2016b) confirmed that a 30° head turn had no detrimental impact on the listeners' lip-reading ability. Therefore, for CI users, the benefits of head orientation and lip-reading could be combined to improve SRT by up to 9 dB.

Grange and Culling (2016b) also tested whether HOB would occur in a typical noisy and reverberant setting. A realistic restaurant simulation was created using binaural room impulse responses from a B&K head-and-torso simulator in a real restaurant. The manikin was sat at 6 different tables with its head in 3 different orientations. Small loudspeakers represented a talker sat across the table and 9 concurrent interferers throughout the restaurant. SRT measurements over headphones showed that NH listeners benefited from a 30° head orientation and, on average, gained ~1.5 dB at the predicted best head orientation. Culling (2016) showed that such a benefit was not due to the acoustic shadow of the head (how head-shadow is most often understood) but instead mostly due to an amplification of the target level at the better ear.

The present study extends those of Grange and Culling (2016a,b) in two ways: First, we investigated head-orientation behaviour of CI users when the head is free, and second, we measured the HOB of hearing-impaired listeners and simulated CI users in the restaurant simulation. Experiment 1 tested (1) whether listeners spontaneously turn their heads when attending to a target talker in noise and (2) whether a simple instruction to explore their HOB leads to better performance. Grange and Culling (2016a) had already showed that in 56% of audio-only trials, listeners spontaneously turned their heads but did not adopt ideal orientations. This may in part be explained by Brimijoin *et al.*'s (2012) finding that asymmetrically impaired listeners tended to optimize target level rather that SNR at their better ear. Experiment 2 tested HOB in the simulated restaurant, for HI listeners and simulated CI users.

EXPERIMENT 1: HEAD-ORIENTATION BEHAVIOURAL TASK

Participants

The same participants as in Grange and Culling (2016b, Expt. 1) were tested according to the rules of our institutional Ethics Committee: 12 young NH listeners, 16 CI users (8 unilateral and 8 bilateral) and 10 NH listeners age-matched to the CI users.

Spatial configurations

The free-head task was run in both the collocated (T_0M_0) and the separated (T_0M_{90} and T_0M_{180}) spatial configurations, so that subtracting a separated-configuration SRT from the collocated-configuration SRT (within a presentation modality) would lead to a measure of SRM. Figure 1 illustrates the Jelfs *et al.* SRM model predictions as a function of head orientation for the separated spatial configurations.

Stimuli

Passages from the *The Wonderful Wizard of Oz* were audio-visually recorded. Each 3-4 s segment of the audio stream was normalised for RMS power. Gaps in speech exceeding 100 ms were excluded from the RMS calculation. Masking noise was filtered to match the long-term spectrum of the voice.

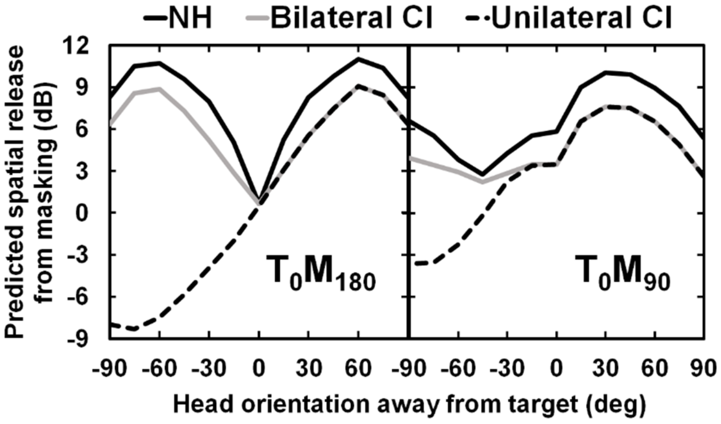


Fig. 1: Predicted spatial release from masking for all listener groups as a function of head orientation and for maskers at 180° or 90° azimuth.

Audio and AV protocol

Listeners were presented with a set of six 6-minute long clips, starting at SNRs of 6 dB for NH and 16 dB for CI users. SNRs diminished at 6 dB/min., so that the SNR would reach the listener's 50% intelligibility point in the collocated condition about two minutes into a clip and no listener could follow a clip all the way to its end. As in Grange and Culling (2016a), listeners were instructed to listen "normally as in a social situation" but "keep your back against the chair's back rest and keep your arms resting on your lap". Listeners were told they would be quizzed on the last 3-5 words they felt they correctly understood in sequence. Presentation stopped when listeners flagged that they had lost track of the thread of the clip, and they recalled the last 3-5 words. The listeners were not told where the target speech would come from, but they spontaneously faced the video monitor at the start of each trial. Next, the listener was informed that head orientation might be beneficial and repeated the free-head test, making use of the same material. The rest of the instructions remained the same as for the first test.

Results

Overhead video recordings were post-processed using MATLAB to recover head orientation. Over two passes, an operator tracked with the mouse pointer the locations of the centre of the listener's head and the then the listener's nose. The two sets of coordinates obtained were combined to extract the listener's head orientation with respect to the target direction. The recalled words were located in the clip's transcripts to estimate SRT: the SNR at which listeners lost track of the clips.

Analysis of the variance of the amounts of head movements (mean, unsigned head orientation) as a function of group, presentation modality, spatial configuration and instruction revealed significant increase of head movements after instruction [$F(1,34) = 179.2$, $p < 0.001$] and inhibition of head movements by AV presentation [$F(1,34) = 91.6$, $p < 0.001$]. AV presentation had a greater inhibiting effect on CI users than NH listeners [$F(3,34) = 221.2$, $p < 0.05$]. Instruction reduced the inhibiting effect of the AV modality [$F(1,34) = 7.1$, $p < 0.05$], particularly for CI users [$F(3,34) = 3.8$, $p < 0.05$]. Differences between spatial configurations were also removed by instruction [$F(2,68) = 3.5$, $p < 0.05$].

Where NH listeners employed head movement, most scanned for intelligibility improvements but few went straight to the predicted most beneficial head orientations. CI users made more conservative head turns and never went straight to the predicted most beneficial head orientations, perhaps because of their poorer sound localisation ability (Kerber and Seeber, 2012). Of NH listeners who scanned for improvement, most settled at sub-optimal orientations, even after passing through optimal orientations. Unilateral CI users mostly turned the correct way after instruction. However, for T_0M_{90}, more than half of age-matched NH listeners and bilateral CI users turned away from the noise, as though they had tried to get away from it, when the optimal strategy was to point their head between speech and noise directions.

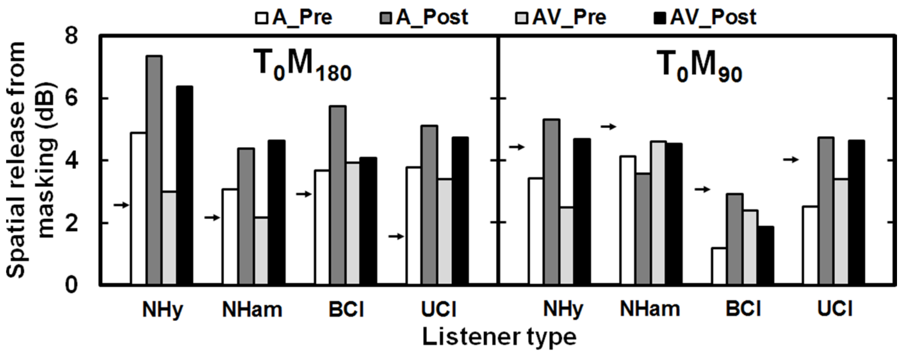


Fig. 2: SRM reached at final head orientation by each group [young NH (NH$_y$) and age-matched NH (NH$_{am}$) listeners; bilateral CI (BCI) and unilateral CI (UCI) users], in each spatial configuration for audio-only (A) and audio-visual (AV) presentation modalities, pre-instruction and post-instruction. Arrows highlight the speech-facing SRMs from Grange and Culling (2016b).

Figure 2 presents the mean pre-instruction and post-instruction SRM for each listener group and in each spatial configurations. Pre-instruction, listeners performed better at T_0M_{180} than if they had remained facing the speech. This is to be expected since for most listeners HOB could be had from any head turn away from the speech. At T_0M_{90}, however, listeners spontaneously performed worse than if they had remained still. Overall, young NH and age-matched NH listeners and CI users all improved as a result of instruction [by 1.6, 0.8 dB and 1.2 dB, $F(1,11) = 7.80$, $p < 0.02$; $F(1,9) = 11.05$, $p < 0.01$ and $F(1,15) = 5.27$, $p < 0.05$, respectively]. Significant correlations between subjective SRMs and SRM predictions at final head orientations were found for each listener type [$r = 0.49$, $t(46) = 3.86$, $p < 0.001$; $r = 0.35$, $t(38) = 2.30$, $p < 0.03$; $r = 0.36$, $\mathrm{t}(60) = 2.96$, $p < 0.005$, respectively], indicating that improved head orientations led to SRM improvement. Despite an overall positive effect of instruction, age-matched NH listeners and bilateral CI users did not improve post-instruction at T_0M_{90}.

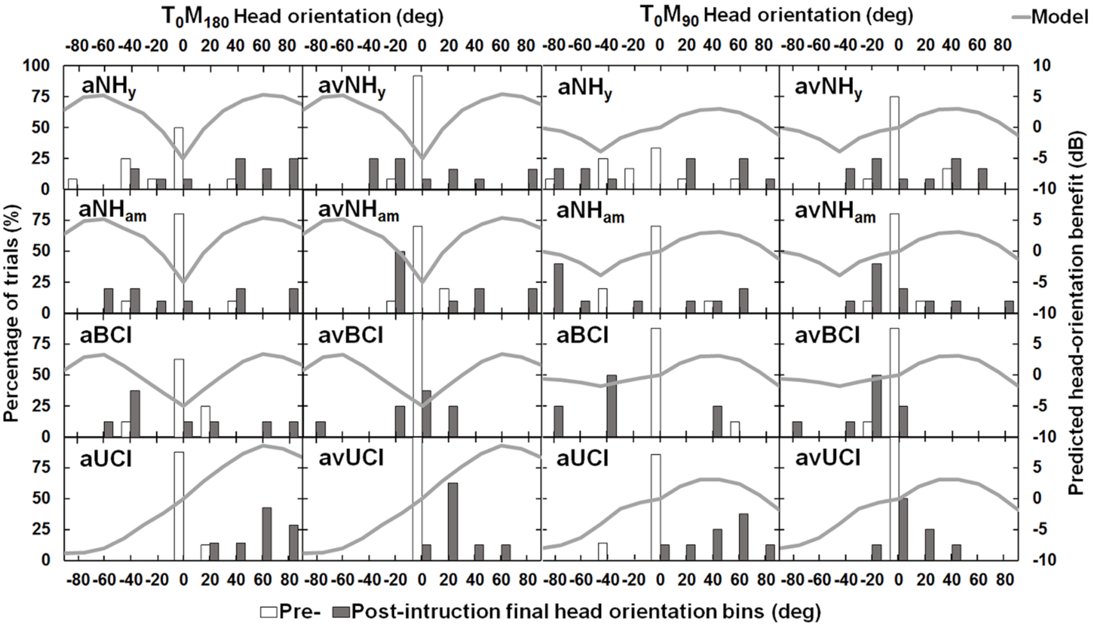


Fig. 3: Histograms of the final head orientations of young NH (NH_y), age-matched NH (NH_{am}), bilateral CI (BCI) and unilateral CI (UCI) listeners for audio-only (a) and audio-visual (av), pre (white bars) and post (dark grey bars) instruction. Predicted SRMs are light grey lines.

Figure 3 shows histograms of final head orientations for each listener group at T_0M_{180} and T_0M_{90}. Model predictions are superimposed to help the reader judge how well listeners discovered optimal HOB pre- and post-instruction. The inhibition of head movements by the presence of visual cues is demonstrated by the tight distribution of final-head-orientations around the speech-facing orientation in the AV modality. At T_0M_{180}, NH and bilateral CI users can get a benefit of turning either way. Unilateral CI users, however, need to turn to present their implanted ear, and it is clear that they all turned the correct way. At T_0M_{90}, all listeners should experience a benefit of pointing their head between speech and noise sources. In only 1 of 16 post-instruction trials, did a unilateral CI user turn the wrong way. In contrast, bilateral listeners settled at detrimental orientations in 12 out of 16 trials, age-matched NH listeners in 11 of 20 trials and young NH listeners in 9 of 24 trials.

EXPERIMENT 2: SIMULATIONS IN A VIRTUAL RESTAURANT

The materials from Grange and Culling's (2016b) second experiment were employed. Listeners and target talkers were sat across the table from each other at each of 6 tables in a virtual restaurant. Interferers came from another 9 tables spread across the restaurant. Interferers were either steady speech-shaped noise or continuous voices. The combination of the 9 interferers produced a spatialized babble or diffuse noise.

Participants

16 young NH adult listeners (mean 21 y.o.) and 14 unaided HI listeners (mean 68 y.o.) with moderate to severe high-frequency loss (40-85 dB HL in at least one ear and increasing from 4 kHz) participated, in accordance with our Ethics Committee rules.

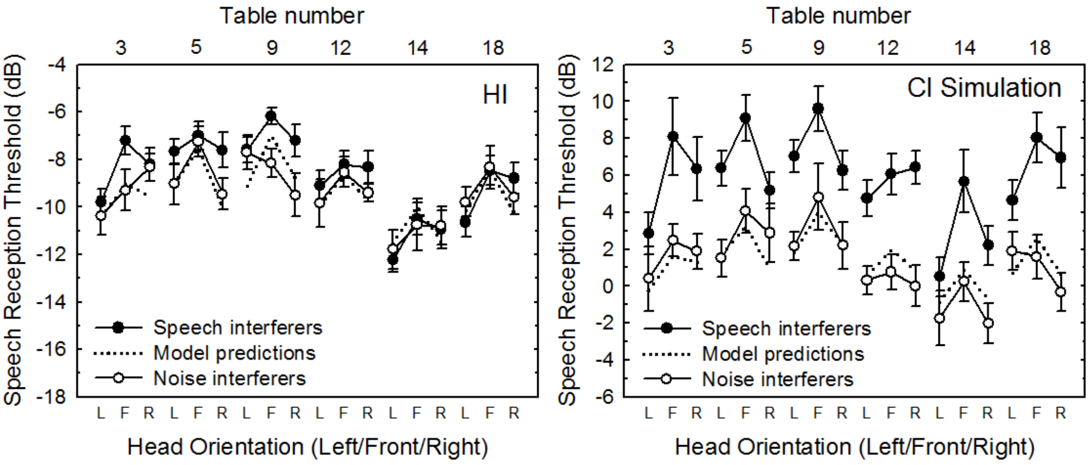


Fig. 4: SRTs obtained for HI listeners (left panel) and simulated CI users in the virtual restaurant, as a function of head orientation (30° to the Left or Right, or Facing the target talker), interferer type (open circles for speech-shaped noise, closed circles for babble) and table. Jelfs *et al.* model predictions (dotted lines), with their mean equalised to mean SRTs in noise, include binaural unmasking for HI listeners, but not for simulated CI users.

Simulation of CI users

The mixed target-interferers signal was passed through SPIRAL, a tone vocoder that incorporates the threshold-elevating effect of current spread (set at 8 dB/oct.) using 80 carrier tones. For details of the vocoder, see Grange *et al.* (2017). NH participants listened to the combined restaurant and CI simulation.

Results with HI listeners and simulated CI users

The left and right panels of Fig. 4 plot SRTs as a function of table number, head orientation (left, front, right) for each interferer type for HI listeners and simulated CI users, respectively. SRTs were ~12 dB higher for simulated CI than HI listeners. The benefit of orienting the head was significant for each listener group [HI, $F(2,26) = 17.4, p < 0.001$; CI, $F(2,36) = 15.1, p < 0.001$]. At the best predicted head orientation and in noise, the magnitude of HOB was 1.2 and 1.7 dB for HI and CI users, respectively. SRTs were significantly higher for speech than noise interferers (by 1 and 5 dB for HI and CI users, respectively). Simulated CI users alone benefitted more from head turns in babble, with 3.2 dB HOB, than in noise.

DISCUSSION

Experiment 1 found that when speech was hard to follow CI users made significantly less spontaneous head turns than NH listeners, particularly with AV present. However, with a simple instruction to explore their HOB, all listener groups could follow the clips to significantly lower SNRs. Our findings suggest that simple training of HI listeners to exploit their HOB could improve their speech understanding in noisy environments. Listeners did not exhibit a clear set of head-orientation behaviours that could be categorised. A study with a much larger sample size would help establish whether behavioural categorisation is justifiable. Such a study could help tailor the design of training programs to each specific listener's needs.

Experiment 2 tested the robustness of HOB with reverberation and multiple interferers. Regardless of the table position within the restaurant or of the interferer type, a HOB of 1.2-3.2 dB could be obtained. Comparing results to Expt. 2 of Grange and Culling (2016b), SRTs are elevated in HI and simulated CI users. In addition, HI and simulated CI listeners exhibited even higher SRTs when immersed in a spatialized babble than in speech-shaped noise. Qin and Oxenham (2003) concluded from their CI simulations that in order to segregate a target voice from background interferers, both F0 segregation and good frequency resolution were required. While our HI listeners may still, via low frequencies, be able to exploit F0 differences, their limited frequency resolution may explain their higher SRTs with interfering voices. For our simulated CI users, not only is their frequency resolution significantly reduced by CI simulation, but they have no access to the F0 cue. This may explain their greater SRT elevation with voiced interferers. What remains unclear is why they appear to benefit more from head orientation (3.2 dB HOB) than their NH or HI counterparts in a spatialized babble. The data suggest that the modulation or informational masking by babble changes faster with head orientation than energetic masking by noise.

To compare our recommendations with current practice, we surveyed 95 CI users and 31 CI clinicians regarding advice given to CI users about head orientation. 89% of clinicians reported frequently or always advising CI users to directly face the talker; 77% of CI users reported the same. A further survey of communication advice available on the internet found 23 "public information" websites. The majority recommended strategies to reduce background noise and stated, in all but one case, that the listener should face the speaker. The clinicians' responses revealed the two assumptions that mostly influenced their advice: (1) that facing the talker leads to better lip-reading and (2) that one must face the talker to benefit from microphone directionality. However, Grange and Culling (2016b, Fig.7) demonstrated that lip-reading was unaffected by head orientation of 30° away from the talker and that maximum sensitivity of a directional microphone on a BTE hearing prosthesis is in fact shifted to 30° to 50° azimuth by the acoustic diffraction of the head.

Overall, these experiments demonstrate that: (1) HOB in a realistic social setting is robust with moderate-to-severe high-frequency hearing loss or for a simulated CI listener; (2) simulated CI users benefit from HOB as much as NH or mild-to-moderate HI individuals, and (3) CI users and other HI listeners would benefit from the development of listening-strategy training that involves appropriate head orientation.

This research was supported by Action on Hearing Loss and the Oticon Foundation.

REFERENCES

Brimijoin, W., Mcshefferty, D., and Akeroyd, M. (**2012**). "Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task," Hear. Res., **283**, 162-168. doi: 10.1016/j.heares.2011.10.009

Bronkhorst, A., and Plomp, R. (**1990**). "A clinical test for the assessment of binaural speech perception in noise," Int. J. Audiol., 29, 275-285. doi: 10.3109/00206099009072858

Culling, J.F. (**2016**). "Speech intelligibility in virtual restaurants," J. Acoust. Soc. Am., **140**, 2418-2426. doi: 10.1121/1.4964401

Grange, J.A., and Culling, J.F. (**2016a**). "The benefit of head orientation to speech intelligibility in noise," J. Acoust. Soc. Am., **139**, 703-712. doi: 10.1121/1.4941655

Grange, J.A., and Culling, J.F. (**2016b**). "Head orientation benefit to speech intelligibility in noise for cochlear implant users and in realistic listening conditions," J. Acoust. Soc. Am., **140**, 4061-4072. doi: 10.1121/1.4968515

Grange, J.A., Culling, J.F., Harris, N.S., and Bergfeld, S. (**2017**). "Cochlear implant simulator with independent representation of the full spiral ganglion," J. Acoust. Soc. Am., **142**, EL484-EL489.

Jelfs, S., Culling, J., and Lavandier, M. (**2011**). "Revision and validation of a binaural model for speech intelligibility in noise," Hear. Res., **275**, 96-104. doi: 10.1016/j.heares.2010.12.005

Kerber, S., and Seeber, B.U. (**2012**). "Sound localization in noise by normal-hearing listeners and cochlear implant users," Ear Hearing, **33**, 445-457. doi: 10.1097/aud.0b013e318257607b

Kock, W. (**1950**). "Binaural localization and masking," J. Acoust. Soc. Am., **22**, 801-804. doi: 10.1121/1.1906692

Qin, M., and Oxenham, A. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am., **114**, 446-454. doi: 10.1121/1.1579009

The role of temporal cues on voluntary stream segregation in cochlear implant users

ANDREU PAREDES GALLARDO*, SARA MIAY KIM MADSEN, TORSTEN DAU, AND JEREMY MAROZEAU

Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

Cochlear implant (CI) listeners experience difficulties in complex listening scenarios, where the auditory system is required to segregate a target signal from the competing sound sources. The present study investigated segregation abilities of CI listeners as a function of temporal cues and examined whether a two-stream percept occurs instantaneously or needs time to build up. CI users participated in a detection task where a sequence of regularly presented bursts of pulses ("B") on a single electrode interleaved with an irregular sequence ("A") presented on the same electrode with a different pulse rate. The pulse rate difference and the duration of the sequences were varied between trials. In half of the trials, a delay was added to the last burst of the regular A sequence and the listeners were asked to detect this delay. As the period between consecutive B bursts was jittered, time judgments between the A and B sequences did not provide a reliable cue to perform the task such that the segregation of A and B should improve performance. The results showed that performance improved with increasing rate differences and increasing sequence duration, suggesting that CI listeners can segregate sounds based on temporal cues and that this percept builds up over time.

INTRODUCTION

The cochlear implant (CI) is probably among the most successful neural prosthesis (Zeng *et al.*, 2008), making it possible for severely hearing-impaired listeners to achieve relatively high levels of speech intelligibility in quiet environments. However, CI listeners experience difficulties when listening in complex listening situations, where the auditory system is required to segregate the target signal from other competing sounds. To understand the role of different factors and cues on the segregation process an "auditory streaming" paradigm has been proposed (e.g., Bregman, 1990; Carlyon, 2004; Moore and Gockel, 2012; Van Noorden, 1975). In this paradigm, two perceptually different sounds (A and B) are presented sequentially to the listener who might fuse them into a single stream or segregate them into two separate streams, depending on the difference between the sounds. In normal-hearing (NH) listeners, large perceptual differences between the sounds facilitate segregation, whereas small differences promote fusion (integration).

*Corresponding author: apaga@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

The duration of the sequence of A and B sounds is another important factor of the auditory streaming paradigm, since the probability of achieving a segregated percept has been reported to increase over time (Anstis and Saida, 1985; Bregman, 1990; Moore and Gockel, 2012). This phenomenon is often referred as the build-up effect.

In electric hearing, perceptual differences can be elicited by varying the place or the rate of stimulation (e.g., Landsberger *et al.*, 2016). Most of the previous studies in CI listeners assessed the role of place cues on stream segregation (Böckmann-Barthel *et al.*, 2014; Chatterjee *et al.*, 2006; Cooper and Roberts, 2009, 2007; Hong and Turner, 2006; Tejani *et al.*, 2017) and little attention has been given to the role of rate or temporal periodicity cues.

Chatterjee *et al.* (2006), Hong and Turner (2009) and Duran *et al.* (2012) assessed the effect of temporal periodicity cues on stream segregation in CI listeners. The results from these studies suggest that larger differences in the temporal envelope or pulse rate between the A and the B sounds facilitate stream segregation. Chatterjee *et al.* (2006) also reported an effect of sequence duration on stream segregation. However, only one listener participated in this preliminary experiment. These studies have presented some evidence that CI listeners can use temporal periodicity cues to segregate sounds. However, there is limited evidence about the effect of the sequence duration on stream segregation (i.e., the build-up effect) in CI listeners, a well-documented phenomenon in NH listeners.

The present study investigated the role of temporal periodicity on stream segregation in CI listeners. Streaming abilities were assessed with a temporal detection task that does not rely on direct reports of perception from the listeners. Temporal periodicity cues were induced by changing the pulse rate at a fixed cochlear location. This was done to evaluate whether CI listeners can use this cue to segregate the streams, as proposed in previous studies. The effect of sequence duration was also investigated to determine whether a two-stream percept builds up over time.

METHODS

Listeners

Seven bilateral CI listeners (one male) participated in this experiment. None of the listeners had residual hearing. All listeners provided informed consent prior to the study and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

Stimuli and conditions

The stimulation paradigm is illustrated in Fig. 1. A sequence of bursts of pulses ("B") presented on a single electrode was interleaved with a sequence ("A") presented on the same electrode with a different pulse rate. The onset-to-onset interval in the B sequence was 340 ms, and a random jitter of ± 220 ms was added to the onset-to-onset in the A sequence. Consecutive A and B sounds were always separated by a minimum interval of 10 ms. In half of the trials, a small delay (Δt) was added to the last burst of the B sequence, which the listeners were asked to detect. To optimize performance, the listener needed to compare the time interval between the last two B-tones to those between previous B-tones, since time judgments between successive A and B tones was an unreliable cue for performing the task. Thus, the task became easier if the A and B sequences were segregated (Micheyl and Oxenham, 2010; Nie *et al.*, 2014; Nie and Nelson, 2015).

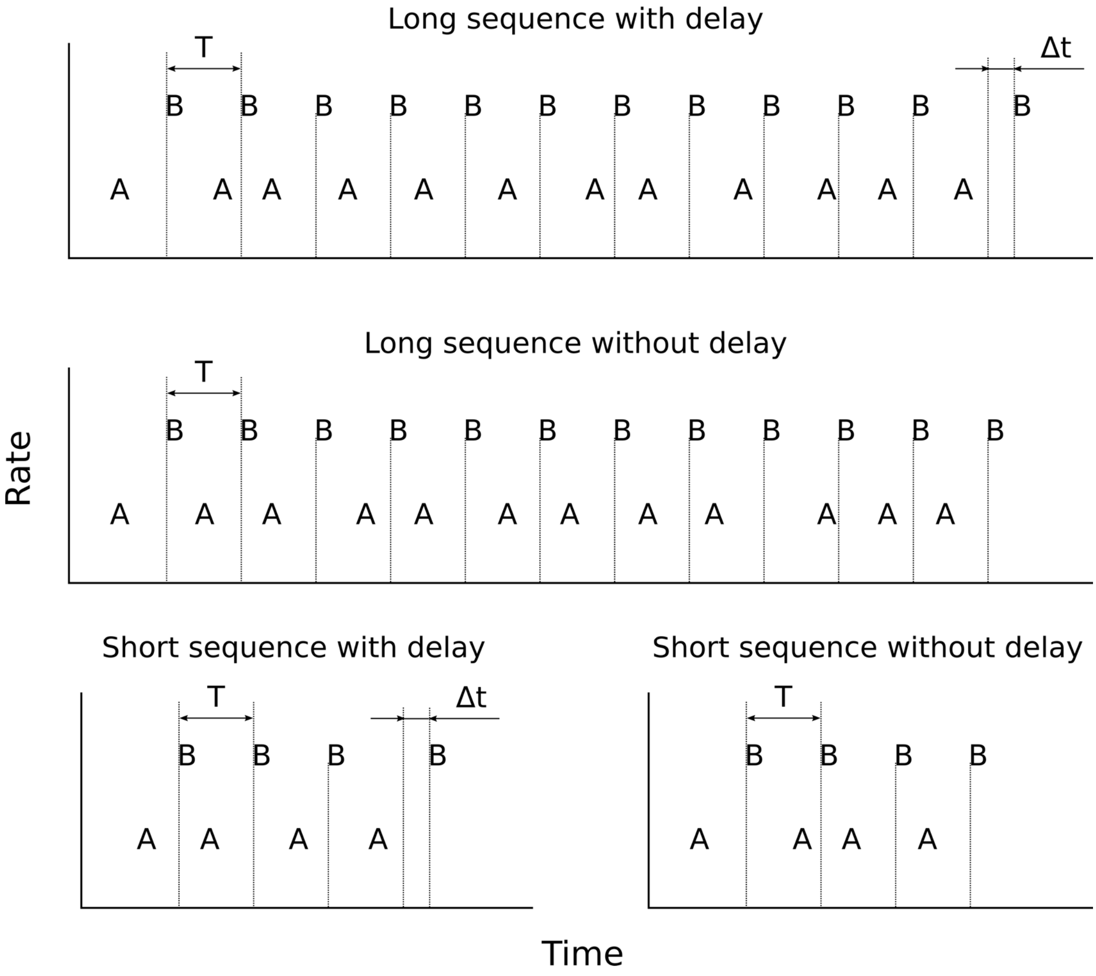


Fig. 1: Graphical representation of the experimental paradigm. The onset-to-onset interval is represented by T and the delay of the last B sound by Δt. The Δrate between A and B sounds varied across conditions.

The B sequence was played with a constant rate of 300 pps, while the A sequence was played with a pulse rate of either 80 or 260 pps, leading to a pulse rate difference (Δrate) between the streams of either 220 or 40 pps. Two sequence durations were tested. The long sequence consisted of 12 AB duplets and the short sequence of 4 AB duplets. All sequences started with the A sequence.

Each A and B sound consisted of a 50-ms biphasic pulse burst presented through electrode 11 with the corresponding rate in monopolar mode. Each biphasic pulse had a phase width of 25 µs and phase gap of 8 µs. The stimuli were presented through the Nucleus Implant Communicator research interface (NIC v2, Cochlear Limited, Sydney).

Temporal detection performance for the long and short sequences was also measured with the B sequence alone. These conditions were easier than the test conditions and, thus, a different (shorter) Δt was used to avoid ceiling effects.

For each combination of rate difference and sequence duration, 60 presentations of the delayed sequence and 60 presentations of the non-delayed sequence were used to calculate the listener's sensitivity (*d'*) to the delayed target.

Loudness balancing

Categorical loudness scaling was performed for each pulse rate in order to find the most comfortable level (MCL) for each listener. Thereafter, all stimuli were loudness matched to the 300 pps stimulus by the listener using a simple user interface (±0.15 dB).

Δt adjustment procedure

Δt were chosen such that all listeners would be equally sensitive to the delayed target in a given condition. The individual delay adjustment procedure was part of a prior study where listeners performed the temporal detection task with sequences consisting of 12 duplets of AB sounds. In that study, A and B were 50-ms bursts of pulses presented at 900 pps to electrodes 19 and 11, respectively. The sensitivity to the delayed target was measured for four different delays based on 60 presentations of each delayed sequence and 60 presentations of the non-delayed sequences. Psychometric functions were fitted to the data of each listener and the individual Δt was defined as the delay leading to *d'* = 2. Individual Δt were always smaller than the jitter applied to each A sound and ranged from 35 to 80 ms.

The same delay adjustment procedure was used to find the individual Δt to be used in the control conditions. In this case, the long sequence without distractor stream was used to fit the psychometric function. The delay leading to *d'* = 3 was chosen as Δt for the control condition. This *d'* value was chosen to keep the control conditions relatively easy while avoiding ceiling effects.

Procedure

A one-interval, two-alternative, forced-choice procedure was used, where the listeners were asked to report whether a given sequence contained a delayed target or not. A total of eight different sequences were presented to the listeners, resulting from the combination of two possible A-sequence pulse rates, two sequence durations and two different Δt (delayed or non-delayed). Short and long sequences were presented in different blocks. In each block, each of the four possible sequences was repeated 12 times in pseudo-random order, ensuring that the distractor electrode alternated from one sequence to the next one. Thus, the first sound of each sequence alternated between a pulse rate of 80 and 260 pps, contributing to reset the build-up of a two-stream percept after each presentation (Roberts *et al.*, 2008). Each block was repeated five times in a random order.

The control conditions were tested in four blocks (two with long sequences and two with short sequences) containing 30 repetitions of the delayed and 30 repetitions of the non-delayed sequences.

Statistical analysis

A mixed-effects linear model was fitted to the computed *d'* scores with the experimental factors as fixed effects terms and the listener-related effects as random effects. The *p*-values for the fixed effects were calculated from *F*-tests based on Sattethwaite's approximation of denominator degrees of freedom and the *p*-values for the random effects were calculated based on likelihood ratio tests (Kuznetsova *et al.*, 2015). Post-hoc analysis was performed through contrasts of least-square means. *p*-values were corrected for multiple comparisons using the Tukey method.

RESULTS AND DISCUSSION

Figure 2 shows the *d'* scores for all combinations of sequence duration and Δrate. Results from the post-hoc analysis are shown with asterisks. Both sequence duration [$F(1,18) = 27.902$, $p < 0.001$], Δrate [$F(2,18) = 13.523$, $p < 0.001$] and their interaction [$F(2,18) = 4.804$, $p < 0.021$] were found to be significant factors in the statistical model.

For the long sequence, greater *d'* scores were achieved for a Δrate of 220 pps than for a Δrate of 40 pps [$t(26.41) = 4.363$, $p = 0.002$, difference estimate = 1.436], implying that CI listeners benefitted from the larger Δrate to perform the temporal detection task. These findings are consistent with earlier work suggesting that larger differences between the temporal periodicity of the A and the B sounds facilitated a segregated percept, both in CI listeners (Chatterjee *et al.*, 2006; Duran *et al.*, 2012; Hong and Turner, 2009) and NH listeners (e.g., Grimault *et al.*, 2002; Roberts *et al.*, 2002; Vliegen *et al.*, 1999; Vliegen and Oxenham, 1999). The effect of Δrate was smaller for the short sequence [$t(26.41) = 2.194$, $p = 0.274$, difference estimate = 0.722], where no significant difference was observed between the *d'* scores achieved for the large and small Δrate conditions. Listeners achieved larger *d'* scores with the long sequence than with the short sequence for the large Δrate condition [$t(18.00) =5.554$, p < 0.001, difference estimate = 1.152] but not for the small Δrate condition [$t(18.00) = 2.113$, p = 0.324, difference estimate = 0.428] or for the no distractor condition [$t(18.00) = 1.482$, p = 0.679, difference estimate = 0.307].

These results suggest that the combination of both a large Δrate and a long sequence facilitated the segregation of the streams, as reflected by the larger *d'* scores obtained for this condition. Results from the "no distractor" condition demonstrated that the sequence duration itself did not affect the temporal detection task. Thus, the difference between the d' scores achieved with the long and the short sequences, for the large Δrate condition, are likely to represent the build-up of a two stream percept.

These findings are consistent with the results from a preliminary experiment by Chatterjee *et al.* (2006) with a single CI listener despite the fact that they relied on direct reports of perception from the listener, which can be problematic with CI listeners (Cooper and Roberts, 2007; Hong and Turner, 2009). The results presented here are also consistent with the findings from Nie and Nelson (2015), who investigated the effect of amplitude modulation (AM) rate and sequence duration in NH listeners. In both studies, a significant interaction was found between AM or pulse rate and the sequence duration, suggesting that CI listeners experience a similar build-up process as NH listeners do (e.g., Anstis and Saida, 1985; Bregman, 1990; Moore and Gockel, 2012).

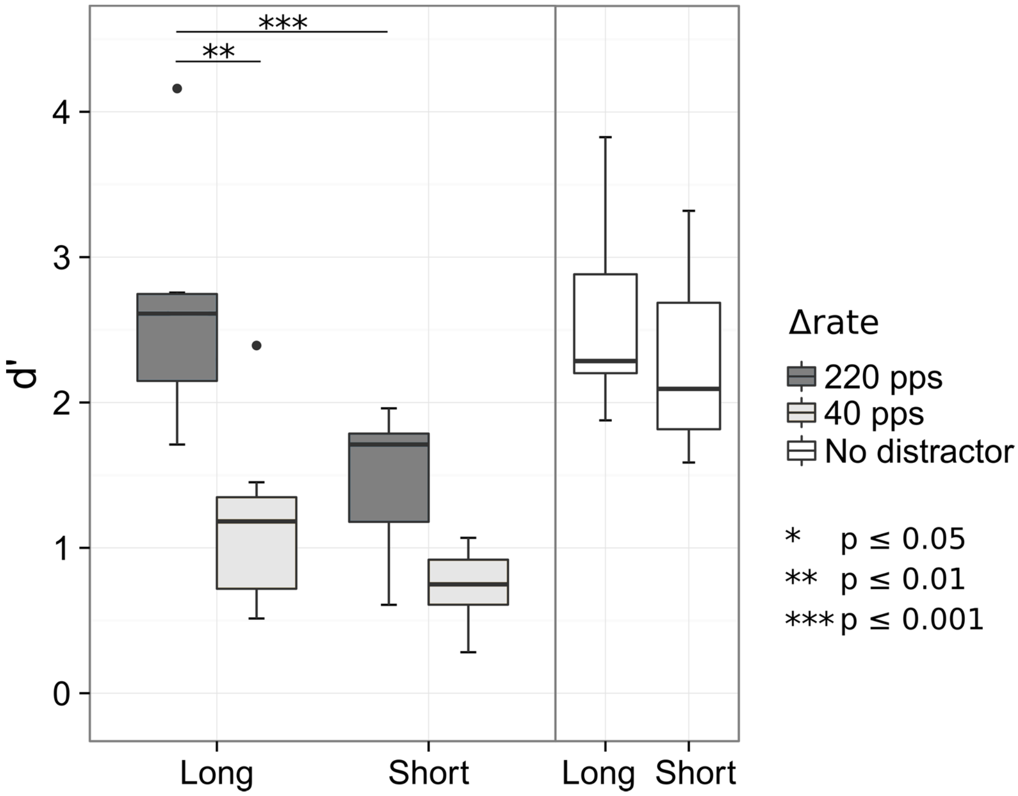


Fig. 2: Sensitivity to the delayed sound (*d'*) for each Δrate and sequence duration. The long and short sequences consisted of 12 and 4 duplets of AB sounds, respectively.

The similarity between the trends observed in NH and CI listeners supports the idea that CI listeners might experience stream segregation in a similar way as NH listeners. However, shorter Δt were needed in the "no distractor" condition to avoid ceiling effects. This reflects the increased difficulty experienced by CI listeners to perform the temporal detection task in the presence of a distractor stream, even when a large Δrate was used. Thus, even though CI listeners seem to be able to achieve a segregated percept and exhibit a similar build-up process as NH listeners, they might not be able to completely ignore a competing stream.

SUMMARY AND CONCLUSIONS

The present study demonstrated that temporal periodicity cues elicited by changes in the stimulus rate can facilitate the segregation of sequential sounds for CI listeners, given that enough time is provided to build up a two-stream percept. Overall, the findings reported here are consistent with earlier work with CI and NH listeners. The similarity in the trends observed between CI and NH listeners suggests that both groups of listeners might experience stream segregation in a similar way. However, these findings are based on the results from a relatively simple task and may not be generalizable to more complex and realistic environments.

ACKNOWLEDGMENTS

Work supported by the Oticon Centre of Excellence for Hearing and Speech Sciences and the Carlsberg foundation. Research equipment provided by Cochlear.

REFERENCES

Anstis, S.M., and Saida, S. (**1985**). “Adaptation to auditory streaming of frequency-modulated tones,” J. Exp. Psychol. Hum. Percept. Perform., **11**, 257-271. doi:10.1037/0096-1523.11.3.257

Bregman, A.S. (**1990**). *Auditory Scene Analysis : The Perceptual Organization of Sound.* The MIT Press.

Böckmann-Barthel, M., Deike, S., Brechmann, A., Ziese, M., and Verhey, J.L. (2014). “Time course of auditory streaming: do CI users differ from normal-hearing listeners?” Front. Psychol., **5**, 775. doi: 10.3389/fpsyg.2014.00775

Carlyon, R.P. (**2004**). “How the brain separates sounds,” Trends Cogn. Sci., **8**, 465-471. doi: 10.1016/j.tics.2004.08.008

Chatterjee, M., Sarampalis, A., and Oba, S.I. (**2006**). “Auditory stream segregation with cochlear implants: A preliminary report,” Hear. Res., **222**, 100-107. doi: 10.1016/j.heares.2006.09.001

Cooper, H.R., and Roberts, B. (**2007**). “Auditory stream segregation of tone sequences in cochlear implant listeners,” Hear. Res., **225**, 11-24. doi: 10.1016/j.heares.2006.11.010

Cooper, H.R., and Roberts, B. (**2009**). “Auditory stream segregation in cochlear implant listeners: measures based on temporal discrimination and interleaved melody recognition,” J. Acoust. Soc. Am., **126**, 1975-1987. doi: 10.1121/1.3203210

Duran, S.I., Collins, L.M., and Throckmorton, C.S. (**2012**). “Stream segregation on a single electrode as a function of pulse rate in cochlear implant listeners,” J. Acoust. Soc. Am., **132**, 3849-3855. doi: 10.1121/1.4764875

Grimault, N., Bacon, S.P., and Micheyl, C. (**2002**). “Auditory stream segregation on the basis of amplitude-modulation rate,” J. Acoust. Soc. Am., 111, 1340-1348. doi: 10.1121/1.1452740

Hong, R.S., and Turner, C.W. (**2006**). "Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients," J. Acoust. Soc. Am., **120**, 360-374. doi: 10.1121/1.2204450

Hong, R.S., and Turner, C.W. (**2009**). "Sequential stream segregation using temporal periodicity cues in cochlear implant recipients," J. Acoust. Soc. Am., 126, 291-299. doi: 10.1121/1.3140592

Kuznetsova, A., Christensen, R.H.B., Bavay, C., and Brockhoff, P.B. (**2015**). "Automated mixed ANOVA modeling of sensory and consumer data," Food Qual. Prefer., **40**, 31-38. doi: 10.1016/j.foodqual.2014.08.004

Landsberger, D.M., Vermeire, K., Claes, A., Van Rompaey, V., and Van de Heyning, P. (**2016**). "Qualities of single electrode stimulation as a function of rate and place of stimulation with a cochlear implant. Ear Hearing, **37**, e149-e159. doi: 10.1097/AUD.0000000000000250

Micheyl, C., and Oxenham, A.J. (**2010**). "Objective and subjective psychophysical measures of auditory stream integration and segregation," J. Assoc. Res. Otolaryngol., **11**, 709-724. doi: 10.1007/s10162-010-0227-2

Moore, B.C.J., and Gockel, H.E. (**2012**). "Properties of auditory stream formation," Philos. Trans. R. Soc. B Biol. Sci., **367**, 919-931. doi: 10.1098/rstb.2011.0355

Nie, Y., Zhang, Y., and Nelson, P.B. (**2014**). "Auditory stream segregation using bandpass noises: evidence from event-related potentials," Front. Neurosci., **8**, 1-12. doi: 10.3389/fnins.2014.00277

Nie, Y., and Nelson, P. (**2015**). "Auditory stream segregation using amplitude modulated bandpass noise," J. Acoust. Soc. Am., **127**, 1809. doi: 10.1121/1.3384104

Roberts, B., Glasberg, B.R., and Moore, B.C.J. (**2002**). "Primitive stream segregation of tone sequences without differences in fundamental frequency or passband," J. Acoust. Soc. Am., **112**, 2074-2085. doi: 10.1121/1.1508784

Roberts, B., Glasberg, B.R., and Moore, B.C.J. (**2008**). "Effects of the build-up and resetting of auditory stream segregation on temporal discrimination," J. Exp. Psychol. Hum. Percept. Perform., **34**, 992-1006. doi: 10.1037/0096-1523.34.4.992

Tejani, V.D., Schvartz-Leyzac, K.C., and Chatterjee, M. (**2017**). "Sequential stream segregation in normally-hearing and cochlear-implant listeners," J. Acoust. Soc. Am., **141**, 50-64. doi: 10.1121/1.4973516

Van Noorden, L.P.A.S. (**1975**). *Temporal Coherence in the Perception of Tone Sequences.* Institute for Perceptual Research.

Vliegen, J., Moore, B.C.J., and Oxenham, A.J. (**1999**). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," J. Acoust. Soc. Am., **106**, 938-945. doi: 10.1121/1.427140

Vliegen, J., and Oxenham, A J. (**1999**). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am., **105**, 339-346. doi: 10.1121/1.424503

Zeng, F.G., Rebscher, S., Harrison, W., Sun, X., and Feng, H. (**2008**). "Cochlear implants: system design, integration, and evaluation," IEEE Rev. Biomed. Eng., **1**, 115-142. doi: 10.1109/RBME.2008.2008250

An improved privacy-aware system for objective and subjective ecological momentary assessment

ULRIK KOWALK*, SVEN KISSNER, PETRA VON GABLENZ, INGA HOLUBE, AND JOERG BITZER

Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, Oldenburg, Germany

The technical components and software features of a new hearing-aid compatible smartphone-based ecological momentary assessment (EMA) system are presented in this paper. EMA is an assessment strategy that seeks to minimise instrumental infliction on the measured entity while data is gathered at multiple points of time. This work builds upon an already developed and deployed smartphone-based system. Objective data is gathered in the form of acoustical features, while subjective data is collected via automatised questionnaires. Since linking objective acoustical measures to subjective assessments is particularly promising with regard to the hearing rehabilitation process, our system has been specially tailored for hearing aid users. The introduction of wireless data transfer has eliminated cable clutter, a customisable questionnaire allows for subjective assessment, and a streamlined user interface complements the design. Like the former version, the current revision ensures the privacy of both participants and third parties. To facilitate cooperative research, source code and custom-built hardware will be released under open source licenses. All additional components are commercially available.

ECOLOGICAL MOMENTARY ASSESSMENT

When conducting a study, practised procedures involve the completion of one or more questionnaires, usually taken in retrospective. The answers given rely heavily on memory, but because memory is of transient nature its distorting effects can strongly influence the results. According to Shiffman *et al.* (2008), ecological momentary assessment (EMA) uses a different approach. Data is recorded at several points during the study, often on sub-hour intervals, resulting in more reliable answers and evaluable evolution of parameters over time. To ensure that measurement does not interfere with the entity being measured, data collection needs to be as ecological as possible. EMA therefore aims at gathering data at the time it is generated without having an influence on the data. For practical reasons, repeated surveys are nowadays usually conducted using digital devices. Galvez *et al.* (2012) incorporated a personal digital assistant-based (PDA) EMA application to explore when and how hearing problems occur throughout the day. As opposed to strictly subjective assessments, a number of

*Corresponding author: ulrik.kowalk@jade-hs.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

audiological studies have implemented recordings of physical variables to correlate with survey results. Banerjee (2011a) used logged data of manual multimemory and volume control adjustments by hearing aid users together with broadband input levels of the hearing aids combined with EMA in order to successfully identify situations in which users desire to modify settings. In a second study, Banerjee (2011b) investigated the correlation between automatic behaviour of hearing aids and EMA surveys to learn more about parameter decisions made by the hearing aids. Timmer *et al.* (2017) verified the validity of EMA according to hearing experiences, using data from an environmental classifier bound to a smartphone-based EMA system by showing high correlation between objective and subjective results as well as an overall high compliance rate. In a recent study published by Wu *et al.* (2017) the authors complemented randomly timed smartphone-based surveys with audio recordings from a portable device the test subjects carried around their neck in order to classify different listening situations.

Our method, in contrast, implements a single application that extracts acoustic features from a live binaural audio stream in real time and combines them with data from randomly or fixedly timed questionnaires. Implementing bluetooth audio transmission from ear-level microphones to a smartphone, it is a singular open-source experiment device that can be used without programming knowledge. High-level privacy-awareness allows for legally unconstrained use in everyday situations.

DESCRIPTION OF THE PREVIOUS SYSTEM

In Kissner *et al.* (2015), a smartphone-based EMA system was presented that included microphones for the purpose of extracting audio features. These microphones were worn like behind-the-ear (BTE) hearing aids and signals were transmitted via cables to an external USB audio adapter connected to the smartphone. Two applications ran simultaneously – one performed extraction of acoustical features and one conducted questionnaires without any internal communication channel between the two applications. Audio features were archived as blocks of bundled data over short periods of time, creating a separate series of blocks for each feature.

ATTRIBUTES OF THE NEW SYSTEM

The technical components and properties of the new system as well as advantages over the old system are described in this section.

Hardware

For the newly developed system, different choices with regard to hardware have been made. The microphones are no longer mounted behind the ears but attached to glasses, right above the ears (see Fig. 1). They are connected to a pocket-sized transmitter box (weight: ca. 36 g) by means of audio cables (length: ca. 0.5 m). The manufactured compartment contains a set of two A/D converters with preamplifiers and voltage offset filters, a lithium-ion polymer (LiPo) battery, and a Bluetooth transmitter for wireless audio transfer to the smartphone. Signals are sent in high resolution via the A2DP protocol. An integrated safeguard circuit monitors the voltage of the LiPo battery and performs automatic shutdown in order to preserve battery life. The transmitter unit has a runtime of more than 8 hours and is charged by either USB, external power supply, or induction coil. RGB-LEDs indicate the current state of the device and transmission power is regulated dynamically when necessary. It can be easily attached to any clothing by an external clip. An outline of the system schematics is shown in Fig. 2.

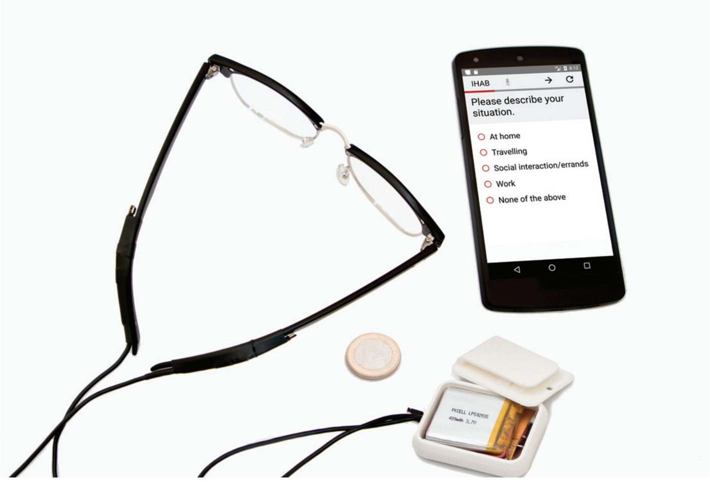


Fig. 1: Prototype of EMA system: Microphones attached to glasses, pocket-sized Bluetooth transmitter, and smartphone with questionnaire

A challenge during development has been the transformation of an Android device to act as a dual channel audio receiver. Since this feature is not provided by the conventional Android system, the device (LG Nexus 5) has been equipped with a modified Android Automotive operating system.

Signal processing

At the current stage, three acoustical features are extracted in realtime by the smartphone from signal blocks $x_m[n] = x[n + m \cdot N]$ of size N, m being the block index. The first measure is the RMS

$$\mathrm{RMS}_m = \sqrt{\sum_{n=0}^{N-1} x_m^2[n]}, \qquad \text{(Eq. 1)}$$

which is calculated in order to obtain binaural loudness levels. As a basic input to voice activity detection, the zero-crossing rate (ZCR) of the signal $x[n]$ and its first derivative $\Delta x[n] = x[n] - x[n-1]$ are recorded as well. The ZCR is calculated

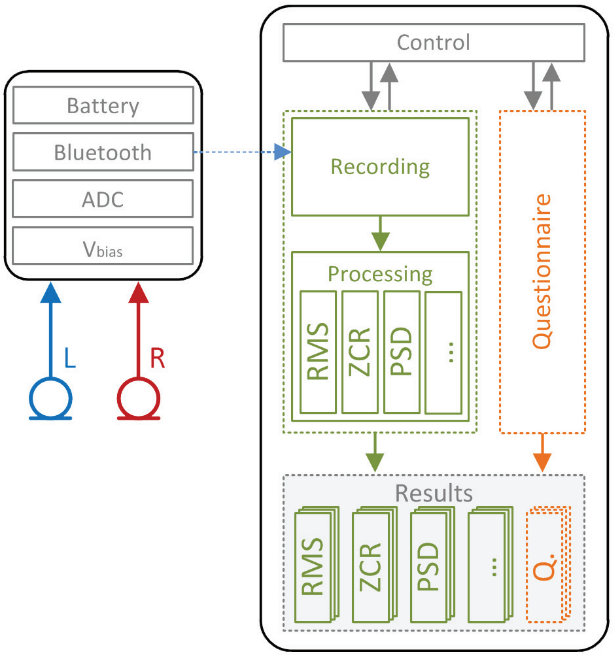


Fig. 2: System layout – dual microphone signals sent to smartphone via Bluetooth transmitter, extraction of acoustical features, and automatic conduction of questionnaire.

according to

$$\mathrm{ZCR}_m = \frac{1}{N-1}\sum_{n=0}^{N-1} s_m[n] \qquad \text{(Eq. 2)}$$

$$\text{with } s_m[n] = \begin{cases} 1 & \text{if } x_m[n]\cdot x_m[n-1] < 0 \\ 0 & \text{else,} \end{cases} \qquad \text{(Eq. 3)}$$

counting the number of zero crossings per signal block. The third feature are power spectral densities Φ, that are calculated for both channels separately, as well as cross-power spectral densities (Φ_{LR}) between left and right channel. In Eq. 4 and Eq. 5, X_m is the Fourier transform of the N-point Hann-windowed signal x_m:

$$\Phi_m[n] = X_m[n] \cdot X_m^*[n] \qquad \text{(Eq. 4)}$$

$$\Phi_{\mathrm{LR},m}[n] = X_{\mathrm{L},m}[n] \cdot X_{\mathrm{R},m}^*[n] \qquad \text{(Eq. 5)}$$

These features were chosen to give an acoustical overview of the participants' daily routine (Bitzer *et al.*, 2016) and will in future be complemented by other objective metrics. To simplify further extension of functionality, the system uses a defined plug-in architecture. Feature extraction is performed on chunks of fixed length (e.g., 60 s) resulting in one time-stamped series per feature.

Privacy-awareness

Two requirements are met by the system: 1) No audio data is stored; 2) No content can be reconstructed from the extracted data. The first criterion is met by the design of the processing engine. The second is implemented in each feature respectively. Special consideration is taken for the PSD feature. In order to prevent reconstruction, PSD time series are smoothed with a time constant of $\tau = 125\,\text{ms}$ and certain frames are omitted. As shown by Kissner *et al.* (2015), no content can thus be retrieved while acoustical information is still usable for study.

Questionnaires

A simplified interface has been developed for adaptation and creation of new questionnaires without the restriction of programming skills. Two software components are essential: the main application in the form of an installable .apk file and at least one questionnaire. The questionnaire is implemented as a formatted, human-readable .xml file with intuitive attributes and optional comments. Time scheduling values are specified as mean interval and randomness margin represented by seconds, the usual case being periodically recurring questionnaires. A margin of 0 seconds yields a steady sampling interval and different questionnaires are selected via a preferences menu. Dynamic structuring helps extract a maximal amount of data through tailored questionnaires by only stating relevant questions, true to a filter attribute. A question is only visible if its criteria are met based on a system of unique answer identifiers (IDs) that come with every answer. Once an answer has been selected, the corresponding ID is saved to memory. Two possible restrictions exist. Either a predefined ID must exist in the memory (positive criterion) or it explicitly must not exist (negative criterion). While a question is shown if one or more positive criteria are satisfied, it will be hidden if one or more negative criteria are met, negative overriding positive. For intuitive assessment, answer formats include radio buttons, checkboxes, emojis, sliders with fixed or arbitrary scales, and free text.

Open source

In order to allow for collaboration, all source code and construction plans will be published under open source licenses. The system uses Nexus 5 smartphones, which are commercially available. The questionnaire management application has been localised across English and German. Future releases will include further languages.

Performance

Several acoustical parameters have been measured in order to asses the technical properties of the current system. For reproduction of test signals, an NTi Audio Talk-Box is used and reference measurements are taken by a G.R.A.S. 40AF free-field microphone pre-amplified by a Brüel & Kjær 2829 type 26TK system. Sound pressure levels are calibrated using a Norsonic Sound Calibrator type 1251 and the experiments are conducted in an anechoic chamber with a volume of approximately $43\,\text{m}^3$.

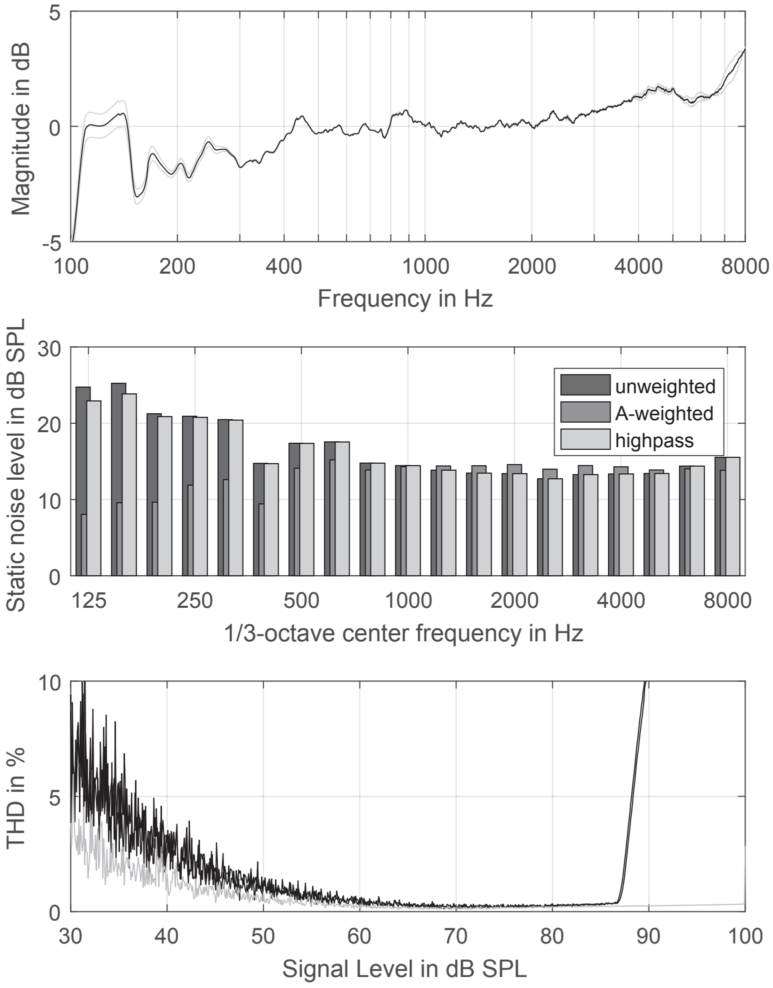


Fig. 3: *Top:* Frequency response of the current system, smoothed in ERB bands. Individual microphone responses are drawn as grey lines, the averaged response is drawn in black. *Middle:* Third-octave noise levels for different pre-processing filters. *Bottom:* Average percentage of total harmonic distortion by the system (black line) and reference microphone (grey line).

Frequency response is measured with calibrated, zero-centred broadband noise played back at a level of 60 dB SPL. The distance between speaker and system microphones is 1.60 m. As shown in Fig. 3 (top), we find a reasonably flat response between 100 Hz and 8 kHz with a slight upward tendency towards higher frequencies. Tolerances between multiple microphones are less than ±0.5 dB between 100 Hz and 200 Hz, less than ±0.1 dB between 200 Hz and 4 kHz and a little higher beyond.

Equivalent input noise levels are examined under the same external conditions as the frequency response. Three different states of internal processing are evaluated: unweighted, A-weighted, and filtered by a second order Butterworth high pass filter with f_0=100 Hz. As depicted in Fig. 3 (middle), unweighted noise levels are relatively equally distributed with emphasis on frequencies lower than 400 Hz.

Total harmonic distortion determines the dynamic range of the system. Measurements are conducted using a Fostex 6301B loudspeaker generating an amplitude sweep signal with a sinusoidal carrier at a frequency of 1 kHz and level ranging from 20 to 100 dB SPL. For a distortion acceptance limit of 2 %, results show dynamic validity ranging from approximately 45 up to 88 dB SPL yielding an estimated dynamic of 43 dB (see Fig. 3, bottom). This renders the system applicable for one of the main acoustical situations of everyday life – conversational speech – which usually lies in the range above 50 dB SPL according to Bitzer *et al.* (2016).

Advantages over the previous system

While being fully functional, the preceding system by Kissner *et al.* (2015) included certain attributes that were updated for the new system. Because a line connection was used to transmit microphone signals to the smartphone, a third party USB audio interface was required at the receiving end. This implied the need for proprietary device drivers, counteracting complete open source publication, and also leading to mechanical instabilities. This system uses wireless transmission, thus eliminating the need for additional hardware. Another advantage is reduced ecological influence on measurements due to omitted inhibitory attributes (e.g., cable clutter). Because the microphones are no longer mounted behind the ears, but are attached to the temple, the concurrent use of behind-the-ear hearing aids is now possible. Integration of signal processing and questionnaire management into one single application has introduced process supervision on a high level, increasing functional security and simplifying usability. New answer formats and scheduling options within the questionnaire editing interface have increased flexibility and have lead to more intuitive assessment.

SUMMARY AND OUTLOOK

A system has been presented that incorporates all instruments to perform privacy-aware EMA of both subjective and objective parameters, while granting the experimenter a high degree of freedom, flexibility, and simplicity. Currently in development is a shared database for swift data exchange, pooling and comparison, which would facilitate international collaboration. Wireless signal transmission over a serial protocol to loosen the constraint on phone brand and model are also being investigated as well as a modulation-based blind estimator of speech quality, the speech to reverberation modulation energy ratio (SRMR, see Falk *et al.*, 2010) to supplement the current feature set. Further options regarding questionnaires are event-based and fixed scheduling. The system will be integrated in a field study scheduled for 2018 (see Meis *et al.*, 2017).

ACKNOWLEDGEMENTS

The authors would like to thank Holger Groenewold and Sven Franz for hardware development and modification of the operating system. This work was supported by the Hearing Industry Research Consortium (IRC). Further support was granted by the governmental funding initiative "Niedersaechsisches Vorab" of the Lower Saxony Ministry for Science and Culture, research focus "Hoeren im Alltag Oldenburg (HALLO)".

REFERENCES

Banerjee, S. (**2011a**). "Hearing Aids in the real world: Use of multimemory and volume controls," J. Am. Acad. Audiol., **22**, 359-374. doi: 10.3766/jaaa.22.6.5

Banerjee, S. (**2011b**). "Hearing aids in the real world: Typical automatic behavior of expansion, directionality, and noise management," J. Am. Acad. Audiol., **22**, 34-48. doi: 10.3766/jaaa.22.1.5

Bitzer, J., Kissner, S., and Holube, I. (**2016**). "Privacy-aware acoustic assessments of everyday life," J. Audio Eng. Soc., **64**, 395-404. doi. 10.17743/jaes.2016.0020

Falk, T.H., Zheng, C., and Chan, W.Y. (**2010**). "A non-intrusive quality and intelligibility masure of reverberant and dereverberated speech," IEEE T. Audio Speech, **18**, 1766-1774. doi: 10.1109/TASL.2010.2052247

Galvez, G., Turbin, M.B., Thielman, E.J., Istvan, J.A., Andrews, J.A., Henry, J.A. (**2012**). "Feasibility of ecological momentary assessment of hearing difficulties encountered by hearing aid users," Ear Hearing, **33**, 497-507. doi: 10.1097/AUD.0b013e3182498c41

Kissner, S., Holube, I., and Bitzer, J. (**2015**). "A smartphone-based, privacy-aware recording system for the assessment of everyday listening situations," Proc. ISAAR, **5**, 445-452.

Meis, M., Krueger, M., Gebhard, M., von Gablenz, P., Holube, I., Grimm, G., and Paluch, R. (**2017**). "Overview of new outcome tools addressing auditory ecological validity: Analyses of behavior in real life listening environments," Proc. ISAAR, **6**, 31-38.

Shiffman, S., Stone, A.A., and Hufford, M.R. (**2008**). "Ecological momentary assessment," Annu. Rev. Clin. Psycho., **4**, 1-32. doi: 10.1146/annurev.clinpsy.3.022806.091415

Timmer, B.H.B., Hickson, L., and Launer, S. (**2017**). "Ecological momentary assessment: Feasibility, construct validity, and future applications," Am. J. Audiol., **26.3S**, 436-442. doi: 10.1044/2017_AJA-16-0126

Wu, Y.H., Stangl, E., Chipara, O., Shabih Hasan, S., Welhaven, A., and Oleson, J. (**2017**). "Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss," Ear Hearing, in press. doi: 10.1097/AUD.0000000000000486

Development and application of a code system to analyse behaviour in real life listening environments

MARKUS MEIS[1,2,*], MELANIE KRUEGER[1,2], MARIA GEBHARD[1,2,3],
PETRA V. GABLENZ[3,2], INGA HOLUBE[3,2], GISO GRIMM[4,2], AND RICHARD PALUCH[1,2,5]

[1] *Hörzentrum Oldenburg GmbH, Oldenburg, Germany*

[2] *Cluster of Excellence Hearing4all, Oldenburg, Germany*

[3] *Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, Oldenburg, Germany*

[4] *HörTech gGmbH, Oldenburg, Germany*

[5] *Carl von Ossietzky University, Oldenburg, Germany*

Numerous studies showed that different hearing aid (HA) algorithms improve speech intelligibility in typical lab situations as measures of clinical efficacy. From the perspective of auditory ecology, it remains obscure to what extent these results really allow for estimating the outcome in listening situations in real life. One promising tool is the observation of participants behaviour induced by different HA settings. We developed an annotation system for coding the behaviour related to the framework of the International Classification of Functioning, Disability and Health (ICF) in iterative steps. The first inputs were derived from a series of lab studies, using virtual acoustics. It was shown that different directional modes of HAs influenced real life behaviour. First indications of activity limitation according to ICF (d3504 'Conversing with many people') were found. Additionally, the behaviour of users in real life was described by means of 'ethnographical walks' outside of the laboratory using field notes. We identified further behaviour patterns addressing spatial awareness. The conversation related ICF sub-categories were validated by analyses of inter-rater reliability (IRR). The outcome of these analyses led to a reformulation of an annotation/coding system for the usage on tablet PCs for instantaneous coding of the test persons behaviour in real life.

INTRODUCTION

In addition to the benefits of hearing aids, as shown in the lab by means of clinical oriented speech tests, hearing specific and generic questionnaires or diaries are used to determine the benefits in every-day life. In several studies it was shown that, without rehabilitation with hearing aids, a hearing loss influences every-day activities negatively and reduces Health-related Quality of Life (HrQoL), particularly in the

*Corresponding author: m.meis@hoerzentrum-oldenburg.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
ISBN: 978-87-990013-6-1.

domains of social functioning and mental well-being (e.g., Chisolm *et al.,* 2007). HrQoL outcome tools are focusing on self-administered questionnaires over a past period of usually four weeks and are mostly filled in retrospectively, and therefore this can be regarded as a measure of long-term HrQoL (L-HrQoL). Summarizing experiences retrospectively, however, is possibly biased by interlocked effects of memory and subjective perception.

Gatehouse *et al.* (1999) proposed an 'auditory ecology' approach, which takes the objective physical characteristics of every-day listening environments and the individual listener's demands in these real listening environments into account. Up to date our knowledge of real life listening environments still lacks resilient empirical and qualitative data. Possible solutions to bridge this gap could be smartphone-based systems to measure the acoustical environment and to combine those objective measures (acoustical information/data) with subjective data of the respective patient in an approach named Ecological Momentary Assessment (EMA; see, e.g., Bitzer *et al.,* 2016; Kowalk *et al.*, 2017; Shiffman *et al.*, 2008). In addition to subjective ratings of everyday situations over a longer period, short items of "momentary" or acute HrQoL, we called it M-HrQoL, are a promising approach to enrich the outcome toolbox of auditory ecology. M-HrQoL items could be included in measurements of the situation-specific self-perception in the actual situation, as a sub-domain of EMA. Self-perception is still one important data source of listening situations, but behaviour, especially the ability to communicate in conversational situations, is a very relevant outcome area too. Paluch *et al.* (2015) showed that communication behaviour changes in relation to different HA modes. They identified two core dimensions of communication behaviour: 'forms of interaction' (Face-to-Face [F-t-F] vs. group communication) and 'interdependence' (symbolic gestures vs. spoken words) based on Strauss (1987). A higher ratio of F-t-F interactions as well as a higher ratio of verbal communication for an adaptive binaural beamformer in contrast to a broader adaptive monaural beamformer was shown in group conversation, but only in a loud super market scene (L_{Aeq_15min} = 67 dB) in contrast to a softer condition of L_{Aeq_15min} = 55 dB. These behaviour descriptions need to be linked to M-HRQoL to assess the user's handicap qualitatively. The framework of the International Classification of Functioning, Disability and Health (ICF) might facilitate an appropriate approach (WHO, 2001). The ICF model allows to describe the dynamic interaction between the components body functions/structure, activities, participation and environment related, as well as person-centered contextual factors. The ICF model has the privilege to provide generic qualifiers of disability/functioning.

SYSTEMATIC DEVELOPMENT OF A BEHAVIOURAL CODE SYSTEM

Lab test: Comparison of directional modes from three HA devices

In total, six male and four female experienced HA users participated in group discussion sessions (mean age=72.6 years, SD = 7.6, $PTA_{4\ (0.5,\ 1.0,\ 2.0,\ 4.0\ kHz)}$ better ear = 49.7 dB HL, SD = 6.7). The participants were divided into two groups of five subjects, which were invited successively. In the experiment, the participants were seated at one table with near and distant communication partners for four group sessions with a duration of 15 minutes each. For the study, three custom-made in-the-ear (ITE) devices from different brands were fitted (first fit) to the test persons (for preliminary data see Latzel *et al.*, 2016). Three devices per ear were built from the identical ear impression for each subject. The power levels of the hearing aids were specified in order to compare the same power levels across test devices. In each HA, a program for speech intelligibility in loud situations was fitted with a directional microphone mode with narrow directionality. The vents of the hearing devices were individually chosen due to the pure tone audiogram and the HA characteristics. The realization of the group discussion procedure was exactly the same as in the study from Paluch *et al.* (2015), but took place only in a noisy scene (L_{Aeq_15min} = 67 dB). After each of the four conversation sessions, a questionnaire was filled out by the participants. For the subjective rating of speech intelligibility a scale from ‘1’ (nothing) to ‘7’ (all) was used. For the analyses of the behavioural data the same annotation scheme as in the Paluch *et al.* (2015) study was applied.

At first glance, the data showed that no differences according the dimensions F-t-F vs. group communication and symbolic gestures vs. spoken words were observed. This pattern of results was contradictory to the subjective ratings of the participants regarding perceived speech intelligibility. Further analysis established statistically significant differences (non-parametrical analyses of repeated measurements) in speech intelligibility ratings for the three devices, indicating that, e.g., device #3 was rated with a median of 2 and device #1 with a median of 3.5. The obtained differences did not reflect the behavioural data; Therefore, a clarification was necessary. A team of three raters inspected once again the whole video material. In an iterative Grounded Theory (Glaser and Strauss, 1967) based process, two further sub-dimensions were striking: different proxemics regarding near vs. distant torso movements (forward-backward) to the dialogue partner and conversations with the distant vs. near dialogue partner. We proposed thus a revised annotation scheme (Meis *et al.*, 2016), as illustrated in Fig. 1.

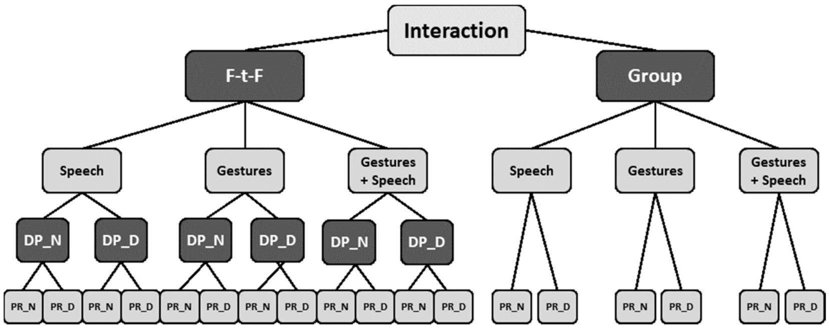


Fig. 1: 18 code annotation scheme of communication / interaction. Face-to-Face: F-t-F Interaction, Group: Group Interaction, DP: Distance Partner: near vs. distant, PR: Proxemics: near vs. distant.

The result was an annotation scheme including in total 18 codes, a hierarchic scheme of interdependent codes, for the four behaviour domains. Using this scheme, 2,939 behaviour units with a time resolution of ~13 s per unit/test person were assessed.

Following the revised annotation scheme, no differences were found regarding the core dimension ‘interdependence’. Regarding the core dimension ‘forms of interaction’ the ratio of F-t-F and F-t-F plus group interaction was highest for device #3, indicating nearly 15% more interactions in contrast to the two other devices. The examination of the F-t-F category ‘Distance to the dialogue partner’ (near vs. distant) showed that test persons using device #3 interacted in > 80% (median) of the assessed interaction units only with the respective near dialogue partner, in contrast to device #1 (Wilcoxon signed rank test, *p*=0.009). Using device #1, the ratio of interactions was balanced regarding the F-t-F communication of the near vs. distant dialogue partner as shown in Fig. 2. Additionally, the analysis of proxemics revealed the significant effect that subjects tended to lean more forward for device #2 (*p*=0.043, Wilcoxon signed rank test) and #3, compared to subjects using device #1.

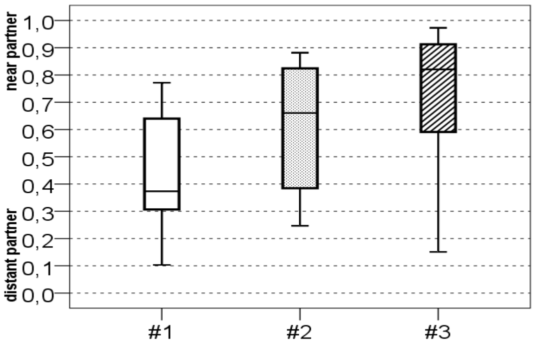


Fig. 2: Communication with near vs. distant dialogue partner for three different devices in %. Boxplots show the distribution of the ratios calculated from F-t-F near partner/near and distant partner communication.

The data regarding ‘Distance to the dialogue partner’ can be interpreted as a limitation of communication activities induced by the microphone mode of the hearing aids. Subjects with sub-optimal HA fittings are rather able to communicate with the near dialogue partner, but not with the respective distant dialogue partner. This result pattern suggested the interpretation of data along theoretical and practical models of HrQoL and the classification of the ICF framework.

Expert review and ethnography

Granberg *et al.* (2014) published a comprehensive ICF core set for hearing loss with the domains ‘body functions’ (‘b’ codes), ‘body structures’ (‘s’ codes), ‘activities and participation’ (‘d codes’), and ‘environmental factors’ (‘e codes’) with in total over 100 codes. This Core Set was used as a basis for a review meeting with six experts in the field of audiology, rehabilitation, and psychology with the main goal to extract codes applicable in *behavior observations* using a 3-point scale. The most prominent ICF codes for behavior analyses were derived from the functions *'activities and participation'*, labeled as 'd' codes and – partly related – 'b' codes.

Codes rated 'moderate appropriate' or 'very appropriate' by the experts were used for external observations in an ethnographic field study (Paluch *et al*., 2017). This ethnographic study was conducted as a stand-alone experiment with 10 test persons (n=2 normal-hearing, n=7 unaided slight to moderate hearing loss, and n=1 moderate hearing loss aided with HA; reference better ear PTA$_4$ according to WHO, 2001). Three test persons classified as unaided were externally observed both unaided and recently fitted with hearing aids. The behavior of the other seven participants was observed in the respective aided or unaided condition. During the external observation the subjects took a 4.5-km walk and visited different locations (cafeteria, several bus stops). A trained observer generated field-notes, based on the methodology of Przyborski and Wohlrab-Sahr (2009). Paluch *et al*. (2017) showed that newly fitted hearing-impaired subjects tended to move their torsos and heads (left-right level) in a significant manner, possibly caused by new auditory input, particularly spatial input.

Based on the results of the reported studies, the expert review, and the ethnographical walks an extended set of ICF categories for behaviour analyses in the field was finally derived (Table 1).

Inter-rater reliability (IRR) of the extended ICF core sets

The future goal is to develop a tool to assess instantaneous behaviour ratings in the field via tablet PC and to combine these external observed behavioural data with objective and subjective data for a multifaceted EMA. Therefore, the proposed extended ICF categories needed a check by IRR procedures, addressing – in a first step – conversation situations.

For the IRR check, a manual with the detailed descriptions of the extended ICF categories was elaborated to provide a clear reference for the evaluation of the different characteristics. IRR was established by three different raters.

- **d160** Focusing attention
    1. Movements torso horizontal axis, frequency <= 45° vs. >45°
    2. Movements head, horizontal axis, frequency <= 45° vs. >45°
- **b140** Attention functions
    1. Sustained attention: face of conversation partner, strong vs. weak
- **d3504** with many people/ **d3503** Conversing with one person
    1. F-t-F vs. group (**only d3504**)
    2. Frequency general verbal communication
    3. Communication partner: distant vs. near (only **d3504**)
    4. Proxemics (torso position) lean forward/backward vertical axis 90°
    5. Frequency change sitting position
    6. Non-understanding gestures, frequency
    7. Speech supporting gestures, frequency

Table 1: Extended ICF categories for behaviour ratings in the field.

Rater	A-B		B-C		A-C	
ICF (sub-) categories/scale	**κ**	r_{Sp}	**κ**	r_{Sp}	**κ**	r_{Sp}
b140_1 Sustained attention face partner: low-medium-high	.39	.58	.32	.56	**.44**	.65
d3504_1 Communication: F-t-F-balanced-group	**.47**	.58	.36	.38	**.57**	.70
d3504_2 Frequency verbal comm.: seldom-sometimes-frequent	**.51**	.72	**.52**	.68	**.43**	.70
d3504_3 Communication partner: near-balanced-distant	**.59**	.73	**.62**	.70	**.72**	.79
d3504_4 Proxemics: forward-balanced-backward	**.57**	.68	.38	.52	**.50**	.59
d3504_5 Change torso position: seldom-sometimes-frequent	.13	.26	.33	.56	.39	.57
d3504_6 Non-understanding gestures: seldom-sometimes-frequent	.07	.29	.35	.40	.16	.32
d3504_7 Speech supporting gestures: seldom-sometimes-frequent	.24	.51	.26	.39	**.46**	.57

Table 2: IRR for extended ICF categories. Cohen's kappa indicating moderate or substantial agreement in bold. A-C = 3 raters; κ = Cohen's kappa; rSp = Spearman's rho. Cohen's kappa agreement: <0 = "poor", 0–0.20 = "slight", 0,21–0,40 = "fair", 0,41–0,60 = "moderate", 0,61–0,80 = "substantial", 0,81–1,00 = "almost perfect"; see Landis and Koch (1977).

The raters had to rate a selection of two video sessions of the ITE benchmark study, presented above. The video material included five subjects and two different devices. In contrast to the study from Latzel *et al*. (2016), the rating referred to 3-min sections (in total 5 ratings in a 15-min conversation) with 5- or 7-point rating scales in order to reduce too frequent annotation activities for the rater in a field situation. We calculated Cohen's Kappa (κ) (Cohen, 1960) and correlations (Spearman's rho r_{Sp}) for pairs of raters to get deeper insight of the rater characteristics and condensed the rating scales into 3-point ordinal scales; see Table 2.

We observed predominantly poor to slight IRR statistical values for the categories b140_1 and d3504_5 to d3504_7. Moderate IRR-values were assessed for the categories F-t-F vs. group (d3504_1), frequency verbal communication (d3504_2), and proxemics (torso position) lean forward/backward vertical axis (d3504_4). Substantial IRR-values were gathered for the interactions with the distant vs. near conversation partner (d3504_3).

It is planned to use category 'd160' only for spatial awareness topics with moving sources, which are not included in the external communication behaviour assessment. Therefore, this category was not included in the IRR procedure.

DISCUSSION AND OUTLOOK

The development and application of a code system to analyse behaviour in real life listening environments was outlined in this paper. Based on the first explorative studies it was shown that ICF categories are related significantly with hearing aid usage, especially signalling *activity limitation* and *participation restriction*. In addition to clinical outcome measures, behavioural data of complex interaction episodes of group conversations in noisy environments offer the possibility to use auditory ecological valid outcome measures, which capture how hearing aids impact behaviour in real life. The approach and the studies presented here should be understood as explorative. They certainly need further theoretical foundation as well as the proof of reproducibility. The IRR-values indicated moderate to substantial inter-rater agreement in relevant ICF categories, but the IRR has to be improved for the usage in the field. The three raters stated that they had difficulties to average different behaviour units inside a three minute section. Moreover, it might be easier to annotate the conversation behaviour directly, e.g., on a tablet PC with a graphical user interface (GUI) for quick and easy tapping, but with a reduced set of codes. In future, we propose to use six hierarchic and interdependent main codes to evaluate group conversation, which include the categories ‘F-t-F vs. group’, ‘near vs. distant dialogue partner’, and ‘near vs. distant proxemics’ plus two codes of non-verbal proxemics ‘near vs. distant proxemics’ during listening. Using instantaneous annotations, the frequency of verbal communication episodes automatically will be recorded. The GUI should be completed with ICF relevant environmental and contextual categories, such as light condition (e240). In the next studies, we are going to combine the proposed eight codes of external behavioural observation with objective acoustical data and subjective items to get a more complete picture of hearing impaired users in real listening environments. In future, the approach reported here, should be combined and/or validated with the automatic assessment of behavioural data, such as head- and eye-tracking procedures.

ACKNOWLEDGMENTS

This work was supported by the Cluster of Excellence EXC 1077/1 *Hearing4all,* funded by the German Research Council (DFG), Matthias Latzel from Sonova, and the Hearing Industry Research Consortium (IRC) 2016 grant.

REFERENCES

Bitzer J., Kissner S., and Holube I. (**2016**). “Privacy-aware acoustic assessments of everyday life,” J. Audio Eng. Soc., **64**, 395-404.

Chisolm T.H., Johnson C.E., Danhauer J.L., Portz L.J.P., Abrams H.B., Lesner S., McCarthy P.A., and Newman C.W. (**2007**). “A systematic review of health-related quality of life and hearing aids: Final report of the American Academy of Audiology task force on the health-related quality of life benefits of amplification in adults,” J. Am. Acad. Audiol., **18**, 151-183.

Cohen, J. (**1960**). “A coefficient of agreement for nominal scales,” Educ. Psychol. Meas., **20**, 37-46.

Gatehouse S., Elberling C., and Naylor G. (**1999**) “Aspects of auditory ecology and psychoacoustic function as determinants of benefits from and candidature for non-linear processing in hearing aids,” Proc. Danavox Symposium, **18**, 221-233.

Glaser, B.G., and Strauss, A.L. (**1967**): *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago: Aldine.

Granberg, S., Möller, K., Skagerstrand, A., Möller, C., and Danermark, B. (**2014**). "The ICF Core Sets for hearing loss: researcher perspective, Part II: Linking outcome measures to the International Classification of Functioning, Disability and Health (ICF)," Int. J. Audiol., **53**, 77-87. doi: 10.3109/14992027.2013.858279.

Kowalk, U., Kissner, S., v. Gablenz, P., Holube, P., and Bitzer, J. (**2017**). "An improved privacy-aware system for objective and subjective ecological momentary assessment," Proc. ISAAR, **6**, 25-30.

Landis, J.R., and Koch, G.G. (**1977**). "The measurement of observer agreement for categorical data," Biometrics, **33**, 159-174.

Latzel, M., Paluch, R., Meis, M., and Krueger, M. (**2016). "**A new tool for subjective assessment of hearing aid performance: Analyses of interpersonal communication—next step(s)," Poster B12, International Hearing Aid Research conference (IHCON), Tahoe City, CA.

Meis, M., Paluch, R., Krueger, M., and Latzel, M. (**2016). "**A new evaluation tool for hearing aids in everyday situations: Video-based analysis of interpersonal communication behavior, part 2," 61st International Congress of Hearing Aid Acousticians, Hannover.

Paluch R., Latzel, M., and Meis, M. (**2015**). "A new tool for subjective assessment of hearing aid performance: Analyses of interpersonal communication" Proc. ISAAR, **5**, 453-460.

Paluch, R., Krueger, M., Grimm, G., and Meis, M. (**2017).** "Moving from the field to the lab: Towards ecological validity of audio-visual simulations in the laboratory to meet individual behavior patterns and preferences," 20. Jahrestagung der Deutschen Gesellschaft für Audiologie, Aalen.

Przyborski, A., and Wohlrab-Sahr, M. (**2009**). *Qualitative Sozialforschung. Ein Arbeitsbuch. 2., korrigierte Auflage.* München: Oldenburg Verlag, 403 Seiten, 978-3-486-59103-3.

Shiffman, S., Stone, A.a., and Hufford, M.R. (**2008**). "Ecological momentary assessment," Ann. Rev. Clin. Psychol., **4**, 1-32. doi: 10.1146/annurev.clinpsy.3.022806.091415

Strauss, A.L. (**1987**). *Qualitative Analysis for Social Scientists.* New York: Cambridge University Press.

WHO (**2001**). *International Classification of Functioning, Disability and Health: ICF.* Geneva: World Health Organization.

Ethnographic research: The interrelation of spatial awareness, everyday life, laboratory environments, and effects of hearing aids

RICHARD PALUCH[1, 3, 4,*], MELANIE KRUEGER[1, 4], MAARTJE M. E. HENDRIKSE[3,4], GISO GRIMM[2, 3, 4], VOLKER HOHMANN[1, 2, 3, 4], AND MARKUS MEIS[1, 4]

[1] *Hörzentrum Oldenburg GmbH, Oldenburg, Germany*

[2] *HörTech gGmbH, Oldenburg, Germany*

[3] *University of Oldenburg, Oldenburg, Germany*

[4] *Cluster of Excellence "Hearing4all", Oldenburg, Germany*

Hearing is multidimensional. It affects the whole body and yet it is still an open question whether and how general factors of everyday life are affected by the use of modern hearing aids (HA) with different signal processing options. This study addressed, therefore, the question to what extent HA may shape the HA users' everyday life. Accordingly, the behavior of N=22 HA users and non-users was observed experimentally using a theory-based ethnographic research design that comprises written reports and several steps of theorizing and reasoning. Data were collected in two specific everyday life situations (road traffic and restaurant) and by three modes (unaided, omnidirectional, and directional microphone mode). The analytical results of the ethnographical studies were summarized and used for testing hypotheses in an advanced laboratory with virtual audio-visual environments reproducing the same everyday life situations. Different typical behavior patterns were identified by means of fieldnotes, indicating that hearing impaired users with the first experience of HA provision showed comparatively expressive orientation reactions towards spatial sound sources. The behavior analyses were partly confirmed by questionnaire data. The analytical results led to first suggestions and improvements for the ongoing (re-)creation of virtual audio-visual scenes.

INTRODUCTION

Audiological research relates primarily to a calculable space and thus refers accordingly to digital space-time assumptions (e.g., Lindemann, 2014; Bentler, 2005; Picou *et al.*, 2014; Ricketts and Henry, 2002). In scientific research of medical devices (e.g., HA), the emphasis is on measuring behavior patterns to explain the benefit of investigated technologies. Therefore, different signal processing options are evaluated by quantified body movements (Brimijoin *et al.*, 2014; Hendrikse *et al.*, 2017).

*Corresponding author: r.paluch@hoerzentrum-oldenburg.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

However, the question of the everyday life benefit of HA and the ecological validity of laboratory settings remains open (see Meis *et al.*, 2017 in this volume). The emphasis is on whether and how HA improve the communication abilities, the social interaction, and participation (Ihde, 2007; 2016; Lindemann, 2014; Plessner, 1975; Zahnert, 2011).

The focus of this mixed methods study was, therefore, the enhancement of the ecological validity of outcome research in audiology and the evaluation of HA in more realistic settings. The goal was to detect differences in user behavior between different everyday life settings to improve the ecological validity in virtual audio-visual laboratory environments. One research method used for this purpose was the ethnographical approach. It is a radically qualitative oriented research method, which helps to understand the behavior of users in acoustically complex everyday environments to develop new outcomes methods and diagnoses in audiology in the long run (Paluch *et al.*, 2015; 2017). These qualitative data were combined with quantitative data.

A further inquiry is planned, e.g., a confrontation with virtual audio-visual scenes in an advanced laboratory (Grimm *et al.*, 2015; 2016). First pilot studies were completed in August 2017. Extensive laboratory evaluations are planned for September and October 2017.

METHOD

For the mixed methods study a specific setup was chosen. Data were gained in (1) a road traffic situation and (2) a restaurant situation in the field (i.e., (1) two different streets in the city of *Oldenburg* and (2) the university cafeteria).

Thus, a street was selected with a high traffic density, i.e., with many pedestrians, cyclists, cars, buses, trucks, etc., on both sides of the road. In addition, environmental sounds such as crows, magpies, dogs, the rustling of trees, etc., were present. The other chosen street was in comparison to the first one a quiet environment with a lower traffic density. The environmental sounds were perceived in the quiet street more clearly, since there was less traffic noise.

The cafeteria situation, on the other hand, was a typical dining situation, where the background noise was characterized by conversations, the rattling of cutlery, and the sounds of the cash desk as well as the kitchen.

Furthermore, for the study three provision conditions were chosen: Subjects were (1) unaided and/or (2) aided with HA with omnidirectional and (3) directional microphone modes. The *Phonak Audéo V90-312* HA were used for all subjects. These were fitted in accordance with the *Adaptive Phonak Digital* fitting formula (Latzel, 2013). All HA were receiver-in-canal (RIC) models with open domes.

The qualitative data were analyzed with relation to the Grounded Theory (GT) approach (Glaser and Strauss, 1967). So the method of data interpretation used here corresponded with a theory-based variant of the GT methodology (Corbin and Strauss, 1990). The interpretation was not based on the traditional GT approach, in which codes should only be interpreted with reference to data (Paluch *et al*., 2015; 2017). In contrast, a GT analysis was carried out with regard to certain theoretical assumptions (Matsuzaki and Lindemann, 2016, p. 503). This approach included positivistic assumptions about auditory spatial awareness and behavior patterns (Blesser and Salter, 2009). Thus, different typical behavior patterns in form of head movements and torso shifts were identified by means of fieldnotes. In addition, quantitative data were collected by questionnaires during the ethnographic walks.

N=22 study participants (age range from 51 to 72 yrs.; mean age = 66.6 ± 4.90 yrs.; 54% female) were recruited for the ethnographical walks. Three groups were involved: Group I included eight listeners with normal hearing (NH) according to WHO (2004); group II were seven unaided listeners with hearing impairment (HI) and a mild hearing loss (HL), who completed the walks in an unaided as well as aided condition during different study trails; and group III were seven aided listeners with mild to moderate HL. Group III only tested the aided condition.

RESULTS

The qualitative outcomes of the ethnographical walks can be summarized as follows: Subjects with NH demonstrated mainly civil inattention in the road traffic situation (Goffman, 1963, pp. 83-88) and were talkative in the restaurant situation. Unaided subjects with HI showed equally unobtrusive behavior patterns in the street and in the cafeteria. However, they had difficulties in understanding questions or sentences during talks. Furthermore, first-time HA users strongly related to the environment via body movements and were reserved in conversations. Experienced HA users, finally, lied between the subjects with NH and the unaided subjects: in the street they behaved as subjects with NH (e.g., civil inattention) and during conversations as unaided subjects (e.g., limited speech intelligibility). Nonetheless, they were also loquacious during conversations. For a detailed qualitative analysis and results of the ethnographical walks see Paluch *et al*. (2017).

Additionally, subjects with NH, unaided subjects with HI, first-time HA users, and experienced HA users were compared via quantitative data. Thereby, the behavior analyses were partly confirmed by questionnaire data. The quantitative results of the everyday life setting questionnaires are presented as box plots (see Figs. 1-3). All items were rated by subjects on a 5-point scale regarding mainly the perception of traffic sources, such as trucks/buses, cars and bicycles, and the perception of speech in the street. The questionnaires related to the cafeteria focused on speech and dining sounds.

Figure 1 shows the results of the road traffic questionnaires regarding volume perception of the subjects. First-time users of HA with omnidirectional microphone modes experienced their environment louder than in the unaided condition. Especially sound sources of objects like trucks, buses, cars, and bikes were experienced louder. The perception of speech was also affected, but not as much as by motor vehicles. This could be an explanation for the strong relation of head movements or torso shifts to sound sources. It could also be an indication of how the adaptation to HA was difficult at first.

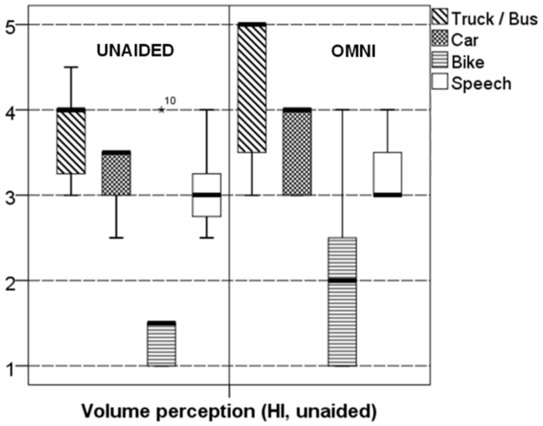


Fig. 1: Results of road traffic questionnaires rating volume perception, scale range from 1 to 5. 1 = too soft, 3 = adequate, 5 = too loud. Group II (N=7). Comparison of unaided conditions and omnidirectional conditions of first-time HA users. The box plots show the median, 25th and 75th quartiles, and outliers.

In other studies (Appleton and König, 2014; Latzel, 2015), it has been pointed out that better speech intelligibility is a crucial aspect of HA. Even though voices were not processed in the same way as technical objects, a louder perception of motor vehicles could lead to distraction towards the understanding of spoken words.

As a remark, it is interestingly to note that in the unaided condition bikes were too quiet for the unaided subjects. The use of HA allowed perceiving sound sources such as bikes more adequate, although subjects reported that bikes had to be quiet.

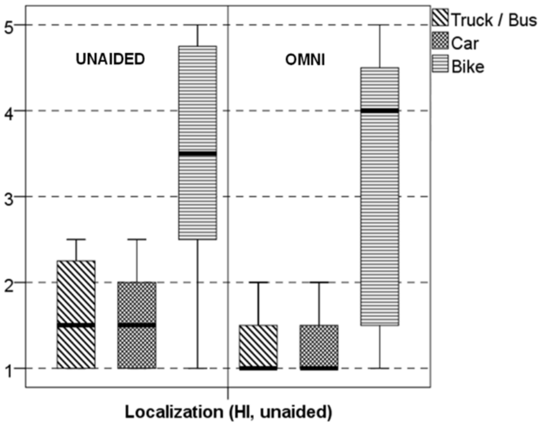


Fig. 2: Results of road traffic questionnaires rating localization, scale range from 1 to 5. 1 = very good, 3 = moderate, 5 = very poor. Group II (N=7). Comparison of unaided conditions and omnidirectional conditions of first-time HA users. The box plots show the median, 25th and 75th quartiles, and outliers.

Moreover, first-time HA users localized sound sources on average better in the aided condition than in the unaided condition. An exception was the localization of bikes, which had been on average better without HA. One explanation could be that sound sources were masked by HA. Besides, bikes were usually experienced in street situations with other traffic participants (e.g., trucks, buses, and cars).

This also confirms the assumption that first-time HA users refer strongly to their environment with body movements. If the direction of sound sources can be localized better, it is likely that this will also be reflected in the movement patterns. Certain sound sources may not have been explicitly perceived for a long time due to increasing hearing loss, so the subjects clearly refer to them with body movements if they hear them appropriately again.

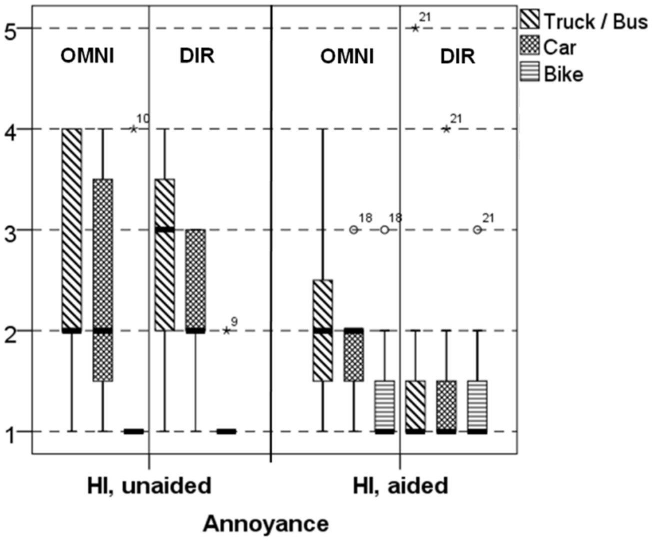


Fig. 3: Results of road traffic questionnaires rating annoyance perception, scale range from 1 to 5. 1 = not at all, 3 = moderately annoyed, 5 = highly annoyed. Group I and II (N=14). Comparison of first-time HA users and experienced HA users with omnidirectional and directional microphone modes. Significant values for annoyance by trucks/buses ($p < 0.05$, Wilcoxon). The box plots show the median, 25th and 75th quartiles, and outliers.

Furthermore, the quantitative data show that experienced HA users were less annoyed by environmental sound sources. Mainly first-time HA users were annoyed by different motor vehicles. This is a further explanation of why an explicit behavior occurred with first-time HA users. They were not only able to locate the sound sources better; they also experienced them too loudly and were thus more annoyed by them. For example, a woman during the walk was referring to a warning signal of a railway crossing gate and a crow with a torso movement when she perceived both. According to her, she was annoyed by the sounds due to the aided condition. Lastly, she looked at a car with a clear movement of her head, because the car was for her almost as loud as a train (Paluch *et al.*, 2017).

Interestingly, omnidirectional microphone modes tend to increase the annoyance of sound sources relative to directional microphone modes. In a further study it is going to be examined in the laboratory whether and how the annoyance manifests itself regarding to different signal processing options. Probably, the directional microphone modes mask more sound sources. It remains open why bikes tend to annoy first-time HA users less.

Significant results, however, were almost not found in the plots ($p > 0.05$, Wilcoxon, see figs. 1-2). Only differences regarding trucks/buses shown in Fig. 3 were significant ($p < 0.05$, Wilcoxon, see Fig. 3). One reason for this is the limited number of subjects that participated in the study. Nonetheless, the tendencies of the box plots are in line with the qualitative results, which showed different experiences of the environment due to usage of HA (Paluch *et al.*, 2017).

CONCLUSIONS

In this paper a mixed methods study was reported, which included both qualitative data and quantitative data. A first outcome of the study was the possibility to show whether and how different microphone modes of HA influence subjects' behavior. It has emerged that quantitative data also support the view that first-time HA users differ notably in their behavior patterns.

The volume perception and the localization of first-time HA users were compared with and without HA in the street situation. In addition, it was shown how the annoyance decreases in regard to sound sources by experienced HA users. The better localization plus the increased sensation of the volume and the annoyance could be a reason why clear body movements of first-time HA users were observed (e.g., strong torso shifts). It should be verified whether these outcomes can also be reproduced in an advanced laboratory with virtual audio-visual scenes.

Finally, the combination of qualitative and quantitative data leads to the assumption that the habituation to loudness decreases the noticeable body movements (Paluch *et al.*, 2017). Probably a habituation to HA contributes to behavior patterns accordingly to shared social expectations (e.g., civil inattention). It would be of further interest to study how long people need to get used to HA and how their behavioral patterns differ over time (e.g., by head movements and torso shifts).

OUTLOOK

Based on the guided walk from phase 1, a virtual audio-visual environment was developed. This virtual environment partly simulates existing areas of the city of *Oldenburg*, and partly adds a fictive area with lower urban density as well as a cafeteria.

In the advanced laboratory the test design is repeated equally to the ethnographic walks. Subjects will experience going along the street in the laboratory and have to answer closed-ended questions at bus stops similar to the first study design. Also in the cafeteria there will be questions relating to the stories told by virtual characters. Thus, the questionnaires can be directly compared with one another.

Ultimately, qualitative field research about the laboratory situations will be conducted and the subjects are going to be recorded on video for analyses of their behavior (VIB-AICRAS©, Paluch *et al.*, 2015). In combination with qualitative interviews this is intended to test the ecological validity of the advanced laboratory and to work out how intensive the immersion by subjects in the laboratory is.

ACKNOWLEDGEMENTS

This work was supported by the Cluster of Excellence EXC 1077/1 “Hearing4all” funded by the German Research Council (DFG).

REFERENCES

Appleton, J., and König, G. (**2014)**. “Improvement in speech intelligibility and subjective benefit with binaural beamformer technology,” Hear. Rev., **21**, 40-42.

Bentler, R.A. (**2005**). “Effectiveness of directional microphones and noise reduction schemes in hearing aids: A systematic review of the evidence,” J. Am. Acad. Audiol., **16**, 473-484. doi: 10.3766/jaaa.16.7.7

Blesser, B., and Salter, L.-R. (**2007**). *Spaces Speak, Are You Listening? Experiencing Aural Architecture*. Cambridge, MA: MIT-Press.

Brimijoin, W.O., Whitmer, W.M., McShefferty, D., and Akeroyd, M.A. (**2014**). “The effect of hearing aid microphone mode on performance in an auditory orienting task,” Ear Hearing, **35**, 204-212. doi: 10.1097/AUD.0000000000000053

Corbin, J., and Strauss, A.L. (**1990**). “Grounded theory research: procedures, canons and evaluative criteria,” Z. Soziol., **19**, 418-427. doi: 10.1007/BF00988593

Glaser, B.G., and Strauss, A.L. (**1967**). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.

Goffman, E. (**1963**). *Behavior in Public Places: Notes on the Social Organization of Gatherings*. New York: Free Press.

Grimm, G., Luberadzka, J., Herzke, T., and Hohmann, V. (**2015**). “Toolbox for acoustic scene creation and rendering (TASCAR): Render methods and research applications,” Proceedings of the Linux Audio Conference, Mainz, Germany. Edited by F. Neumann.

Grimm, G., Kollmeier, B., and Hohmann, V. (**2016**). “Spatial acoustic scenarios in multichannel. Loudspeaker systems for hearing aid evaluation,” J. Am. Acad. Audiol., **27**, 557-566. doi: 10.3766/jaaa.15095

Hendrikse, M.M.E., Llorach, G., Grimm, G., and Hohmann, V. (**2017**). “Influence of visual cues on head and eye movements during listening tasks in multi-talker audiovisual environments with animated characters,” Manuscript submitted for publication.

Ihde, D. (**2016**). *Acoustic Technics*. Lanham: Lexington Books.

Ihde, D. (**2007**). *Listening and Voice: Phenomenologies of Sound*. 2nd edition. Albany: State University of New York Press.

Latzel, M. (**2013**). "Compendium 4 – Adaptive Phonak Digital (APD)," Phonak Compendium (http://www.phonakpro.com/com/b2b/en/evidence.html).

Latzel, M. (**2015**). "Adaptive StereoZoom – adaptive behavior improves speech intelligibility, sound quality and suppression of noise," Phonak Field Study News.

Lindemann, G. (**2014**). *Weltzugänge. Die mehrdimensionale Ordnung des Sozialen*. Weilerswist: Velbrück Wissenschaft.

Matsuzaki, H., and Lindemann, G. (**2016**). "The autonomy-safety-paradox of service robotics in Europe and Japan: a comparative analysis," AI & Soc., **31**, 501-517. doi: 10.1007/s00146-015-0630-7

Meis, M., Krueger, M., Gebhard, M., v. Gablenz, P., Holube, I, Grimm, G., and Paluch, R. (**2017**). "Development and application of a code system to analyse behaviour in real life listening environments." in Proc. ISAAR, **6**, 31-38.

Paluch R., Latzel, M., and Meis, M. (**2015**). "A new tool for subjective assessment of hearing aid performance: Analyses of interpersonal communication," Proc. ISAAR, **5**, 453-460.

Paluch, R., Krueger, M., Grimm, G., and Meis, M. (**2017**). "Moving from the field to the lab: Towards ecological validity of audio-visual simulations in the laboratory to meet individual behavior patterns and preferences," in 20. Jahrestagung der Deutschen Gesellschaft für Audiologie. CD-ROM. ISBN 978-3-9813141-7-5.

Picou, E.M., Aspell, E., and Ricketts, T.A. (**2014**). "Potential benefits and limitations of three types of directional processing in hearing aids," Ear Hearing, **35**, 339-352. doi: 10.1097/AUD.0000000000000004

Plessner, H. (**1975**). *Die Stufen des Organischen und der Mensch. Einleitung in die philosophische Anthropologie*. 3rd edition. Berlin/New York: de Gruyter.

Ricketts, T.A., and Henry, P. (**2002**). "Evaluation of an adaptive, directional-microphone hearing aid," Int. J. Audiol., **41**, 100-11. doi: 10.3109/14992020209090400.

WHO (**2004**). *International Statistical Classification of Diseases and Related Health Problems*. World Health Organization, 2004.

Zahnert, T. (**2011**). "The differential diagnosis of hearing loss," Dtsch. Arztebl. Int., **108**, 433-444. doi: 10.3238/arztebl.2011.0433

Preliminary investigation of the categorization of gaps and overlaps in turn-taking interactions: Effects of noise and hearing loss

A. JOSEFINE SØRENSEN[1,2,*], ADAM WEISSER[2], AND EWEN N. MACDONALD[1]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *GN ReSound A/S, Ballerup, Denmark*

Normal conversation requires interlocutors to monitor the ongoing acoustic signal to judge when it is appropriate to start talking. Categorical thresholds for gaps and overlaps in turn-taking interactions were measured for normal-hearing and hearing-impaired listeners in both quiet and multitalker babble (+6 dB SNR). The slope of the categorization functions were significantly shallower for hearing impaired listeners and in the presence of background noise. Moreover, the categorization threshold for overlaps increased in background noise.

INTRODUCTION

In normal conversation, talkers take turns speaking in a manner that is flexible (i.e., not organized in advance) and often rapid. Over the decades since the seminal paper by Sacks *et al.* (1974) was published, many models have been proposed to explain why the switching between interlocutors remains fluid with rapid transitions between speakers. In general, all of these models involve interlocutors monitoring aspects of the ongoing acoustic signal to judge when the current talker will stop or has stopped talking (for further discussion and review including the distribution of acoustic overlaps and gaps in normal turn taking see Heldner and Edlund, 2010; Levinson and Torreira, 2015). Background noise and hearing impairments are known to reduce many aspects of auditory perception (e.g., speech intelligibility) and can reduce cognitive spare capacity (i.e. resources for higher-level processing of speech; Rudner and Lunner, 2014). Thus, it is possible that noise and/or hearing loss may alter normal communication dynamics by altering the perception of acoustic cues monitored by interlocutors (for a discussion of these cues, see Gravano and Hirschberg, 2011).

As a first step towards investigating this, the present study focused on the perception of turn-taking interactions in a manner similar to Heldner (2011). Specifically, normal-hearing and hearing-impaired listeners were asked to listen to pre-recorded turn-taking interactions and categorize whether the interactions were perceived as overlaps (i.e., the second talker started before the first had finished), gaps (i.e., there was

[*] Corresponding author: ajs@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

silence/pause between when the first talker stopped and the second talker started), or neither, when the acoustic interval between the speech offset of the first talker and the speech onset of the second talker was systematically varied from −500 ms to 500 ms.

METHOD

The participants in this study were 24 normal-hearing (mean age 36 years) and 7 hearing-impaired (mean age 72 years) native Danish listeners. The hearing impaired group had primarily moderate hearing loss (see Fig. 1). The procedure was approved by the Science-Ethics Committee for the Capital Region of Denmark. A speech corpus was developed based on the dialogue from a Danish translation of the play "Educating Rita" (Author: Willy Russell, translator: Riri Ianke Firing), consisting of a conversation between a man and a woman. In addition to portions of the script, 11 turn-taking interactions were created from everyday conversation topics. Overall, the corpus consisted of 85 turn-taking interactions, 43 where the woman starts and the man takes over and 42 in the opposite order. Of these turn-taking interactions, 44 were in the form of a question-answer, 7 in the form of statement-question, 29 in the form statement-answer, and 5 in the form of statement-statement.

The corpus was recorded in an audiometric sound booth, with the male and female native Danish talkers standing approximately 1 m apart from each other. The speech was recorded using Cubase Elements 7 (Steinberg) with microphones (AKG C451B) with pop filters placed approximately 10 cm from each talker and connected to a Fireface UFX (RME) soundcard. The talkers were instructed to read the sentences as naturally as possible, but to pause for a second or two during the turn-taking interaction to avoid crosstalk. Each turn-taking interaction was repeated twice. The recordings were reviewed to select the best utterances for each turn-taking interaction. These were the utterances that were relatively constant in level and for which the interaction sounded the most natural to the first author. After selecting the best utterances, they were segmented by hand using Praat (Boersma and Weenink, 2002). The mean duration of each utterance was 1.35 s (standard deviation 0.49 s) and varied between 0.5 and 2.95 s. After editing, the sentences were normalized to have the same RMS level.

Categorization of overlaps, gaps, and no-gap-no-overlap was obtained using a 3-alternative forced-choice paradigm for 17 acoustic intervals, ranging from -500 to 500 ms. Here, we use the convention that a negative acoustic interval corresponds to an overlap (i.e., the onset of speech from the second talker occurs before the offset of speech from the first talker) and a positive interval corresponds to a silence. We used a three-category procedure rather than conducting a two-category procedure twice (i.e., gap vs. no gap; overlap vs. no overlap) to ensure listeners would maintain the same internal criteria they use when listening to regular conversations. On each trial, participants listened to a turn-taking interaction that was composed from an edited pair of utterances that were shifted in time and then mixed to create the desired

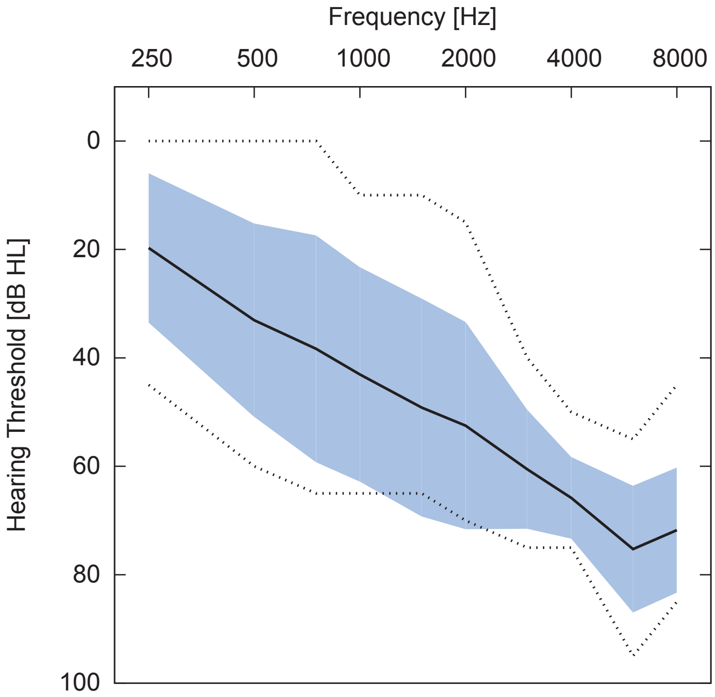


Fig. 1: Audiometric thresholds of the hearing-impaired listeners. The solid black line indicates the mean hearing threshold, the shaded region indicates one standard deviation, and the dotted lines indicate minimum and maximum measured thresholds.

acoustic interval. After listening to a turn-taking interaction, the participant pressed one of three buttons on a computer screen. The buttons were labeled with Danish text corresponding to the English terms overlap, gap, and no-gap-no-overlap. For each listener, five judgments of each interval in both quiet and in a background of seven-talker English babble[1] (mixed to achieve an SNR of +6 dB) were obtained. The order in which the acoustic intervals were presented was randomized across trials. Thus, the utterances judged for each acoustic interval varied across listeners. Half of the listeners judged turn-taking interactions in quiet before judging turn-taking interactions in babble. The other half judged the two conditions in the opposite order.

Stimuli were presented over headphones (Sennheiser HD 650) in a sound booth. For the normal hearing listeners, the stimuli were presented at 65 dB SPL. To compensate for reduced audibility, the stimuli (i.e., speech and noise in the babble condition and speech alone in the quiet condition) was further amplified using the Cambridge Formula (Moore and Glasberg, 1998) for each individual hearing impaired listener. Overall, it took each listener approximately 15-20 minutes to complete the experiment.

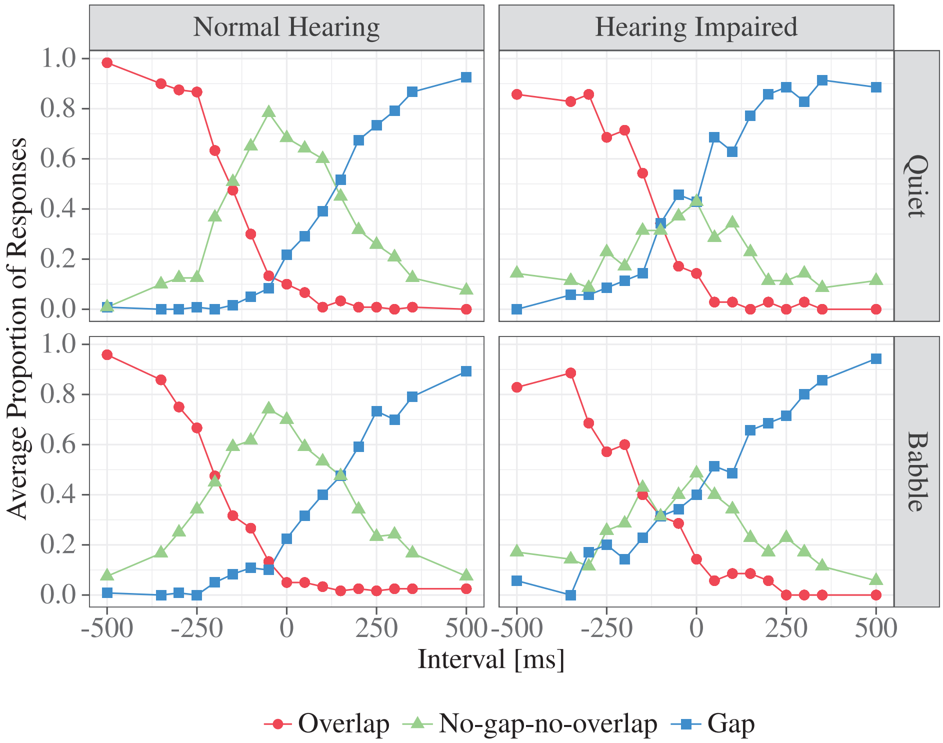


Fig. 2: The average proportion of responses of overlap (squares), gap (circles), and no-gap-no-overlap (triangles) as a function of acoustic interval. A negative acoustic interval indicates acoustic overlap (i.e., an interval where both talkers are speaking) while a positive interval indicates an acoustic gap. The top and bottom panels present the results for the normal and impaired listeners, respectively. The left and right panels present results when the turn-taking stimuli were presented in quiet and in multitalker babble (with an SNR of +6 dB), respectively.

RESULTS

The average proportion of responses of overlap, gap, and no-gap-no-overlap as a function of acoustic interval is plotted in Fig. 2. To estimate categorical thresholds, cumulative Gaussian functions were fitted to individuals' data after smoothing using a simple moving average of responses from three neighbouring acoustic intervals. The average mean (which is used as the estimated category threshold for that individual) and standard deviation of the fitted Gaussian functions are plotted in Fig. 3. Note that in this figure, the sign of the means for overlap results have been inverted to better compare their magnitudes with the results for gap categorization.

Repeated measures ANOVAs were conducted on the fitted means of categorizing overlaps and gaps, with background condition (quiet vs. babble) as within-subject

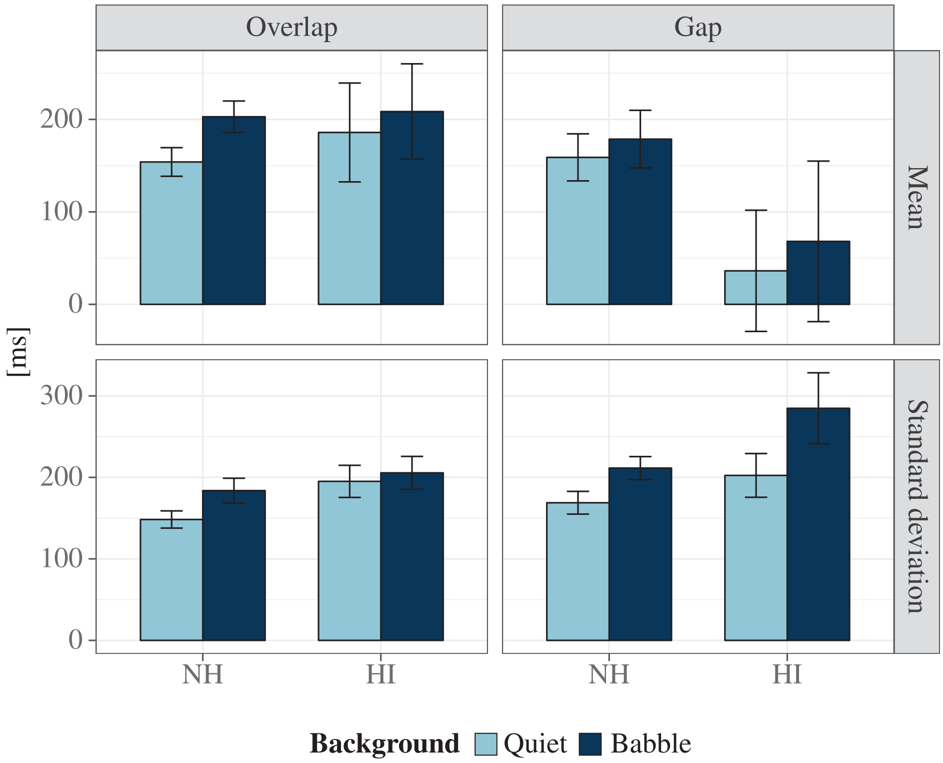


Fig. 3: Average mean (top panels) and standard deviation (bottom panels) of cumulative Gaussian functions fitted to individual's proportion of responses. The left and right panels present results for overlap and gap categorizations averaged within groups of normal hearing (NH) and hearing impaired (HI) listeners respectively. Here, the sign of the means for overlap results have been inverted to better compare their magnitudes with the results for gap perception. The bars indicate one standard error.

and hearing status (normal hearing vs hearing impaired) as between-subjects factors. For the overlap categorization, only the main effect of background condition was significant [$F(1,29) = 5.782, p < 0.023$]. For the gap categorization there was no significant effect of either background condition, or hearing status, but there was a trend for the effect of hearing status [$F(1,29) = 3.329, p < 0.078$]. None of the interactions were significant.

A repeated measures ANOVA on the fitted standard deviations (slopes) of the categorization functions for overlaps and gaps was conducted with background condition (quiet vs. babble) and category (gap vs. overlap) as within-subject factors and hearing status (normal hearing vs hearing impaired) as between-subjects factor. Main effects of background condition [$F(1,29) = 8.455, p < 0.01$], category [$F(1,29) =$

$11.565, p < 0.01$], and hearing status were significant [$F(1,29) = 4.524, p < 0.05$]. None of the interactions were significant.

When compared with the results from the normal hearing listeners, the proportion of responses for the no-gap-no-overlap from the hearing-impaired listeners is greatly reduced (see Fig. 2). To quantify this, the average proportion of responses of no-gap-no-overlap for acoustic intervals ranging from -150 to 150 ms (i.e., acoustic intervals between the average categorical thresholds of overlap and gap in quiet by normal-hearing listeners) were calculated for each individual in each condition. A repeated measures ANOVA with background condition as within- and hearing status as between-subjects factors confirmed a main effect of group [$F(1,29) = 9.174, p < 0.005$] such that the hearing-impaired listeners used the no-gap-no-overlap category less frequently than the normal-hearing listeners.

DISCUSSION

Virtually all models of turn taking in conversation involve interlocutors monitoring the ongoing acoustic signal. Thus, it is possible that the presence of background noise or hearing loss could influence conversational dynamics by disrupting the perception of acoustic cues monitored by interlocutors. In the present study, categorical thresholds for perceiving a turn-taking interaction with an acoustic overlap as an overlap increased when listening in the presence of a background noise. This effect was observed, even though the level of the multitalker babble was relatively low (SNR of +6 dB) and should not have decreased the intelligibility of the turn-taking interaction utterances. The slopes of the categorization functions of the hearing-impaired listeners were shallower than those of the normal-hearing listeners.

In a previous study, Heldner (2011) measured gap and overlap thresholds separately using a two-alternative forced-choice procedure (i.e., turn-taking interactions with positive acoustic intervals were judged as either gap or no-gap, whereas turn-taking interactions with negative acoustic intervals were judged as either overlap or no-overlap). In the present study, the results of the normal-hearing listeners suggested they perceived three clear categories (see Fig. 2) and the categorical thresholds were consistent with the detection thresholds reported by Heldner (2011). The hearing-impaired listeners exhibited a much lower proportion of responses for no-gap-no-overlap than the normal hearing listeners for small acoustic intervals (i.e., between -150 and 150 ms). Thus, it is not clear if this between-group difference is because the hearing-impaired listeners perceived only two categories (i.e., either gap or overlap) or the task was too difficult (i.e., choosing between three rather than two categories). It should be noted that the hearing impaired listeners were much older than the normal hearing listeners. Thus, the differences observed could also be due to aging effects (both in auditory perception and cognition) rather than or in addition to hearing loss. The hearing-impaired group's small detection thresholds for gaps was therefore not attributed to a higher sensitivity, but rather an effect of poor categorization as indicated by their significantly shallower slopes and their inability to use the no-gap-no-overlap category. A less cognitively demanding follow up study using two 2-alternative forced-choice paradigms, as in Heldner (2011), is needed to help disentangle these effects.

The corpus used in this study was recorded from scripted dialog rather than the spontaneous conversations used in previous studies into the perception of gaps in turn-taking interaction (e.g., Walker and Trimboli, 1982; Heldner, 2011). Further, in the present study, the recorded turn-taking interactions were edited to systematically vary the acoustic interval. In contrast, previous studies have presented the original acoustic recordings, which contained a wide range of acoustic intervals. Nevertheless, the thresholds measured in this study are consistent with these previous studies. For normal-hearing listeners, the average threshold for categorizing a gap in quiet was approximately 160 ms. This is close to the thresholds of 120 ms reported by Heldner (2011) and 180 ms estimated[2] from the data published by Walker and Trimboli (1982). Similarly, for normal-hearing listeners, the average threshold for categorizing an overlap in quiet was approximately 155 ms, which is close to the 120 ms reported by Heldner (2011) and indicates the same symmetry between the gap and overlap thresholds (i.e., a duration of about one syllable in order to perceive it as either a gap or an overlap).

CONCLUSION

In the present study, thresholds for perceiving overlaps and gaps were obtained for both normal-hearing and hearing-impaired listeners in both quiet and noise. The results indicated that the threshold for perceiving an overlap increased in the presence of background noise. Furthermore, the categorization functions for both gaps and overlaps were shallower in the presence of background noise and for hearing impaired compared to normal hearing listeners. The gap categorization was very different for the hearing-impaired group, and it is not clear if it is an effect of the paradigm or some higher order processing. Thus it is suggested that a follow-up study using a different paradigm is needed to explore this. In conclusion the presence of background noise influences the perception of acoustic cues used to judge turn-taking interaction. This suggests that comparing turn taking in quiet and noisy conditions may be useful for validating proposed models of turn taking, particularly in relation to language processing (e.g., Levinson and Torreira, 2015). Furthermore, it suggests that parts of the language processing system is affected even at SNRs well above the typical SRT, and that these effects are not captured by standard speech intelligibility tests.

ACKNOWLEDGEMENTS

This experiment was conducted as a bachelor project by the first author in collaboration with GN ReSound. The authors would like to thank Andreas Schousboe from GN ReSound A/S for his help in facilitating the project, as well as Johan Gabriel Selby, Lotte Hagen Hernvig, and Brent Kirkwood, all from GN ReSound A/S, for their help with the experiment.

ENDNOTES

[1] Seven-talker babble with both male and female voices was created from an unpublished dialogue and trialogue corpus created previously by the second author. The levels of all talkers in the babble were adjusted to have the same RMS.

[2] The categorical threshold corresponding to 50% proportion of responses was estimated from a linear interpolation of the proportion of responses for 100 and 200 ms (42% and 52%, respectively).

REFERENCES

Boersma, P., and Weenink, D. (**2002**). "Praat, a system for doing phonetics by computer," Glot Int., **5**, 341-345.

Gravano, A., and Hirschberg J. (**2011**). "Turn-taking cues in task-oriented dialogue," Comput. Speech Lang., **25**, 601-634. doi: 10.1016/j.csl.2010.10.003

Heldner, M., and Edlund, J. (**2010**). "Pauses, gaps, and overlaps in conversations," J. Phonetics, **38**, 555-568. doi: 10.1016/ j.wocn.2010.08.002

Heldner, M. (**2011**). "Detection thresholds for gaps, overlaps, and no-gap-no-overlaps," J. Acoust. Soc. Am., **130**, 508-513. doi: 10.1121/1.3598457

Levinson, S.C., and Torreira, F. (**2015**). "Timing in turn-taking and its implications for processing models of language," Front. Psychol., **6**, 731. doi: 10.3389/fpsyg.2015.00731

Moore, B.C.J., and Glasberg, B.R. (**1998**). "Use of loudness model for hearing-aid fitting. I. Linear hearing aids," Br. J. Audiol., **32**, 317-335. doi: 10.3109/03005364000000083

Rudner, M., and Lunner, T. (**2014**). "Cognitive spare capacity and speech communication: A narrative overview," BioMed Res. Int., 869726. doi: 10.1155/2014/869726

Sacks, H., Schegloff, E.A., and Jefferson, G. (**1974**). "A simplest systematics for the organization of turn-taking for conversation," Language, **50**, 696-735. doi:10.2307/412243

Walker, M.B., and Trimboli, C. (**1982**). "Smooth transitions in conversational interactions," J. Soc. Psychol., **117**, 305-306. doi: 10.1080/00224545.1982.9713444

Using fNIRS to study audio-visual speech integration in post-lingually deafened cochlear implant users

XIN ZHOU[1,2,*] HAMISH INNES-BROWN[1,2], AND COLETTE MCKAY[1,2]

[1] *Bionics Institute, Melbourne, Australia*

[2] *Medical Bionics Department, University of Melbourne, Melbourne, Australia*

The aim of this experiment was to investigate differences in audio-visual (AV) speech integration between cochlear implant (CI) users and normal hearing (NH) listeners using behavioural and functional near-infrared spectroscopy (fNIRS) measures. Participants were 16 post-lingually deafened adult CI users and 13 age-matched NH listeners. Participants' response accuracy in audio-alone (A), visual-alone (V), and AV modalities were measured with closed-set /aCa/ non-words and with open-set CNC words. AV integration was quantified by using a probability model and a cue integration model that predicted participants' AV performance given minimal or optimal integration. Using fNIRS, brain activation was measured when listening to or watching A, V, or AV speech with or without multi-talker babble. For fNIRS, evidence of AV integration was measured using the *principle of inverse effectiveness (PoIE)* model (comparing the difference in activation in two brain regions between A and AV modalities in quiet and noise conditions). Behavioural AV integration was similar in the two groups for CNC words but poorer in the CI group compared to NH group for consonant perception. Our fNIRS data did not demonstrate any AV integration in either NH listeners or CI users, by testing the PoIE.

INTRODUCTION

Neuroplasticity and changes in speech processing strategies have been reported in cochlear implant (CI) users (see review by Anderson *et al.*, 2016). These changes are thought to be due to hearing loss and increased reliance on lip-reading before implantation, and the introduction of distorted hearing input after cochlear implantation. In this study, the audio-visual (AV) integration ability of CI users was of special interest. Rouger *et al.* (2007) used a cue integration model to quantify AV integration ability and claimed that CI users had better AV integration ability than normal listening (NH) listeners when the latter were listening to vocoded speech. Using electroencephalography (EEG) measures, Schierholz *et al.* (2015) investigated changes in response in auditory cortex of CI users and NH listeners when visual-alone (V) cues were added to audio-alone (A) object stimuli compared to the response in A condition. Changes of response in auditory cortex in that study were interpreted as the amount of AV integration. Compared to the older NH listeners, Schierholz *et al.*

*Corresponding author: xzhou@bionicsinstitute.org

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
ISBN: 978-87-990013-6-1.

(2015) found that older CI users had larger AV integration responses in auditory cortex. Results from the above two studies suggested that CI users may have better AV integration ability and more neural AV integration response than NH listeners.

We investigated whether functional near-infrared spectroscopy (fNIRS) could reveal AV integration in experienced CI users and we hypothesized that CI users have increased AV speech integration compared to age-matched NH listeners, using both behavioural and fNIRS measures. To reveal AV integration in fNIRS measures, we used the principle of inverse effectiveness (PoIE) first found in a study of Meredith and Stein (1983). This rule assumes that when V cues are added to A stimuli, enhancement of neural responses should be greater when the effectiveness of stimuli in each modality is low compared to high. The PoIE was derived using the dynamic response of multisensory neurons to stimuli of different effectiveness levels (Perrault *et al.*, 2005) and has also been applied to in functional magnetic resonance imaging (fMRI) and EEG studies (Holmes, 2007; James *et al.*, 2012). In this study, we investigated two regions of interest (ROIs), i.e., left superior temporal sulcus (LSTS) and left occipital cortex (LOC) where the PoIE has been previously demonstrated in NH listeners using fMRI (Laurienti *et al.*, 2005; Stevenson *et al.*, 2009). Using A, V, and AV speech stimuli, we tested the PoIE of fNIRS responses in the 2 ROIs of NH listeners and CI users, separately. We hypothesised that compared to NH listeners, CI users would show larger fNIRS measures of AV integration activation in at least one of the two ROIs.

METHOD

Participants

Sixteen post-lingually deafened adult CI users and 13 aged-matched NH listeners were recruited for this study. All the participants were native English speakers, with no history of diagnosed neurological disorder, and with normal or corrected-to-normal vision. All CI users had a right-ear implant and experience of using the CI for more than 12 months. The ages of participants in the CI and old NH group ranged from 45 to 82 (mean ± SD: 69.0 ± 9.1) and 52 to 76 years (mean ± SD: 64.9 ± 7.1), respectively, with no significant mean difference in age ($t = 1.38$, $p = 0.179$). To develop the cue integration model for AV speech integration, an additional 16 young NH listeners were also recruited, with ages ranging from 21 to 39 years (mean ± SD: 28.7 ± 5.3). All participants provided their written informed consent.

Speech stimuli

Two types of speech stimuli were used to measure AV integration ability. The first type were 12 consonant tokens in the form of /aCa/, with the 12 consonants being 'B', 'D', 'F', 'G', 'K', 'M', 'N', 'P', 'S', 'T', 'V', 'Z'. The second type were Consonant-Nucleus-Consonant (CNC) words (Peterson and Lehiste, 1962). For all the consonant and CNC word stimuli, the A and V components of video recordings were separated. The levels of all the auditory consonant/CNC stimuli were normalized to the same root mean square (RMS) level.

Speech tests and AV integration ability

Speech tests were conducted in A, V, and AV modalities, using software Max/Msp (https://cycling74.com). Visual stimuli were presented on an LCD monitor at a 1.5-m distance and in front of the participant. Auditory stimuli were delivered to the right-ear processor of CI users via direct audio input accessory or the right-side insert earphone of NH listeners. The level of sound directly input to the CI processor or earphone was set equivalent to 65 dBA (F_{max}). Speech sounds in the A and AV modalities were presented with babble noise at a participant-dependent signal-to-noise ratio (SNR), at which each participant could achieve 50% of the consonants or 50% of the phonemes in the CNC words correct in the A condition (denoted SNR50%). For each individual participant, SNR50% was first determined using an adaptive procedure. During the consonant discrimination task, 12 consonants in the same modality were presented sequentially in a pseudo-random order with four repeats. In total, 48 consonants in A, V, and AV modality were presented. Participants responded using a touch-screen with 12 buttons corresponding to the 12 consonants. No feedback about response accuracy was provided. For the CNC word identification task, 60 different CNC words in each modality were presented in a pseudo-random order in blocks of 20 stimuli. Participants were required to verbally repeat back the word they recognised each time. For both types of speech stimuli, the order of A, V, and AV modalities was randomly chosen.

AV integration for each participant was quantified using a probability model (Blamey *et al.*, 1989) and a cue integration model (Rouger *et al.*, 2007). The probability model estimates participants' AV performance P_{AV}^{est} when auditory and visual speech processing are independent, i.e., *minimum integration* happens (Eq. 1), where, P_A and P_V are response accuracies in A and V, respectively.

$$P_{AV}^{est} = P_A + P_V - P_A * P_V \qquad \text{(Eq. 1)}$$

The cue integration model predicts AV performance when *optimal cue integration* happens between the two modalities. The cue integration model assumes that to be able to understand speech information, we need to recognize at least a certain number (T) of cues correctly. Further, our perception of the cues has a Poisson distribution (Eq. 1), where λ is the average number of cues that we recognise.

$$P(n > T) = \Sigma_{k=n} \left(\lambda^k e^{-\lambda}\right)/k! \qquad \text{(Eq. 2)}$$

Threshold T depends on the type of speech stimuli, regardless of modality. When optimal integration happens, the number of cues recognised in the AV modality (λ_{AV}) equals the sum of those recognised in A (λ_A) and V (λ_V) modalities, i.e. $\lambda_{AV} = \lambda_{\mathrm{A}} + \lambda_{\mathrm{V}}$. Based on participants' performance in A (P_A) and V (P_V) modalities, λ_{A} and λ_{V} can be estimated using Eq. 2.

To apply the cue integration model, we tested a group of young NH listeners to obtain the stimulus-dependent T values which best fit the data for young NH listeners' AV performance, i.e. $P_{AV}^{est} = P_{\mathrm{AV}}$, with $\lambda_{AV}^{est} = \lambda_{\mathrm{A}} + \lambda_{\mathrm{V}}$. We then applied these T values to old NH listeners and CI users to assess whether the older NH listeners or CI users had better or worse AV integration than the young NH listeners.

fNIRS imaging

Data collection

In this study, a continuous-wave fNIRS device (NIRScout, NIRX medical technologies, LLC) with 16 LED illumination sources and 16 photodiode detectors was used. fNIRS measures the concentration changes of oxygenated (HbO) and deoxygenated (HbR) haemoglobin in the blood. To (partly) remove the signals recorded from extracerebral tissue, two 1.3-cm 'short' channels that were located in the anterior temporal cortex of each side were used. For fNIRS imaging, data were recorded from the two ROIs, i.e. LSTS and LOC, as shown in Fig. 1.

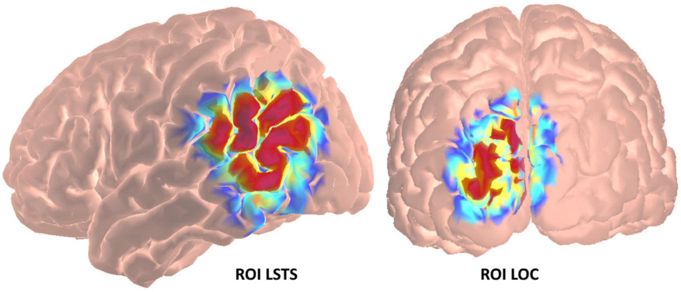


Fig. 1: ROIs where fNIRS responses were measured, i.e., LSTS and LOC.

A block-design was used for fNIRS data collection, with the length of a stimulus block being 14.5 s. Each stimulus block was preceded and followed by a 25-s white fixation cross on the black screen of the CRT monitor. To ensure participants remained focused on the experiment, they were asked to perform a recognition task at the end of each block. Seven blocks of stimuli in each modality were presented.

Six testing periods of fNIRS data were collected, with the first three testing periods using consonant stimuli, and the second three using CNC word stimuli. For each type of speech stimuli, the first testing period used blocks of A and AV stimuli in quiet, and the second testing period used blocks of A and AV stimuli with babble noise. When A stimuli were presented, there was a static picture of the female speaker on the monitor. For these two testing periods, 7 blocks of A and AV stimuli were played in pseudo-random order. The SNR of the babble noise was presented at participant-dependent levels (SNR50%) previously determined for behavioural speech tests. In the third testing period, 7 blocks of stimuli in V modality were presented, with no auditory input through the earphone or CI processor. The recording of response in V modality is supplementary, to check that responses in the ROIs in two A and V modalities correlate with responses in AV modality.

Data analysis

fNIRS data analysis consisted of signal pre-processing and signal processing. Signal pre-processing included 1) identifying and removing step-like artefacts that were caused by sudden loss of contact between optodes and skin, 2) excluding channels that had poor data quality, 3) estimation of haemodynamic response and band-pass filtering to remove environmental noises. Short-channel-separation was further conducted to remove the extracerebral response from the long channels within the ROIs. This was done by first extracting the first principal component (PC_1) of HbO or HbR from the 2 short channels by using principal component analysis (PCA). Channels with short distance were assumed to only measure responses from the extracerebral tissues, and PC_1 was assumed to be the systemic response that would exist globally. A general linear model (GLM) was then used, as shown in Eq. 3, to remove the PC_1 signal from the response in each long channel (Y_L). Within Eq. 3, HRF was the experimental specific haemodynamic response function model for different types of stimuli. β was the coefficient of HRF and α was the coefficient of PC_1 estimated from GLM; ε was the residual noise.

$$Y_L = [HRF_1, HRF_2] * [\beta_{1,}\, \beta_2]' + \alpha * PC_1 + \varepsilon \qquad \text{(Eq. 3)}$$

After short-channel-separation in the long channels, the averaged hemodynamic response across the 7 blocks was then estimated for stimuli of each modality. Outlier blocks of response were excluded. Only the HbO response were used for further statistical analysis. To test our hypothesis that the fNIRS data in old NH listeners would show the PoIE, the inequality in Eq. 4 was used. The left and right sides of Eq. 4, represent the differences between HbO responses in the AV and A modalities when the auditory background was quiet (Q) and with noise (N), respectively.

$$\left(A_Q V - A_Q\right) < (A_N V - A_N) \qquad \text{(Eq. 4)}$$

RESULTS

AV integration: Behavioural performance

Figure 2 plots the speech test results in three modalities for young, old NH listeners, and CI users when responding to consonants (first row) and CNC words (second row). Black dashed lines and magenta dash-dot lines plot the probability model and the cue integration model predicted AV performance, respectively, in each group. For the cue integration model, the stimulus-dependent T thresholds of 1 and 3 for consonant and CNC word stimuli, respectively, which were obtained based on the best fit for young NH listeners' performance, were applied to old NH listeners and CI users. Figure 2 shows that when responding to consonant stimuli (first row), the cue integration model (magenta dash-dot line), fits old NH listeners' AV performance (red dots) well but CI users' performance was lower than predicted by the young NH based model. In contrast, the probability model (black dash line) fits CI users' performance (red dots) well, i.e., CI users showed essentially independent use of A and V cues in AV mode. These results showed that when responding to consonant stimuli, old NH listeners had comparable AV integration with young NH listeners (optimal cue integration), while our experienced CI users had less AV integration ability than NH listeners. When responding to CNC word stimuli, as shown in Fig. 2 (second row), the cue integration model (magenta dash-dot line) fits the AV performance (red dots) of both CI users and old NH listeners well, i.e., both old NH listeners and CI users had optimal integration compared to young NH listeners.

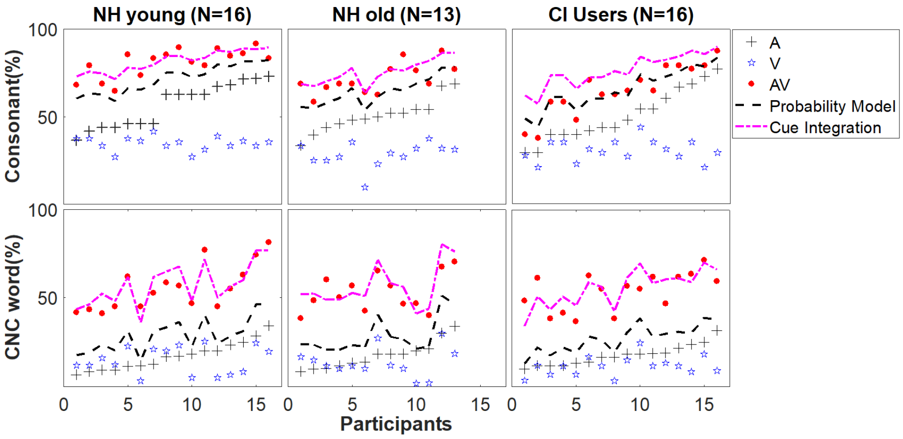


Fig. 2: Audio-visual (AV) speech perception of consonants and CNC words in NH listeners and CI users.

AV integration: fNIRS imaging

Figure 3 shows the fNIRS response in ROI LOC of age-matched NH listeners (first row) and CI users (second row) when responding to consonant stimuli. Red lines and shaded areas plot the mean and standard error of mean (SEM) of HbO; blue plots HbR response. Vertical dashed lines indicate the stimulus onset and offset. From left to right, each column plots $(A_QV - A_Q)$, $(A_NV - A_N)$, and $(A_NV - A_N) - (A_QV - A_Q)$ measures, respectively. A pairwise running one-tailed t-test was performed on the HbO response between quiet and noisy conditions, using Eq. 4. Permutation *t*-tests (Groppe *et al.*, 2011) were done to control familywise error rate for multiple comparisons. No significantly larger response was found in the noisy condition than in quiet, i.e., no occurrence of the PoIE in the fNIRS responses in ROI LOC, in either CI users or NH listeners. The same statistical analysis was done for responses in two ROIs and to two types of speech stimuli. The PoIE of AV integration was not significantly demonstrated for NH listeners or CI users for either speech stimulus type, in either of the ROIs.

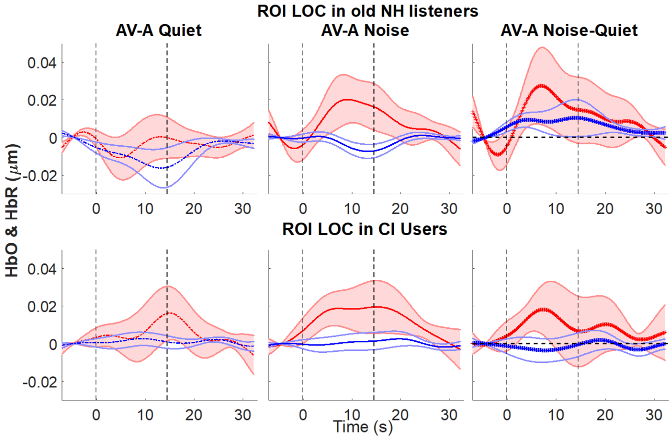


Fig. 3 fNIRS response of the old NH listeners and CI users in the ROI LOC when responding to consonant stimuli.

DISCUSSION AND CONCLUSION

This study examined AV speech integration in CI users and old NH listeners using behavioural and fNIRS measures. Using behavioural measures, CI users had poorer AV integration compared to old NH listeners when responding to consonant stimuli, but had comparable AV integration ability when responding to CNC word stimuli. For fNIRS imaging, no PoIE of AV integration was observed in either of the two ROIs for either CI users or age-matched NH listeners.

Our behavioural results that CI users had comparable or poorer AV speech integration ability than NH listeners could be because, first, they were CI users who have years of experience of using their implant and no longer relied on lip-reading for speech perception. Thus, these CI users showed no super-normal lip-reading ability or AV integration ability than NH listeners. Further, when responding to consonant stimuli, CI users' performance in AV modality was mainly dependent on their performance in A modality. As shown by the cue integration model that, participants only needed to recognise more than one cue from the consonant stimuli to make a correct response. When responding to consonant in AV modality, CI users selectively attended to A cues and ignored V cues. This selective attention maladaptively affected their AV integration.

Our fNIRS results that no PoIE being observed in either group could be because, first, large variance of response existed in each group, which derived from both experimental measures and individual's difference in fNIRS response. As to reveal this inverse effectiveness of AV integration, fNIRS measures were estimated from responses recorded in four different conditions, resulting in too much noise in the data. Also, largely variant AV integration responses have been reported in old NH listeners, due to their wider AV integration window (Diederich *et al.*, 2008). Further, because of the limited spatial resolution of fNIRS compared to that of fMRI, the ROIs in this study were larger and less focussed than those in the fMRI studies that showed the PoIE. All these reasons make it challenging to reveal the PoIE of AV integration in our old NH listeners and CI users.

REFERENCES

Anderson, C.A., Lazard, D.S., and Hartley, D.E.H. (**2016**). "Plasticity in bilateral superior temporal cortex: effects of deafness and cochlear implantation on auditory and visual speech processing," Hear. Res., **343**, 138-149.

Blamey, P.J., Cowan, R.S., *et al.* (**1989**). "Speech perception using combinations of auditory, visual, and tactile information." J. Rehabil. Res. Dev., **26**, 15-24.

Diederich, A., Colonius, H., *et al.* (2008). "Assessing age-related multisensory enhancement with the time-window-of-integration model," Neuropsychologia, **46**, 2556-2562.

Groppe, D.M., Urbach, T.P., *et al.* (**2011**). "Mass univariate analysis of eventrelated brain potentials/fields I: A critical tutorial review," Psychophysiology, **48**, 1711-1725.

Holmes, N.P. (**2007**). "The law of inverse effectiveness in neurons and behaviour: multisensory integration versus normal variability," Neuropsychologia, **45**, 3340-3345.

James, T.W., Stevenson, R.A., *et al.* (**2012**). "Inverse effectiveness and BOLD fMRI," *The New Handbook of Multisensory Processes* (Stein BE, ed.), pp. 207-222.

Laurienti, P.J., Perrault, T.J., *et al.* (**2005**). "On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies," Exp. Brain Res., **166**, 289-297. doi: 10.1007/s00221-005-2370-2

Meredith, M.A., and Stein, B.E. (**1983**). "Interactions among converging sensory inputs in the superior colliculus," Science, **221**, 389-391.

Perrault, T.J., Vaughan, J.W., *et al.* (**2005**). "Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli," J. Neurophysiol., **93**, 2575-2586.

Peterson, G.E., and Lehiste, I. (**1962**). "Revised CNC lists for auditory tests," J Speech Hear. Disord., **27**, 62-70.

Rouger, J., Lagleyre, S., *et al.* (**2007**). "Evidence that cochlear-implanted deaf patients are better multisensory integrators," Proc. Natl. Acad. Sci. USA, **104**, 7295-7300. doi: 10.1073/pnas.0609419104

Schierholz, I., Finke, M., *et al.* (**2015**). "Enhanced audio–visual interactions in the auditory cortex of elderly cochlear-implant users," Hear. Res., **328**, 133-147.

Stevenson, R.A., and James, T.W. (**2009**). "Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition," NeuroImage, 44, 1210-1223. doi: 10.1016/j.neuroimage.2008.09.034

Speech processing using adaptive auditory receptive fields

ASHWIN BELLUR AND MOUNYA ELHILALI*

Laboratory for Computational Audio Perception, Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA

The auditory system exhibits a remarkable ability to adapt to its listening environment, driven both by sensory-based cues and goal-directed processes. Here, we focus on the role of attentional feedback in facilitating processing of speech sounds in presence of nonstationary noises. We examine a theoretical formulation for retuning of cortical-like receptive fields to enable robust detection of speech sounds in presence of interference. The framework employs modulation-tuned filters aimed at emulating tuning characteristics of neurons at the level of auditory cortex. This bank of filters is then modulated based on goal-directed feedback to enhance separability between the feature representation of speech and nonspeech sounds. We hypothesize that this retuning procedure results in an emphasis of unique speech and nonspeech modulations in a high-dimensional space. We discuss the implications of this retuning on the fidelity of encoding speech sounds in presence of seen and novel noise conditions, and discuss implications of such plasticity in facilitating listening in challenging acoustic environments, hence opening the door to adaptive and intelligent audio technology that can emulate the biological system.

INTRODUCTION

When engaged in a conversation in a noisy cafeteria, our brain relies on cognitive processes particularly attention to help navigate the challenging acoustic stimulus impinging on its ears and detect sounds of interest. Attention acts as information bottleneck that sifts through acoustic cues and helps boost the signal-to-noise representation of targets relative to interferers in order to ultimately facilitate processing of these sounds of interest. An increasing body of work suggests that attending to a target sound induces profound but rapid adaptation effects in brain responses. Magnetoencephalography (MEG) recordings in listeners attending to a target speech in presence of competing talkers showed selective enhancement of neural phase-locking to the attended stream resulting in improved and robust reconstruction of the attended speech regardless of the signal-to-noise relative to the interferer (Akram *et al.*, 2016; Ding and Simon, 2012; Puvvada and Simon, 2017). The readout of the attended speech appears to also be in synchrony with enhancement in brain oscillations (particularly alpha rhythm), which selectively modulates the neural representation of the attended stimulus resulting in improved segregation

*Corresponding author: mounya@jhu.edu

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

(Wostmann *et al.*, 2016). Similar results have been reported using electroencephalography (EEG) where selective attention in noisy environments (e.g., competing talkers, reverberation) also improve neural encoding of the speech envelope of the attended stream (Fuglsang *et al.*, 2017; O'Sullivan *et al.*, 2014). A refined look at neural activity at the single neuron level has also corroborated these findings using high-density intracranial electrode arrays in human participants (Mesgarani and Chang, 2012). Results show that neural responses in non-primary auditory cortex (posterior superior and middle temporal gyrus) are driven almost solely by the attended speaker.

A natural question that arises is how does the auditory system balance a stable sensory encoding and perceptual decoding in presence of such profound adaptation effects (Seriès *et al.*, 2009). Given the distributed neural circuitry underlying this attention-induced modulation, one interpretation of these effects is at the perceptual stage whereby adaptation of perceptual estimates implies refining the *interpretation* of sensory encoding for different tasks/environments. This account is often favored in engineering solutions which employ similar forms of adaptation (e.g., domain adaptation, model adaptation) in machine learning to adapt to specific targets or classes or generalize models across conditions of the data (Ben-David *et al.*, 2010; Gauvain and Lee, 1994; Leggetter and Woodland, 1995; Siohan *et al.*, 2001).

An alternative interpretation is that observed effects are in fact due to adaptation of the sensory mapping itself. This form of adaptation implies that cognitive processes might receive inconsistent or suboptimal encoding (Seriès *et al.*, 2009). If feature maps themselves are retuning, they are altering the representation of the incoming stimulus hence requiring perceptual processes to compensate for this warped mapping or at least take it into account. Electrophysiological recordings in single neurons as early as auditory cortex put forth evidence in support of adaptation of sensory feature maps. Cortical activity in animals engaged in various behavioural tasks shows that tuning characteristics of these neurons exhibit rapid tuning shifts in line with the behavioural task at hand (Elhilali *et al.*, 2007; Fritz *et al.*, 2003; Lu *et al.*, 2017; Winkowski *et al.*, 2017). Effects of this adaptation can be gleaned through their neural spectro-temporal receptive fields (STRFs). An STRF is a measure that characterizes the steady state response properties of auditory neurons, spanning their temporal dynamics and spectral selectivity (Elhilali *et al.*, 2013). At the level of auditory cortical areas, these very receptive fields reflect the inherent properties of individual neurons which reshape their tuning to reflect task demands and relevant targets or backgrounds in an auditory scene (Atiani *et al.*, 2014; David *et al.*, 2012; Engineer *et al.*, 2014; Fritz *et al.*, 2005).

In this work, we examine the theoretical underpinnings of the attention-driven receptive field plasticity in shaping neural encoding of incoming sound signals, and effectively enhancing detection of target sounds in complex scenes. We focus this question in the case of listening to speech sounds in presence of noise interferers or distortions such as reverberation. Here, we review recent work which leverages STRF plasticity in models for robust detection of speech in presence of background noise (Bellur and Elhilali, 2017; Carlin and Elhilali, 2015b). We comment on implications of observed changes from both models in interpreting observed changes in the biological system.

MODELING RECEPTIVE FIELD PLASTICITY

The transformations undertaken along the auditory system can be emulated by a multistage process whereby the incoming acoustic waveform is mapped from a one-dimensional signal representation along time to various feature dimensions that highlight characteristics of the acoustic waveform along both time, frequency, and spectrotemporal modulations. These transformations – achieved through a variety of analysis maps – act as feature detectors to extract cues relevant for processing and interpretation of incoming signals (Eggermont, 2001; Nelken and Bar-Yosef, 2008).

In the current work, we ask the question: How would the system behave if a sensory mapping stage, specifically at the level of cortical processing, would receive feedback that induces changes in its properties in a direction dictated by the feedback signal, and within constrains imposed by the system? We contrast two approaches to achieve such optimization, a linearized vs. nonlinear approach, as discussed next.

A linearized optimization of receptive field plasticity

In a first study, we examine a framework for such feedback defined in a discriminative fashion (Carlin and Elhilali, 2015b). In this setup, the cortical stage is retuned to contrast the mapping of speech and non-speech stimuli. The model starts by transforming all incoming signals into a time-frequency spectrogram, by employing a model of the auditory periphery (Chi *et al.*, 2005). This stage maps the acoustic waveform $x(t)$ through a series of stages including an array of asymmetric, constant-Q band-pass filters, first order derivative, half-wave rectification and spectral derivative, before smoothing the responses using a short time window $w(t,\tau) = e^{-t/\tau}u(t)$ to mimic the loss of phase locking observed at the level of the midbrain. The auditory spectrogram $s(t,f)$ is next processed by an adaptable feature extraction framework, based on the processes of the cortical regions and task-driven plasticity observed in the auditory pathway. Carlin and Elhilali (2015b) propose using an ensemble of adaptable STRFs to extract frequency and spectro-temporal dynamics information from the auditory spectrogram. STRFs used in this work are neurophysiologically-recorded function obtained from non-behaving ferrets (recorded in studies by Elhilali *et al.* (2004) and Fritz *et al.* (2003). These biological STRFs are used as initial spectro-temporal filters upon which attentional feedback will be applied to induce plastic changes in line with the discriminative framework. Since the approach employs biologically-obtained filters in a non-parametric form, it uses a linear model using logistic regression to retune these filters in a manner that enhances the ability of the system to detect speech in a noisy environment.

The adaptive framework is formulated as maximizing the conditional likelihood of labels y with respect to the weighted ensemble response E, as defined below

$$p(Y = y|\boldsymbol{E}, \boldsymbol{w}) \equiv \sigma(y\boldsymbol{w}^T\boldsymbol{E}) \qquad \text{(Eq. 1)}$$

where $\sigma(\gamma) = {}^{1}/_{(1+\exp(-\gamma))}$ is the logistic function and $y \in \{+1,-1\}, y = +1$ denotes speech and $y = -1$ denotes non-speech. Let $r_k(t,f)$ be the firing rate of the k^{th} neuron:

$$r_k(t,f) = h_k(t,f) *_{tf} s(t,f) \qquad \text{(Eq. 2)}$$

where $h_k\,(t\,,f)$ is the transfer function of the k^{th} STRF and $*_{tf}$ is the 2D convolution over time and frequency axes. The corresponding modulation domain representation can be determined as

$$|R_k(\omega,\Omega)| = |H_k(\omega,\Omega)|.\,|S(\omega,\Omega)| \qquad \text{(Eq. 3)}$$

where $R_k(\omega,\Omega)$, $H_k(\omega,\Omega)$ and $S(\omega,\Omega)$ are the 2D discrete Fourier transforms of the firing rate, STRF and stimulus spectrogram, respectively. ω represents temporal modulations or rates (in Hz) and Ω represents spectral modulations or scale (in cycles/octave). The ensemble response $\boldsymbol{E}$ in Eq. 1 defined as

$$\boldsymbol{E} = [1, \textstyle\sum_{\boldsymbol{\omega\Omega}}|R_1(\omega,\Omega)|\,, \ldots, \sum_{\boldsymbol{\omega\Omega}}|R_K(\omega,\Omega)|] \in \mathbb{R}^{K+1} \qquad \text{(Eq. 4)}$$

is a supervector of responses of the K neurons to a stimulus. $\boldsymbol{w} = [w_0, w_1, \ldots, w_k]$ in equation 1 is the vector of regression coefficients for the K neurons of the ensemble.

Throughout this framework, the model mimics common experimental paradigms whereby neurons are characterized with a ‘default’ tuning transfer function H_0. These are typically obtained when the auditory system is not engaged in any active task, but is in a passive state. Once the system is engaged in a task, these filter parameters H_0 are retuned, yielding adapted receptive fields H_a. In the proposed framework by Carlin and Elhilali (2015b), the adaptation problem is cast as an optimization with goal to minimize the cost function $J(w,\mathcal{H}_a)$ defined as

$$J(w,\mathcal{H}_a) = \tfrac{1}{2}\,\|\boldsymbol{w}\|_2^2 - \tfrac{C}{M}\textstyle\sum_m log\big(\sigma(y_m\boldsymbol{w}^T\boldsymbol{E}_m)\big) + \tfrac{\lambda}{2}\sum_k\|\Delta_k\|_F^2 \qquad \text{(Eq. 5)}$$

where $\mathcal{H}_a = \{|H_k^a(\omega,\Omega)|\}_{k=1}^K$ and $\Delta_k = |H_k^a(\omega,\Omega)| - |H_k^0(\omega,\Omega)|$. $H_k^0(\omega,\Omega)$ is the default tuning of the k^{th} neuron and $H_k^a(\omega,\Omega)$ its adapted tuning. By formulating the adaption process in this manner, the framework seeks to obtain a weighted set of retuned neural ensemble that maximizes the conditional probability averaged over all stimuli (*M*). The Δ_k term ensures that each individual neuron retunes marginally from its default tuning, consistent with the observation that cortical neurons maintain stable properties while adapting marginally to behavioral tasks (Elhilali *et al.*, 2007).

In order to determine the regression parameters $\boldsymbol{w}$ and retuned STRF ensemble $\mathcal{H}_a$, block coordinate descent is employed, alternating between the 2 convex problems

$$argmin\, J(w,\mathcal{H}_a) \quad s.t.\ \ |H_k^a(\omega,\Omega)| \geq 0\ \forall k,\omega,\Omega$$

$$argmin\, J(w,\mathcal{H}_a) \quad s.t.\ \ w_k > 0$$

Upon convergence, the solution to these two convex problems can be written as

$$|H_k^a(\omega,\Omega)|= |H_k^0(\omega,\Omega)| + \frac{C}{\lambda}. w_k. \frac{1}{M}\sum_m y_m\big(1-\sigma(y_m \boldsymbol{w}^T \boldsymbol{E}_m)\big)S_m(\omega,\Omega) \qquad \text{(Eq. 6)}$$

$$\boldsymbol{w} = \frac{C}{M}\sum_m y_m\big(1-\sigma(y_m \boldsymbol{w}^T \boldsymbol{r}_m)\big)\boldsymbol{r}_m \qquad \text{(Eq. 7)}$$

where M denotes the number of stimuli used for the adaptation process.

It can be seen from the constraints and solution equations (Eqs. 6 and 7) that by enforcing the weights to be positive and using the labels $y_m = +1$ for speech and $y_m = -1$ for non-speech, the adaptation process seeks to enhance speech modulation while suppressing non-speech content. Another interesting observation relates to the impact of the stimulus. By interpreting $1-\sigma(y_m \boldsymbol{w}^T \boldsymbol{E}_m)$ as prediction error, certain stimuli that are too difficult to predict have a stronger impact on the adaptation process. Furthermore, it can be seen in Eq. 7 that neurons that are task-relevant receive larger weights in contrast to the task-irrelevant neurons.

A nonlinear parametric optimization of receptive field plasticity

In contrast to the approach described above, Bellur and Elhilali (2017) explore an alternate framework to model task-driven plasticity, focusing on 3 broad different takes to the optimization problem: First, the approach in Bellur and Elhilali (2017) employs parameterized Gabor filters to encode spectrotemporal dynamics, instead of physiologically recorded receptive fields. By employing parameteric functions to emulate cortical receptive fields, Gabor filters can be re-tuned to achieve a non-linear transformation in contrast to the linear adaptation of filter patches as used in Carlin and Elhilali (2015b). Second, instead of assigning fixed class labels $y_m = \pm 1$ to distinguish speech from non-speech tokens, the approach in Bellur and Elhilali (2017) employs a generative probabilistic model using Gaussian mixture models (GMMs) to serve as *object* representations of clean speech and non-speech classes (Duda *et al.*, 2000). In this case, the optimization seeks to retune the Gabor filters in a manner that enhances the ability of the GMMs to discriminate between noisy speech and nonspeech, thereby adapting the feature extraction process to work even under novel noise conditions. Third, the optimization process employs a Genetic algorithm (Michalewicz, 1996). This approach differs from the convex optimization formulated in Carlin and Elhilali (2015b) and allows to search the parameter space for the Gabor filters to ensure improved discrimination between the two classes with respect to the fixed GMMs.

This approach follows the same general framework as presented earlier. A time-domain waveform is first mapped through a model of the auditory periphery to derive an auditory spectrogram $s(t,f)$. Then, a bank of 2D Gabor filters are applied to analyze the spectral and temporal modulations in the spectrogram. Such filters are considered a reasonable approximation of cortical receptive fields observed in the mammalian auditory system (Ezzat *et al.*, 2007; Theunissen *et al.*, 2000). The filters are parametrized as:

$$g_k(t,f) = \frac{\alpha_k}{2\pi\sigma_{tk}\sigma_{fk}} e^{-\frac{1}{2}\left(\frac{t_1^2}{\sigma_{tk}^2}+\frac{f_1^2}{\sigma_{fk}^2}\right)} e^{2\pi j(\omega_k t+\Omega_k f)} \qquad \text{(Eq. 8)}$$

where $t_1 = tcos(\theta_k) + fsin(\theta_k)$ and $f_1 = -tsin(\theta_k) + fcos(\theta_k)$. σ_{tk} and σ_{fk} denote the temporal and spectral bandwidths of the Gaussians of the k^{th} Gabor filter, respectively. θ_k specifies the orientation of the main lobe of the Gabor filter and α_k is a gain term. ω_k and Ω_k are the rate and scale of the k^{th} Gabor filter.

The auditory spectrogram is convolved with a bank of Gabor filters $\mathcal{g} = \{g_1, g_2, \ldots, g_K\}$ spanning the spectrotemporal space set by the chosen parameters (Eq. 9). The output is then collapsed along the time axis to obtain the spectrotemporal dynamics and frequency information as shown in equation Eq. 10.

$$C_k(t,f) = |s(t,f) *_{tf} g_k(t,f)| \qquad \text{(Eq. 9)}$$

$$T_k(f) = \int C_k(t,f)dt \qquad \text{(Eq. 10)}$$

Like the regression approach, the Gabor filter model in Bellur and Elhilali (2017) starts with a default set of parameters $\mathcal{g}^0 = \{g_1^0, g_2^0, \ldots, g_K^0\}$ analogous to the passive receptive fields used in Carlin and Elhilali (2015b). The Gabor parameters are then retuned for robust speech activity detection, to obtain an adapted filter bank of Gabor filters denoted as $\mathcal{g}^a = \{g_1^a, g_2^a, \ldots, g_K^a\}$. These adapted filters are derived based on statistical models of speech and non-speech data; Gaussian mixture models of clean speech and nonspeech estimated based on their spectrotemporal modulation features (Eq. 10). A held out set of noisy speech and nonspeech data is then used adapt the filters in manner that enhances the ability of the GMMs to discriminate between noisy speech and nonspeech even in mismatched conditions. The hypothesis at the center of this work is that this retuning process will lead to highlighting the *discriminable regions* of the spectrotemporal modulation space as represented by the GMMs, hence resulting in robust speech activity detection under novel noise conditions.

The Gabor filters are retuned using a genetic algorithm which scans the parameter space. It employs a fitness measure to gauge the suitability of the parameter choice. In Bellur and Elhilali (2017), the fitness measure used is d-prime, defined as:

$$d' = \frac{\mu_{ns} - \mu_s}{\sqrt{\frac{1}{2}(\sigma_{ns}^2 + \sigma_s^2)}} \qquad \text{(Eq. 11)}$$

μ_c and σ_c denote the mean and standard deviation respectively of the log likelihood ratio (LLR) values estimated using the GMMs trained on clean speech (c=s) and nonspeech (c=ns) data. The genetic algorithm is initialized with the default parameters ($\mathcal{g}^0$) as a member of the first generation. The algorithm then propagates through multiple generations to find the fittest member ($\mathcal{g}^a$) as defined by the equation Eq. 11.

OPTIMIZED MAPPING OF SPEECH AND NONSPEECH SOUND CLASSES

Figure 1A shows results of the adaptation process in terms of the average difference between the modulation profiles of the STRF after and before adaptation; That is $\langle|H_k^a(\omega,\Omega)| - |H_k^0(\omega,\Omega)|\rangle_k$ where $\langle.\rangle_k$ denotes averaging. The figure illustrates that the neural ensemble tends to emphasize slower modulations especially for positive rates (which correspond to downward modulations), which are commensurate with modulations in speech sounds (Elliott and Theunissen, 2009). Given the choice of label values and the fact that the weights $\boldsymbol{w}$ are set to be positive, the adaptation framework also leads to suppression of responses to faster modulations, hence diminishing the response to non-speech regions of the spectrotemporal space.

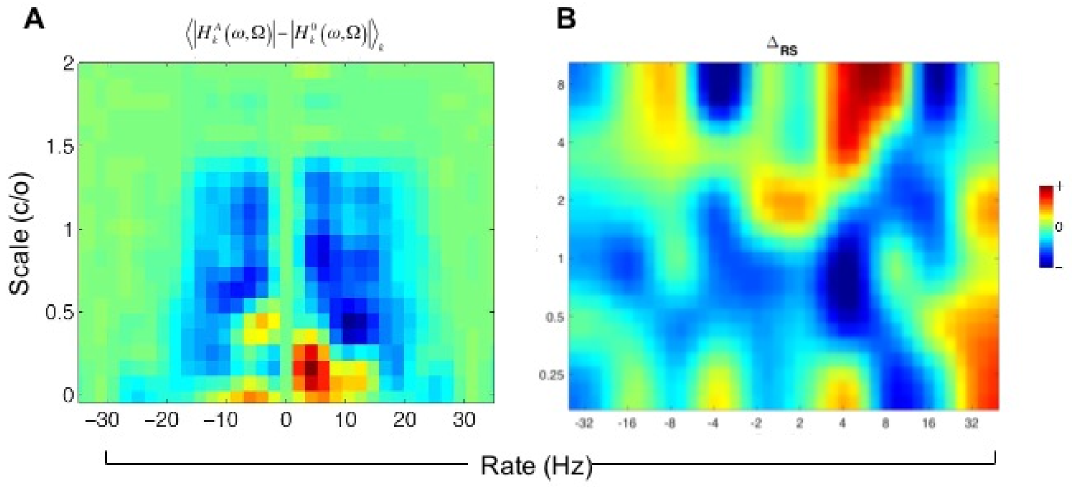


Fig. 1: (A) Average difference in the responses of STRFs before and after adaptation using linearized regression. The difference is measured as $\langle |H_k^a(\omega,\Omega)| - |H_k^0(\omega,\Omega)| \rangle_k$ where $\langle . \rangle_k$ denotes the average operation [Figure reproduced from (Carlin and Elhilali, 2015a) with permission from IEEE]. **(B)** Δ_{RS} difference between the energies in the rate scale space on using $\mathscr{g}^a$ and $\mathscr{g}^0$ filter banks in the nonlinear optimization approach using Gabor filters.

Figure 1B shows the difference between the energies in the rate scale space on using $\mathscr{g}^a$ and $\mathscr{g}^0$ filter banks. Δ_{RS} depicted in this figure is estimated as:

$$\Delta_{RS} = \langle \textstyle\sum_f \sum_t |s_m(t,f) *_{tf} \mathscr{g}^a| - \sum_f \sum_t |s_m(t,f) *_{tf} \mathscr{g}^0| \rangle_m \qquad \text{(Eq. 12)}$$

where $\langle . \rangle$ denotes the average over all stimuli, both noisy speech and nonspeech. Δ_{RS} illustrates the difference in energies on projecting the stimuli on to the spectrotemporal modulation space using the 2 sets of Gabor filter banks. It can be seen that while slower modulations are emphasized, broadband fast modulations are also emphasized, as well fast spectral modulation at 4-Hz rate. The figure also suggests that greater discriminability is attained on adapting the filters because sparse non-overlapping regions of speech and nonspeech are emphasized on adaptation, while overlapping regions are suppressed.

Further insight into the behavior of the Gabor model can be gleaned from contrasting the log-likelihood estimates with respect to both speech and non-speech data. Figure 2A shows the histogram of the log likelihood ratio values of noisy speech and non-speech stimuli estimated, before ($\mathscr{g}^0$) and after adaptation ($\mathscr{g}^a$) of the Gabor filters. As can be seen from the plots, the classes are more separable on using the retuned filter bank. It is interesting to note that on adaptation, the LLR values for the 2 classes do not necessarily move in the opposite directions, rather they become narrower owing to the fact that the *d*-prime measures reward lesser spread of the LLR values for a class. Figures 2B and 2C show a schematic summarizing the impact of the different optimization approaches on the resulting representation of speech and nonspeech classes.

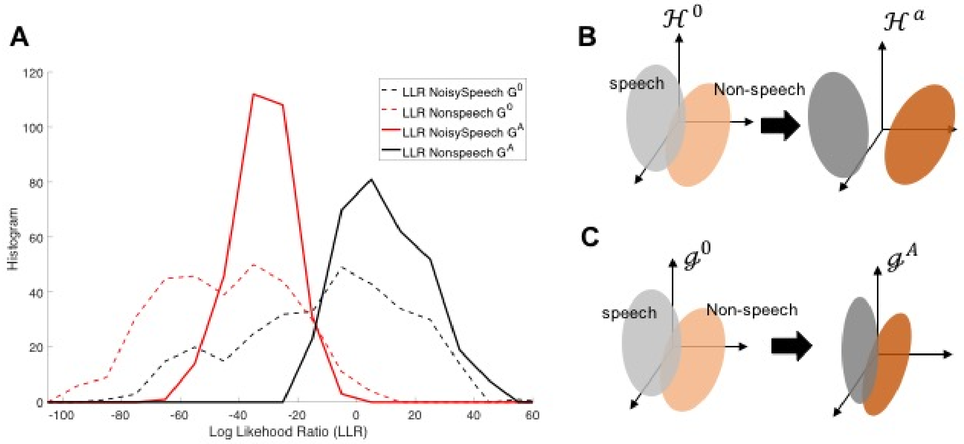


Fig. 2: (A) Histogram of the log likelihood ratio values of noisy speech and nonspeech stimuli before ($\mathcal{G}^0$) and after adaptation ($\mathcal{G}^a$) of the Gabor filters. **(B, C)** Schematic of changes in mapping of speech and non-speech classes using linearized regression vs. nonlinear optimization.

CONCLUSIONS

The models reviewed here shed light on two possible strategies that improve speech detection in noise: (i) An approach that pushes the perceptual maps of speech and nonspeech further apart from each other (Fig. 2B). This is achieved by reshaping the feature maps to emphasize acoustic cues unique to speech and de-emphasize characteristics of nonspeech. As shown in Fig. 1A, putting more emphasis on slow temporal modulations in the region around ~4 Hz results in highlighting areas known to correlate well with characteristics of speech signals (e.g., syllabic rate, Elliott and Theunissen, 2009). This outcome is achieved through a linearized optimization of cortical receptive fields that allows minor tweaks to their response properties in a linear way.

In contrast, a parametrized approach that exhaustively searches the space of cortical filters represented as Gabor functions combined with statistical modeling of the perceptual decision space results in a different outcome by tightening the perceptual maps of speech vs. nonspeech classes (Fig. 2C). This outcome is an equally acceptable solution to the stated problem, and in fact has been shown to yield superior performance of speech detection in noise, especially when contrasted with novel noisy speech and nonspeech conditions.

Overall, either strategy (or combined) offers a robust biomimetic approach to adaptive signal processing to improve sound perception in noise. It remains to be seen which approach is more in line with scheme underlying neural plasticity in the brain. As more advanced experimental techniques emerge and paradigms are able to train animals on more sophisticated behavioral tasks, it will be possible to tease apart the theoretical underpinnings of attention-driven neural plasticity in the auditory system.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health under grant R01HL133043 and the Office of Naval research under grant ONR N000141612045.

REFERENCES

Akram, S., Presacco, A., Simon, J.Z., Shamma, S.A., and Babadi, B. (**2016**). “Robust decoding of selective auditory attention from MEG in a competing-speaker environment via state-space modeling,” Neuroimage, **124**, 906-917. doi: 10.1016/j.neuroimage.2015.09.048

Atiani, S., David, S.V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S.A., *et al.* (**2014**). “Emergent selectivity for task-relevant stimuli in higher-order auditory cortex,” Neuron, **82**, 486-499. doi: 10.1016/j.neuron.2014.02.029

Bellur, A., and Elhilali, M. (**2017**). “Feedback-driven sensory mapping adaptation for robust speech activity detection,” IEEE T. Audio Speech, **25**, 481-492. doi: 10.1109/TASLP.2016.2639322

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., and Vaughan, J.W. (**2010**). “A theory of learning from different domains,” Mach. Learn., **79**, 151-175. doi: 10.1007/s10994-009-5152-4

Carlin, M.A., and Elhilali, M. (**2015a**). “A framework for speech activity detection using adaptive auditory receptive fields,” IEEE T. Audio Speech, **23**, 2422-2433. doi: 10.1109/TASLP.2015.2481179

Carlin, M.A., and Elhilali, M. (**2015b**). “Modeling attention-driven plasticity in auditory cortical receptive fields,” Front. Comput. Neurosci., **9**, 106. doi: 10.3389/fncom.2015.00106

Chi, T., Ru, P., and Shamma, S.A. (**2005**). “Multiresolution spectrotemporal analysis of complex sounds,” J. Acoust. Soc. Am., **118**, 887-906.

David, S.V., Fritz, J.B., and Shamma, S.A. (**2012**). “Task reward structure shapes rapid receptive field plasticity in auditory cortex,” Proc. Natl. Acad. Sci. USA, **109**, 2144-2149. doi: 10.1073/pnas.1117717109

Ding, N., and Simon, J.Z. (**2012**). “Emergence of neural encoding of auditory objects while listening to competing speakers,” Proc. Natl. Acad. Sci. USA, **109**, 11854-11859. doi: 10.1073/pnas.1205381109

Duda, R.O., Hart, P.E., and Stork, D.G. (**2000**). *Pattern Classification.* Wiley.

Eggermont, J.J. (**2001**). “Between sound and perception: reviewing the search for a neural code,” Hear. Res., **157**, 1-42.

Elhilali, M., Fritz, J.B., Klein, D.J., Simon, J.Z., and Shamma, S.A. (**2004**). "Dynamics of precise spike timing in primary auditory cortex," J. Neurosci., **24**, 1159-1172. doi: 10.1523/JNEUROSCI.3825-03.2004

Elhilali, M., Fritz, J.B., Chi, T.-S., and Shamma, S.A. (**2007**). "Auditory cortical receptive fields: Stable entities with plastic abilities," J. Neurosci., **27**, 10372-10382. doi: 10.1523/JNEUROSCI.1462-07.2007

Elhilali, M., Shamma, S.A., Simon, J.Z., and Fritz, J.B. (**2013**). "A linear systems view to the concept of STRF," in *Handbook of Modern Techniques in Auditory Cortex*. Eds. D. Depireux and M. Elhilali (Nova Science Pub Inc), 33-60.

Elliott, T.M., and Theunissen, F.E. (**2009**). "The modulation transfer function for speech intelligibility," PLoS Comput. Biol., **5**, e1000302.

Engineer, C.T., Perez, C.A., Carraway, R.S., Chang, K.Q., Roland, J.L., and Kilgard, M.P. (**2014**). "Speech training alters tone frequency tuning in rat primary auditory cortex," Behav. Brain Res., **258**, 166-178. doi: 10.1016/j.bbr.2013.10.021

Ezzat, T., Bouvrie, J.V, and Poggio, T. (**2007**). "Spectro-temporal analysis of speech using 2-d Gabor filters," Proc. Interspeech, 506-509.

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. doi: 10.1038/nn1141

Fritz, J.B., Elhilali, M., and Shamma, S.A. (**2005**). "Rapid task-dependent plasticity in primary auditory cortex," in *Auditory Cortex-Towards a Synthesis of Human and Animal Research*. Wds. P. Heil, R. Konig, E. Budinger, and H. Scheich (Mahwah, NJ: Lawrence Erlbaum Associates), 445-466.

Fuglsang, S.A., Dau, T., and Hjortkjær, J. (**2017**). ”Noise-robust cortical tracking of attended speech in real-world acoustic scenes," Neuroimage, **156**, 435-444. doi: 10.1016/j.neuroimage.2017.04.026

Gauvain, J.-L., and Lee, C.-H. (**1994**). "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," IEEE T. Speech Audio, **2**, 291-298.

Leggetter, C.J., and Woodland, P.C. (**1995**). "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," Comput. Speech Lang., **9**, 171-185.

Lu, K., Xu, Y., Yin, P., Oxenham, A.J., Fritz, J.B., and Shamma, S.A. (**2017**). "Temporal coherence structure rapidly shapes neuronal interactions," Nat. Commun., **8**, 13900. doi: 10.1038/ncomms13900

Mesgarani, N., and Chang, E.F. (**2012**). "Selective cortical representation of attended speaker in multi-talker speech perception," Nature, **485**, 233-236. doi: 10.1038/nature11020

Michalewicz, Z. (**1996**). *Genetic Algorithms + Data Structures = Evolution Programs*. Springer Science & Business Media.

Nelken, I., and Bar-Yosef, O. (**2008**). "Neurons and objects: The case of auditory cortex," Front. Neurosci., **2**, 107-113. doi: 10.3389/neuro.01.009.2008

O’Sullivan, J.A., Power, A.J., Mesgarani, N., Rajaram, S., Foxe, J.J., Shinn-Cunningham, B.G., *et al.* (**2014**). “Attentional selection in a cocktail party environment can be decoded from single-trial EEG,” Cereb. Cortex., 1697-1706. doi: 10.1093/cercor/bht355

Puvvada, K.C., and Simon, J.Z. (**2017**). “Cortical representations of speech in a multitalker auditory scene,” J. Neurosci., **37**, 9189-9196. doi: 10.1523/ JNEUROSCI.0938-17.2017

Seriès, P., Stocker, A.A., and Simoncelli, E.P. (**2009**). “Is the homunculus “aware” of sensory adaptation?” Neural Comput., **21**, 3271-3304.

Siohan, O., Chesta, C., and Lee, C.-H. (2001). “Joint maximum a posteriori adaptation of transformation and HMM parameters,” IEEE T. Speech Audio, **9**, 417-428. doi: 10.1109/89.917687

Theunissen, F.E., Sen, K., and Doupe, A.J. (**2000**). “Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds,” J. Neurosci., **20**, 2315-2331.

Winkowski, D.E., Nagode, D.A., Donaldson, K.J., Yin, P., Shamma, S.A., Fritz, J.B., *et al.* (**2017**). “Orbitofrontal cortex neurons respond to sound and activate primary auditory cortex neurons,” Cereb. Cortex, 1-12. doi: 10.1093/cercor/bhw409

Wostmann, M., Herrmann, B., Maess, B., and Obleser, J. (**2016**). “Spatiotemporal dynamics of auditory attention synchronize with speech,” Proc. Natl. Acad. Sci. USA, **113**, 3873-3878. doi: 10.1073/pnas.1523357113

Fluctuation contrast and speech-on-speech masking: Model midbrain responses to simultaneous speech

LAUREL H. CARNEY[1,2,*]

[1] *Departments of Biomedical Engineering, Neuroscience, and Electrical & Computer Engineering, Del Monte Institute for Neuroscience, University of Rochester, Rochester, NY, USA*

[2] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

At the level of the auditory midbrain, low-frequency fluctuations within each frequency channel drive neurons with band-pass modulation transfer functions (MTFs). The amplitude of low-frequency fluctuations in ascending neural signals is affected by stimulus amplitude due to the gradual saturation of the inner hair cells (IHCs) beginning at moderate sound levels. This level dependence of low-frequency fluctuation amplitudes results in contrast cues at the level of the midbrain: Spectral peaks result in lower responses of cells with bandpass-MTFs, whereas spectral valleys result in higher responses. Here, we focus on model population midbrain responses with different best-modulation frequencies (BMFs) to simultaneous speech. Midbrain responses were simulated for single hearing-in-noise (HINT) sentences and for a pair of simultaneous sentences, spoken by a male and a female. Correlations between population responses to individual male (or female) sentences and responses to simultaneous sentences vary with BMF in the range of the male (or female) fundamental frequencies. The pattern of fluctuation contrast across frequency in the midbrain representation provides a framework for studying speech-on-speech masking for listeners with normal hearing and sensorineural hearing loss.

INTRODUCTION

Neural responses to complex sounds such as speech convey a number of interacting cues to the central nervous system. A better understanding of which cues are most significant for encoding speech information should improve the ability to restore or enhance these cues for listeners with hearing loss. Furthermore, it is important to understand how candidate cues are affected by hearing loss as well as by typical challenges such as masking. In this study, we use computational models to explore fluctuation contrast cues in response to spoken sentences. In particular, we show examples of how these cues are affected by a competing voice, in anticipation of future tests that will take advantage of listener performance data collected using the Competing Voices Test (CVT; Bramsløw *et al.*, 2014) with the Danish Hearing-in-Noise (HINT) sentences (Nielsen and Dau, 2009). We also explore examples of model

*Corresponding author: laurel.carney@rochester.edu

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
ISBN: 978-87-990013-6-1.

responses that illustrate how sensorineural hearing loss affects fluctuation contrast cues in response to single speakers and competing voices.

One goal of this work is to test the hypothesis that across-frequency contrasts in fluctuation provide cues for the locations of formants (Carney *et al.*, 2015) and other spectral features in speech. Fluctuation contrasts are set up in the auditory periphery, and ultimately provide a robust representation in the responses of midbrain neurons. Neurons in the auditory midbrain (inferior colliculus, IC) are sensitive to low-frequency fluctuations, or periodicities, in their inputs. Auditory-nerve (AN) fibres phase-lock to the low-frequency fine structure in stimuli, and they simultaneously phase-lock to low-frequency fluctuations in response to complex sounds (Joris and Yin, 1992; review: Joris *et al.*, 2004). The low-frequency fluctuations in AN responses to speech include both the fundamental frequency of voiced sounds and the features of cochlear-filter induced envelopes in response to noisy sounds (Joris, 2003) such as fricative consonants.

Low-frequency fluctuations in peripheral responses are not limited to low characteristic frequencies (CFs) but occur across the entire AN population. These fluctuations in the neural responses are conveyed via the brainstem to the midbrain. AN fibres that convey fluctuations may have saturated average discharge rates, especially in the case of the sensitive high-spontaneous-rate (HSR) AN fibres, but their temporal patterns still convey substantial information in the form of phase-locking to fine-structure and low-frequency fluctuations. These low-frequency fluctuations are effective in exciting (or suppressing) midbrain neurons (Krishna and Semple, 2000; Nelson and Carney, 2007; Kim *et al.*, 2015). In the IC, average discharge rates depend on the amplitudes of low-frequency fluctuations on their inputs. Thus, changes across frequency in the amplitude of peripheral fluctuations carried by AN fibres result in a profile of midbrain rates that vary across frequency.

The above description is focused on how AN fibres carry the fluctuations in complex sounds in to the central nervous system (CNS). However, in the healthy ear, the representation of spectral features by these fluctuations is strongly shaped by two nonlinearities in the inner ear. First, near spectral peaks, such as formants in voiced sounds, the saturation of inner hair cell (IHC) transduction results in a “flattening” of the low-frequency fluctuations, or envelope-related features. As a result, AN fibres near spectral peaks have relatively low-amplitude low-frequency fluctuations, and instead have temporal responses that are dominated by a single harmonic closest to the spectral peak. This phenomenon is referred to as “synchrony capture” because it has been quantified on the basis of fine-structure phase-locking (Miller *et al.*, 1997), but it could equally well be referred to as a suppression of the low-frequency fluctuation. Importantly, the saturation of IHCs in frequency channels near spectral peaks results in changes in fluctuation amplitude across frequency channels, referred to here as fluctuation contrasts.

The second inner-ear nonlinearity that plays a role in shaping fluctuation contrasts is compressive cochlear amplification, which determines the sensitivity of the organ of corti response, and thus the set-point of the IHC transduction nonlinearity. Because (in the healthy ear) cochlear sensitivity is controlled in a frequency-dependent manner, fluctuation contrasts can occur at each spectral peak in a complex sound. For example, synchrony capture occurs for multiple formant peaks in AN vowel responses, not only at the highest magnitude peak (Delgutte and Kiang, 1984).

Contrasts in the amplitude of fluctuations in AN responses across frequency channels provide a code for spectral peak frequencies (Carney *et al.*, 2015), but this code requires both frequency-dependent cochlear amplification, driven by the outer hair cells (OHCs), and sensitive transduction by IHCs. Therefore, sensorineural hearing loss involving reduced sensitivity of OHCs and/or IHCs will result in decreased fluctuation contrasts. The impact of modelled sensorineural hearing loss on model IC population responses is explored here.

Previous studies of the fluctuation contrast cues have reported model IC and physiological responses for gaussian-noise maskers in normal-hearing rabbits (Carney *et al.*, 2015). Responses of models with sensorineural hearing loss to speech with additive Gaussian noise have also been described (Carney *et al.*, 2016). Here we extend this exploration to competing-voice maskers. We hypothesize that the ability to segregate speakers with different fundamental frequencies requires a) valid fluctuation contrast cues set up in peripheral responses, and b) midbrain neurons tuned to modulation frequencies in the range of voice pitch (e.g., Langner and Schreiner, 1988). Using a simple model for band-enhanced modulation tuning in the IC, we can study the ability to separate responses of speakers with different F0s based on midbrain responses.

MODEL METHODS

The example responses illustrated here were created using the Zilany *et al.* (2014) AN model as the input to a simple modulation filter model for IC neurons (Mao *et al.*, 2013). This IC model is a simplification of the same-frequency inhibition-excitation model of Nelson and Carney (2004); The modulation filter model allows more flexible selection of the IC best modulation frequency (BMF). Modulation filter responses were passed through a first-order low-pass filter (Fc = 500 Hz) to approximate the frequency limit of temporal following in the IC (Joris *et al.*, 2004).

Model responses with sensorineural hearing loss shown here were simulated by reducing the sensitivity of the AN model IHC by setting the parameter CIHC to 0.2 for all CFs, and by reducing the cochlear amplification of the IHCs by setting the parameter CIHC to 0.2 for all CFs. This simple strategy for simulating sensorineural hearing loss results in a model with sloping hearing loss that ranges from ~15-20 dB at low frequencies up to ~40 dB at high frequencies. More detailed audiometric configurations for comparison to individual listener results can be modelled by fine tuning these parameters as a function of frequency.

RESULTS

Responses of model populations of IC neurons driven by fluctuations in the normal-hearing AN model are shown in Fig. 1. Four population responses to HINT sentences

(Nilsson *et al.*, 1994) are illustrated. Figure 1A shows the response of a population of model IC cells to a HINT sentence spoken by a male. The IC model neurons have CFs from 200-6000 Hz, and all cells have BMF=100 Hz, in the range of the F0 for a male speaker. IC modulation filters are broad (quality index, Q=1); therefore, small changes in F0 over the course of a sentence do not cause large changes in the responses of these filters. This population of IC cells are most strongly driven by features associated with a male voice.

Figure 1B shows responses of IC model cells with BMF = 200 Hz, in the range of female voice pitch, in response to a HINT sentence spoken by a female. In Figs. 1A-1B, the formant frequencies during voiced portions of the sentences are indicated by white circles, which represent frequencies at which the responses to fluctuations are suppressed due to IHC saturation. The yellow (white) regions represent frequency channels that respond strongly to low-frequency fluctuations that are set up in the periphery. Note that the AN-fibre average rates are saturated across all of the low to mid frequencies during the voiced sounds, but the differences in the amplitudes of the low-frequency fluctuations result in strong formant-related features in the model IC responses. During the consonants (e.g., cyan square), fluctuations are strongest in frequency bands associated with the rising slope of the stimulus spectra, rather than at the peak frequencies where IHCs are often saturated and thus fluctuation amplitudes are reduced.

Figures 1C and 1D show responses of the same IC population models to simultaneous presentation of the two sentences from Figs. 1A and 1B. In both panels, fluctuation contrast features from both sentences are apparent. However, the IC population tuned to BMFs in the range of male pitch (Fig. 1C) is dominated by features in the sentence spoken by the male (e.g., white circles). Likewise, in the model cells tuned to an F0 in the female range (Fig. 1D), the fluctuation contrast features are most similar to those in the response to the female sentence (e.g., orange circles).

To quantify the degree of similarity of the response to the masked sentence to the response to single sentences, correlations between the two-dimensional (CF × time) images were computed. Similar to conventional studies of masking, a comparison between the response to the target plus masker and the response to the masker alone provides an indication of how well the target can be segregated from the masker based on this representation. In this case, a lower correlation between two images indicates better separability of the target from the masker based on the fluctuation contrast features in the IC responses.

Figure 2 shows the correlations between T+M and M-alone as a function of the BMF of the IC neurons. Each point in Fig. 2A represents the correlation of two 2D images of IC population responses, such as those shown in Fig. 1. IC cells with BMFs in the range of the male voice have the largest differences (lowest correlations) between the response to the male speaker alone and the response to the simultaneous sentences. Similarly, the responses of IC cells with BMFs near the female F0 would be best able to segregate the female-alone response from the competing-voice masker.

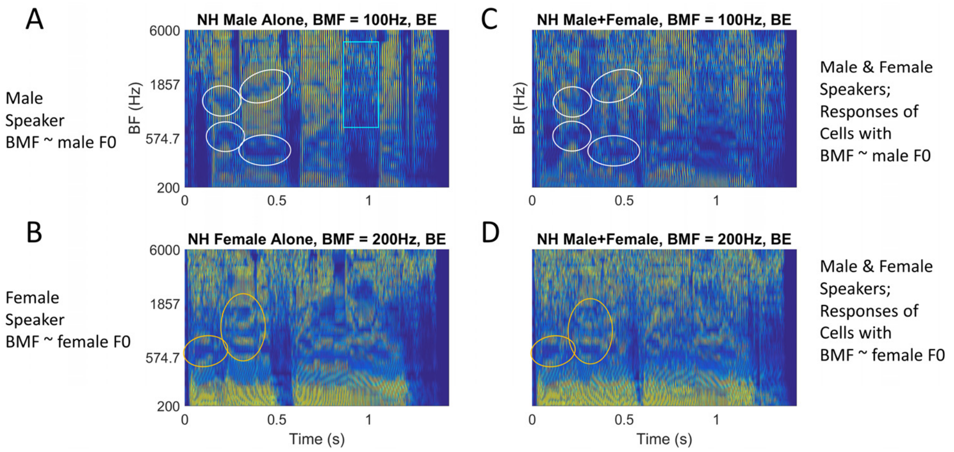


Fig. 1: Examples of population IC model responses to HINT sentences. Model responses here received inputs a population of normal-hearing (NH) high-spontaneous-rate AN fibres with characteristic frequencies ranging from 200-6 kHz. Each IC model is a bandpass modulation filter; best modulation frequencies are in the range of F0s for male (100 Hz, A, C) or female (200 Hz, B, D) speakers. A) Responses of a 100-Hz BMF IC model population to a male saying "Strawberry jam is sweet." White circles highlight response features for F1 and F2 in the first 2 words. The cyan square highlights the response to a fricative. B) Responses of a 200-Hz BMF IC population to a female of saying "They heard a funny noise." Orange circles highlight a few response features. C) 100-Hz BMF population response to simultaneous presentation of the sentences in A and B. D) 200-Hz BMF population response to the simultaneous sentences from A and B. Qualitatively, the response to the combined sentences contain fluctuation features from both sentences, but the neurons with 100-Hz BMF represent features that are more similar to the male-alone responses (e.g., white circles), and the 200-Hz responses are more similar to the female-alone response (e.g., orange circles). Sentences are from the English HINT test (Nilsson *et al.*, 1994). (color online)

Figure 2B shows similar calculations for IC models that include sensorineural hearing loss (SNHL) in the model AN inputs. At first glance, the more linear models for SNHL ear result in responses to the voices that appear to be more separable; However, the representation of many features in the sentences are reduced by decreased sensitivity. Figure 3 shows population responses for the models with SNHL; These responses include fewer cross-frequency contrasts associated with vowel formants, and very little response to consonants, as compared to responses of the normal-hearing model (cf. Fig. 1). The dotted lines in Fig. 2B show the same calculations after linear amplification of the speech inputs by 20 dB. This amplification restores the representation of some features in the populations responses (not shown), but decreases the separation of the correlations shown in Fig. 2B.

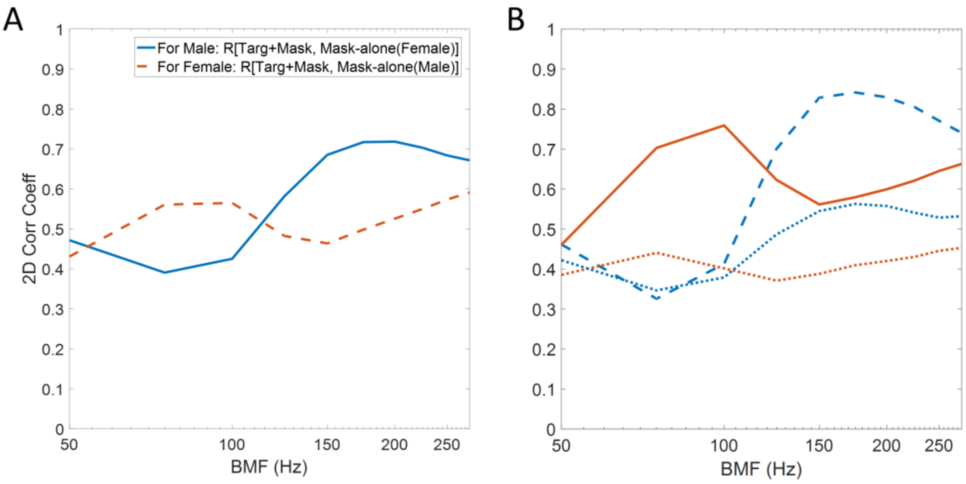


Fig. 2: Correlations between IC responses at each BMF to Target + Masker (i.e. two simultaneous sentences) and Masker alone (one sentence). Sentences are the same as those illustrated in Fig. 1. A) Blue: For the segregation of the male voice, the correlation considered is that between simultaneous sentence (Target + Masker) and Female (Masker alone). This correlation is lowest for IC population responses that have BMFs near the male speaker's F0. Red: Similarly, for segregation of the female voice, the lowest correlations between (Target + Masker) and Male (Masker alone) occur for IC cells with BMFs in the range of the female F0. B) Solid and dashed: Same as A, but for simulations with SNHL in AN model. Dotted lines: Same as above, except that a simple linear gain of 20 dB was included to increase audibility of speech features in the model population response. (color online)

SUMMARY

This presentation described initial steps towards describing cross-frequency fluctuation contrasts in midbrain responses and their potential for representing sentences in the presence of competing-voice maskers. The fluctuation contrast cues vary across populations of IC neurons with different BMFs, allowing segregation of speech with different F0s. Sensorineural hearing loss reduces the contrasts. Future work will apply these models and correlations to a larger set of sentences for which listener performance data has been collected. The effects of practical amplification schemes on the competing-voice maskers can also be explored for models with different audiometric configurations.

Further work is required to quantitatively evaluate the correspondence between changes in fluctuation contrast cues and changes in listener performance with competing voice maskers. This work could also be extended to studies of competing voices with similar F0s (same gender; e.g., Bramslow *et al.*, 2017).

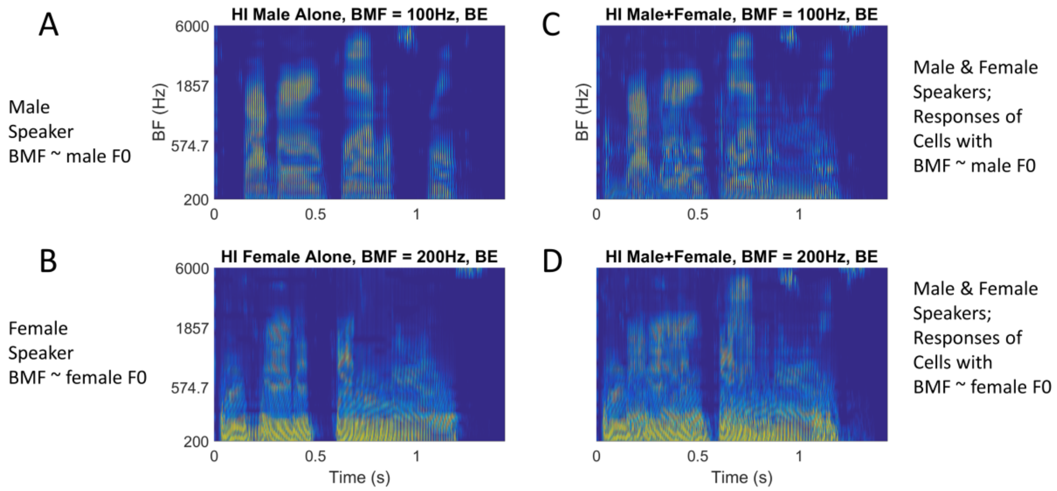


Fig. 3: Examples of population IC model responses to HINT sentences. Model responses here received inputs from a population of HSR AN fibres with sensorineural hearing loss (sloping loss with thresholds elevated by ~10 dB at low CFs to ~40 dB at high CFs). HINT sentences are the same as in Fig. 1. A) Responses of a 100-Hz BMF IC model population to male sentence. B) Responses of a 200-Hz BMF IC population to a female sentence. C) 100-Hz BMF population response to simultaneous sentences. D) 200-Hz BMF population response to the simultaneous sentences. (color online)

ACKNOWLEDGEMENTS

This study was initiated during a sabbatical stay as a Visiting Professor in the Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark. Supported by NIH grants DC001641 and DC010813, and by a fellowship from the Hanse Wissenschaftskolleg, Delmenhorst, Germany.

REFERENCES

Bramsløw, L., Vatti, M., Hietkamp, R.K., and Pontoppidan, N.H. (**2014**). "Design of a competing voices test," Poster presented at International Hearing Aid Conference (IHCON).

Bramsløw, L., Vatti, M., Rossing R., and Pontoppidan, N.H. (**2017**). "An improved competing voices test for test of attention," Proc. ISAAR, **6**, 279-286.

Carney, L.H., Li, T., and McDonough, J.M. (**2015**). "Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations," Eneuro, **2**, ENEURO-0004.

Carney, L.H., Kim, D.O., and Kuwada, S. (**2016**). "Speech coding in the midbrain: Effects of sensorineural hearing loss," in *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing* (Springer), Adv. Exp. Med. Biol., **894**, 427-435. PMID: 27080684

Delgutte, B., and Kiang, N.Y. (**1984**). “Speech coding in the auditory nerve: I. Vowel-like sounds,” J. Acoust. Soc. Am., **75**, 866-878.

Joris, P.X., and Yin, T.C. (**1992**). “Responses to amplitude-modulated tones in the auditory nerve of the cat,” J. Acoust. Soc. Am., **91**, 215-232.

Joris, P.X. (**2003**). “Interaural time sensitivity dominated by cochlea-induced envelope patterns,” J. Neurosci., **23**, 6345-6350.

Joris, P.X., Schreiner, C.E., and Rees, A. (**2004**). “Neural processing of amplitude-modulated sounds,” Physiol. Rev., **84**, 541-577.

Kim, D.O., Zahorik, P., Carney, L.H., Bishop, B.B., and Kuwada, S. (**2015**). “Auditory distance coding in rabbit midbrain neurons and human perception: monaural amplitude modulation depth as a cue,” J. Neurosci., **35**, 5360-5372.

Krishna, B.S., and Semple, M.N. (**2000**). “Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus,” J. Neurophysiol., **84**, 255-273.

Langner, G., and Schreiner, C.E. (**1988**). “Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms,” J. Neurophysiol., **60**, 1799-1822.

Mao, J., Vosoughi, A., and Carney, L.H. (**2013**). “Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues,” J. Acoust. Soc. Am., **134**, 396-406.

Miller, R.L., Schilling, J.R., Franck, K.R., and Young, E.D. (**1997**). “Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers,” J. Acoust. Soc. Am., **101**, 3602-3616.

Nelson, P.C., and Carney, L. H. (**2004**). “A phenomenological model of peripheral and central neural responses to amplitude-modulated tones,” J. Acoust. Soc. Am., **116**, 2173-2186. PMCID: PMC1379629

Nelson, P.C., and Carney, L.H. (**2007**). “Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus,” J. Neurophysiol., **97**, 522-539.

Nielsen, J.B., and Dau, T. (**2009**). “Development of a Danish speech intelligibility test,” Int. J. Audiol., **48**, 729-741.

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.

Zilany, M.S.A., Bruce, I.C., and Carney, L.H. (**2014**). “Updated parameters and expanded simulation options for a model of the auditory periphery,” J. Acoust. Soc. Am., **135**, 283-286. PMCID: PMC3985897

Can long-term exposure to non-damaging noise lead to hyperacusis or tinnitus?

MARTIN PIENKOWSKI[*]

Osborne College of Audiology, Salus University, Philadelphia, PA, USA

Hearing loss triggers changes in the central auditory system, some maladaptive. A region of primary auditory cortex (A1) deprived of input responds more strongly to cochlear lesion-edge frequencies, and its spontaneous firing rate (SFR) increases. This spontaneous and sound-evoked hyperactivity has been associated with tinnitus and hyperacusis, respectively. Regional increases in A1 spontaneous and sound-evoked activity are also observed after long-term exposure to non-damaging levels of noise. Adult cats exposed to such noise bands had suppressed SFR and evoked activity in the A1 region mapped to the noise band, but had increased SFR and evoked activity in A1 regions above and below the band. We hypothesized that, post-exposure, frequencies within the noise band should for some time be perceived as softer than before (hypoacusis), whereas frequencies outside of the noise band might be perceived as louder than before (hyperacusis), and might even be internalized as tinnitus. To investigate this possibility, adult CBA/Ca mice were exposed for >2 months to 8–16 kHz bandpass noise at 70 dB SPL, and tested for hypo/hyperacusis and tinnitus using prepulse inhibition (PPI) of the acoustic startle reflex (ASR), and gap-PPI of the ASR (GPIAS), respectively. ABRs and DPOAEs showed that the 70 dB SPL exposure was indeed non-damaging, whereas the same noise band at 75 dB SPL appeared to cause cochlear synaptopathy. Contrary to hypothesis, long-term exposure to non-damaging noise had no significant effect on PPI ASR and GPIAS testing. These negative findings nevertheless have important implications for PPI and GPIAS testing, and for the mechanisms of tinnitus and hyperacusis.

INTRODUCTION

Loud noise exposure can destroy cochlear outer and inner hair cells (OHCs and IHCs) and nerve fibers (ANFs), resulting in permanent hearing loss, typically at higher sound frequencies (Kujawa and Liberman, 2015). This can trigger a number of changes, some maladaptive, at various levels of the central auditory system. For example, the high-frequency area of primary auditory cortex (A1), when deprived of cochlear input, becomes more active spontaneously, and responds more strongly to input from the better-preserved mid-frequency turn of the cochlea, which becomes over-represented in A1 (Eggermont, 2017). Although this can enhance aspects of hearing at the over-represented frequencies, it could also lead to tinnitus – phantom ringing or hissing perceived as originating in the ears or head (Eggermont, 2012), and to hyperacusis – a reduced tolerance of moderate to loud sounds (Tyler *et al.*, 2014; Pienkowski *et al.*, 2014).

*Corresponding author: mpienkowski@salus.edu

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

Spontaneous central hyperactivity has been linked with behavioral evidence of tinnitus in animals with permanent noise-induced hearing loss (Kaltenbach *et al.*, 2004; Engineer *et al.*, 2011; Li *et al.*, 2013; Ropp *et al.*, 2014; Coomber *et al.*, 2014; Basura *et al.*, 2015; Longenecker and Galazyuk, 2016; Wu *et al.*, 2016; Sturm *et al.*, 2017), and with temporary hearing loss induced by salicylate (Eggermont and Kenmochi, 1998). Likewise, sound-evoked central hyperactivity has been linked with animal models of hyperacusis after noise trauma (Sun *et al.*, 2012; Chen *et al.*, 2013; Hickox and Liberman, 2014), salicylate injection (Turner and Parish, 2008; Sun *et al.*, 2009), and hereditary progressive hearing loss (Carlson and Willott, 1996; Ison *et al.*, 2007; Xiong *et al.*, 2017). Still, aspects of these animal data are puzzling (see Discussion), and evidence from human brain imaging studies linking spontaneous central hyperactivity with tinnitus (Elgoyhen *et al.*, 2015), and sound-evoked hyperactivity with hyperacusis (Gu *et al.*, 2010), remains more preliminary.

Regional increases in A1 spontaneous and sound-evoked activity are also observed after long-term exposure to non-damaging levels of noise (Pienkowski and Eggermont, 2011). In a series of studies on adult cats exposed to various tone pip ensembles and noise bands at ~70 dB SPL for weeks to months at a time, it was shown that A1 responses were strongly suppressed at frequencies within the exposure band (particularly at its edges), but were generally enhanced at frequencies above and/or below the exposure band (Pienkowski and Eggermont, 2009; 2010a; 2010b; Pienkowski *et al.*, 2011; 2013; note: the seminal study by Noreña *et al.*, 2006, used an exposure of ~80 dB SPL). We attributed the suppression to a homeostatic reduction in central gains in response to the persistent sound stimulus, and the enhancement to decreased lateral inhibition from the suppressed region (Pienkowski and Eggermont, 2012). These changes slowly reversed (also over weeks or months) after the end of the exposure (Pienkowski and Eggermont, 2009; 2010a; 2010b). Interestingly, the spontaneous hyperactivity was generally seen in the enhanced regions of A1 (Noreña *et al.*, 2006; Pienkowski and Eggermont, 2009; 2010b; Munguia *et al.*, 2013), not in the deprived region as is the case with permanent hearing loss (Eggermont, 2017). Given these data, we wondered whether long-term exposure to non-damaging noise could lead to hyperacusis or tinnitus. Specifically, we hypothesized that, post-exposure, frequencies within the noise band should for some time be perceived as softer than before (hypoacusis), whereas frequencies outside of the noise band might be perceived as louder than before (hyperacusis), and might even be internalized as tinnitus.

To investigate this possibility, adult CBA/Ca mice were exposed for >2 months to 8–16 kHz bandpass noise at ~70 dB SPL, and tested for hypo/hyperacusis and tinnitus using prepulse inhibition (PPI) of the acoustic startle reflex (ASR), and gap-PPI of the ASR (GPIAS), respectively. Auditory brainstem responses (ABRs) and distortion product otoacoustic emissions (DPOAEs) were used to show that the 70 dB SPL exposure was indeed non-damaging, whereas the same stimulus at 75 dB SPL appeared to cause cochlear synaptopathy. Contrary to hypothesis, long-term exposure to non-damaging noise had no significant effect on PPI ASR and GPIAS testing. As will be discussed, these negative findings nevertheless have important implications for PPI and GPIAS testing, and for the mechanisms of tinnitus and hyperacusis.

METHODS

Animals and noise exposure

This work was approved by the Institutional Animal Care and Use Committee of Salus University. Nine normal-hearing male CBA/Ca mice (Jackson Laboratories) served as subjects in the main experiment, and were exposed bilaterally for 2 months continuously to sharply-filtered 8–16 kHz noise at ~70 dB SPL, beginning at about 3 months of age. Six mice served as unexposed controls, and another six were exposed to the same 8–16 kHz noise at ~75 dB SPL. The noise was synthesized in Adobe Audition, and played out by a free-field loudspeaker (Tucker Davis Technologies [TDT], Model MF1), which was mounted just above the cages housing the mice. All mice were kept on a 12-h light/dark schedule (light 8 am–8 pm) and were given free access to food and water. There were no signs of long-term distress in any of the noise-exposed mice.

Assessment of loudness perception and tinnitus using the acoustic startle reflex

The ASR is a protective reflex elicited by an intense sound (Koch, 1999). In mice, it involves a whole-body flinch and jump, the force of which was measured using a motion-sensitive platform in an anechoic foam-lined, sound-attenuating startle chamber (San Diego Instruments, SR-LAB). ASR amplitudes can be reduced by preceding the intense, startling sound with a less-intense, non-startling “prepulse”, known as prepulse inhibition (PPI) of the ASR. The degree of ASR reduction, termed the magnitude of the PPI, is related to the behavioral salience of the prepulse. For example, the greater the perceived loudness of the prepulse, the greater the magnitude of the PPI (e.g., Carlson and Willott, 1996). Thus, an estimate of the rodent’s loudness function (i.e., perceived loudness vs. sound intensity) can be obtained by measuring the magnitude of the PPI as a function of the prepulse intensity, at a given prepulse frequency. The GPIAS variant of PPI (Galazyuk and Hébert, 2015) substitutes a silent gap in a narrowband noise (NBN) background for the tone prepulse. It is believed that a ringing tinnitus with a similar pitch to the NBN background reduces the salience of the gap, decreasing the magnitude of the PPI. Figure 1 illustrates these ideas, including the stimulus parameters used in the present study. For PPI and GPIAS testing, startle stimuli were 20-ms bursts of broadband noise (BBN) at 105 dB SPL. Tone prepulses were also 20 ms long (including 1 ms on/off $\cos^2$ ramps), preceded the startle noise by 100 ms (onset-to-onset), and were presented at 50 or 70 dB SPL. For GPIAS testing, silent gaps 20 or 50 ms long were embedded in 1/3-octave NBN at 65 dB SPL, and also preceded the startle burst by 100 ms. These sound stimuli were synthesized using Adobe Audition, and played out by a HiVi Isodynamic Tweeter (Model RT2C-A). Stimulus levels were calibrated with a 1/4 inch ACO Pacific microphone (Model 7017) placed inside the startle chamber.

Figure 2A shows the experiment timeline, and Fig. 2B a block diagram of a single startle session. Each session consisted of 362 startle trials with an average inter-startle interval of 5 s (range 3–7 s), for a total session time of ~30 min. GPIAS testing was conducted at NBN frequencies of 6, 8, 11, 16, 23 and 32 kHz, and in BBN. PPI testing was conducted using tone prepulses at frequencies of 4, 6, 8, 11, 16, 23 and 32 kHz. Each gap-in-noise or prepulse frequency was presented in a block of 21 trials in pseudorandom order, with 7 startle-only trials, and 7 trials each for 20 and 50-ms gaps, or for 50 and 70 dB SPL prepulses. The ratio of the ASR amplitude with a gap or a prepulse to that without a gap

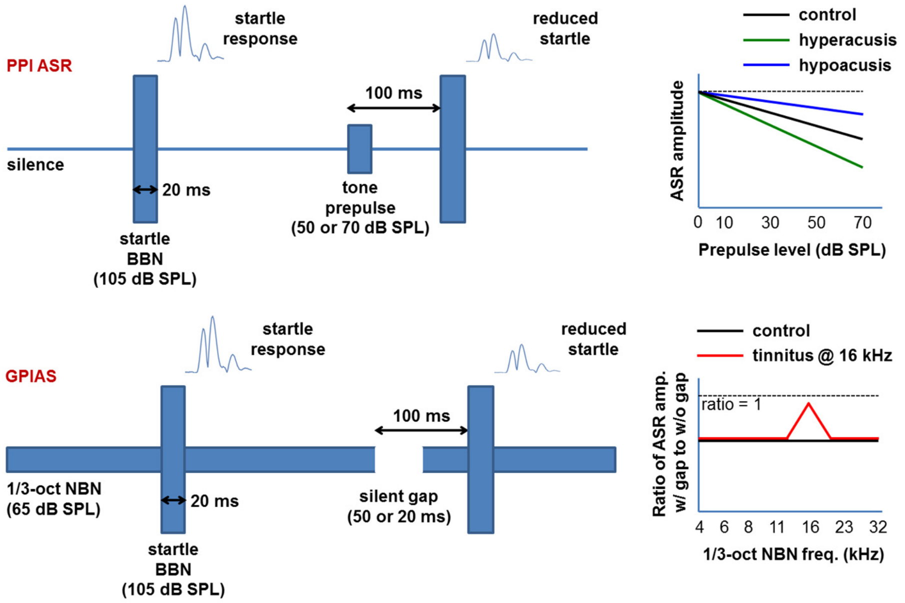


Fig. 1: Schematic diagrams of PPI ASR (top) and GPIAS (bottom), including the stimulus parameters used in the present study. Also shown are hypothetical PPI functions in animals with hypo- and hyperacusis (top-right), and hypothetical GPIAS results that are positive for tinnitus at 16 kHz (bottom-right).

or prepulse was calculated for each block, and constituted the “raw data” for the session. In addition, 3 blocks of “I/O functions” were run, in which startle-only amplitudes were measured in response to 20 ms-long tones (including 1 ms on/off ramps) at 4, 6, 8, 11, 16 and 23 kHz, and to BBN, at both 85 and 105 dB SPL. Finally, one block of startle-only trials to 105 dB SPL BBN was measured at the beginning and end of the session to track within-session adaptation of the startle response. Each mouse completed many such sessions (see below), with the order of the GPIAS and PPI blocks interchanged and randomized between sessions to offset adaptation effects. All sessions were conducted during the day, but in darkness, with the lights off inside the startle box.

Prior to the first ASR test session, the 9 mice were gradually acclimated to the startle chamber and test stimuli over a period of 2 weeks. Each mouse was then tested during 12 sessions, as described above, and the final results were averaged across sessions. Each mouse was limited to one session per day, and completed the 12 sessions over a period of 3–4 weeks. This was followed by the 2 month noise exposure, and then another 3–4 weeks of ASR testing after a short startle re-acclimation period. During this post-exposure ASR testing, the noise stimulus was left on for 12 h each night (8 pm–8 am). The mice were tested in random order during the day, but no earlier than 10am, 2 h after noise offset for the day. Maintaining the noise exposure at night eliminated the potential confound on post-exposure testing of the gradual reversal of noise-induced changes after the cessation of exposure (Pienkowski and Eggermont, 2009; 2010a; 2010b).

Startle response analysis was automated using custom software written in Mathematica. Reliable responses were identified using a template-matching algorithm similar to that

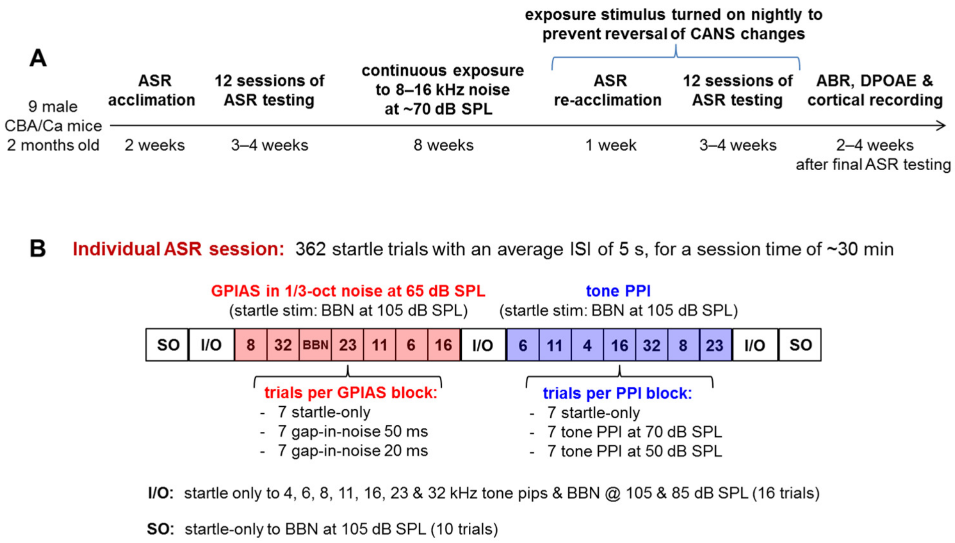

A
9 male
CBA/Ca mice
2 months old
ASR
acclimation
2 weeks
12 sessions of
ASR testing
3–4 weeks
continuous exposure
to 8–16 kHz noise
at ~70 dB SPL
8 weeks
exposure stimulus turned on nightly to
prevent reversal of CANS changes
ASR
re-acclimation
1 week
12 sessions of
ASR testing
3–4 weeks
ABR, DPOAE &
cortical recording
2–4 weeks
after final ASR testing
B
Individual ASR session: 362 startle trials with an average ISI of 5 s, for a session time of ~30 min
GPIAS in 1/3-oct noise at 65 dB SPL
(startle stim: BBN at 105 dB SPL)
tone PPI
(startle stim: BBN at 105 dB SPL)
SO
I/O
8
32
BBN
23
11
6
16
I/O
6
11
4
16
32
8
23
I/O
SO
trials per GPIAS block:
- 7 startle-only
- 7 gap-in-noise 50 ms
- 7 gap-in-noise 20 ms
trials per PPI block:
- 7 startle-only
- 7 tone PPI at 70 dB SPL
- 7 tone PPI at 50 dB SPL
I/O: startle only to 4, 6, 8, 11, 16, 23 & 32 kHz tone pips & BBN @ 105 & 85 dB SPL (16 trials)
SO: startle-only to BBN at 105 dB SPL (10 trials)

Fig. 2: A. Experiment timeline. **B.** Block diagram of a single startle session. See Methods for a detailed description.

described by Grimsley *et al.* (2015). Thousands of startle trials were checked by eye and the performance of the algorithm was found to be excellent. Startle amplitude was taken as the largest peak in each trace. Individual mouse and group-averaged ASR results were compared pre- and post-exposure using two-way ANOVAs with post-hoc Bonferroni tests.

DPOAE and ABR recording

DPOAEs and ABRs were measured from the left ears and mastoid areas 2 weeks following the completion of post-exposure (or control) ASR sessions. Mice were anesthetized with a mixture of 50 mg/kg ketamine and 10 mg/kg xylazine, injected intraperitoneally, and were topped up with half doses of this mixture as needed to maintain a state of areflexia. They were placed on a homeothermic blanket which kept their body temperature at 36.5°C, inside a single-walled sound-attenuating chamber (ETS-Lindgren). Following the completion of DPOAE and ABR testing, extracellular recordings were attempted from A1, after which the mice were sacrificed. However, the cortical data are too preliminary to report here.

ABRs were always recorded first. Stimuli were tone bursts at 4, 6, 8, 11, 16, 23 and 32 kHz, and were 3 ms in duration including 1-ms $\cos^2$ on and off ramps. Stimuli were synthesized using TDT software (SigGen), and played out by a TDT MF1 speaker coupled to the animal’s left ear canal with a 10 cm tube and sealed probe. Sound levels were calibrated with the probe coupled to a 1/4 inch microphone (ACO Pacific, 7017) with an additional 7 mm-long plastic tube, intended to approximate the length of the mouse ear canal. Stimuli were presented at 10–90 dB SPL at each frequency, in 10 dB steps, with 512 repetitions per level and a presentation rate of 21.1 /s. The ABR was recorded differentially between the left mastoid area and vertex (ground electrode in the nape of the neck) using subdermal needle electrodes (Rochester Electro-Medical, Inc., Model S83018-R9). Potentials were amplified, digitized, and filtered between 100 and 3,000 Hz under the control of TDT software (BioSig). At low stimulus levels, measurements were repeated twice, and ABR threshold was defined as the lowest SPL that yielded a reproducible ABR, minus 5 dB (half step size). Peak-to-trough amplitudes were then determined for mouse ABR waves 1–4 at all supra-threshold SPLs, if the wave was distinct.

Following ABR recording, DPOAEs were measured from the left ear using an OAE probe coupled to a pair of TDT MF1 speakers and to an Etymotic Research microphone (ER-10B+). Stimuli were synthesized using TDT software (SigGen). The frequency of the higher primary tone (f_2) was again set to 4, 6, 8, 11, 16, 23 or 32 kHz, and the frequency of the lower primary tone (f_1) was given by $f_1 = f_2/1.2$. Levels of f_1 (L_1) ranged from 20 to 80 dB SPL in 10 dB steps, with $L_2 = L_1 – 10$ dB. DPOAEs at frequency $2f_1–f_2$ were amplified and digitized under the control of TDT software (BioSig). DPOAE amplitudes are reported in units of dB V, and DPOAE threshold was defined as the lowest level of L_1 (again minus the step size of 5 dB) at which the DPOAE amplitude was above the 99% confidence interval for the microphone noise floor, averaged across the six frequency bins adjacent to $2f_1–f_2$.

RESULTS

ABRs and DPOAEs

Nine normal-hearing adult male CBA/Ca mice were exposed 24 h/day for 2 months and then 12 h/day for 1 month to sharply-filtered 8–16 kHz noise at 70 dB SPL. ABRs and DPOAEs were measured 2 weeks after the end of the 3 month exposure. They were compared to measurements made at the same age in six unexposed control mice, and in another six mice which were exposed to the same 8–16 kHz noise at 75 dB SPL. Group-averaged ABR results are shown in Fig. 3A (± 1 SE or standard error). There were no significant differences in ABR thresholds between the three groups (p=0.56 for the main effect of group across frequency; two-way ANOVA). ABR wave 1 input-output functions were not affected after exposure to 70 dB SPL noise, but were significantly reduced after exposure to 75 dB SPL at frequencies of 8 (p=0.031), 11 (p=0.004), and 16 kHz (p=0.007) (i.e., only at frequencies within the noise band; all other frequencies were p>0.05 as indicated). These p-values reflect the main effect of group across ABR stimulus level (two-way ANOVA), and were not corrected for multiple comparisons at the various stimulus SPLs. Note that none of the differences at individual stimulus SPLs were significant at the p=0.05 level after post-hoc Bonferroni correction, only the main effects. Importantly, no significant differences were found between groups in the amplitudes of ABR waves 2–4 (data not shown). Also, there were no significant differences between groups in DPOAE thresholds and DPOAE input-output functions at any primary tone frequency (Fig. 3B). As will be discussed, these results are consistent with mild noise-induced "hidden hearing loss" or cochlear synaptopathy (Kujawa and Liberman, 2015) following exposure at 75 dB SPL, but no noise-induced cochlear damage following exposure at 70 dB SPL.

PPI ASR

ASR results are compared pre- and post-exposure for the 9 mice exposed to 8–16 kHz noise at the non-damaging level of 70 dB SPL. It was hypothesized that this exposure would cause sound frequencies within the noise band to be perceived as softer than before (hypoacusis), whereas frequencies above and/or below the noise band would be perceived as louder than before (hyperacusis).

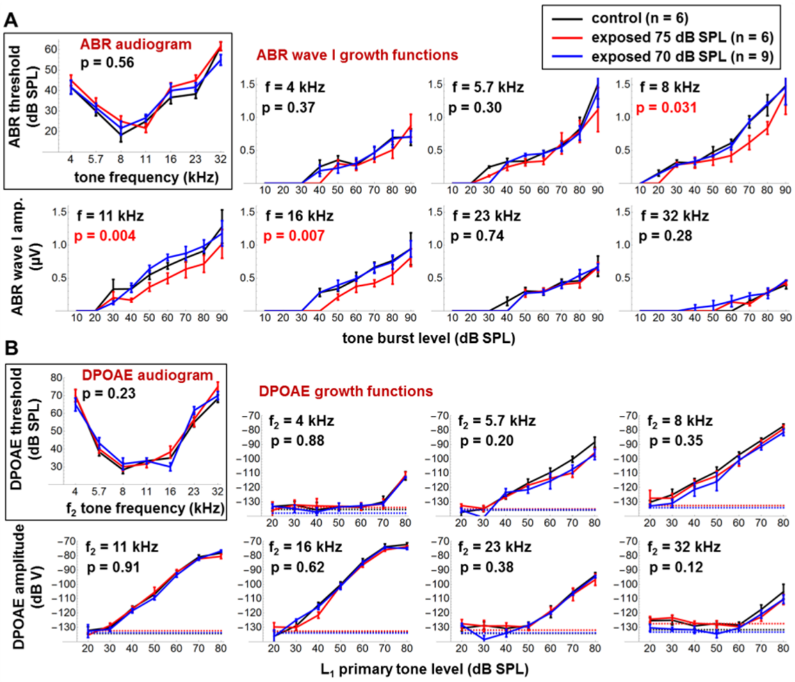


Fig. 3: A. ABR audiograms, and wave 1 growth functions at different stimulus frequencies, for unexposed control mice, and for mice exposed for 3 months to 8–16 kHz noise at 70 dB SPL and 75 dB SPL. Error bars show ±1 SE. **B.** As 3A but showing DPOAE audiograms and growth functions at different stimulus frequencies, for unexposed and exposed mice.

Figure 4A shows group-averaged ASR amplitudes (± 1 SE) in response to BBN and tonal startle stimuli presented at 85 dB SPL (black traces) and 105 dB SPL (grey traces). There were no significant differences pre- vs. post-exposure at any startle frequency or level (p>0.05, as indicated, for the main effect of group across frequency; separate two-way ANOVAs were run at each SPL). Group-averaged PPI results (± 1 SE) are shown in Fig. 4B, for prepulse levels of 50 dB SPL (black traces) and 70 dB SPL (grey traces). Note that the PPI effect was highly significant (p<<0.05) at the group level for all prepulse frequencies and SPLs (i.e., ASR amplitude ratios with vs. without prepulse are all <<1). However, again there were no significant differences post-exposure (p>0.05, as indicated). Individual animal results (not shown) also do not support the hypothesis that at least some of the mice may have developed frequency-specific hypo- or hyperacusis post-exposure.

GPIAS

GPIAS results are also reported pre- and post-exposure for the 9 mice exposed to 8–16 kHz noise at the non-damaging level of 70 dB SPL. Figure 5 shows group-averaged ASR amplitude ratios (± 1 SE) with and without 20-ms (black traces) and 50-ms (grey traces) silent gaps embedded in 1/3-octave NBN at a range of frequencies, and in BBN. As will be discussed further, GPIAS testing was performed using both 50 ms and 20 ms gaps because previous auditory cortical ablation studies have suggested that cortex is not essential for GPIAS with gaps of 50 ms or longer, but is required at gap durations of <30 ms (Ison *et al.*, 1991; Bowen *et al.*, 2003; Weible *et al.*, 2014). However, there

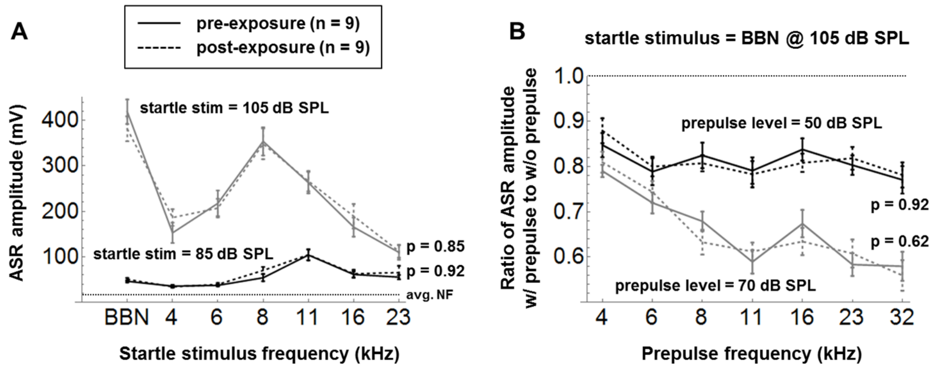


Fig. 4: A. Group-averaged ASR amplitudes (± 1 SE) in response to BBN and tonal startle stimuli presented at 85 dB SPL (black traces) and 105 dB SPL (grey traces), pre- and post-exposure. The average startle system noise floor is drawn in. **B.** Group-averaged PPI results (± 1 SE) for prepulse levels of 50 dB SPL (black traces) and 70 dB SPL (grey traces).

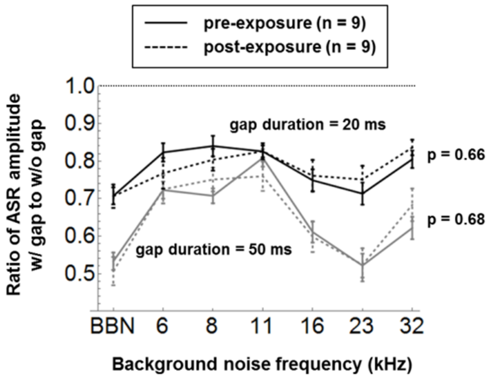


Fig. 5: Group-averaged GPIAS results (± 1 SE) for gap durations of 20 ms (black traces) and 50 ms (grey traces).

were no significant differences post-exposure with either 50 or 20-ms gaps, despite the fact that the GPIAS effect itself was highly significant (*p*<<0.05) at the group level for both gap durations in all NBN backgrounds (i.e., ASR amplitude ratios with vs. without gap are all <<1). Again, individual animal results (not shown) do not support the idea that at least some of the mice may have developed tinnitus post-exposure.

DISCUSSION

Effects of exposure to moderately loud noise on the auditory periphery

There was evidence of cochlear synaptopathy in CBA/Ca mice following a 3-month exposure in adulthood to 8–16 kHz noise at 75 dB SPL, but not at 70 dB SPL. ABR wave 1 amplitudes at suprathreshold SPLs were significantly reduced 2 weeks after the end of the 75 dB SPL exposure, and this was specific to stimuli at 8, 11 and 16 kHz (i.e., frequencies within the exposure band; Fig. 5A), while DPOAEs were not affected at any stimulus frequency (Fig. 5B), nor were ABR wave 2–4 amplitudes (data not shown). This pattern of damage differs to some extent from that observed following shorter exposures to louder noise. A study by Fernandez *et al.* (2015) compared the effects on CBA/Ca mice of 2 hour exposures to 8–16 kHz noise at 100 and 91 dB SPL. By 2 weeks post-exposure, DPOAEs had returned to pre-exposure baselines in both cases, while ABR wave 1 amplitudes were reduced (indicative of synaptopathy) only after the louder, 100 dB SPL exposure. However, the pattern of DPOAE and ABR temporary threshold shifts (TTS), measured at 1 day post-exposure, was instructive: After the 100 dB exposure, maximum TTS was observed at the highest frequencies tested, >30 kHz, but after the 91 dB exposure, maximum TTS developed at around 20 kHz, only slightly above the 8–16 kHz exposure band (Fernandez *et al.*, 2015). These results can likely be explained by the greater "spread of excitation" to more basal regions of the cochlea at higher exposure levels. At the more moderate level of 75 dB SPL (present study), this spread of excitation is small, and the damage appears to be limited to the cochlear region mapped to the exposure band. Maison *et al.* (2013) exposed CBA/Ca mice to 8–16 kHz noise at 84 dB SPL for 1 week, and found reduced ABR wave 1 amplitudes and IHC synapse counts 1 week post-exposure; as in the present study, the reduction was greatest at 8–16 kHz, although some reduction was also seen above the exposure band.

The present study appears to be the first to suggest that cochlear synaptopathy can occur after prolonged exposure to noise at levels as low as 75 dB SPL. At least in CBA/Ca mice, this is close to the threshold for a damaging exposure, as mice exposed to 70 dB SPL did not show any ABR-based evidence of synaptopathy. A recent opinion piece argued that humans are less susceptible than rodents to TTS and by extension to cochlear synaptopathy, and that the bandpass exposures used in the animal studies are not representative of real-world noise (Dobie and Humes, 2017). While these are fair points, the present data suggest that long-term exposure at currently permissible occupational levels (e.g., 85 dB A for 8 h/day; NIOSH, 1998; OSHA, 2002) may not in fact be safe for the ear.

Effects of exposure to moderately loud noise on the central auditory system

Preliminary data, not presented here, suggest similar noise-induced changes in mouse A1 to those reported previously in cats. These changes are also likely exhibited at the level of the thalamic medial geniculate body (MGB), as inferred from cortically-recorded local field potentials (Pienkowski and Eggermont, 2011). Lau *et al.* (2015) performed whole brain functional magnetic resonance imaging following long-term BBN exposure of adult rats at 65 dB SPL, and found reduced noise-evoked activation of A1 and MGB, but no change in the inferior colliculus (IC) and lower brainstem. Thus, it seems that the effects of non-damaging noise are mostly limited to the thalamocortex, while the effects of damaging noise are already prominent at the level of the IC and lower brainstem (Eggermont, 2017).

Negative PPI ASR and GPIAS findings after exposure to non-damaging noise

There were no significant exposure-induced changes in ASR amplitudes measured in response to tones and BBN at 85 and 105 dB SPL (Fig. 4A), and no changes in ASR amplitude ratios with and without tone prepulses at 50 and 70 dB SPL (Fig. 4B), and with and without 20 and 50-ms silent gaps embedded in NBN and BBN backgrounds (Fig. 5). Thus, there was no evidence of hypo/hyperacusis or tinnitus, as assessed by the ASR, in mice exposed to non-damaging noise.

A possible reason for why hypo/hyperacusis was not detected in the present study is that exposure to moderately loud noise causes changes mainly at the thalamocortical level (see above), whereas PPI of the ASR appears sensitive to changes mainly at the (pre-attentive) brainstem level, as suggested by previous studies on decerebrate rats (Davis and Gendelman, 1977; Fox, 1979; Li and Frost, 2000). Thus, PPI of the ASR should not be used to study the behavioral correlates of neural changes that are observed mainly at the thalamocortical but not at the brainstem level.

This caveat does not necessarily apply to the GPIAS form of PPI, which at present is widely adopted for tinnitus screening in rodents. In the present study, GPIAS testing was performed using both 50 and 20 ms gaps, as previous work has suggested that auditory cortex is not essential for GPIAS with gap durations of 50 ms or longer, but is required for gaps less than ~30 ms (Ison *et al.*, 1991; Bowen *et al.*, 2003; Weible *et al.*, 2014). Nevertheless, results here were negative for both gap durations, implying that moderately noise-exposed mice did not develop tinnitus, at least as assessed by GPIAS, in spite of our previous work showing that adult cats exhibited noise-induced regional increases in A1 spontaneous activity. If GPIAS screening for tinnitus can indeed be trusted (see below), the negative findings suggest that auditory cortical hyperactivity is not sufficient to cause tinnitus.

Recent studies on rats (Ropp *et al.*, 2014), guinea pigs (Coomber *et al.*, 2014), and mice (Longenecker and Galazyuk, 2016) found increased spontaneous firing rates (SFRs) in the IC of all noise-exposed animals, *including those testing negative for tinnitus using GPIAS.* Interestingly, Hickox and Liberman (2014) also failed to find GPIAS-based evidence of tinnitus in CBA/Ca mice after a damaging noise dose (8–16 kHz at 100 dB SPL for 2 h). This is especially surprising in light of more recent data showing that the 100 dB SPL noise dose increased IC SFRs by an even greater margin than the same noise at 105 dB SPL, which of course caused more damage to the cochlea (Hesse *et al.*, 2016).

Several studies have reported evidence of cochlear damage in people with tinnitus and clinically normal audiograms (Weisz *et al.*, 2006; Schaette and McAlpine, 2011; Gu *et al.*, 2012; Paul *et al.*, 2017), who comprise about 10% of all tinnitus cases. A large study of Danish workers reported that the prevalence of tinnitus increased with occupational noise exposure level and duration in workers with hearing loss, but was not associated with the noise dose in workers with clinically normal hearing (Rubak *et al.*, 2008). On the other hand, Guest *et al.* (2017) found a link between tinnitus and noise exposure history in young adults with normal audiograms, but no evidence of synaptopathy or other cochlear damage. Thus, it remains possible that despite the negative GPIAS results reported here, tinnitus could be induced by prolonged exposure to non-damaging noise even in the absence of hearing loss.

REFERENCES

Basura, G.J., Koehler, S.D., and Shore, S.E. (**2015**). "Bimodal stimulus timing-dependent plasticity in primary auditory cortex is altered after noise exposure with and without tinnitus," J. Neurophysiol., **114**, 3064-3075.

Bowen, G.P., Lin, D., Taylor, M.K., *et al.* (**2003**). "Auditory cortex lesions in the rat impair both temporal acuity and noise increment thresholds, revealing a common neural substrate," Cereb. Cortex., **13**, 815-822.

Carlson, S., and Willott, J.F. (**1996**). "he behavioral salience of tones as indicated by prepulse inhibition of the startle response: relationship to hearing loss and central neural plasticity in C57BL/6J mice," Hear. Res., **99**, 168-175.

Chen, G., Lee, C., Sandridge, S.A., *et al.* (**2013**). "Behavioral evidence for possible simultaneous induction of hyperacusis and tinnitus following intense sound exposure," J. Assoc. Res. Otolaryngol., **14**, 413-424.

Coomber, B., Berger, J.I., Kowalkowski, V.L., *et al.* (**2014**). "Neural changes accompanying tinnitus following unilateral acoustic trauma in the guinea pig," Eur. J. Neurosci., **40**, 2427-2441.

Davis, M., and Gendelman, P.M. (**1977**). "Plasticity of the acoustic startle response in the acutely decerebrate rat," J. Comp. Physiol. Psychol., **91**, 549-563.

Dobie, R.A., and Humes, L.E. (**2017**). "Commentary on the regulatory implications of noise-induced cochlear neuropathy," Int. J. Audiol., **56**, 74-78.

Eggermont, J.J., and Kenmochi, M. (**1998**). "Salicylate and quinine selectively increase spontaneous firing rates in secondary auditory cortex," Hear. Res., **117**, 149-160.

Eggermont, J.J. (**2012**). *The Neuroscience of Tinnitus.* Oxford University Press.

Eggermont, J.J. (**2017**). "Acquired hearing loss and brain plasticity," Hear. Res., **343**, 176-190.

Elgoyhen, A.B., Langguth, B., De Ridder, D., *et al.* (**2015**). "Tinnitus: perspectives from human neuroimaging," Nat. Rev. Neurosci., **16**, 632-642.

Engineer, N.D., Riley, J.R., Seale, J.D., *et al.* (**2011**). "Reversing pathological neural activity using targeted plasticity," Nature, **470**, 101-104.

Fernandez, K.A., Jeffers, P.W., Lall, K., *et al.* (**2015**). "Aging after noise exposure: Acce-leration of cochlear synaptopathy in "recovered" ears," J. Neurosci., **35**, 7509-7520.

Fox, J.E. (**1979**). "Habituation and prestimulus inhibition of the auditory startle reflex in decerebrate rats," Physiol. Behav., **23**, 291-297.

Galazyuk, A., and Hébert, S. (**2015**). "Gap-prepulse inhibition of the acoustic startle reflex (GPIAS) for tinnitus assessment: Current status and future directions," Front. Neurol., **6**, 88.

Grimsley, C.A., Longenecker, R.J., Rosen, *et al.* (**2015**). "An improved approach to separating startle data from noise," J. Neurosci. Meth., **253**, 206-217.

Gu, J.W., Halpin, C.F., Nam, E.C., *et al.* (**2010**). "Tinnitus, diminished sound-level tolerance, and elevated auditory activity in humans with normal hearing sensitivity," J. Neurophysiol., **104**, 3361-3370.

Gu, J.W., Herrmann, B.S., Levine, R.A., *et al.* (**2012**). "Brainstem auditory evoked potentials suggest a role for the ventral cochlear nucleus in tinnitus," J. Assoc. Res. Otolaryngol., **13**, 819-833.

Guest, H., Munro, K.J., Prendergast, G., *et al.* (**2017**). "Tinnitus with a normal audiogram: Relation to noise exposure but no evidence for cochlear synaptopathy," Hear. Res., **344**, 265-274.

Hesse, L.L., Bakay, W., Ong, H.C., *et al.* (**2016**). "Non-monotonic relation between noise expo-sure severity and neuronal hyperactivity in the auditory midbrain," Front. Neurol., **7**, 133.

Hickox, A.E., and Liberman, M.C. (**2014**). "Is noise-induced cochlear neuropathy key to the generation of hyperacusis or tinnitus?" J. Neurophysiol., **111**, 552-564.

Ison, J.R., O'Connor K., Bowen, G.P., *et al.* (**1991**). "Temporal resolution of gaps in noise by the rat is lost with functional decortication," Behav. Neurosci., **105**, 33-40.

Ison, J.R., Allen, P.D., and O'Neill, W.E. (**2007**). "Age-related hearing loss in C57BL/6J mice has both frequency-specific and non-frequency-specific components that produce a hyperacusis-like exaggeration of the acoustic startle reflex," J. Assoc. Res. Otolaryngol., **8**, 539-550.

Kaltenbach, J.A., Zacharek, M.A., Zhang, J., *et al.* (**2004**). "Activity in the dorsal cochlear nucleus of hamsters previously tested for tinnitus following intense tone exposure," Neurosci. Lett., **355**, 121-125.

Koch, M. (**1999**). "The neurobiology of startle," Prog. Neurobiol., **59**, 107-128.

Kujawa, S.G., and Liberman, M.C. (**2015**). "Synaptopathy in the noise-exposed and aging cochlea: Primary neural degeneration in acquired sensorineural hearing loss," Hear. Res., **330**, 191-199.

Lau, C., Zhang, J.W., McPherson, B., *et al.* (**2015**). "Functional magnetic resonance imaging of the adult rat central auditory system following long-term, passive exposure to non-traumatic acoustic noise," Neuroimage, **107**, 1-9.

Li, L., and Frost, B.J. (**2000**). "Azimuthal directional sensitivity of prepulse inhibition of the pinna startle reflex in decerebrate rats," Brain Res. Bull., **51**, 95-100.

Li, S., Choi, V., and Tzounopoulos, T. (**2013**). "Pathogenic plasticity of Kv7.2/3 channel acti-vity is essential for the induction of tinnitus," Proc. Natl. Acad. Sci. USA., **110**, 9980-9985.

Longenecker, R.J., and Galazyuk, A.V. (**2016**). "Variable effects of acoustic trauma on behavioral and neural correlates of tinnitus in individual animals," Front. Behav. Neurosci., **10**, 207.

Maison, S.F., Usubuchi, H., and Liberman, M.C. (**2013**). "Efferent feedback minimizes cochlear neuropathy from moderate noise exposure," J. Neurosci., **27**, 5542-5552.

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.

NIOSH (**1998**). "Criteria for a recommended standard: Occupational noise exposure," National Institute for Occupational Safety and Health Publication No: 98-126.

Noreña, A.J., Gourévitch, B., Aizawa, N., *et al.* (**2006**). "Spectrally enhanced acoustic environ-ment disrupts frequency representation in cat auditory cortex," Nat. Neurosci., **9**, 932-939.

OSHA (**2002**). *Hearing Conservation.* Occupational Safety and Health Administration, U.S. Department of Labor, Publication No: OSHA 3074.

Paul, B.T., Bruce, I.C., and Roberts, L.E. (**2017**). "Evidence that hidden hearing loss underlies amplitude modulation encoding deficits in individuals with and without tinnitus," Hear. Res., **344**, 170-182.

Pienkowski, M., and Eggermont, J.J. (**2009**). "Long-term, partially-reversible reorganization of frequency tuning in mature cat primary auditory cortex can be induced by passive exposure to moderate-level sounds," Hear. Res., **257**, 24-40.

Pienkowski, M., and Eggermont, J.J. (**2010a**). "Intermittent exposure with moderate-level sound impairs central auditory function of mature animals without concomitant hearing loss," Hear. Res., **261**, 30-35.

Pienkowski, M., and Eggermont, J.J. (**2010b**). "Passive exposure of adult cats to moderate-level tone pip ensembles differentially decreases AI and AII responsiveness in the exposure frequency range," Hear. Res., **268**, 151-162.

Pienkowski, M., and Eggermont, J.J. (**2011**). "Cortical tonotopic map plasticity and behaviour," Neurosci. Biobehav. Rev., **35**, 2117-2128.

Pienkowski, M., Munguia, R., and Eggermont, J.J. (**2011**). "Passive exposure of adult cats to bandlimited tone ensembles or noise leads to long-term response suppression in auditory cortex," Hear. Res., **277**, 117-126.

Pienkowski, M., and Eggermont, J.J. (**2012**). "Reversible long-term changes in auditory processing in mature auditory cortex in the absence of hearing loss induced by passive, moderate-level sound exposure," Ear. Hearing, **33**, 305-314.

Pienkowski, M., Munguia, R., and Eggermont, J.J. (**2013**). "Effects of passive, moderate-level sound exposure on the mature auditory cortex: Spectral edges, spectrotemporal density, and real-world noise," Hear. Res., **296**, 121-130.

Pienkowski, M., Tyler, R.S., Roncancio, E.R., *et al.* (**2014**). "A review of hyperacusis and future directions: Part II. Measurement, mechanisms, and treatment," Am. J. Audiol., **23**, 420-436.

Ropp, T.J., Tiedemann, K.L., Young, E.D., *et al.* (**2014**). "Effects of unilateral acoustic trauma on tinnitus-related spontaneous activity in the inferior colliculus," J. Assoc. Res. Otolaryngol., **15**, 1007-1022.

Rubak, T., Kock, S., Koefoed-Nielsen, B., *et al.* (**2008**). "The risk of tinnitus following occupational noise exposure in workers with hearing loss or normal hearing," Int. J. Audiol., **47**, 109-114.

Schaette, R., and McAlpine, D. (**2011**). "Tinnitus with a normal audiogram: physiological evi-dence for hidden hearing loss and computational model," J. Neurosci., **31**, 13452-13457.

Sturm, J.J., Zhang-Hooks, Y.X., Roos, H., *et al.* (**2017**). "Noise trauma induced behavioral gap detection deficits correlate with reorganization of excitatory and inhibitory local circuits in the inferior colliculus and are prevented by acoustic enrichment," J. Neurosci., **37**, 6314-6330.

Sun, W., Lu, J., Stolzberg, D., *et al.* (**2009**). "Salicylate increases the gain of the central auditory system," Neuroscience, **159**, 325-334.

Sun, W., Deng, A., Jayaram, A., *et al.* (**2012**). "Noise exposure enhances auditory cortex responses related to hyperacusis behaviour," Brain Res., **1485**, 108-116.

Turner, J.G., and Parish, J. (**2008**). "Gap detection methods for assessing salicylate-induced tinnitus and hyperacusis in rats," Am. J. Audiol., **17**, S185-S192.

Tyler, R.S., Pienkowski, M., Roncancio, E.R., *et al.* (**2014**). "A review of hyperacusis and future directions. Part I. Definitions and manifestations," Am. J. Audiol., **23**, 402-419.

Weible, A.P., Moore, A.K., Liu, C., *et al.* (**2014**). "Perceptual gap detection is mediated by gap termination responses in auditory cortex," Curr. Biol., **24**, 1447-1455.

Weisz, N., Hartmann, T., Dohrmann, K., *et al.* (**2006**). "High-frequency tinnitus without hearing loss does not mean absence of deafferentation," Hear. Res., **222**, 108-114.

Wu, C., Martel, D.T., and Shore, S.E. (**2016**). "Increased synchrony and bursting of dorsal cochlear nucleus fusiform cells correlate with tinnitus," J. Neurosci., **36**, 2068-2073.

Xiong, B., Alkharabsheh, A., Manohar, S., *et al.* (**2017**). "Hyperexcitability of inferior colli-culus and acoustic startle reflex with age-related hearing loss," Hear. Res., **350**, 32-42.

The unique role of the non-lemniscal pathway on stimulus-specific adaptation (SSA) in the auditory system

GUILLERMO V. CARBAJAL[1,2] AND MANUEL S. MALMIERCA[1,2,3,*]

[1] *Auditory Neuroscience Laboratory, Institute of Neuroscience of Castilla y León (INCYL), Salamanca, Spain*

[2] *Salamanca Institute for Biomedical Research (IBSAL), Salamanca, Spain*

[3] *Department of Cell Biology and Pathology, Faculty of Medicine, University of Salamanca, Salamanca, Spain*

Stimulus-specific adaptation (SSA) is a special type of adaptation that allows neurons to cease responding only to repetitive, background stimuli, while preserving its responsiveness for other, new upcoming deviant stimuli. It emerges subcortically in non-lemniscal neurons of the inferior colliculi, propagating and evolving throughout the auditory pathway, until reaching its uppermost manifestation in the non-lemniscal areas of the auditory cortex. In this review, we will discuss the fundamental role of the non-lemniscal pathway in the generation of SSA, which is usually disregarded in cortical SSA research, despite being a major anatomical source of the mismatch negativity (MMN).

INTRODUCTION

Stimulus-specific adaptation (SSA) was firstly found in the auditory system by Ulanovsky and colleagues (2003) using mostly multi-unit activity recordings in the cat. In this pioneering study, they proposed that SSA in the primary areas of the auditory cortex (A1) could be the neuronal correlate of the mismatch negativity (MMN), an scalp-recorded evoked potential elicited by rare events that has demonstrated being a great tool for neurocognitive research (Näätänen *et al.*, 2007), with potential clinical applications (Näätänen *et al.*, 2012). They also assumed that SSA had to be a purely cortical activity, like the MMN, since their original recordings in the auditory thalamus failed to show SSA. However, these inceptive suppositions were later proven to be incomplete, inasmuch as (1) there were some notable discrepancies between the dynamics and sources of the MMN and the SSA recorded in A1, and (2) there was SSA being generated subcortically, actually as early as at the midbrain level. Both limitations could be accounted for by the same missing aspect: the fundamental role of the non-lemniscal auditory pathway in the generation of SSA, as we will discuss in the following. With the exception of two classical papers (Irvine and Huebner, 1979; Schreiner and Cynader, 1984), the role of non-lemniscal auditory cortex in adaptation still remains somewhat overlooked as of date, with very few SSA

*Corresponding author: msm@usal.es

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.

ISBN: 978-87-990013-6-1.

studies going beyond A1 (Nieto-Diego and Malmierca, 2016; Parras *et al.*, 2017). In this review, we will illustrate the tight relation between SSA emergence and the non-lemniscal auditory pathway in order to stimulate its inclusion in future SSA research.

SSA AS A HIGHER-ORDER TYPE OF ADAPTATION

Adaptation is an omnipresent property of neurons in the auditory system. It allows neurons to stop responding to redundant stimulation, thus exerting a protective role by avoiding an overload of the processing systems (Megela and Teyler, 1979). Most types of adaptation can be understood as rather basic physiological mechanisms, governed by activity-dependent cellular properties operating at the level of the neuron's output (Gutfreund, 2012; Pérez-González & Malmierca, 2014). SSA defines a higher level type of adaptation, depending more on the history of stimulation of the neuron rather than on its intrinsic properties (Ulanovsky *et al.*, 2004; 2003). Neurons showing SSA are able to adapt to frequently occurring stimuli (standards) selectively, while strongly resuming their firing whenever a rare stimulus (deviant) appears into the scene (Nelken, 2014). In other words, what makes SSA a unique kind of adaptation is that it is based on the input of the neuron, rather than its output, hence constituting an integrative endeavour observable at cellular level. An endeavour that must be critical for survival. With every repetition, a standard stimulus loses informative power. By selectively diminishing the resources devoted to process these standard sounds and dampening its perceptual representation, more resources are available for those novel sounds that are potentially more informative (Malmierca *et al.*, 2015). Consequently, deviant stimuli are automatically more salient and perceptually advantaged, giving rise to psychophysical effects such as attention capture (Tiitinen *et al.*, 1994) or pop-outs (Diliberto *et al.*, 2000), and it could be even at the base of the assembling of perceptual objects (Nelken, 2004).

THE NON-LEMNISCAL PATHWAY PERFORMS A HIGHER-ORDER TYPE OF SENSORY PROCESSING

Auditory information is transmitted along a series of several nuclei organised in a hierarchical manner, where different auditory features are progressively extracted at each level. Along the auditory neuraxis, two parallel pathways can be distinguished marking each station they cross with structural and functional characteristic features. Almost half a century ago, Ann Graybiel (1973) coined and defined the so-called *"lemniscal line system"* and *"lemniscal adjunct system"* as general categorisation of sensory conduction routes referred to the lemniscus. Since then, the distinction between "lemniscal" (also referred as "core" or "primary") and "non-lemniscal" (also referred as "belt" or "nonprimary") pathways have been widely used in auditory research (Hu, 2003; Jones, 2003; Lee and Winer, 2008). Making this simple distinction, we can easily classify and understand the role of the multiple subdivisions present in the inferior colliculus (IC), the medial geniculate body of the thalamus (MGB) and the auditory cortex (AC; Fig. 1).

The lemniscal pathway represents a core of neurons in every auditory nucleus that tend to be sharply tuned and organised in rather clear tonotopic fashion made of

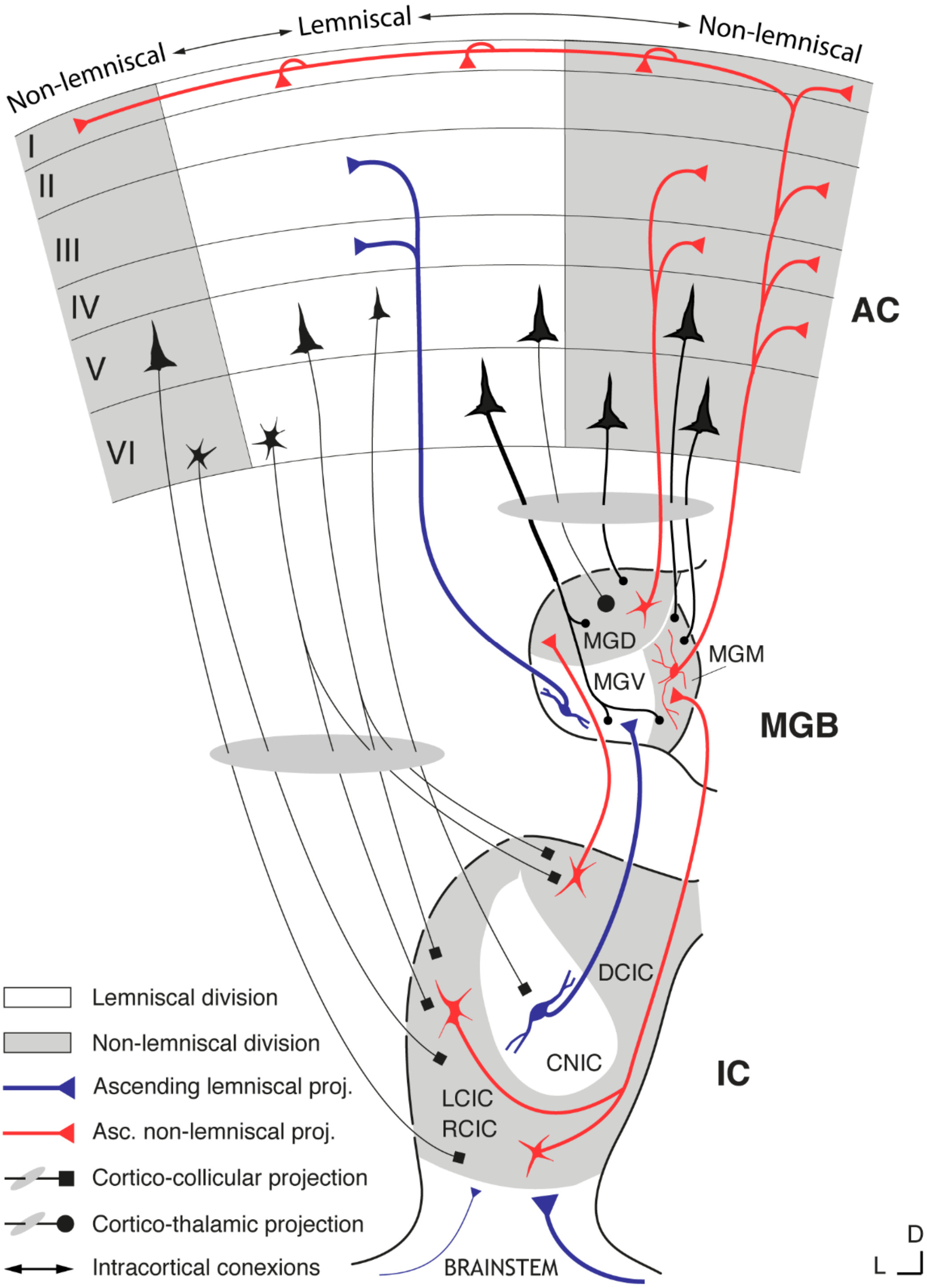


Fig. 1: Schematic diagram of the auditory pathway, showing the major stations and projections that constitute the lemniscal and non-lemniscal pathways. Note that divisions in subcortical nuclei are well preserved across species, while AC fields vary markedly (Malmierca, 2003; Malmierca and Hackett, 2010). Adapted from Malmierca *et al.* (2015).

anatomical laminae or bands. The majority of the neurons in each band project to their homologous band in the next station of the lemniscal pathway (Malmierca *et al.*, 2015). In addition to the precise tuning of their frequency-response areas (FRA; Fig. 2A), lemniscal neurons also show in general a better consistency in their response to the sound, including shorter latencies, greater firing rates, more overall spikes fired per stimulus and higher spontaneous activity than their non-lemniscal counterparts (Malmierca *et al.*, 2015). In other words, the response of these very tonographically organised neurons is fundamentally driven by the physical features of the sound, receiving mostly ascending inputs from lower lemniscal stations in the auditory neuraxis. Because of these characteristics, lemniscal divisions are considered to be part of a first-order stage of processing, forming a primary system more engaged in building up an accurate perceptual representation of the stimulus, disregarding its context or other abstract relations between sounds. The rat lemniscal pathway consists of the central nucleus of the IC (CNIC), the ventral division of the MGB (MGV), and the primary auditory cortex which includes the A1 field, the anterior auditory field (AAF) and the ventral auditory field (VAF) of the AC.

Parallel to the lemniscal pathway, another system referred to as the non-lemniscal pathway lies in which any trace of tonotopical distribution is at its best diffuse. The non-lemniscal pathway constitutes a belt of broadly-tuned neurons that gets inputs from the lemniscal core they are wrapping, and from other non-lemniscal stations: Subcortical non-lemniscal neurons send ascending projections to the next non-lemniscal station (Loftus *et al.*, 2008) while cortical neurons from belt areas send descending projections mostly (albeit not exclusively) to the non-lemniscal divisions of the MGB and the IC (Fig. 1) (Malmierca and Ryugo, 2011). The fact that non-lemniscal neurons shape this loop-like connectivity network with heavy cortical modulation, in addition to their comparatively longer response latencies, the broadness of their FRAs (Fig. 2B) and their adjunct anatomical position relative to the lemniscal stream, strongly indicates that they must exert an integrative function in the auditory system. Consequently, non-lemniscal divisions are part of a higher order stage of processing, constituting a secondary system capable of processing more complex aspects of the auditory scene analysis and tracking the history of stimulation, as required to account for the generation of SSA. The rat non-lemniscal pathway includes the rostral (RCIC), lateral (LCIC) and dorsal (DCIC) cortices of the IC, the dorsal (MGD) and medial (MGM) divisions of the MGB, and the suprarhinal auditory field (SRAF) and the posterior auditory field (PAF) of the AC.

SSA FIRSTLY EMERGES IN THE SUBCORTICAL NON-LEMNISCAL PATHWAY

As mentioned previously, Ulanovsky *et al.* (2003) initially suggested a cortical origin of SSA, since in their original work they could not find any signs of SSA in the auditory thalamus, most probably because they recorded very few neurons, most likely from the ventral (lemniscal) division of the MGB (although no details of the anatomical location of the recordings are given in their study). But this exclusively cortical nature of SSA was soon revisited and conceptualized after the discovery of

SSA in the IC (Ayala et al., 2015; Ayala & Malmierca, 2015, 2017; Duque & Malmierca, 2015; Duque *et al.*, 2012, 2016; Malmierca *et al.*, 2009; Parras *et al.*, 2017; Patel *et al.*, 2012; Pérez-González *et al.*, 2005, 2012; Pérez-González & Malmierca, 2012; Valdés-Baizabal *et al.*, 2017; Zhao *et al.*, 2011) and in the MGB (Anderson & Malmierca, 2013; Anderson *et al.*, 2009; Antunes & Malmierca, 2014; Antunes *et al.*, 2010; Duque *et al.*, 2014; Parras *et al.*, 2017). Significant and strong SSA appeared in the IC cortices, the MGD and intensely in the MGM, so sharply distributed exclusively in the non-lemniscal stations that the mere measurement of population SSA in a subcortical nucleus could provide enough evidence to distinguish between the lemniscal and non-lemniscal divisions of it.

The corticocentric interpretation of SSA was not completely dismissed after proving the existence of SSA in subcortical stations, probably due to the already strong established connection between SSA and MMN. It was suggested then that subcortical SSA might be “imposed” by the cortex (Nelken and Ulanovsky, 2007) given the massive corticocolicular projections that the IC cortices receive, and the impressively dense corticothalamic projections, that outmatch the thalamocortical output by a factor of ten (Malmierca *et al.*, 2015). Descending projections must necessarily exert at least a considerable modulatory function, but the prime source of SSA cannot be pinned down just by investigating connectivity. In order to address this question, studies of reversible deactivation of the AC using a cooling technique were conducted while recording the MGB (Antunes and Malmierca, 2011) and the IC (Anderson and Malmierca, 2013). The general results demonstrated that indeed the AC clearly modulated the firing rate of the non-lemniscal neurons in a gain-control manner (Malmierca *et al.*, 2015; Pérez-González *et al.*, 2012), helping to increase the contrast between standard and deviant stimuli by affecting the discharge rate to both proportionally (Ayala *et al.*, 2016; Duque *et al.*, 2015; Pérez-González *et al.*, 2012).

Nevertheless, the overall subcortical SSA levels and dynamics were mostly unaffected by cortical deactivation, with only about half of the adapting IC neurons and almost none in the MGB showing some change in their SSA sensitivity. In light of these results, it would be more plausible that SSA in A1 were actually inherited from subcortical non-lemniscal structures than viceversa. Although the possibility of SSA being generated *de novo* at the intrinsic microcircuitry of each station cannot be ruled out, it is reasonable to suggest that SSA must be a detection property that firstly emerges in the non-lemniscal IC, given that SSA has not being detected earlier in the auditory pathway (Ayala and Malmierca, 2013; Ayala *et al.*, 2013). From the IC cortices, SSA is transmitted downstream through the non-lemniscal subcortical pathway towards the cortex, where AC neurons work in complex integration of stimulus properties across multiple time scales and are less specialized for feature detection (Nelken, 2004), including the feature of novelty.

SSA IN NON-LEMNISCAL CORTICAL AREAS CAN BETTER ACCOUNT FOR THE GENERATION OF THE MMN

Despite the initial general acceptance of SSA as being the best candidate for the neuronal generator of MMN, there was still a time breach between the relatively long

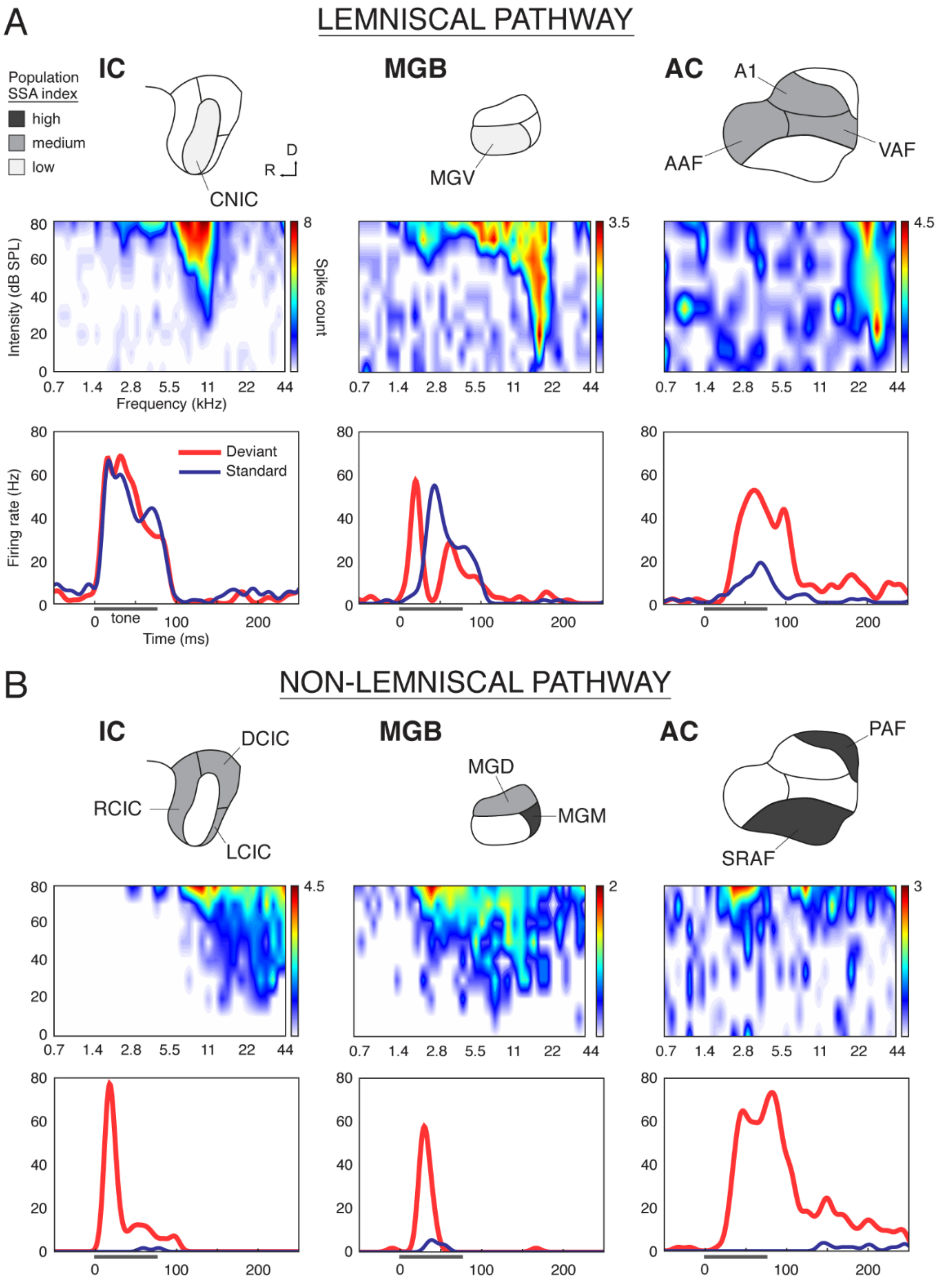


Fig. 2: Distribution of SSA along the rat auditory neuraxis. In the first row of each block, lemniscal (A) and non-lemniscal (B) subdivisions of the main post-lemniscus auditory nuclei are shaded indicating the strength of the population SSA present in it. In the second row, the FRA of a representative neuron of that subdivision is displayed, followed below (third row) by the corresponding responses of that neuron to a certain tone when presented in conditions of high probability (standard) or low probability (deviant).

peak latencies of the MMN and the swift cortical SSA reported in Ulanovsky *et al.* (2003), which sees it maximum rather close to the stimulus onset. Most importantly, the anatomical location of the reported SSA did not fit well with the topography of the change-detection MMN either, whose alleged generators are pinned down in the region of the secondary auditory cortex in humans (Alho, 1995), cats (Pincze *et al.*, 2001) and rats (Shiramatsu *et al.*, 2013). In spite of these considerable limitations, most of SSA research conducted in AC as of date is confined to A1.

Only in two recent studies (Nieto-Diego and Malmierca, 2016; Parras *et al.*, 2017), the lack of detailed studies on SSA beyond A1 is finally addressed by thoroughly recording of single-unit, multi-unit activity and local-field potentials in each of the auditory cortical fields of the rat. Besides confirming SSA presence in lemniscal AC, evidence provided demonstrates that SSA is even more robust in non-lemniscal AC fields. SSA properties differ substantially between lemniscal (primary) and non-lemniscal (nonprimary) fields. Cortical SSA distribution creates a topographic gradient that segregates the highest SSA levels to non-lemniscal fields in a sharp fashion, remarkably paralleling SSA subcortical organisation. Thereby, the continuity of the lemniscal and non-lemniscal pathways in the cortex is reflected by SSA distribution. Within non-lemniscal fields, SSA is much stronger and develops faster due to the more intense suppression and longer delay it produces on the responses to standard stimuli, which is not rare to find utterly obliterated. Levels of SSA within non-lemniscal regions are much higher around the beginning of the response than in the lemniscal ones, remaining strong up to 200 ms after the stimulus onset (Fig. 3A). Therefore, it can be argued that the non-lemniscal cortical regions are more suitable candidates for being mayor contributors in the MMN generation than their lemniscal homologues in the cortex.

Regarding local-field potentials, their difference wave correlated in time and strength with the SSA observed in single-unit and multi-unit activity recordings, confirming greater levels in non-lemniscal fields (Nieto-Diego and Malmierca, 2016; Parras *et al.*, 2017). These difference waves showed the same morphology in all cortical fields, with a fast negative deflection (Nd) followed by a positive one (Pd). On the one hand, the Nd occurred earlier and tended to be larger in lemniscal fields than in the non-lemniscal ones, suggesting a lemniscal origin (Fig. 3B). This early deflection could be related with the modulations of the scalp-recorded middle latency responses that correspond to the first response of the primary AC to a deviant event, which take place previous to the occurrence of the MMN (Escera and Malmierca, 2014). On the other hand, the Pd peaked homogeneously along the AC, so its generation must hinge on intracortical processing and reciprocal interaction between lemniscal and non-lemniscal fields, further suggesting a bottom-up propagation of SSA. Most importantly, the Pd tended to peak between 60 and 80 ms (Fig. 3B), well within the range of MMN-like potentials recorded in the rat (50-100 ms) (Harms *et al.*, 2016). This synchronicity finally allows to overcome the discrepancies in the time course and anatomical source of the SSA and the MMN, thus setting a bridge between both in which the cornerstone is the non-lemniscal contribution.

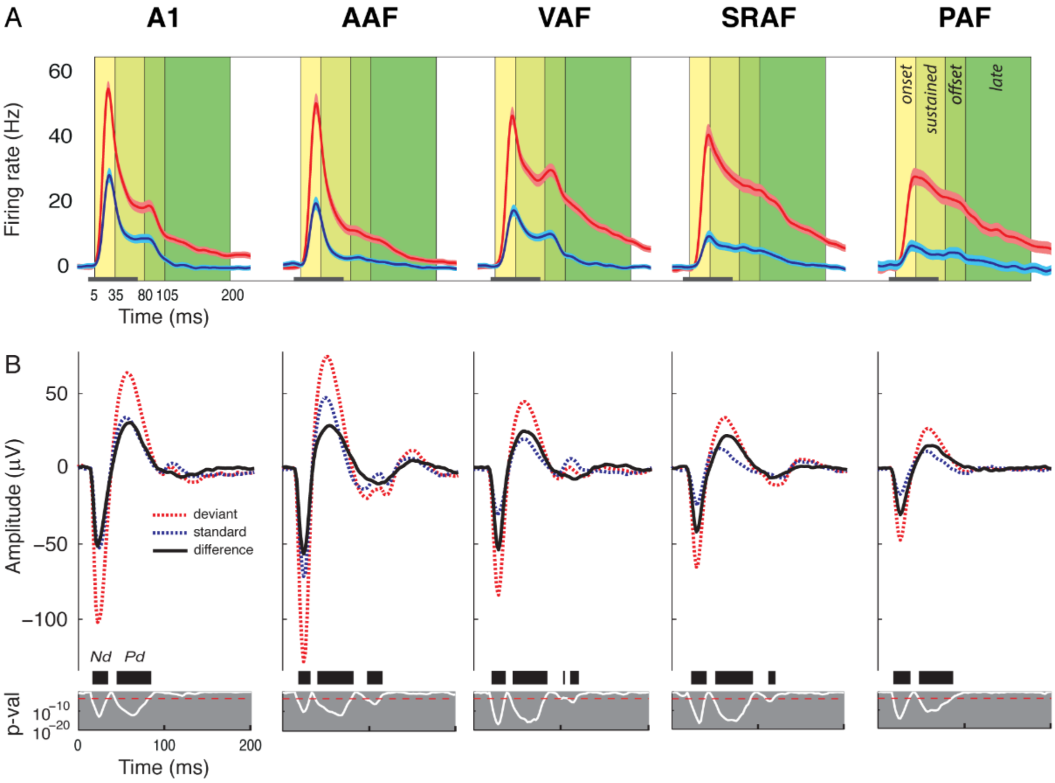


Fig. 3: Grand-average of the responses to standard (STD) and deviant (DEV) tones recorded as multi-unit activity (A) and local-field potentials (B) within each AC cortical field. Adapted from Nieto-Diego and Malmierca (2016).

CONCLUSION

Whether generated in situ in lemniscal AC or just inherited subcortically, the fact is that lemniscal neurons in the cortex show SSA, so it would be imprecise to say that SSA is a purely non-lemniscal property. However, the inverse can definitely be asserted. The SSA is a defining feature of the non-lemniscal auditory pathway, with prevailing presence all along it. The appearance of SSA as early as the level of the midbrain in the cortices of the IC suggests it is an emerging property of the non-lemniscal subcortical structures, while in non-lemniscal cortical areas SSA achieves its most refine manifestation. All this reaffirms the notion of the non-lemniscal pathway as a parallel higher-order stage of sensory processing that goes beyond the faithful representation of auditory stimuli predominant in the lemniscal pathway, being able to extract more complex features in auditory events, like novelty. Thus, it can be argued that regularity encoding and deviance detection are capabilities of the auditory brain that have a non-lemniscal foundation, essential in the generation of SSA and MMN.

ACKNOWLEDGEMENTS

This work has been funded by Spanish MINECO (SAF2016-75803-P).

REFERENCES

Alho, K. (**1995**). "Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes," Ear Hearing, **16**, 38-51.

Anderson, L.A., Christianson, G.B., and Linden, J.F. (**2009**). "Stimulus-specific adaptation occurs in the auditory thalamus," J. Neurosci., **29**, 7359-7363. doi: 10.1523/JNEUROSCI.0793-09.2009

Anderson, L.A., and Malmierca, M.S. (**2013**). "The effect of auditory cortex deactivation on stimulus-specific adaptation in the inferior colliculus of the rat," Eur. J. Neurosci., **37**, 52-62. doi: 10.1111/ejn.12018

Antunes, F.M., Nelken, I., Covey, E., and Malmierca, M.S. (**2010**). "Stimulus-specific adaptation in the auditory thalamus of the anesthetized rat," PLoS ONE, **5**. doi: 10.1371/journal.pone.0014071

Antunes, F.M., and Malmierca, M.S. (**2011**). "Effect of auditory cortex deactivation on stimulus-specific adaptation in the medial geniculate body," J. Neurosci., **31**, 17306-17316.

Antunes, F.M., and Malmierca, M.S. (**2014**). "An overview of stimulus-specific adaptation in the auditory thalamus," Brain Topogr., **27**, 480-499. doi: 10.1007/s10548-013-0342-6

Ayala, Y.A., and Malmierca, M.S. (**2013**). "Stimulus-specific adaptation and deviance detection in the inferior colliculus," Front. Neural Circuit., **6**, 1-16. doi: 10.3389/fncir.2012.00089

Ayala, Y.A., and Malmierca, M.S. (**2015**). "Cholinergic modulation of stimulus-specific adaptation in the inferior colliculus," J. Neurosci., **35**, 12261-12272. doi: 10.1523/JNEUROSCI.0909-15.2015

Ayala, Y.A., Pérez-González, D., Duque, D., Nelken, I., and Malmierca, M.S. (**2013**). "Frequency discrimination and stimulus deviance in the inferior colliculus and cochlear nucleus," Front. Neural Circuit., **6**, 119. doi: 10.3389/fncir.2012.00119

Ayala, Y.A., Pérez-González, D., and Malmierca, M.S. (**2016**). "Stimulus-specific adaptation in the inferior colliculus: The role of excitatory, inhibitory and modulatory inputs," Biol. Psychol., **116**, 10-22. doi: 10.1016/j.biopsycho.2015.06.016

Ayala, Y.A., and Malmierca, M.S. (**2017**). "The effect of inhibition on stimulus-specific adaptation in the inferior colliculus," Brain Struct. Funct., 1-17. doi: 10.1007/s00429-017-1546-4

Ayala, Y.A., Udeh, A., Dutta, K., Bishop, D., Malmierca, M.S., and Oliver, D.L. (**2015**). "Differences in the strength of cortical and brainstem inputs to SSA and non-SSA neurons in the inferior colliculus," Sci. Rep., **5**, 10383. doi: 10.1038/srep10383

Diliberto, K.A., Altarriba, J., and Neill, W.T. (**2000**). "Novel popout and familiar popout in a brightness discrimination task," Percept. Psychophys., **62**, 1494-1500. doi: 10.3758/BF03212149

Duque, D., Perez-Gonzalez, D., Ayala, Y.A., Palmer, A.R., and Malmierca, M.S. (**2012**). "Topographic distribution, frequency, and intensity dependence of stimulus-specific adaptation in the inferior colliculus of the rat," J. Neurosci., **32**, 17762-17774. doi: 10.1523/jneurosci.3190-12.2012

Duque, D., Malmierca, M.S., and Caspary, D.M. (**2014**). "Modulation of stimulus-specific adaptation by GABA(A) receptor activation or blockade in the medial geniculate body of the anaesthetized rat," J. Physiol., **592**, 729-743. doi: 10.1113/jphysiol.2013.261941

Duque, D., and Malmierca, M.S. (**2015**). "Stimulus-specific adaptation in the inferior colliculus of the mouse: anesthesia and spontaneous activity effects," Brain Struct. Funct., **220**, 3385-3398. doi: 10.1007/s00429-014-0862-1

Duque, D., Wang, X., Nieto-Diego, J., Krumbholz, K., and Malmierca, M.S. (**2016**). "Neurons in the inferior colliculus of the rat show stimulus-specific adaptation for frequency, but not for intensity;" Sci. Rep., **6**, 24114. doi: 10.1038/srep24114

Escera, C., and Malmierca, M.S. (**2014**). "The auditory novelty system: An attempt to integrate human and animal research," Psychophysiology, **51**, 111-123. doi: 10.1111/psyp.12156

Graybiel, A.M. (**1973**). "The thalamo-cortical projection of the so-called posterior nuclear group: A study with anterograde degeneration methods in the cat," Brain Res., **49**, 229-244. doi: 10.1016/0006-8993(73)90420-4

Gutfreund, Y. (**2012**). "Stimulus-specific adaptation, habituation and change detection in the gaze control system," Biol. Cybern., **106**, 657-668. doi: 10.1007/s00422-012-0497-3

Harms, L., Michie, P.T., and Näätänen, R. (**2016**). "Criteria for determining whether mismatch responses exist in animal models: Focus on rodents," Biol. Psychol., **116**, 28-35. doi: 10.1016/j.biopsycho.2015.07.006

Hu, B. (**2003**). "Functional organization of lemniscal and nonlemniscal auditory thalamus," Exp. Brain Res., **153**, 543-549. doi: 10.1007/s00221-003-1611-5

Irvine, D.R., and Huebner, H. (**1979**). "Acoustic response characteristics of neurons in nonspecific areas of cat cerebral cortex," J. Neurophysiol., **42**, 107-122. doi: 10.1152/jn.1979.42.1.107

Jones, E.G. (**2003**). "Chemically defined parallel pathways in the monkey auditory system," Ann. NY Acad. Sci., **999**, 218-233. doi: 10.1196/annals.1284.033

Lee, C.C., and Winer, J.A. (**2008**). "Connections of cat auditory cortex: I. Thalamocortical system;" J. Comp. Neurol., **507**, 1879-1900. doi: 10.1002/cne.21611

Loftus, W.C., Malmierca, M.S., Bishop, D.C., and Oliver, D.L. (**2008**). "The cytoarchitecture of the inferior colliculus revisited: A common organization of the lateral cortex in rat and cat," Neuroscience, **154**, 196-205. doi: 10.1016/j.neuroscience.2008.01.019

Malmierca, M.S. (**2003**). "The structure and physiology of the rat auditory system: an overview," in R.J. Bradley, R.A. Harris, and P. Jenner (Eds.), *International Review of Neurobiology* Vol. 56 (Academic Press), pp. 147–212.

Malmierca, M.S., Cristaudo, S., Pérez-González, D., and Covey, E. (**2009**). "Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat," J. Neurosci., **29**, 5483-5493. doi: 10.1523/jneurosci.4153-08.2009

Malmierca, M.S., and Hackett, T.A. (**2010**). "Structural organization of the ascending auditory pathway," in D.R. Moore, A. Rees, and A.R. Palmer (Eds.), *The Oxford Handbook of Auditory Science: The Auditory Brain* (Oxford University Press), pp. 9-42.

Malmierca, M.S., and Ryugo, D.K. (**2011**). “Descending connections of auditory cortex to the midbrain and brain stem,” in *The Auditory Cortex* (Boston, MA: Springer), pp. 189-208. doi: 10.1007/978-1-4419-0074-6_9

Malmierca, M.S., Anderson, L.A., and Antunes, F.M. (**2015**). “The cortical modulation of stimulus-specific adaptation in the auditory midbrain and thalamus: a potential neuronal correlate for predictive coding,” Front. Sys. Neurosci., **9**, 1-14. doi: 10.3389/fnsys.2015.00019

Megela, A.L., and Teyler, T.J. (**1979**). ”Habituation and the human evoked potential,” J. Comp. Physiol. Psychol., **93**, 1154-1170. doi: 10.1037/h0077630

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (**2007**). “The mismatch negativity (MMN) in basic research of central auditory processing: A review,” Clin. Neurophysiol., **118**, 2544-2590. doi: 10.1016/j.clinph.2007.04.026

Näätänen, R., Kujala, T., Escera, C., Baldeweg, T., Kreegipuu, K., Carlson, S., and Ponton, C. (**2012**). “The mismatch negativity (MMN) – A unique window to disturbed central auditory processing in ageing and different clinical conditions,” Clin. Neurophysiol., **123**, 424-458. doi: 10.1016/j.clinph.2011.09.020

Nelken, I. (**2004**). “Processing of complex stimuli and natural scenes in the auditory cortex,” Curr. Opin. Neurobiol., **14**, 474-480. doi: 10.1016/j.conb.2004.06.005

Nelken, I., and Ulanovsky, N. (**2007**). Mismatch negativity and stimulus-specific adaptation in animal models,” J. Psychophysiol., **21**, 214-223. doi: 10.1027/0269-8803.21.34.214

Nelken, I. (**2014**). “Stimulus-specific adaptation and deviance detection in the auditory system: experiments and models,” Biol. Cybern., **108**, 655-663. doi: 10.1007/s00422-014-0585-7

Nieto-Diego, J., and Malmierca, M.S. (**2016**). “Topographic distribution of stimulus-specific adaptation across auditory cortical fields in the anesthetized rat,” PLoS Biol., **14**, 1-30. doi: 10.1371/journal.pbio.1002397

Parras, G.G., Nieto-Diego, J., Carbajal, G.V., Valdés-Baizabal, C., Escera, C., and Malmierca, M.S. (**2017**). “Neurons along the auditory pathway exhibit a hierarchical organization of prediction error,” Nat. Commun., **8**. doi: 10.1038/s41467-017-02038-6

Patel, C.R., Redhead, C., Cervi, A.L., and Zhang, H. (**2012**). “Neural sensitivity to novel sounds in the rat’s dorsal cortex of the inferior colliculus as revealed by evoked local field potentials,” Hear. Res., **286**, 41-54. doi: 10.1016/j.heares.2012.02.007

Pérez-González, D., Hernández, O., Covey, E., Malmierca, M.S., and Schulze, H. (**2012**). “GABAA-mediated inhibition modulates stimulus-specific adaptation in the inferior colliculus,” PLoS One, e34297. doi: 10.1371/journal.pone.0034297

Pérez-González, D., Malmierca, M.S., and Covey, E. (**2005**). “Novelty detector neurons in the mammalian auditory midbrain,” Eur. J. Neurosci., **22**, 2879-2885. doi: 10.1111/j.1460-9568.2005.04472.x

Pérez-González, D., and Malmierca, M. S. (**2012**). “Variability of the time course of stimulus-specific adaptation in the inferior colliculus,” Front. Neural Circuit., **6**, 107. doi: 10.3389/fncir.2012.00107

Pérez-González, D., and Malmierca, M.S. (**2014**). "Adaptation in the auditory system: an overview," Front. Integ. Neurosci., **8**, 1-10. doi: 10.3389/fnint.2014.00019

Pincze, Z., Lakatos, P., Rajkai, C., Ulbert, I., and Karmos, G. (**2001**). "Separation of mismatch negativity and the N1 wave in the auditory cortex of the cat: a topographic study," Clin. Neurophysiol., **112**, 778-784. doi: 10.1016/S1388-2457(01)00509-0

Schreiner, C.E., and Cynader, M.S. (**1984**). "Basic functional organization of second auditory cortical field (AII) of the cat," J. Neurophysiol., **51**, 1284-1305. doi: 10.1152/jn.1984.51.6.1284

Shiramatsu, T.I., Kanzaki, R., Takahashi, H., Sams, M., and Näätänen, R. (**2013**). "Cortical mapping of mismatch negativity with deviance detection property in rat," PLoS One, **8**, e82663. doi: 10.1371/journal.pone.0082663

Tiitinen, H., May, P., Reinikainen, K., and Näätänen, R. (**1994**). "Attentive novelty detection in humans is governed by pre-attentive sensory memory," Nature, **372**, 90-92. doi: 10.1038/372090a0

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

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (**2004**). "Multiple time scales of adaptation in auditory cortex neurons," J. Neurosci., **24**.

Valdés-Baizabal, C., Parras, G.G., Ayala, Y.A., and Malmierca, M.S. (**2017**). "Endocannabinoid modulation of stimulus-specific adaptation in inferior colliculus neurons of the rat," Sci. Rep., 6997. doi: 10.1038/s41598-017-07460-w

Zhao, L., Liu, Y., Shen, L., Feng, L., and Hong, B. (**2011**). "Stimulus-specific adaptation and its dynamics in the inferior colliculus of rat," Neuroscience, **181**, 163-174. doi: 10.1016/j.neuroscience.2011.01.060

The neural processing of phonemes is shaped by linguistic analysis

TOBIAS OVERATH[1,2,3,*] AND JACKSON C. LEE[1]

[1] *Duke Institute for Brain Sciences, Duke University, Durham, NC, USA*

[2] *Center for Cognitive Neuroscience, Duke University, Durham, NC, USA*

[3] *Department of Psychology and Neuroscience, Duke University, Durham, NC, USA*

Speech perception entails the mapping of the acoustic waveform to its linguistic representation. For this transformation to succeed, the speech signal needs to be tracked across multiple temporal scales in order to decode linguistic units ranging from phonemes to sentences. Here, we investigate how linguistic knowledge, and the temporal scale of linguistic analysis, influence the neural processing of a fundamental linguistic unit, the phoneme. To obtain control over the linguistic scale of analysis, we use a novel speech-quilting algorithm (Overath *et al.*, 2015) to control the acoustic structure available at different linguistic units (phoneme, syllable, word). To obtain control over the linguistic content, independent of the temporal acoustic structure, we construct speech quilts from both familiar (English) and foreign (Korean) languages. We recorded electroencephalography in healthy participants and show that the neural response to phonemes, the phoneme-related potential, is shaped by linguistic context only in a familiar language, but not in a foreign language. The results suggest that the processing of the acoustic properties of a fundamental linguistic unit, the phoneme, is already shaped by linguistic analysis.

INTRODUCTION

Speech is an intrinsically temporal signal with a rich temporal structure: Its linguistic constituents, such as phonemes, syllables, words, or sentences, all have characteristic durations, ranging from tens of milliseconds (in the case of phonemes) to hundreds (words) or thousands of milliseconds (sentences) (Rosen, 1992; Stevens, 2000; Poeppel, 2003). Our understanding of the neural architecture supporting speech perception has increased substantially over the last two decades (Hickok and Poeppel, 2007; Friederici and Gierhan, 2013). However, where and how the acoustic analysis of temporal speech structure interfaces with linguistic representations (such as syntax, lexicon, or semantics) is still poorly understood. While there is evidence that speech is analyzed at different temporal analysis scales, which are instantiated via a hierarchical organization across auditory and frontal cortices (Hasson *et al.*, 2008; Lerner *et al.*, 2011), these apply to relatively long temporal analysis windows

[*] Corresponding author: t.overath@duke.edu

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
ISBN: 978-87-990013-6-1.

commensurate with words, sentences, and paragraphs. In contrast, less is known about the neural representation of smaller linguistic units, in particular phonemes and syllables, which form the ‘building blocks’ upon which longer linguistic structures are built.

Previous studies have typically used isolated phonemes, consonant-vowel transitions, or words to investigate the underlying neural processes (Phillips *et al.*, 2000; Sanders and Neville, 2003; Tremblay *et al.*, 2003; Martin *et al.*, 2008). However, by presenting linguistic units in isolation, the role of predictive, top-down linguistic processes such as learned phonological, morphological or syntactical rules (Park *et al.*, 2015; Kocagoncu *et al.*, 2017), which are ubiquitous and automatic in natural speech perception, remained unclear. More recently, Khalighinejad *et al.* (2017) used continuous, natural speech to demonstrated that different categories of phonemes (e.g., vowels, nasals, fricatives, or plosives) have distinct neural correlates, or phoneme-related potentials (PRP). However, this approach is unable to differentiate between acoustic and linguistic processes, since listening to natural speech in a familiar language automatically recruits both.

What is needed, therefore, is an experimental approach that dissociates acoustic from linguistic processes during the analysis of temporal speech structure. Such an approach requires two essential features: (1) control over the linguistic structure, or units at which analysis occurs; (2) control over the linguistic content. We propose the following paradigm that allows the dissociation of acoustic and linguistic speech processes: To obtain control over the linguistic units of analysis, we modify a novel sound-quilting algorithm (Overath *et al.*, 2015) to control acoustic structure at the level of different linguistic units (phonemes, syllables, words) by shuffling and then stitching them together. This approach yields new ‘speech quilts’ that preserve the natural temporal speech structure only up to the linguistic unit, but not beyond. To obtain control over the linguistic content, independent of the temporal acoustic structure of linguistic units, we construct speech quilts from both familiar (English) and foreign (Korean) languages. This approach ensures that any changes at the signal-acoustics level affect both languages identically, while manipulating the linguistic percept differently. Thus, neural responses that vary as a function of the size of the linguistic unit (phoneme, syllable, word) will imply the presence of linguistic processing, while neural responses that are unaffected by linguistic unit will imply aspects of acoustic processing.

In this study, we investigated how acquired linguistic knowledge influences the neural processing of a fundamental linguistic unit, the phoneme, in different contexts. We recorded electroencephalography (EEG) from participants while they listened to speech quilts carrying information at the level of phonemes, syllables, or words, as well as natural speech, in either a familiar (English) or foreign language (Korean). We hypothesized that the PRP would be modified as a function of linguistic context only in a familiar language, due to linguistic processes, but not in a foreign language.

METHODS

Participants

The 18 right-handed participants (mean age = 23, range = 18-31, 10 females) were native speakers of American English, with no knowledge of Korean. All reported to have normal hearing and no history of neurological or psychiatric diseases. Participants provided written consent prior to participating in the study, in accordance with the Duke University Institutional Review Board.

Stimuli

The stimuli were derived from mono recordings (44100-Hz sampling rate, 16-bit resolution) of four female bilingual English/Korean speakers reading from a book in either language (native English and native Korean speakers judged the recordings as coming from native speakers, respectively). The recordings were then segmented into phonemes, syllables, and words using the Penn Phonetics Lab Forced Aligner Toolkit (Yuan and Lieberman, 2008) for English, and the Korean Phonetic Aligner Program Suite (Yoon and Kang, 2013) for Korean. The alignment was then manually checked for eventual segmentation errors by a native English and Korean speaker, respectively. Korean is a phonetic language that shares no etymological roots with English and has a different grammatical structure (Sohn, 1999).

We placed a number of constraints on the quilted stimuli. 1) Phonemes had to be between 20-80 ms in duration, syllables between 100-240 ms, and words between 300-600 ms to be included in the phoneme, syllable, or word quilts, respectively. 2) Syllables that were also words were excluded from consideration in the syllable quilts. 3) Two identical phonemes could not be next to each other, since this does not happen in normal English or Korean speech. The relative phoneme distribution (frequency of occurrence of a given phoneme across conditions) was not affected by these constraints: Phoneme frequency profiles within a language were significantly correlated ($0.85 < \rho < 0.99$, all $p < 0.001$).

The stimuli are based on a slight modification of the quilting algorithm introduced in Overath *et al.* (2015), such that instead of quilting equal-length segments, here we quilt linguistic units. Briefly, a source signal is divided into linguistic units (here either phonemes, syllables, or words), which are then pseudorandomly rearranged and stitched together to create a new speech quilt signal. By using an L^2 norm when choosing adjacent linguistic units to approximate the original unit-to-unit change in the original speech signal, and by using pitch-synchronous overlap-add (PSOLA) (Moulines and Charpentier, 1990) to avoid sudden frequency jumps at unit boundaries, the quilting algorithm ensures that low-level acoustic attributes (e.g., amplitude modulation rate, frequency spectrum) in the speech quilt are similar to those in the original speech signal. All stimuli are 6 s long and are speech quilts made up of phonemes, syllables, and words, as well as original, unaltered excerpts from the recordings.

Experimental procedure

Participants were familiarized with recordings of the four different speakers, and then performed a behavioral task in which they listened to brief recordings in English or Korean and were asked to identify the speaker for each trial (irrespective of language). Participants responded by pressing one of keys 1, 2, 3, or 4 for speakers 1-4.

In the EEG experiment, each condition of a 2 Language (English, Korean) × 4 Linguistic unit (phoneme-, syllable-, word-quilt, natural speech) design was presented a total of 48 times (12 exemplars per speaker) over the course of 4 runs. The inter-trial-interval was 2 s. Participants performed the same speaker identification task as in the prior behavioral experiment (irrespective of language and linguistic unit).

Stimuli were presented at a comfortable listening level (~60 dB) through high-fidelity Sennheiser HD-25 on-ear headphones via a low-latency Fireface UC USB sound card, using Psychophysics Toolbox Version 3 (Brainard, 1997) running in Matlab.

EEG recording and analysis

EEG data were recorded on a 63-channel active electrode system (Brain Vision ActiChamp, Brain Products) using a customized, extended coverage, elastic electrode cap (EASYCAP, Herrsching, Germany) (Woldorff *et al.*, 2002). This cap provides extended coverage of the head from just above the eyebrows to below the inion posteriorly and has electrodes that are equally spaced across the cap. Two fronto-lateral electrodes track horizontal eye movements, while an additional external electrode just underneath the left eye tracks vertical eye movements. Data are recorded at a 1000-Hz sampling rate (with a DC to 260 Hz bandpass) referenced to the right mastoid, and are then re-referenced off-line to the average of the left and right mastoids.

Data were analyzed using EEGLAB (Delorme and Makeig, 2004) and custom-written Matlab scripts. Standard artifact rejection algorithms and independent component analysis (ICA) implemented in EEGLAB were used to remove eye-blink and physiological noise artifacts. The PRP analysis largely followed that outlined in Khalighinejad et al. (2017): Data were z-scored, epoched between −100 and 600 ms relative to phoneme onset, baseline corrected (−100 to 0 ms), and bandpass filtered between 2-15 Hz. To determine time windows of interest for our subsequent analyses, we centered 50-ms windows around the P50, N100, and P200 peaks derived from the PRP across languages and all electrodes. Results are shown for a region-of-interest (ROI) containing 9 fronto-central electrodes around electrode FCz.

RESULTS

Behavioral performance in the speaker identification task improved with linguistic unit length, and interacted with language familiarity (Fig. 1): A repeated measures (RM) ANOVA revealed main effects of Linguistic unit ($F_{(3,51)} = 12.87$, $p < 0.001$, $\eta^2_p = 0.43$), Language ($F_{(1,17)} = 9.49$, $p = 0.007$, $\eta^2_p = 0.36$), and an interaction ($F_{(3,51)} = 4.51$, $p = 0.007$, $\eta^2_p = 0.21$). Post-hoc pairwise comparisons (Bonferroni corrected) revealed that performance was significantly better in English than in Korean, except in the phoneme quilt condition ($p > 0.05$).

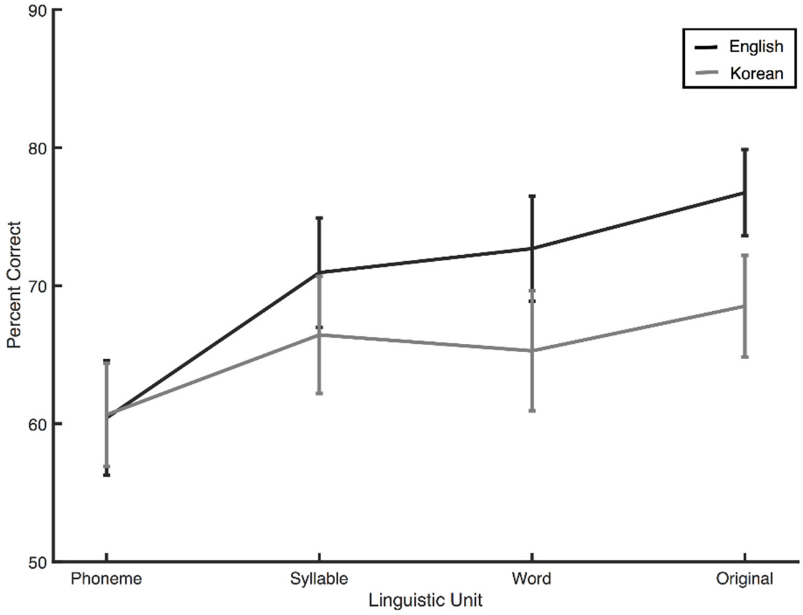


Fig. 1: Mean percent correct performance (±SEM) in the speaker identification task. Performance was well above chance (25%).

Next, we computed the grand-average PRP across all phonemes. As in Khalighinejad *et al.* (2017), the PRP showed a clear succession of P50 and N100 components, as well as a weak P200, in both English and Korean (Fig. 2). For each component, we ran RM ANOVAs with factors Language (English, Korean) and Linguistic Unit (phoneme-, syllable- , word-quilts, and natural speech; Table 1). The P50 component revealed main effects for both factors, as well as an interaction. The N100 component showed main effects for both factors, while the P200 component revealed a main effect of Linguistic Unit and an interaction.

	P50 (35-85 ms)			**N100 (90-140 ms)**			**P200 (180-230 ms)**		
	Unit	**Lang.**	**Inter.**	**Unit**	**Lang.**	**Inter.**	**Unit**	**Lang.**	**Inter.**
F-value	15.79	5.53	4.37	3.71	12.43	n.s.	3.23	n.s.	3.21
p-value	< 0.001	0.031	0.008	0.017	0.003	n.s.	0.03	n.s.	0.031
η^2_p	*0.48*	*0.25*	*0.2*	*0.18*	*0.42*	n.s.	*0.16*	n.s.	*0.16*

Table 1: RM ANOVA with factors Linguistic Unit and Language. F-value degrees of freedom are (3,51) for Linguistic Unit and (1,17) for Language factors.

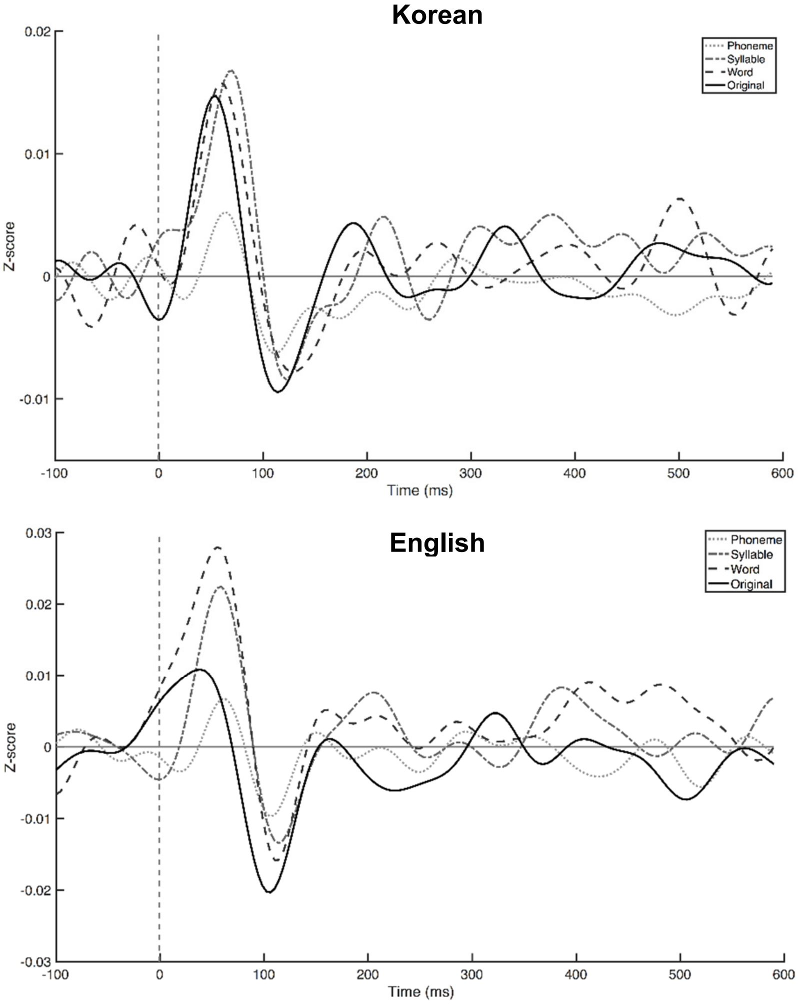


Fig. 2: The phoneme-related potential (PRP) for Korean (top) and English (bottom) conditions (phoneme-, syllable-, and word quilts, as well as natural speech). Note that in Korean the N100 component of the PRP is not significantly affected by the linguistic context; In English, the strength of the N100 component increases from phoneme-quilts to natural speech.

To investigate in more detail the effect of linguistic unit in English and Korean, we computed RM ANOVAs separately for each language. In English, P50, N100, and P200 differed in magnitude as a function of Linguistic Unit ($F_{(3,51)} = 12.44, p < 0.001$, $\eta^2_p = 0.42$; $F_{(3,51)} = 3.86$, $p = 0.015$, $\eta^2_p = 0.19$; and $F_{(3,51)} = 3.75$, $p = 0.016$, $\eta^2_p =$ 0.16, respectively). The N100 revealed the clearest effect of linguistic unit, whereby its absolute magnitude increased monotonically from phoneme quilts to natural speech. In Korean, only the early P50 component was affected by linguistic unit ($F_{(3,51)} = 5.05$, p = 0.004, $\eta^2_p = 0.23$); However, post-hoc pairwise comparisons revealed that this was driven by a more categorical, rather than graded, effect, whereby the P50 in phoneme quilts was significantly different from any of the other conditions.

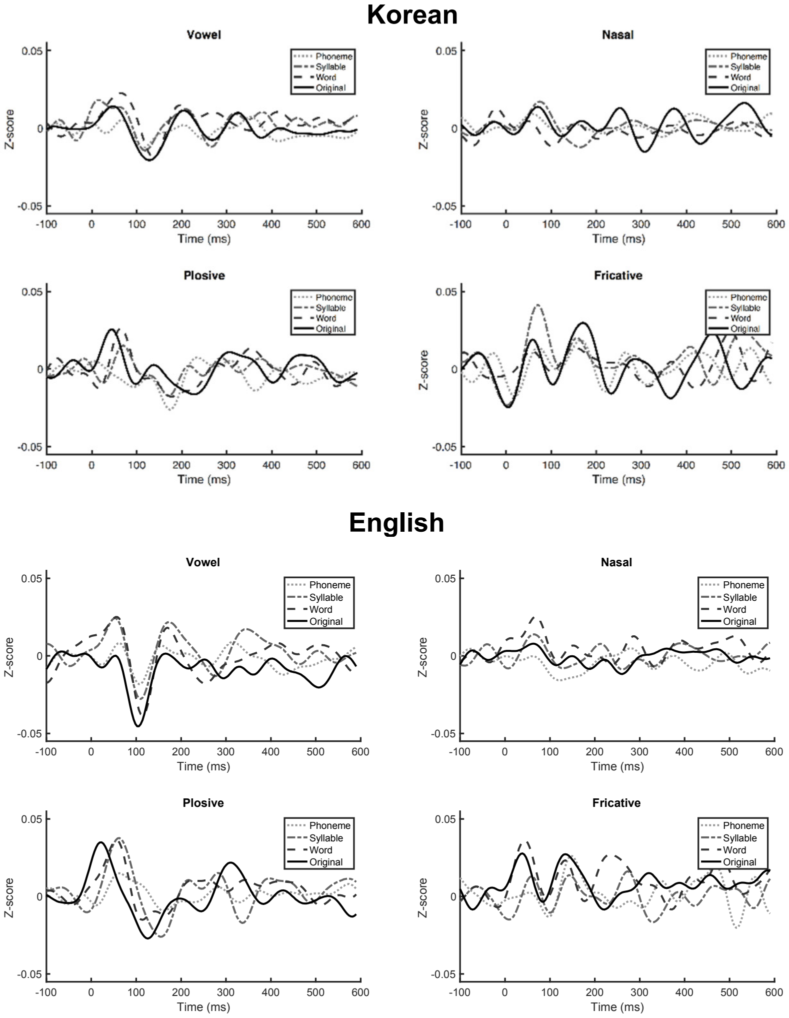


Fig. 3: Category-specific (vowel, nasal, plosive, fricative) PRPs as a function of linguistic unit (phoneme-, syllable-, word-quilt, natural speech) in Korean (top) and English (bottom). Note the overall similarity between languages for the original speech conditions. In English, the N100 component of the vowel PRP shows the clearest gradated PRP.

The analyses so far have treated all phonemes the same; however, phonemes can be classified by their manner of articulation (Ladefoged and Johnson, 2010), and we next investigated the main classes of plosives, fricatives, nasals, and vowels, to determine whether different phoneme categories are differentially affected by the linguistic context as a function of language. Figure 3 shows the PRP for each phoneme category in each language as a function of linguistic unit (phoneme-, syllable-, or word-quilt, as well as natural speech). We focus here on the N100 component, which showed a significant effect of linguistic unit in English. In English, the vowel PRP showed a clear graded response ($F_{(3,51)} = 4.45$, $p = 0.007$, $\eta^2_p = 0.21$); Post-hoc pairwise comparisons (Bonferroni corrected) revealed significant differences between the phoneme-quilt and natural speech conditions ($p = 0.025$), as well as a tendency between phoneme-quilt and word-quilt conditions ($p = 0.055$). The other three phoneme categories did not show a graded response as a function of linguistic unit. In Korean, no phoneme category was affected by linguistic unit size in any systematic way.

DISCUSSION

The preliminary results reported here demonstrate that the processing of a fundamental linguistic unit, the phoneme, is already shaped by linguistic analysis, but only if a linguistic repertoire is available. In the familiar language, the phoneme-related potential showed a graded N100 response as the size of the linguistic unit increased (from phoneme quilts to normal speech); This was most pronounced for vowels. In contrast, the PRP was generally unaffected by linguistic context in a foreign language.

The design of directly comparing the effect of linguistic unit size in two languages allowed the dissociation of acoustic and linguistic neural processes. Acoustic processing would be shared between languages, while linguistic processing would be indicated by a differentiation of the response (e.g., PRP) as a function of linguistic context. In the current study, the similarity of the PRP in natural English and Korean speech (e.g., Fig. 2) therefore reveals a shared mechanism for processing acoustic properties that are common to phonemes in both languages. For example, the vowel PRP in both English and Korean displayed a characteristic N100 similar to that in Khalighinejad *et al.* (2017) (Fig. 3). In contrast, the linguistic context within which phonemes appeared influenced the PRP in a systematic manner only in the familiar, but not the foreign language. This suggests that a linguistic repertoire (e.g., syntax, lexicon, or semantics), when available, shapes the processing of acoustic properties of temporal speech structure, even at a fundamental level such as the phoneme.

The results have implications for our understanding of how acoustic and linguistic representations interface already at an early level of speech processing. For example, difficulties in speech perception in children with developmental dyslexia (Molinaro *et al.*, 2016), or older adults with hidden hearing loss (Plack *et al.*, 2014), might arise from a compromised acousto-linguistic transformation at fast temporal scales such as those of phonemes. More generally, the results inform speech and language models that need to explain a fundamental question in speech perception: Where and how the analysis of the acoustic speech signal is transformed into linguistic representations that enable speech comprehension.

ACKNOWLEDGEMENTS

The authors would like to thank Frankie Pennington and Joon Hyun Paik for manually checking phoneme, syllable, and word onsets and offsets for English and Korean recordings, respectively.

REFERENCES

Brainard, D.H. (**1997**). “The psychophysics toolbox,” Spat. Vis., **10**.

Delorme, A., and Makeig, S. (**2004**). “EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics,” J. Neurosci. Methods, **134**, 9-21.

Friederici, A.D., and Gierhan, A.M.E. (**2013**). “The language network,” Curr. Opin. Neurobiol., **23**, 250-254.

Hasson, U., Yang, E., Vallines, I., Heeger, D.J., and Rubin, N. (**2008**). “A hierarchy of temporal receptive windows in human cortex,” J. Neurosci., **28**, 2539-2550.

Hickok, G., and Poeppel, D. (**2007**). “The cortical organization of speech processing,” Nat. Rev. Neurosci., **8**, 393-402.

Khalighinejad, B., da Silva, G.C., and Mesgarani, N. (**2017**). “Dynamic encoding of acoustic features in neural responses to continuous speech,” J. Neurosci., **37**, 2176-2185.

Kocagoncu, E., Clarke, A., Devereux, B.J., and Tyler, L.K. (**2017**). “Decoding the cortical dynamisc of sound-meaning mapping,” J. Neurosci., **37**, 1312-1319.

Ladefoged, P., and Johnson, K. (**2010**). *A Course in Phonetics*. Boston: Wadsworth.

Lerner, Y., Honey, C.J., Silbert, L.J., and Hasson, U. (**2011**). “Topographic mapping of a hierarchy of temporal receptive windows using a narrated story,” J. Neurosci., **31**, 2906-2915.

Martin, B.A., Tremblay, K.L., and Korczak, P. (**2008**). “Speech evoked potentials: from the laboratory to the clinic,” Ear Hearing, **29**, 285-313.

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., and Carreiras, M. (**2016**). “Out-of-synchrony speech entrainment in developmental dyslexia,” Hum. Brain Mapp., **37**, 2767-2783.

Moulines, E., and Charpentier, F. (**1990**). “Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones,” Speech Commun., **9**, 453-467.

Overath, T., McDermott, J.H., Zarate, J.M., and Poeppel, D. (**2015**). “The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts,” Nat. Neurosci., **18**, 903-911.

Park, H., Ince, R.A., Schyns, P.G., Thut, G., and Gross, J. (**2015**). “Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners,” Curr. Biol., **25**, 1649-1653.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., and Roberts, T. (**2000**). “Auditory cortex accesses phonologcal categories: an MEG mismatch study,” J. Cogn. Neurosci., **12**, 1038-1055.

Plack, C.J., Barker, D., and Prendergast, G. (**2014**). “Perceptual consequences of "hidden" hearing loss,” Trends Hear., **18**, 1-11.

Poeppel, D. (**2003**). “The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time',” Speech Commun., **41**, 245-255.

Rosen, S. (**1992**). “Temporal information in speech: acoustic, auditory and linguistic aspects,” Philos. Trans. R. Soc. Lond. B. Biol. Sci., **336**, 367-373.

Sanders, L.D., and Neville, H.J. (**2003**). "An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers," Brain Res. Cogn. Brain Res., **15**, 228-240.

Sohn, H.-M. (**1999**). *The Korean Language*. Cambridge: Cambridge University Press.

Stevens, K.N. (**2000**). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Tremblay, K.L., Friesen, L., Martin, B.A., and Wright, R. (**2003**). "Test-retest reliability of cortical evoked potentials using naturally produced speech sounds," Ear Hearing, **24**, 225-232.

Woldorff, M.G., Liotti, M., Seabolt, M., Busse, L., Lancaster, J.L., and Fox, P.T. (**2002**). "The temporal dynamics of the effects in occipital cortex of visual-spatial selective attention," Brain Res. Cogn. Brain Res., **15**, 1-15.

Yoon, T.-J., and Kang, Y. (**2013**). "The Korean Phonetic Aligner Program Suite," http://korean.utsc.utoronto.ca/kpa/

Yuan, J., and Lieberman, M. (**2008**). "Speaker identification on the SCOTUS corpus," J. Acoust. Soc. Am., **123**, 3878.

Processing of fundamental frequency changes, emotional prosody and lexical tones by pediatric CI recipients

MONITA CHATTERJEE[1,*], MICKAEL L. D. DEROCHE[2], SHU-CHEN PENG[3], HUI-PING LU[4], NELSON LU[3], YUNG-SONG LIN[4,5], AND CHARLES J. LIMB[6]

[1] *Auditory Prostheses & Perception Laboratory, Center for Hearing Research, Boys Town National Research Hospital, Omaha, NE, USA*

[2] *Center for Research on Brain, Language and Music, McGill University, Montreal, QC, Canada*

[3] *Center for Devices and Radiological Health, United States Food and Drug Administration, Silver Spring, MD, USA*

[4] *Chi-Mei Medical Center, Department of Otolaryngology, Tainan, Taiwan*

[5] *Taipei Medical University, Taipei, Taiwan*

[6] *Department of Otolaryngology-Head & Neck Surgery, University of California San Francisco School of Medicine, San Francisco, CA, USA*

As cochlear implants (CIs) do not provide adequate representation of the harmonic structure of complex sounds, the perception of the voice fundamental frequency (F0) is severely limited in CI users. As F0 plays an important role in speech prosody and in lexical tones, this deficit has a negative impact on communication. Here we focus on the pediatric CI population, most of whom were prelingually deaf and were implanted before three years of age, within the most adaptive period of the brain's development. Our results suggest that, relative to their normally-hearing peers, school-age children with CIs have significant deficits in their sensitivity to both static and dynamic F0-changes. In addition, children with CIs also have deficits in their identification of emotional prosody and in lexical-tone recognition.

INTRODUCTION

Present-day cochlear implants (CIs) transmit acoustic-phonetic cues for speech recognition with sufficient fidelity for good-to-excellent speech recognition performance in quiet by the average patient, particularly when context cues are available and top-down reconstruction of the intended message can compensate for the degraded input. However, voice pitch, which conveys prosodic information and provides important cues for lexical tone recognition in tonal languages, is not transmitted appropriately by CIs. The voice fundamental frequency and its harmonics are lost in CI processing. Only the temporal envelopes extracted from the

*Corresponding author: monita.chatterjee@boystown.org

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

broader channel filters, which are modulated periodically when multiple harmonics of the fundamental frequency (F0) fall within the bandpass region of the filter, carry some voice pitch information. Unfortunately for CI patients, this envelope periodicity does not result in a salient enough percept to support music perception. For the recognition of prosody or lexical tones, however, the high level of pitch perception demanded by music may not be necessary. However, the extent to which the degraded pitch transmitted by CIs supports the recognition of these cues is not well understood.

For children who are implanted within the first months or years of life and are developing with the device in place, the implications of prosodic cue perception may be more profound than for post-lingually deaf adults, as language acquisition is thought to be mediated by prosodic information in infants and toddlers. These two populations provide an interesting comparison, bringing two different brains to the table. The post-lingually deaf adults developed auditory, cognitive and linguistic skills with good sensory and phonological representations. In contrast, pediatric CI recipients must develop all three with the degraded input of the device; Further, these very same cognition and language skills must then be harnessed in top-down repair of the degraded input. In studying this population, important questions regarding neural plasticity related factors (age at implantation, device experience, developmental factors) must be considered alongside other factors (sensitivity to the sensory input). For tone language speakers, we are additionally interested in the possible role of the tonal environment in the development of harmonic pitch sensitivity in pediatric CI recipients and in their normally hearing peers.

Here, we review recent work from our collaborative group on the recognition of emotional prosody and lexical tones by children with CIs in the context of the broader literature, and discuss implications for the development of CIs and rehabilitation of CI patients.

SENSITIVITY TO HARMONIC PITCH BY PEDIATRIC CI RECIPIENTS

Deroche *et al.* (2014) investigated harmonic pitch perception in pediatric CI users in the US and in Taiwan. They measured F0 difference limens in a 3-interval, 3-alternative forced-choice procedure, using the method of constant stimuli. The level of the stimuli was roved to avoid loudness cues. Psychometric functions were obtained, and thresholds for F0 discrimination were derived from these data. As expected, they reported large deficits in the children with CIs, both for a 100-Hz F0 reference and a 200-Hz F0 reference. No differences were observed between the children in the US and Taiwan, for both hearing status, suggesting that the tonal language environment did not influence static F0 discrimination in the normal-hearing (NH) or the CI children. There were small but significant age effects explaining about 10% of the variance across subjects, but no effects of age at implantation. In addition, they also reported that, in a task in which the participants had to discriminate between amplitude-modulation rate differences in sinusoidally amplitude-modulated broadband noise, the NH children's performance was the same as that of the CI children's (both being slightly worse than CI children's performance in the F0 discrimination task). Thus, when the NH children were compelled to use the same envelope periodicity cues available to CI children, their performance was not significantly different from that of the CI children's. This suggests further that the deficits in the CI children's performance may not be due to other differences between the populations (cognition, etc.), at least in these tasks. The CI children's performance in the F0 discrimination task was significantly correlated with that in the amplitude modulation (AM) rate discrimination task, supporting the idea that similar cues were used in the two tasks (despite small changes in the exact shape of the temporal envelopes and their coherence across channels). On the other hand, the NH children's thresholds in the two tasks were not correlated, supporting the idea that the NH children used very different cues in the two tasks (i.e., fine structure periodicity in the F0 discrimination task and temporal envelope periodicity in the AM rate discrimination task).

In a second study, Deroche *et al.* (2016) investigated the sensitivity to dynamic (swept F0) pitch in children with CIs in the US. They used two tasks, a direction labelling task and a direction discrimination task. In the first task, listeners heard a swept F0 harmonic complex and indicated whether it was rising or falling in a single-interval, 2-alternative forced-choice procedure. In the second task, they heard three swept F0 complexes: The first served as the reference, one of the remaining two intervals was identical to the reference in sweep range and direction while the other interval was identical in range but was swept in the opposite direction. The initial F0 was roved to avoid spectral-edge-related pitch cues, and level was roved to avoid any loudness cues. Results were similar to those obtained in the 2014 study, in that large deficits were observed in the CI children in both tasks. Significant age effects were also found (and corroborated by differences between children and adults in both populations), but no effects of age at implantation were observed.

We are presently conducting parallel studies in children in Taiwan in the swept-F0 tasks described above. In preliminary data, we observe a significant advantage in NH children in Taiwan over NH children in the US in both the direction labelling and the direction discrimination task (Fig. 1). This advantage is dominated by differences between the groups in the early developmental years, when the NH children in Taiwan appear to be already adult-like in both tasks, while the NH children in the US show a significant developmental effect, converging on the Taiwanese children's performance in their later teenage years. The data with child CI recipients shows no significant differences between the groups and no effects of age, although the difference between the US and Taiwanese CI users in the labelling task is marginally significant at p=0.051. Further data collection and analyses are ongoing to verify this trend.

To what extent are the F0 sweep rates in our studies relevant to those in lexical tones? He *et al.* (2016) attempted to relate the sensitivity to dynamic F0 sweeps to that of realistic F0 change rates in Mandarin tones. They adaptively modified the F0 range over which complex tones or recordings of lexical tones could be identified (as falling, rising, flat, dip, or peak). They found that NH adults could recognize 400-ms lexical tones with an F0 range only 40-cents wide, resulting in F0 change

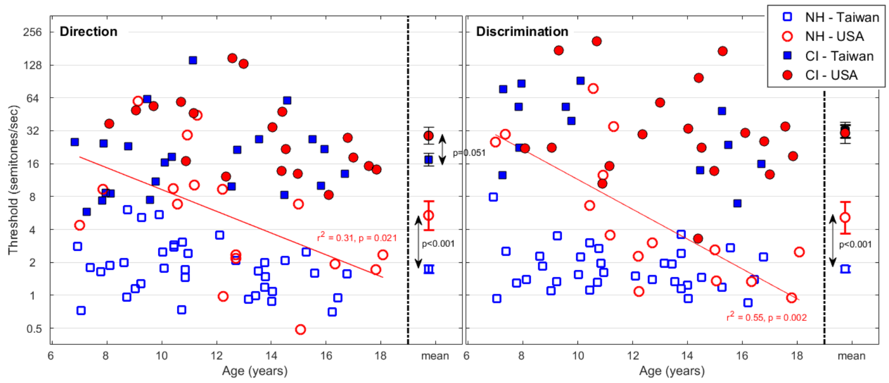


Fig. 1: Thresholds (in semitones/sec) obtained in an F0-sweep direction-labeling task (left-hand panel) and in an F0-sweep direction-discrimination task (right-hand panel) plotted against participants' age (in years). Blue and red symbols indicate results obtained in native speakers of American English and Mandarin, respectively. Open and filled symbols indicate results obtained in children with normal hearing and CIs, respectively. The right-hand side of each plot shows group means and standard deviations. (color online)

rates of only 1 semitone/s (or 2 semitones/s for dips). In contrast, CI adults needed F0 change rates of 2 octaves/s. These results and the size of the deficits exhibited by CI adults are very consistent with the dynamic F0 sensitivities reported by Deroche *et al.* (2016).

To summarize, our results indicate significant deficits in children with CIs (relative to their NH peers) in F0 discrimination of both static changes in F0 and swept-F0, dynamic changes. The data show evidence for an effect of linguistic environment, with the tone-language environment benefiting children with NH in both tasks, particularly in the early years of development. This finding is supported by other related studies in adults using both perceptual and physiological measures showing differences in the mechanisms of F0 coding and FM-sweep-coding in native speakers of Mandarin (Krishnan *et al.*, 2015; Krishnan *et al.*, 2011; Luo *et al.*, 2007a).

LEXICAL TONE RECOGNITION BY PEDIATRIC CI USERS

Lexical tones are relatively rapid changes in the F0 contour of syllables that signal the meaning of a word. In Mandarin, tones fall into four categories: Tone 1 consists of a high, flat F0, Tone 2 is predominantly rising in F0, Tone 3 is more complex, with a low initial F0, dipping to a minimum before rising again, and Tone 4 is a short, falling F0 contour. Similar to other studies, Peng *et al.* (2017) showed significant deficits in Taiwanese pediatric CI recipients' tone recognition (compared to NH peers) using a single-interval, 4-alternative forced-choice task in which participants pointed to the appropriate meaning depicted in a picture on the screen.

Performance was significantly predicted by the children's age at implantation but not by duration of device experience or by age at testing. This suggests that the ability to identify tones develops within a short period post-implant-activation, but does not change thereafter.

In a second task, Peng *et al.* (2017) asked whether CI children could utilize a secondary acoustic cue (duration) to their advantage in a tone recognition task. To do this, they used stimuli consisting of manipulated versions of an original utterance of the disyllabic word *"yanjing"*, in which the first syllable *yan* was kept constant but the second syllable *jing* was manipulated to have 8 steps of F0 contour, varying from Tone 1 (flat) to Tone 4 (falling). For each of these variations, six levels of duration of the second syllable were applied orthogonally. When the 2nd syllable is spoken in Tone 1, the word *yanjing* means "eyes", and when spoken in Tone 4, it means "eye glasses". The listeners heard each resynthesized utterance (in randomized order) and indicated which of the two meanings they thought it was associated with, in a single-interval, 2-alternative forced-choice task. The data were analysed using logistic regression to investigate the extent to which the children used F0 or duration cues to perform the task. It is to be noted that duration is not a strong cue for lexical tone recognition in Mandarin: It is a weak secondary cue at best.

Results (Fig. 2) showed that NH children used F0 cues extensively, but ignored the duration cue. On the other hand, the CI children made significant use of the F0 cue, but with less certainty than the NH children, while they used the duration cue significantly as well, much more so than the NH children. This suggests that even though CI children have significant deficits in F0 processing, they do rely on the cue in lexical tone recognition. Secondly, these results indicate that even though duration is not a strong acoustic cue for tones, the CI children implicitly develop the knowledge to use it appropriately to support their performance when it is explicitly provided. Age at implantation predicted their use of the duration cue, again suggesting an early development of tone-recognition skills.

Peng *et al.* (2017) also found that the use of the F0 cue in the yan jing task was a strong predictor of performance in lexical tone recognition with naturally uttered words. This is interesting, given the weakness of F0 sensitivity in this population, and underscores the dominance of F0 as a cue for lexical tones. The use of the duration cue, however, did not predict performance with the lexical tone recognition task, possibly because the cue is not reliable in natural Mandarin speech. However, the possibility remains that appropriate training with the secondary cue might help the children optimize their use of duration in natural speech recognition.

To summarize, children with CIs showed significant deficits in lexical tone recognition with naturally uttered words. They also showed a reliance on both F0 contour and duration cues in a Tone 1-Tone 4 cue-weighting task, while their NH peers only relied on the F0 contour. Lexical tone recognition was significantly predicted by the listeners' age at implantation, but not by their device experience.

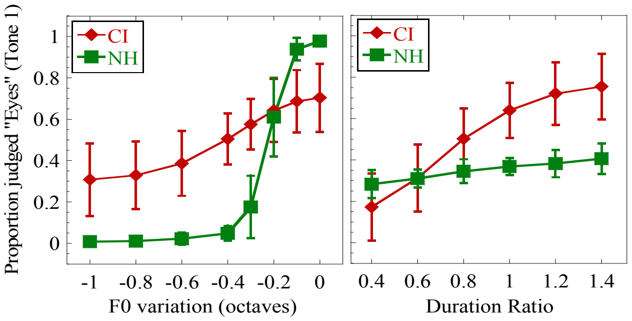


Fig. 2: The left hand panel shows the proportion of times the *yanjing* stimulus was associated with "Eyes" (Tone 1), plotted as a function of F0 variation of the second syllable by listeners with NH (squares) and with CIs (diamonds). The right hand panel shows the same, but plotted as a function of duration values of the stimuli. Error bars show +/- 1 s.d. from the mean. [Adapted from Peng et al. (2017)]

VOICE EMOTION RECOGNITION BY CHILDREN WITH CIS

Emotional prosody is critical for social communication. In children, appropriate emotional communication is important for the development of social interactions, both with their peers and their caregivers, and for social cognitive development in general. Luo *et al.* (2007b) showed that adult CI users had significant impairments in voice emotion perception. As many adult CI users are post-lingually deaf, it is reasonable to assume that they developed relatively normal understanding of emotions and emotional communication prior to their hearing loss and cochlear implantation. The literature on emotion understanding in children with CIs is relatively sparse, but suggests that while basic facial emotion recognition is developed in this population by the time they are school-aged, deficits in voice emotion recognition remain. We studied emotion recognition by school-age children with CIs using child-directed speech (i.e., exaggerated prosody), and compared performance with NH peers, adults with NH and adults with CIs (Chatterjee *et al.*, 2015). We found that, consistent with previous work in emotion recognition by CI recipients, the CI users showed significant deficits in voice emotion recognition. Interestingly, the post-lingually deaf adults and the pre-lingually deaf children with CIs showed very similar performance, suggesting a device-limitation in the task rather than factors related to cognition or language. However, this remains to be demonstrated. It is possible that similar performance can be achieved by different mechanisms.

Many of the participants in this study also participated in the studies on F0 processing by Deroche *et al.* (2014) and Deroche *et al.* (2016) described earlier. Based on correlational analyses, wee find that F0 sensitivity in all our tasks is a significant predictor of voice emotion recognition by CI listeners. This is noteworthy, given that CI patients show such large deficits in F0 processing, and also given that voice emotion is conveyed by numerous cues, including intensity, speaking rate, and timbre. This underscores the critical importance of F0 in the recognition of emotional prosody.

We also note that the stimuli in the Chatterjee *et al.* (2015) study contained exaggerated prosody, and that NH adults and children performed at ceiling in the task. This suggests that we underestimated the true deficit experienced by CI children in everyday communication, which consists of far more muted emotional cues.

Chatterjee *et al.* (2015) reported that the children with CIs showed a weak effect of device experience in their emotion recognition, but did not show a significant effect of developmental age. The children with NH showed a significant effect of age with full-spectrum stimuli, although their performance in this condition was near-ceiling, not significantly different from the adult NH listeners' performance, and significantly better than the performance of the children with CIs in the same condition. The children with NH also performed the task with 8-channel noise vocoded speech. While adults with NH showed performance similar to the mean CI performance in this condition, the children with NH showed significant deficits. Further analyses revealed a strong developmental effect (Fig. 3), with older children with NH showing significantly better performance than younger children, who clearly struggled to identify the emotions in the 8-channel vocoded condition. These findings underscore an important problem. The children with NH had experience

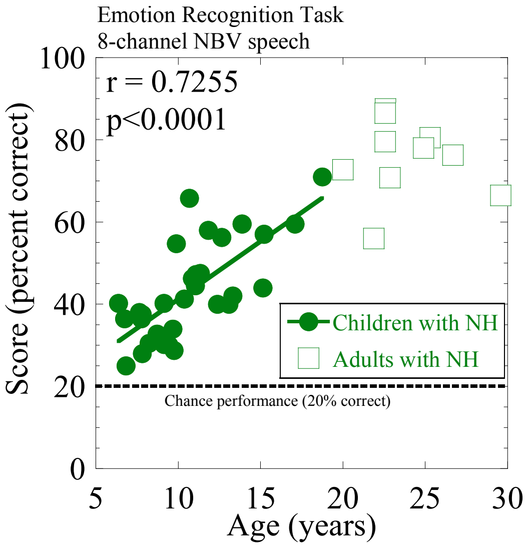


Fig. 3: Developmental effects in voice emotion recognition by children with NH when attending to 8-channel NBV speech. These children showed near-ceiling performance and smaller developmental effects in the same task with full-spectrum (unprocessed) speech. [Adapted from Chatterjee *et al.* (2015)]

with acoustic inputs, and different forms of degraded acoustic inputs (e.g., listening in reverberation or noise). They had presumably also developed language and cognition normally. Still, they struggled in this first encounter with CI-simulated speech. This speaks to the considerable difficulties faced by young children with CIs who do not have experience with acoustic inputs and have not developed the cognition or language skills that might help them to make sense of the new sensory signals. The advantage they have is the earlier implantation age and the greater plasticity of the brain within the first years of development. It seems critical, therefore, to ensure that rehabilitation efforts are maximized in the early months with the CI.

Unlike the lexical tone recognition data, the voice emotion recognition study did not show an effect of age at implantation on performance. Instead, duration of experience with the device (which was correlated with developmental age) was a significant predictor. This suggests that learning to read emotions continues as children develop, and that rehabilitation strategies may help by emphasizing prosodic cues even after the most sensitive period has passed.

SUMMARY AND CONCLUSIONS

To summarize, our results indicate that pediatric recipients of CIs must cope with large deficits in sensitivity to both static and dynamic changes in F0. In our studies, age at implantation is not predictive of performance in these psychophysical tasks, but the children with CIs show improvements in task performance as their duration of experience with the device (correlated with their chronological age at testing) increases. In preliminary work, we observe an advantage in dynamic-F0 sensitivity in native speakers of Mandarin over native speakers of American English. Sensitivity to voice emotion also shows deficits in children with CIs, and also shows age-related improvements. Although vocal emotions are represented by multiple acoustic cues, the dominance of F0 cues is demonstrated by the fact that F0 sensitivity is significantly correlated with performance in voice emotion recognition in CI users, even though this information is greatly degraded in this population. Native speakers of Mandarin developing with CIs showed significant deficits in lexical tone recognition relative to their NH peers, but also showed an ability to appropriately rely on a weak secondary acoustic cue, duration, in lexical tone contrast judgments. The age at implantation was a significant predictor of performance in lexical tone recognition, but increased duration of experience with the device did not correlate with performance in these tasks, suggesting a limit to adaptive advantages in sensitivity to tones early in life.

Overall, we conclude that the device-limitations placed on F0 coding in CIs are not sufficiently overcome by neural plasticity (age at implantation or duration of experience) as the child acquires experience with the device. Linguistic experience (tone language environment) may confer advantages to the NH population, but significant advantages have not been observed in our preliminary data with CI users as yet. The fact that the Mandarin-speaking children with CIs were able to utilize the secondary acoustic cue in tone identification suggests a role for training in tone recognition. We further conclude that F0-coding-limitations play a role in CI users' sensitivity to vocal emotion processing. These findings have implications for social communication, language acquisition, and rehabilitation efforts for the pediatric CI population and for device development strategies in the future.

ACKNOWLEGEMENTS

This work was supported by the following grants from the National Institutes of Health: R01 DC004786-08S1, R21 DC011905, R01 DC014233 (MC), and by support from the Chi-Mei Research Foundation and Taiwan National Science Council Grant 100-2314-B-038-037 (YSL).

REFERENCES

Chatterjee, M., Zion, D.J., Deroche, M.L., Burianek, B.A., Limb, C.J., Goren, A.P., Kulkarni, A.M., and Christensen, J.A. (**2014**) "Voice emotion recognition by cochlear-implanted children and their normally-hearing peers," Hear. Res., **322**, 151-162. doi: 10.1016/j.heares.2014.10.00

Deroche, M.L., Lu, H., Limb, C.J., Lin, Y. and Chatterjee, M. (**2014**). "Deficits in the pitch sensitivity of cochlear-implanted children speaking English or Mandarin," Front. Neurosci. **8**, 282. doi: 10.3389/fnins.2014.00282

Deroche, M.L.D., Kulkarni, A.M., Christensen, J.A., Limb, C.J., and Chatterjee, M. (**2016**) "Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants," Front. Neurosci., **10**, 0007. doi:10.3389/ fnins.2016.00073

He, A., Deroche, M.L.D., Doong, J., Jiradejvong, P., and Limb, C.J. (**2016**). "Mandarin tone identification in cochlear implant users using exaggerated pitch contours," Otol. Neurotol., **37**, 324-331. doi: 10.1097/MAO.0000000000000980

Luo, H., Boemio, A., Gordon, M., and Poeppel, D. (**2007a**). "The perception of FM tones by Chinese and English listeners," Hear. Res., **224**, 75-83. doi: 10.1016/j.heares.2006.11.007

Luo, X., Fu, Q.J., and Galvin, J.J. 3rd. (**2007b**) "Vocal emotion recognition by normal-hearing listeners and cochlear implant users," Trends Amplif., **11**, 301-315. doi: 10.1177/1084713807305301

Krishnan, A., Gandour, J.T., Ananthakrishnan, S., Bidelman, G.M., and Smalt, C.J. (**2011**) "Linguistic status of timbre influences pitch encoding in the brainstem," Neuroreport, **22**, 801-803. doi: 10.1097/WNR.0b013e32834b2996

Krishnan, A., Gandour, J.T., Ananthakrishnan, S., and Vijayaraghavan, V. (**2015**). "Language experience enhances early cortical pitch-dependent responses," J. Neurolinguistics, **33**, 128-148. doi: 10.1016/j.jneuroling.2014.08.002

Peng, S.C., Lu, H.P., Lu, N., Lin, Y.S., Deroche, M.L., and Chatterjee, M. (**2017**) "Processing of acoustic cues in lexical tone identification by pediatric cochlear implant recipients," J. Speech Lang. Hear. Res., **60**, 1223-1235. doi: 10.1044/2016_JSLHR-S-16-0048

Data-driven hearing care with time-stamped data-logging

NIELS HENRIK PONTOPPIDAN[1,*], XI LI[1], LARS BRAMSLØW[1], BENJAMIN JOHANSEN[1,2], CLAUS NIELSEN[1], ATEFEH HAFEZ[1], AND MICHAEL KAI PETERSEN[1, 2]

[1] *Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark*

[2] *Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark*

Modern hearing aids holds significant personalization potentials while the processes associated with the administration do not fully accommodate the dialogue for finding the optimized and personalized settings. The hearing aids presented here use a connected smartphone to log a snapshot of 21 sound environment parameters every minute, e.g., sound pressure level in low, mid, and high frequencies and broadband, the estimate of the signal-to-noise ratio in the same 4 bands, the sound environment detector, etc. This data stream shows the sound environments that the user of the hearing aids experiences. The continuous stream of sound environment data is supplemented by the user's operation of the hearing aid, e.g., which program is chosen when, and how is the volume control adjusted as well. Whenever the user changes program or volume, the change is logged with the time stamp. Together, the continuous and event based data logging reveals in which situations the user prefers a given program and on the bigger time-scale, which program that should be the default program. The close integration of the hearing aid, the mobile phone, and cloud services turning the hearing aid into an Internet of Things device not only enable the learning and adaptation but also supplementing the dialogue between user and audiologist with objective data about the actual use of the hearing aids.

INTRODUCTION

How can a hearing-aid user describe what they did not hear? How can a hearing aid user describe situations where hearing is difficult to the extent where the audiologist becomes sufficiently certain about the validity of the description? This dialogue based trial and error procedure can be rather time consuming, and may prevent the fitting process from exploring sufficient number of alternatives, and thus may prevent the fitting of the hearing aids from reaching the degree of personalization which modern hearing aids support.

We present a prototype hearing aid that can provide information about sound environments and use. With this hearing aid, we ask, can data logging enhance the hearing aid fitting? Moreover, if so, how can analysis of logged data enhance hearing

*Corresponding author: npon@eriksholm.com

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

aid fitting? The prototype hearing aid is developed for the H2020 project EVOTION (http://h2020evotion.eu) and transmits program changes, volume adjustments, and continuous sound environment data to a connected smartphone for buffering, local storage, and upload to cloud services utilizing the Internet of Things services of Oticon Opn (Oticon, 2016), the hearing aid which EVOTION hearing aid is based on.

The perspectives for use of such hearing aids in future hearing-care are very diverse; ranging from empowering the hearing-aid user, empowering the dialogue between hearing-aid user and audiologist, learning based adaptation of selected hearing-aid features, providing developers feedback on use of prospective hearing-aid features to clinical trials with new hearing-aid features. Additionally, the scenarios are not limited to hearing-aid features, but can also evaluate if auditory training increases the ability to hear in difficult situations or to cope with increasingly difficult situations. In fact, in the H2020 project EVOTION the hearing aids will contribute to research on public health policy modelling (Prasinos *et al.*, 2017).

METHOD

Internet services

Hearing aids connect to internet and internet services through an accompanying service installed on a smartphone and connected to the hearing aid using Bluetooth Low Energy (Oticon, 2016). This enables interaction with internet enables services, such as If This Then That (IFTTT, http://www.ifttt.com). Through IFTTT the individual user can define so-called recipes, where defined actions on the hearing aid can trigger another service and vice versa. Such recipe could connect the low-battery alert on the hearing aid to a text-messaging service, alerting a parent that the child's hearing aid is running low on battery. Another recipe connects an IFTTT enabled doorbell to the hearing aid, and alerts the wearer when someone presses the doorbell. The EVOTION hearing aids use the same communication methods between hearing aid and mobile as the IFTTT service.

The present data logging system use the NRF Connect app available on Google Play and App Store to intercept the data transmitted from the hearing aid to the smartphone. It also requires that the hearing aid user remembers to save the data stored in the NRF Connect app every night to a file. In this study with only a few hearing-aid users, the hearing-aid users must remember to save and upload the logging data every night. However, this is only true for the hearing-aid users taking place in this pilot study, the hearing-aid users in EVOTION will have a special app that acts as a remote control for the EVOTION hearing aids (like the Oticon On app does for Oticon Opn).

Sound environment data

Every minute an interrupt triggers the EVOTION hearing aid to transmit the current value of 21 sound environment parameters. The first 20 parameters are Sound Pressure Level, Noise Floor Level, Modulation Index, Modulation Envelope, and Signal to Noise Ratio measured in dB in four frequency ranges: 0-1.3 kHz, 1.3-4.1 kHz, 4.1-10 kHz, and 0-10 kHz. The last parameter is Sound Environment, which can take four values Quiet, Speech, Speech-in-Noise, and Noise. The 21 parameters are merely a small subset of the estimators running continuously in the hearing aid to characterize the sound environment and adapt the automatic systems to the current sound environment. When the connected mobile phone receives the message with the 21 sound environment parameters the values are stored and time-stamped.

Personalization

The personalization explored with the EVOTION hearing aids are the settings of OpenSound Navigator™ (Le Goff *et al.*, 2016). As shown in Fig. 1, the settings of OpenSound Navigator have different thresholds where OpenSound Navigator increases the amount of processing. The labels: High, Medium, and Low refer to how much help, e.g., how often the hearing aid attenuates a sound labelled as noise. In the Low setting, the signal to noise ratio required to trigger a certain amount of attenuation is higher than for the High setting.

The usual fitting practice is to assign each hearing-aid user to one of the settings: High, Medium, and Low, based on their preferences assessed in a questionnaire. The fitting of the EVOTION hearing aid gives the user access to four OpenSound Navigator settings as four programs. Allowing the hearing-aid user to shift between the four programs, provides the hearing-aid user access to a meta-parameter (Schum and Beck, 2006), which was previously only available to the audiologist.

As the hearing-aid user explores the different programs in different sound environments, the data logging will show which program the hearing aid user prefers, measured as the program used the most in such situation. Obviously, this requires the hearing-aid user to try out the different programs in different sound environments. When instructed to do so, hearing-aid users do vary the used programs (Johansen *et al.*, 2017), however, in that study the hearing aids did not transmit the sound environment data as is the case with the EVOTION hearing aids.

The personalization takes place the moment that the analysis of the logged data enables the selection of a single profile as the preferred profile. Either as an overall preference or as the preferred profile in a given situation characterized by the 21 sound environment parameters. The overall preference can be obtained by modifying the default program of the hearing aid, while the situation based preference require the mobile phone to use the remote control interface to select the preferred program.

With the EVOTION hearing-aids we collect preferences in a different way than the popular ecological momentary assessment (EMA) method (Wu *et al.*, 2015; Kissner *et al.*, 2015). The data-logging supports EMA and the two methods can be used in parallel, such that an EMA event is logged on the phone similar to the program shifts and volume adjustments. It is evident that some kind of hearing-aid user input must be logged, but whether asking the hearing-aid user to rate settings at certain intervals or if usage patterns are sufficient remains to be investigated.

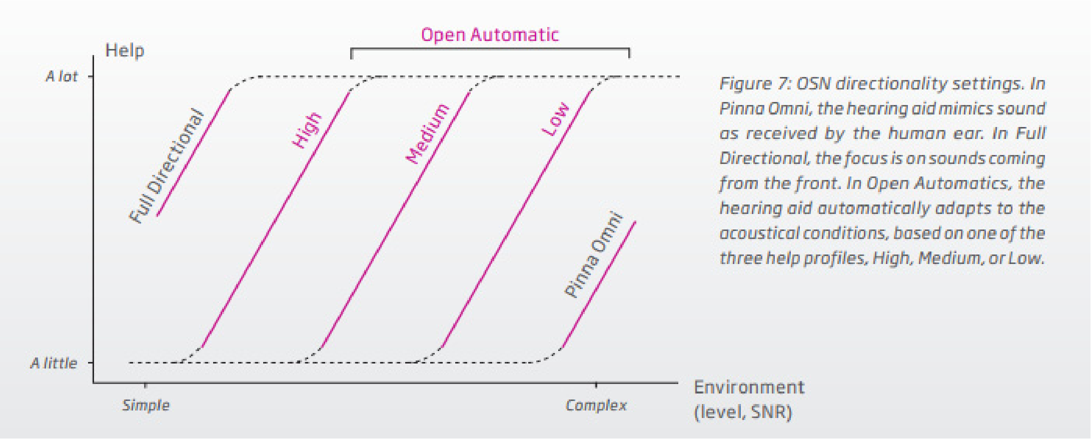

Open Automatic
Help
A lot
A little
Full Directional
High
Medium
Low
Pinna Omni
Simple
Complex
Environment
(level, SNR)
Figure 7: OSN directionality settings. In Pinna Omni, the hearing aid mimics sound as received by the human ear. In Full Directional, the focus is on sounds coming from the front. In Open Automatics, the hearing aid automatically adapts to the acoustical conditions, based on one of the three help profiles, High, Medium, or Low.

Fig. 1: OpenSound Navigator settings as function of environment.

RESULTS

Figures 2 and 3 show the first data logged by a hearing-aid user with the EVOTION hearing aids and acts to demonstrate the that hearing aids can log continuous sound environment data and event based program changes.

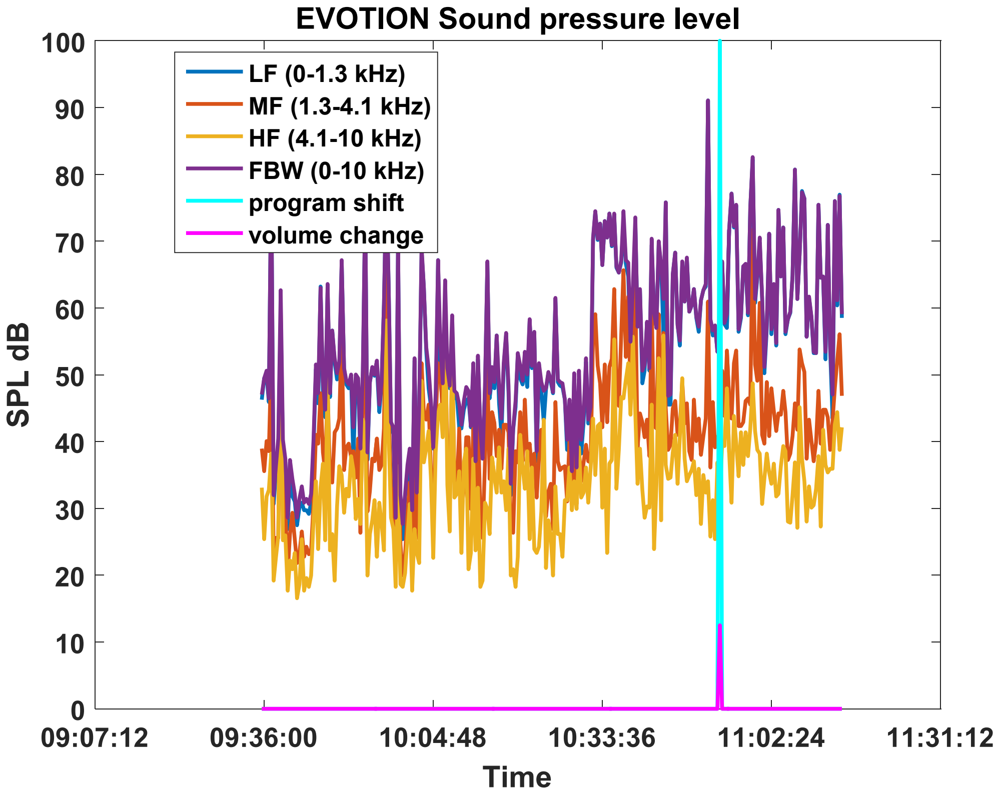


Fig. 2: Logged Sound Pressure Level data and program shift.

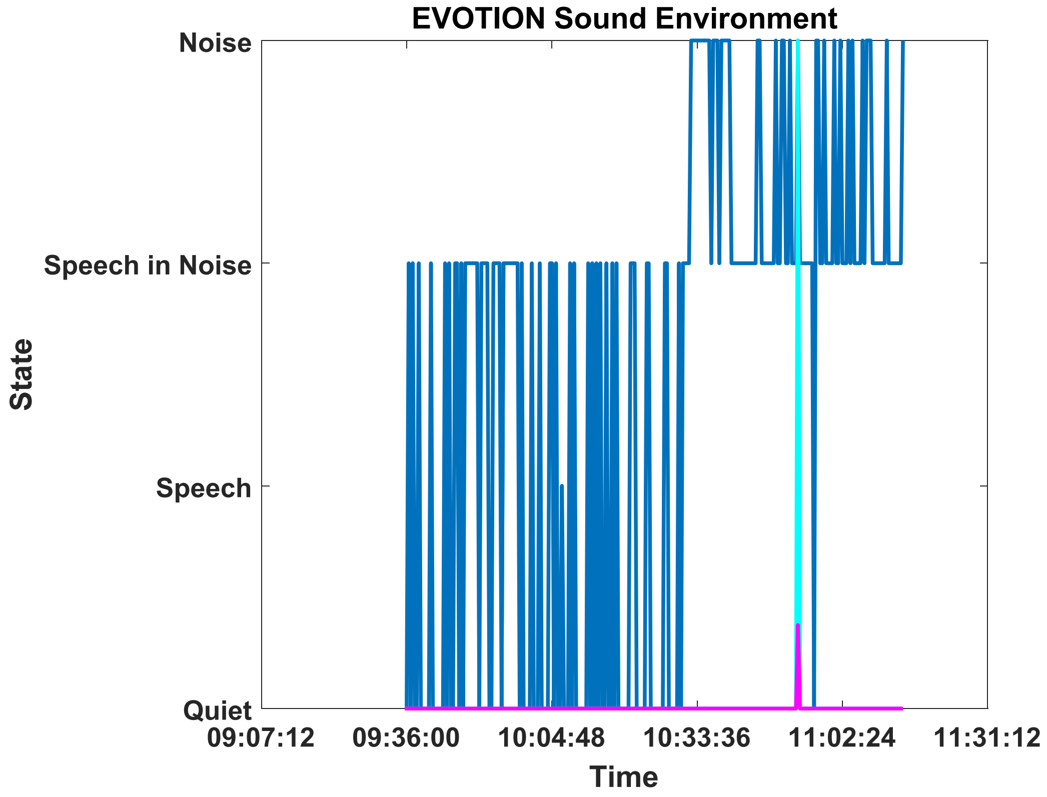


Fig. 3: Logged Sound Environment data.

Figure 2 shows a hearing-aid user who spends most of the morning in a relatively quiet environment, and Figure 3 shows that even if the levels are not loud, the sound environment shifts between quiet, and speech in noise. However, at 10:30 something seems to happen, as the sound levels go up, and also the sound environment detector suddenly shifts between speech in noise and noise.

The EVOTION hearing aids may also provide valuable data even if the user is not using all of the four programs. The logging of an event indicates that the hearing aid is in use, and thus an app or a cloud based data analysis tool can tell hearing-aid users for how long they use their hearing aid, if they are adhering to the agreed usage target, as well as providing a much more detailed overview of when the hearing aids were in use. In previous studies using objective measure of hearing aid use (Laplante-Lévesque *et al.*, 2014) was based on aggregated tables of hearing aid use, and did not give access to as detailed information about usage patterns as the current study allows.

CONCLUSIONS

We have presented the EVOTION hearing aid and showed the first sound environment data, logged by a hearing-aid user.

ACKNOWLEDGEMENTS

This work was partly funded by European Union's Horizon 2020 Research and Innovation Program under grant agreement 727521 EVOTION (h2020evotion.eu). Benjamin Johansen's PhD project is funded by Oticon Foundation and Copenhagen Centre for Health Technology (www.cachet.dk). The authors are grateful to the hearing-aid users using the EVOTION hearing aids even if the functionality of the NRF connect app far from matches the performance and stability of the mobile application developed for Oticon Opn.

REFERENCES

Johansen, B., Flet-Berliac, Y.P.R., Korzepa, M.J., Sandholm, P., Pontoppidan, N.H., Petersen, M.K., and Larsen, J.E. (**2017**). "Hearables in hearing care: Discovering usage patterns through IoT devices," in *International Conference on Universal Access in Human-Computer Interaction*, Springer, pp. 39-49.

Kissner, S., Holube, I., and Bitzer, J. (**2015**). "A smartphone-based, privacy-aware recording system for the assessment of everyday listening situations," Proc. ISAAR, **5**, 445-452.

Laplante-Lévesque, A., Nielsen, C., Jensen, L.D., and Naylor, G. (**2014**). "Patterns of hearing aid usage predict hearing aid use amount (data logged and self-reported) and overreport," J. Am. Acad. Audiol., **25**, 187-198.

Le Goff, N., Jensen, J., Pedersen, M.S., and Callaway, S.L. (**2016**). "An Introduction to OpenSound Navigator," retrieved from https://www.oticon.com/~/media/Oticon US/main/Download Center/White Papers/15555-9950 - OpnSound Navigator.pdf

Oticon (**2016**). "Opn hearing aid offers 'open sound' experience," The Hearing Journal, **69**, 44. doi: 10.1097/01.HJ.0000489201.24832.e3

Prasinos, M., Spanoudakis, G., and Koutsouris, D. (**2017**). "Towards a model-driven platform for evidence based public health policy making," 29th International Conference on Software Engineering and Knowledge Engineering, Pittsburgh, USA. doi: 10.18293/SEKE2017-035

Schum, D., and Beck, D. (**2006**). "Meta controls and advanced technology amplification," AudiologyOnline, March 27, retrieved from http://www.audiologyonline.com/articles/meta-controls-and-advanced-technology-989.

Wu, Y.-H., Stangl, E., Zhang, X., and Bentler, R.A. (**2015**). "Construct validity of the ecological momentary assessment in audiology research," J. Am. Acad. Audiol., **26**, 872-884.

Steering of audio input in hearing aids by eye gaze through electrooculography

ANTOINE FAVRE-FÉLIX[1,2,*], RENSKJE K. HIETKAMP[1], CARINA GRAVERSEN[1], TORSTEN DAU[2], AND THOMAS LUNNER[1,2,3]

[1] *Eriksholm Research Centre, Snekkersten, Denmark*

[2] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[3] *Swedish Institute for Disability Research, Linnaeus Centre, HEAD, Linköping University, Linköping, Sweden*

The behavior of a person during a conversation typically involves both auditory and visual attention. Visual attention implies that the person directs his/her eye gaze towards the sound target of interest, and hence the detection of the gaze may provide a steering signal for future hearing aids. Identification of the sound target of interest could be used to steer a beamformer or select a specific audio stream from a set of remote microphones. We have previously shown that in-ear electrodes can be used to identify eye gaze through electrooculography (EOG) in offline recordings. However, additional studies are needed to explore the precision and real-time feasibility of the methodology. To evaluate the methodology we performed a test with hearing-impaired subjects seated with their head fixed in front of three targets positioned at −30°, 0°, and +30° azimuth. Each target presented speech from the Danish DAT material, which was available for direct input to the hearing aid using head related transfer functions. Speech intelligibility was measured in three conditions: a reference condition without any steering, an ideal condition with steering based on an eye-tracking camera, and a condition where eye gaze was estimated from EarEOG measures to select the desired audio stream. The capabilities and limitations of the methods are discussed.

INTRODUCTION

Due to more advanced signal processing algorithms developed in recent years (Puder, 2009), the performance of hearing aids (HA) has greatly increased. Nevertheless, many HA users still report difficulties to communicate in acoustically challenging conditions with various sound sources and in the presence of reverberation. One of the reasons for the difficulties may be that the HAs are not sensitive to the user's attention and therefore cannot "react" accordingly while normal-hearing listeners typically are able to selectively attend to the target of interest. The European project COCOHA (COgnitive COntrol of a Hearing Aid) attempts to track the attention of the

*Corresponding author: afav@eriksholm.com

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

HA user via electroencephalographic (EEG) brain activity as well as via electrooculography (EOG). EOG is strongly correlated with eye movements, such that the user's attention could be estimated by eye gaze, assuming that the user is looking at the target of interest. Manabe *et al.* (2013) and Favre-Félix *et al.* (2017) showed that it is possible to reliably measure EOG using in-ear electrodes. Favre-Félix *et al.*, (2017) even used a solution with electrodes integrated in the hearing aid mold designed by the Eriksholm research centre (Fiedler *et al.*, 2016, Pedersen *et al.*, 2014). Thus, advanced hearing devices may in the future be able to measure the eye gaze and steer the amplification of attended versus unattended audio input accordingly. However, even though EOG is closely correlated to eye movements, real-time steering from an EOG signal may be difficult to implement. Here, we designed an experiment to evaluate the potential of steering audio signals through EOG. Hearing-impaired participants were presented one target talker and two competing talkers. Different steering conditions were considered: a control condition where the eyes did not provide any steering; a condition where the eye gaze was estimated using the EOG signal and the audio was steered accordingly, and an ideal condition where the eye gaze was accurately detected through an eye tracker and the audio was steered accordingly.

METHODS

Participants

Eleven hearing-impaired participants took part in the study. Their average age was 75 years, with a standard deviation of 8.9 years. Their audiograms showed moderate to moderately-severe sensorineural, symmetrical hearing loss; the maximum difference between the left and right ear (averaged between 125 and 8000 Hz) was 10 dB and the frequency pure-tone average of thresholds at 500, 1000, 2000, and 4000 Hz ranged from 45 to 69 dB HL (average 55 dB HL). The participants were wearing state-of-the-art behind-the-ear devices fitted with the NAL-NL2 rationale with directionality and noise reduction features turned off.

Stimuli and experimental setup

The participants were presented speech from the Danish DAT material (Nielsen and Dau, 2013), an open-set speech corpus with target words embedded in a carrier sentence. The material consists of sentences in the form of "Dagmar/Asta/Tine tænkte på en *skjorte* og en *mus* i går" ("Dagmar/Asta/Tine thought of a *shirt* and a *mouse* yesterday"). "Skjorte" and "mus" are two target words that change between each sentence and between each talker. All sounds were presented diotically via direct audio input. The participants were asked to direct their gaze at the talker indicated by a light-emitting diode (LED) and to repeat the two target words after the sentence was presented.

The experimental setup is shown in Fig. 1. The participants' head was fixed by a chin-rest. In front of the participant, at a distance of 72 cm, the voices of three talkers (one target talker, two interferers) were presented from the locations −30°, 0° and +30°

azimuth relative to the chin-rest. The sounds were generated via generic head-related transfer functions (HRTFs) corresponding to the three directions. The level of the target talker was initially 6 dB higher than the level of each of the interfering maskers. This was done since hearing-impaired listeners typically have a speech reception threshold (corresponding to 50% correct speech intelligibility) at a target-to-masker ratio (TMR) of +6 dB (Nielsen and Dau, 2013).

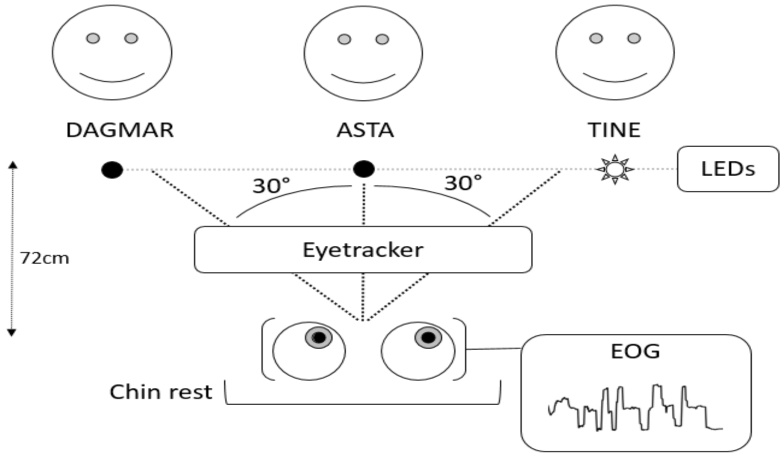


Fig. 1: Representation of the experimental setup. There were three talkers (one target talker (Tine in this example), indicated by an active LED, and two interfering talkers) in front of the participant. The head was fixed with a chin-rest and the eye gaze was measured with an eye tracker and estimated via EOG.

In the control condition (without steering), the behavior of the participant had no impact on the presentation of the audio signal. In the "EOG steering" condition, the EOG signal was used to estimate the eye gaze and to amplify the audio coming from the estimated target talker. In the "eye-tracking" steering condition, the eye gaze of the participant was detected through an eye tracker. In the "EOG steering" and the "eye-tracking" conditions, the audio signal coming from the visually attended talker at was amplified by additional 12 dB to ensure that the participant could clearly identify the target source while still perceiving the interferers (McShefferty *et al.*, 2016).One training list of 20 sentences and three test lists of 20 sentences were used for each condition. The target switched randomly between each sentence such that each talker was presented at least six times per list and each possible transition occurred at least twice.

In the eye-tracking condition, the gaze was estimated at a rate of 30 Hz using an Eyetribe eye tracker (The Eye Tribe ApS, Copenhagen, Denmark). For practical reasons the calibration of the eye tracker was set once and was not adjusted to each individual participant. For the EOG signal, the bioelectric potentials were measured with a g.tec biosignal amplifier (medical engineering GmbH, Schiedlberg, Austria) sampling at 256 Hz, using three in-ear electrodes in each ear, an electrode on each temple and a reference and a ground electrode on the arm. The EarEOG signal studied was from the cleanest electrode in the right ear re-referenced to the cleanest electrode in the left ear, the EOG signal studied was from the electrode on the right temple re-referenced to the electrode on the left temple. Originally, the goal was to use in-ear electrodes (EarEOG) to steer the amplification instead of surface EOG on the temples. However, during testing the EarEOG signal was considered to be too noisy to be reliably used for steering at this stage. Therefore, the EOG signal, which is less affected by noise interference, was considered instead.

EOG steering algorithm

The main challenge of using EOG in real-time is a direct current (DC) drift that is created by the interface between the skin and the electrodes (Huigen *et al.*, 2002; Favre-Félix *et al.*, 2017). Therefore, it is not straightforward to accurately determine the eye gaze relative to the nose from these measurements, whereas eye movements indicative of attention switch can easily be detected. In order to extract meaningful information, a bandpass filter with cut-off frequencies of 0.1 and 4 Hz was applied to the EOG signal. This filtering is effective when the eyes move rapidly (i.e. when the eyes stay less than two seconds on a target), but not when the eyes are fixated on a target. When the eyes are fixated, low-frequency components appear in the EOG, which are then filtered out such that the signal approaches zero. The algorithm used in this study was designed to detect the changes in eye gaze, to estimate when the eyes switch from one target to another and to anticipate this modification of the EOG signal caused by the filtering. According to the positioning of the electrodes that were used to measure the EOG, the filtered EOG signal was positive when the eyes moved to the right and the filtered EOG signal was negative when the eyes moved to the left Since there are three possible targets, five patterns of potential movements can occur: no movement, switching to a target on the right, switching to a target on the left, switching to two targets on the right and switching to two targets on the left.

For this continuous classification, two thresholds were set. The first threshold differentiated between a movement and no movement. The second threshold, higher than the first one, differentiated between switching by one or two targets as illustrated in Fig. 2. The sign of the EOG signal indicated whether the eyes were moving to the left or to the right. A target change was detected when the signal remained above the threshold for 500 ms. This allowed the system to be robust against brief noises, such as eye blinks and jaw movements. Once a target change was detected, the EOG signal was reset to zero to anticipate the modification caused by the filtering. Using this classification system, a mistake could potentially propagate over several sentences Therefore, the algorithm was reset to the middle target in the beginning of each list of 20 sentences. When the participant repeated the words they heard, the algorithm was locked because jaw movements are known to have a strong influence on the in-ear electrode signal.

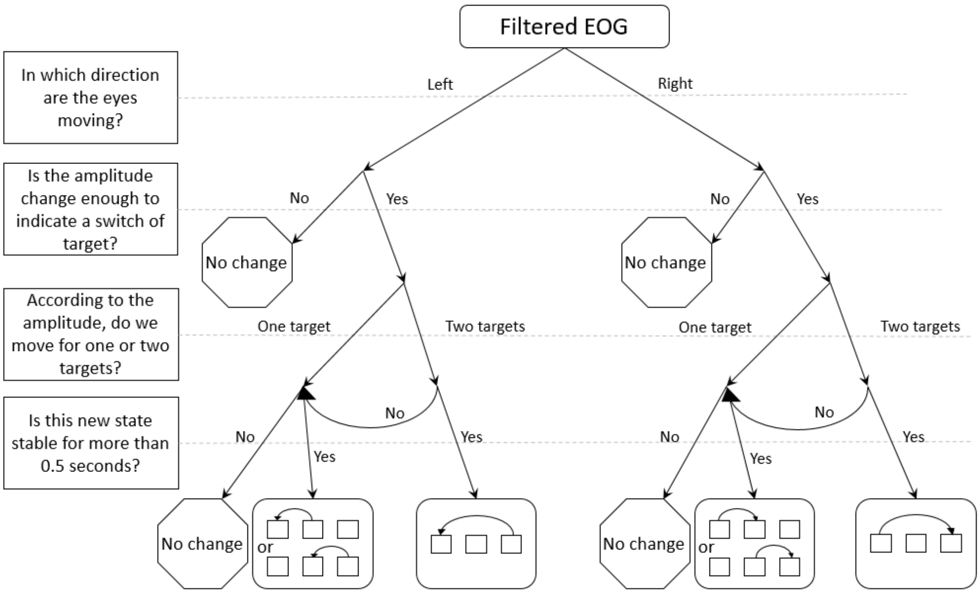

Filtered EOG
In which direction are the eyes moving?
Left
Right
Is the amplitude change enough to indicate a switch of target?
No
Yes
No change
According to the amplitude, do we move for one or two targets?
One target
Two targets
Is this new state stable for more than 0.5 seconds?
No
Yes
or

Fig. 2: Decision tree representing the decisions taken by the algorithm to estimate attention shift for the EOG steering system. First, the algorithm evaluates the sign of the filtered EOG to determine the direction the eyes are moving. Then the signal is compared to thresholds values to decide if the estimated eye movement is large enough to change the target source and, if so, to decide to switch to a target. Finally, the algorithm controls that the signal change is not caused by a transient noise.

Analysis

The analysis of the data focused on the results obtained from the seven participants for whom EOG data were measured. For the other four participants, for whom only EarEOG signals were recorded, the data were too noisy for the algorithm to detect the attended target reliably. The scoring of the correctly repeated words per sentence from the DAT material was measured. Two aspects of that score were considered: the score of individual words that were correctly repeated, and the score of full sentences that were correctly repeated. A t-test analysis was applied to compare these scores between conditions, averaged across participants.

The accuracy of the eye-gaze detection algorithm was estimated throughout the whole duration of the experiment, including during the “no steering” and the “eye-tracker steering” conditions. For the duration of each sentence, the estimated target was compared to the target the participant was supposed to attend. Analysis showed two types of errors: when the algorithm changed the target while the sentence was playing, representing a “switch error”, and when the algorithm was fixed on the wrong target, in which case it was possible to estimate how much the algorithm deviated from the attended target.

RESULTS

In terms of word scoring, in the “no steering” condition the participants repeated the word correctly 58.1% of the time, on average, with a standard deviation of 19.3%. In the “EOG steering” condition, the percentage of correct responses was 66.2%, with a standard deviation of 21.7%. In the “eye-tracker steering” condition, the percentage correct was 84.9% with a standard deviation of 11.9%. There was a significant difference between the “no steering” and “eye-tracker steering” conditions (*p*<0.00001) and between the “eye-tracker steering” and “EOG steering” conditions (*p*<0,001), but no significant difference between the “no steering” and the “EOG steering” conditions as illustrated in the left panel in Fig. 3.

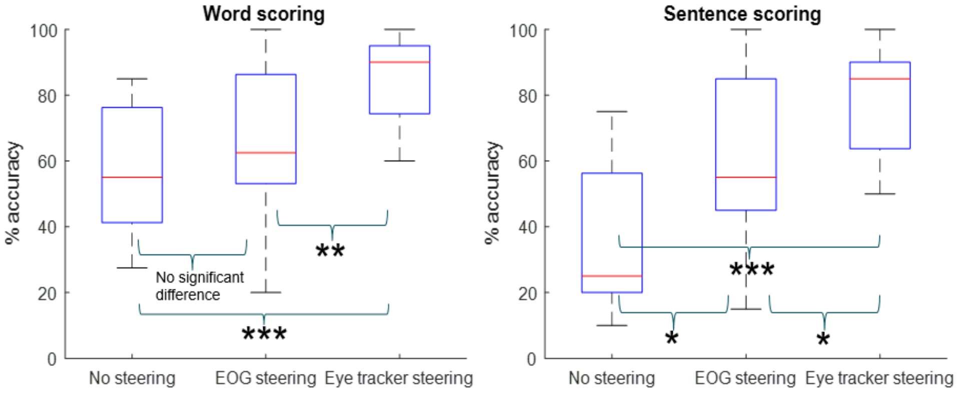


Fig. 3: Average word scoring (left panel) per condition and average sentence scoring (right panel) per condition (* *p*<0.01; ** *p*<0.001; *** *p*<0.00001) in the three conditions with “no steering”, “EOG steering” and “Eye-tracker steering”.

In terms of sentence scoring, in the “no steering” condition the participants, on average, repeated the whole sentence correctly 38.8% of the time, with a standard deviation of 22%. In the “EOG steering” condition, the corresponding percentage correct was 61.7%, with a standard deviation of 23.4%. In the “eye-tracker steering” condition, the percentage amounted to 78.1% (standard deviation 15.8%). There was a significant difference between the “no steering” and “eye-tracker steering” conditions (*p*<0.00001), between the “EOG steering” and “eye-tracker steering” conditions (*p*<0,01) and between the “no steering” and the “EOG steering” conditions (*p*<0,01) as illustrated in the right panel in Fig. 3.

The algorithm to estimate the attended target through EOG had an accuracy of 65%. The algorithm erroneously detected a change in the middle of a sentence 8.5% of the time and selected the wrong target 26.5% of the time. Specifically, 13.5% of the time the left neighbour was selected, 7% the right neighbour, 3% the left target when it actually was the one to the right, and 3% the right target when it actually was the one to the left. This error distribution is illustrated in Fig. 4. Since there are three targets, a random selection of the target would result in 33% accuracy, or less if the change during the sentence was taken into account. Therefore, the target estimation algorithm used in this experiment was considered effective.

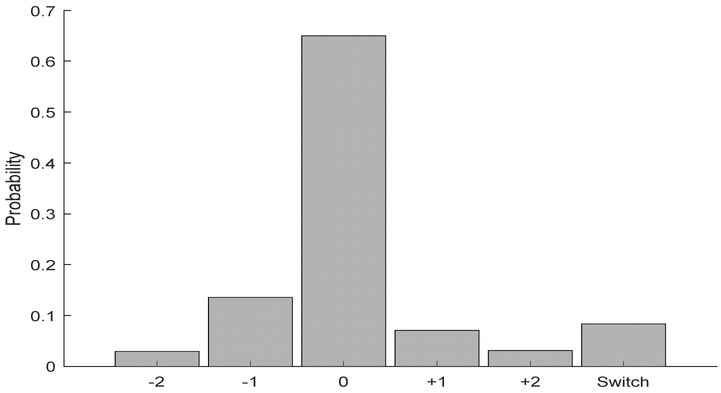


Fig. 4: Histogram representing the accuracy of the algorithm representing the distribution of correct (“0”) and incorrect decisions (“+/−1”, “+/−2”, switch).

DISCUSSION

In this study, we evaluated the potential of steering audio signals through EOG. When estimating EOG from surface electrodes, a significant improvement was seen in sentence performance but not for word scoring compared to the no steering condition. However, the performance of the EOG was in both cases significantly less compared to the ideal condition using an eye tracker to estimate gaze direction.

The EOG estimates were based on surface electrodes on the temples, although the original goal was to use EarEOG. Unfortunately, the EarEOG were too noisy to run the algorithm. This phenomenon was an unexpected challenge, as previous recordings of EarEOG did not display such noise issues (Favre-Félix *et al.*, 2017). Furthermore, the setup in this experiment was tested in a pilot experiment before the actual study presented in this paper was conducted. It is possible that the participants of the present study felt less comfortable with the experimental setup than those involved during pilot testing. Further studies both with normal-hearing and hearing-impaired listeners will clarify the origin of the noise.

The results obtained with both word and sentence scoring using the eye-tracker steering demonstrated the potential of a device that is steered via eye gaze. There were still some errors in this condition, mostly resulting from the calibration of the eye tracker that was not adjusted to the individual participant. Based on the results obtained in this study, a future technology to separate voices in a “cocktail-party” like situation may be based on gaze steering. This could for example be utilized by using a remote microphone for each talker. These results demonstrate that, in such a scenario, a steering device controlled by the eyes may greatly benefit the user.

When the EOG steering algorithm selects the correct target, the whole sentence is amplified by 12 dB. Therefore, if one of the target words is repeated correctly it is likely that the other target word will also be correct. This is not the case in the "no steering" condition. This may explain the significant difference between the two conditions in the case of sentence scoring. The EOG steering algorithm is helpful and increases performance. However, the error rate needs to be minimized before the algorithm can be considered in a realistic implementation.

Moreover, in the current system, classification errors may propagate over several sentences. Since only EOG was used for that steering condition, no additional information was provided for a better error estimation, e.g. via Kalman filtering. Information provided by e.g., head movements and an eye-gaze behavior model that takes head movements into account may allow a better estimation of the visual attention and, thus, may reduce the number of errors to a satisfying degree.

Indeed, in the tests of the present study, the participants had their head fixed, which is unnatural during a conversation. Further studies will take head movement into account. Head movement can be estimated using an accelerometer, a gyroscope and a magnetometer, which can easily be implemented inside a hearing aid.

In conclusion, this study has demonstrated the advantage of applying a steering interface to hearing impaired persons to increase speech intelligibility. However, the study also pointed out challenges of noise reduction from in-ear sensors and the need for additional studies allowing free movements of the head.

REFERENCES

Favre-Felix, A., Graversen, C., Dau, T., and Lunner, T. (**2017**). "Real-time estimation of eye gaze by in-ear electrodes," IEEE EMBC.

Fiedler, L., Obleser, J., Lunner, T., and Graversen, C. (**2016**). "Ear-EEG allows extraction of neural responses in challenging listening scenarios – a future technology for hearing aids?," IEEE EMBC.

Huigen, E., Peper, A., and Grimbergen, C.A. (**2002**). "Investigation into the origin of the noise of surface electrodes," Med. Biol. Eng. Comput., **40**, 332-338. doi: 10.1007/BF02344216

McShefferty, D., Whitmer, W.M., and Akeroyd, M.A. (**2016**). "The just-meaningful difference in speech-to-noise ratio", Trends Hear. doi: 10.1177/2331216515626570

Manabe, H., Fukumoto, M., and Yagi, T. (**2013**). "Conductive rubber electrodes for earphone-based eye gesture input interface," ISWC. doi: 10.1145/2493988.2494329

Nielsen, J.B., and Dau T. (**2013**). "A Danish open-set speech corpus for competing-speech studies," J. Acoust. Soc. Am., **135**, 407-420. doi: 10.1121/1.4835935

Pedersen, E.B., and Lunner, T. (**2014**) "Cognitive hearing aids? Insights and Possibilities", Mechanics of Hearing. doi: 10.1063/1.4939399

Puder, H. (**2009**), "Hearing aids: An overview of the state-of-the-art, challenges, and future trends of an interesting audio signal processing applications", IEEE ISPA. doi: 10.1109/ISPA.2009.5297793

A method to analyse and test the automatic selection of hearing aid programs

HENDRIK HUSSTEDT[1,*], SIMONE WOLLERMANN[1], AND JÜRGEN TCHORZ[2]

[1] *German Institute of Hearing Aids, Lübeck, Germany*

[2] *University of Applied Sciences Lübeck, Lübeck, Germany*

Digital hearing aids usually provide different hearing aid programs. This means different settings can be selected to adapt the signal processing to different hearing situations. Furthermore, advanced devices often include a classification algorithm that continuously analyses the acoustic environment and automatically selects a hearing aid program accordingly. However, there exists no method to analyse this adaptive feature. Therefore, we present a possibility to analyse and test which hearing aid program is active in a specific hearing situation. To proof the concept, hearing aids of two different manufacturers are analysed. These results give insights into the differences between classification strategies and classification quality among hearing aid manufacturers. Moreover, it shows that some signals, which humans can easily classify, are difficult to classify for hearing aids. Furthermore, the result of one device is compared with the classification entries of the data logging feature, which shows good agreement and verifies the new method. In addition, this comparison shows that the new method allows for a more comprehensive analysis so that using the data logging is no reasonable alternative.

INTRODUCTION

Digital hearing aids usually provide different hearing aid programs. This means the hearing aid can store different set of parameters, defining the signal processing, which is useful to adapt the signal processing to different hearing situations (Schaub, 2008; Husstedt, 2016). For some devices, the user manually selects the desired hearing aid program (see Fig. 1a). For more advanced devices, a classification algorithm continuously analyses the acoustic environment and selects a hearing aid program accordingly (see Fig. 1b). However, neither for the hearing aid user nor for the hearing aid professional is it clear what program the hearing aid actually selects in a specific hearing situation. Manufacturers pursuing different strategies so that different hearing aids may classify the same hearing situation differently. Furthermore, the hearing aid does not always select the proper hearing aid program, since the classification of hearing situations is still a difficult task (Tchorz *et al.*, 2016). A false classification causes an improper signal processing and thus may reduce speech intelligibility, comfort, and user satisfaction.

*Corresponding author: h.husstedt@dhi-online.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
ISBN: 978-87-990013-6-1.

In this work, we present a method that allows one to analyse and test the automatic selection of hearing aid programs. With this method, measurement results show which hearing aid program is active in a specific hearing situation. This gives insights into the classification strategy of hearing aid manufacturers and helps to evaluate the performance and quality of the applied classification algorithms. The rest of the paper is organized as follows. First, the measurement procedure is explained in detail. Then, it is demonstrated how the method is applied to two state-of-the-art hearing aids. Moreover, to verify the new method, the result of one device is compared with the entries of the data logging feature. Finally, the results are summarized and a conclusion is drawn.

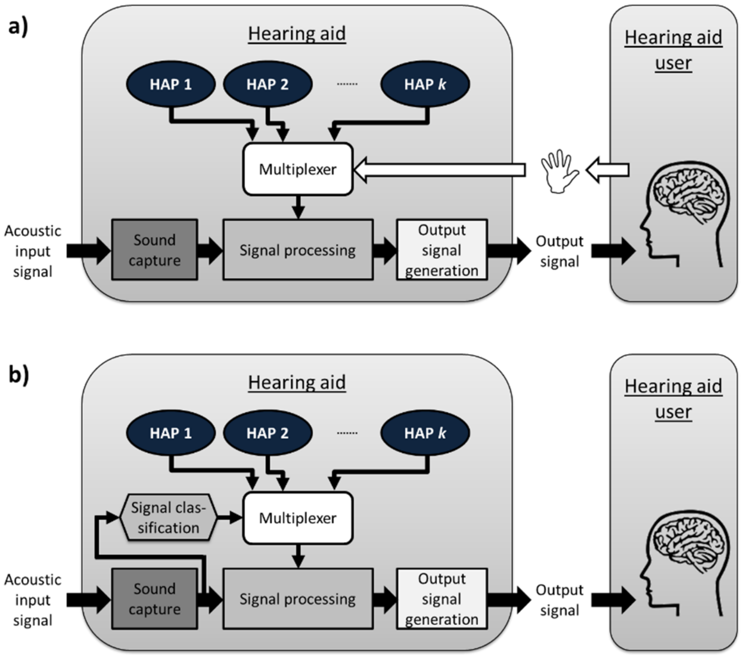


Fig. 1: Visualisation of the manual (a) and automatic (b) selection of hearing aid programs (HAP).

MEASUREMENT PROCEDURE

Preliminary part

In order to analyse the automatic selection of hearing aid programs, it is necessary to choose test signals representative for every hearing situation considered (e.g., music for the music program). Then, every of the n hearing aid programs is configured in an arbitrary way as reference, e.g., with reference test gain (RTG). However, it is not important to have equal configurations for different hearing aid programs. In a next step, every of the test signals is successively presented to the hearing aid and each time the output signal is measured and saved as reference (see upper left part of Fig. 2).

Measurement part

During the measurement part, n measurement cycles are performed where a marker is set just to one hearing aid program at a time. In this context, setting a marker means changing the signal processing of the corresponding hearing aid program so that a change of the output signal is noticeable. For instance, setting a marker could be realized by simply changing the gain for the corresponding hearing aid program. During each measurement cycle, all test signals are presented to the hearing aid, and each time the output signal is measured (see Fig. 2). These signals are then compared with the reference signals saved during the preliminary part. The comparison clearly shows which signal is affected by the marker. In the ideal case, a difference only occurs for that signal where a maker is set to the corresponding hearing aid program. If for example a marker is set to the speech program, a difference should only occur for speech signals. However, if we consider that the output signal is affected when speech is present and a marker is set to the music program, one can conclude that the speech signal is classified as music.

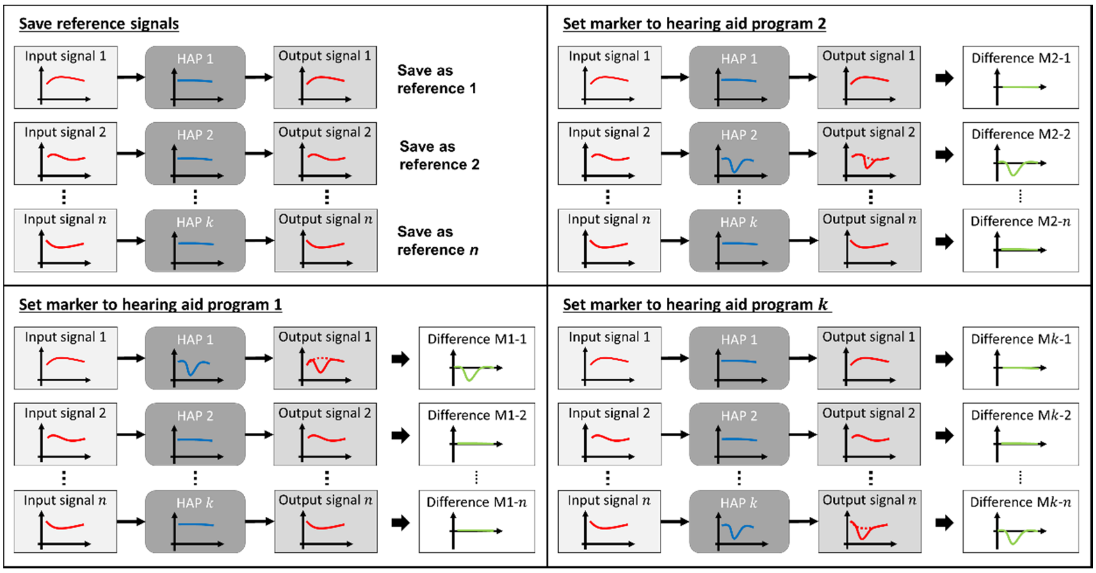


Fig. 2: Visualisation of the preliminary measurement and saving of the reference signals (upper left part). The other parts of the figure visualise the measurement cycles where the maker is set to hearing aid program 1, 2, and k. In this visualisation, only one input signal is presented for each hearing aid program (HAP) so that the number of test signals n is equal to the number of HAPs k. Moreover, the input signal and the corresponding HAP have the same index – e.g., if input signal 1 represents a speech signal, HAP 1 is the speech program.

MEASURMENT

Test signals

A comparison of several hearing aid manufacturers shows that most of the upper class models can at least distinguish between the following listening situations: speech, speech in noise, music, and noise. Thus, these listening situations are considered in the following (see Table 1). Furthermore, it is important to mention that many devices can also classify the situation "quiet". However, without any input signal, the method is not applicable so that this situation is not considered. Nevertheless, this is no significant limitation, because in quiet, there is no external input signal that can be processed so that it is of minor interest for user what signal processing is active.

Index	Listening situation	Test signal
1	Speech	ISTS 65 dB
2	Speech + Noise	ISTS 65 dB + IFnoise 60 dB
3	Speech + Noise	ISTS 65 dB + IFnoise 50 dB
4	Speech	Audio book 65 dB
5	Speech + Noise	Audio book 65 dB + IFnoise 60 dB
6	Speech + Noise	Audio book 65 dB + IFnoise 50 dB
7	Music	Piano 65 dB
8	Music	Violine 65 dB
9	Noise	IFnoise 65 dB
10	Noise	Gravel sieving 65 dB

Table 1: List of test signals used for the measurements. The index indicates in which order the signals are presented to the hearing aid, and the loudness is given as sound pressure level (SPL) in decibel.

For each of the four listening situations at least two test signals are chosen (see Table 1). For speech, the International Speech Test Signal (ISTS; Holube *et al.*, 2010) and the German audio book "Abendlied" are used. For speech in noise, both speech signals are mixed with the International Female Noise (IFnoise) with signal-to-noise ratios (SNR) of +5 dB and +15 dB. The IFnoise was generated by using multiple overlapping of the speech material of the ISTS so that it has the same long-term average spectrum as the ISTS (EHIMA, 2016). As test signals for music, a piano and a violin track without any voices are used. Finally, as noise, the IFnoise and an industry noise caused by gravel sieving are considered.

Study design

Two upper class hearing aids of two different manufacturers are analysed with the new method. During the measurement, the ten test signals of Table 1 are presented in a free field, and the output signal is recorded with an ear simulator according to IEC 60318-4. To program the devices, in the fitting software, the "first fit" function is used for a hearing loss of type N3 according to IEC 60118-15. As marker, a reduction of the gain of approx. 20 dB between 1 kHz and 3 kHz is programed. For the comparison of each output signal with the reference, several measures are possible, e.g., simply comparing the overall sound pressure level (SPL). However, it turned out that a more robust and more sensitive method is using the 1/3 octave levels of both signals. Hence, the differences for all 1/3 octave levels between 500 Hz and 8 kHz are computed and then, the root mean square (RMS) of these differences is calculated. This RMS of all 1/3 octave level differences is shortly denoted as Δ with $[\Delta]$ = dB, and used for all results presented in the following. Figures 3 and 4 show the results of hearing aid I and II, respectively. In these figures, the darkness of the pixels represents the values of Δ. Repeatability measurements show that the impact of measurement tolerances on Δ is below 0.4 dB. Therefore, the darkness map begins at 0.5 dB so that all values below 0.5 dB result in white pixels. Moreover, values of Δ between 0.5 dB and 2 dB are coloured by a grey scale to indicate small differences. Values of Δ above 2 dB are represented by a black pixel, since a clear effect of the marker can be recognized.

The focus of this study is on the final result of the classification algorithms, rather than on the transient behaviour. Therefore, both hearing aids have 55 s time to adopt to a test signal, and the output signal between 55 s to 60 s is evaluated only. Furthermore, Figs. 3 and 4 also include the expected classification, which is indicated by crosses. Nevertheless, since no standardized definitions exist for hearing situations, it is not clear what signal-to-noise ratio (SNR) separates speech, speech in noise, and noise. Consequently, the crosses especially for speech, and speech in noise should not be interpreted as correct or ideal classification.

Results

If we have a look at the results of device I and II, we can notice that the same input signal triggers more than one hearing aid program. An explanation for this effect can be that the hearing aid is switching back and forth between two hearing aid programs, or that the signal processing of two hearing aid programs is superposed. A reason for superposition could be that hearing aids do not hardly switch between different hearing situations, but allow for a smooth transition. An analysis of the transient behaviour can give deeper insights, but is not the focus of this work.

If we look at the results for test signals 1 to 6 with speech and speech in noise, we see that both devices mainly detect speech or speech in noise. Device I more often detects speech and not speech in noise, e.g., if we look at the results for test signal "Audio book 65 dB + IFnoise 50 dB". However, as explained in the foregoing, there is no clear definition of what SNR separates speech and speech in noise, so that both results can be seen as appropriate classification. However, device I classifies "ISTS 65 dB + IFnoise 60 dB", and device II classifies "ISTS 65 dB" partly as noise. If we assume this effect to be stronger, it might be a problem for the hearing aid user, because speech is processed as noise so that may be the gain for speech is reduced. On the other hand, we see that device I partly detects the IFnoise, which has the same long-term average spectrum as the ISTS, as speech in noise. This could lead to discomfort, because the noise is processes as speech so that the gain for the noise could be elevated. As another peculiarity, device II classifies the test signal "Piano 65 dB" as speech. This is astonishing, since the test signal does not include any voices and no human would classify this track as speech.

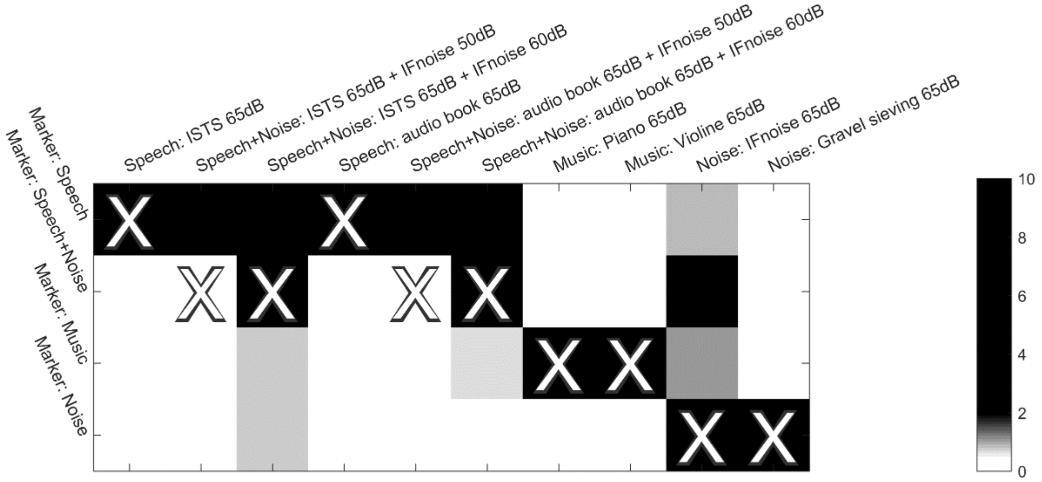


Fig. 3: Measurement results for device I evaluated in the time between 55 s to 60 s. The crosses indicate the expected classification.

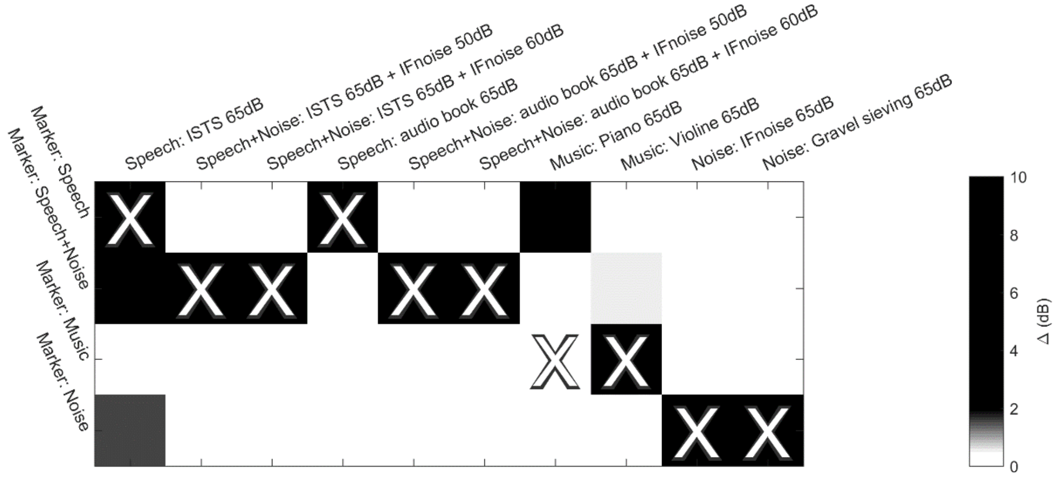


Fig. 4: Measurement results for device II evaluated in the time between 55 s to 60 s. The crosses indicate the expected classification.

COMPARISON WITH RESULTS OF THE DATA LOGGING FEATURE

Most modern hearing aids provide a feature commonly denoted as data logging. This feature shows the hearing aid professional information about the use of the hearing aid so that the fitting can better be adapted to the individual needs of the patient. As one type of information, many hearing aids log the hearing situation experienced by the user. In the fitting software this results is often depicted as relative time data in percentage, e.g., 30 % of the time the user experienced noise, etc.

Exactly these data are used to verify the new method. To this end, one test signal is presented to device II for 1 h, and afterwards, each time the result of the data logging is read out. A long presentation time is necessary, since the data logging does not store signals presented for a few minutes only. Figure 5 depicts the results of the data logging feature in a format similar to the results of Fig.3 and Fig. 4. The only difference is that the colour map represents the relative time in percentage.

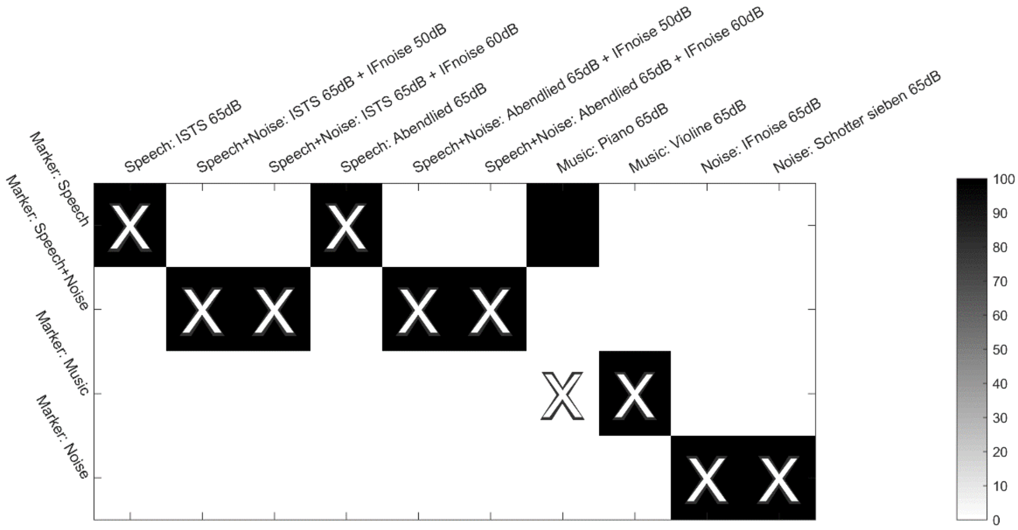


Fig. 5: Results of the data logging feature for device II (see also Fig. 4). One signal is presented for 1 h, and afterwards, each time the result of the data logging is read out.

If we compare the results of Fig. 4 with the results of Fig. 5, we see almost no difference. Only the result for the test signal "ISTS 65 dB" does not completely agree. Both figures show a classification as speech whereas in Fig. 4 the signal is additionally classified as speech in noise, and noise. There are multiple possible reasons for this difference, e.g., the classification has not reached the steady state after 55 s as in Fig. 4 so that the result is different in Fig. 5 where 1h is considered. Another reason could be that the hearing aid switches between multiple hearing situation, but speech in noise and noise is not detected often enough to be stored in the data logging feature.

CONCLUSION

The method presented allows one to analyse what hearing aid program is automatically selected by the hearing aid in a specific hearing situation. This gives insights into the classification strategy and quality among different hearing aid manufacturers – e.g., the results show that the SNR at which speech is separated from speech in noise varies for different manufacturers. Furthermore, there are some signals such as the IFnoise or the piano track, which are easy to classify for humans, but can be difficult to classify for hearing aids.

In addition, a comparison of the results of one hearing aid with results of the data logging feature shows good agreement and verifies the new method. Nevertheless, using the data logging is no reasonable alternative, because entries in the data logging are stored only after a long time (usually > 30 min). Thus, measurements take multiple hours. Moreover, the data logging only shows what hearing situation has been detected, but not if the corresponding signal processing is really active. Finally, another advantage of the new method over the data logging is that also the transient behaviour of the automatic selection of hearing aid programs can be analysed. This is very useful, since not only the reliability but also the time until a new situation has been classified is very important for the hearing aid user. Therefore, this will be subject of future work.

REFERENCES

European Hearing Instrument Manufacturers Association (EHIMA) (**2016**). "Description and Terms of Use of the IFFM and IFnoise signals," (Available at: http://www.ehima.com/wp-content/uploads/2016/06/IFFM_and_IFnoise.zip)

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. doi: 10.3109/14992027.2010.506889

Husstedt, H., (**2016**). "Definition und Nachweis von Hörprogrammen bei Hörsystemen," 61st International Congress of Hearing Aid Acousticians, Hannover, Germany.

Schaub. A., **(2008)**. *Digital Hearing Aids.* Edited by Birgitta Brandenburg (Thieme Medical Publishers. Inc., New York), pp. 107-123.

Tchorz, J., Wollermann, S., and Husstedt, H. **(2017)**. "Classification of Environmental Sounds for Future Hearing Aid Applications," Proceedings of the 28th Conference on Electronic Speech Signal Processing (ESSV 2017), Saarbrücken, Germany, pp. 294-299.

Adapting bilateral directional processing to individual and situational influences

TOBIAS NEHER[1,2,*] KIRSTEN C. WAGENER[3], AND MATTHIAS LATZEL[4]

[1] *Medizinische Physik and Cluster of Excellence "Hearing4all", Oldenburg University, Oldenburg, Germany*

[2] *Institute of Clinical Research, University of Southern Denmark, Odense, Denmark*

[3] *Hörzentrum Oldenburg GmbH, Oldenburg, Germany*

[4] *Phonak AG, Stäfa, Switzerland*

This study examined differences in benefit from bilateral directional processing. Groups of listeners with symmetric or asymmetric audiograms <2 kHz, a large spread in the binaural contribution to speech-in-noise reception (i.e., the binaural intelligibility level difference, BILD), and no difference in age or overall degree of hearing loss took part. Aided speech reception was measured using virtual acoustics together with a simulation of a linked pair of closed-fit behind-the-ear hearing aids. Five processing schemes and three acoustic scenarios were used. The processing schemes differed in the trade-off between signal-to-noise ratio (SNR) improvement and binaural cue preservation. The acoustic scenarios consisted of a frontal target talker and two lateral speech maskers or spatially diffuse noise. For both groups, a significant interaction between BILD, processing scheme and acoustic scenario was found. This interaction implied that, for lateral speech maskers, users with BILDs >2 dB profited more from low-frequency binaural cues than from greater SNR improvement, while for smaller BILDs the opposite was true. Audiometric asymmetry reduced the BILD influence. In spatially diffuse noise, the maximal SNR improvement was beneficial. Moreover, binaural tone-in-noise detection performance (N_0S_π threshold) at 500 Hz predicted the benefit from low-frequency binaural cues effectively. These results provide a basis for adapting bilateral directional processing to the user and the scenario.

INTRODUCTION

Although hearing-impaired listeners can differ substantially in their speech-in-noise abilities, the responsible factors are yet to be fully understood. As a consequence, ways of addressing these differences with hearing devices remain scarce. The current study aimed to shed more light on the factors driving benefit from binaural information for speech-in-noise reception, and to identify ways of tailoring directional hearing aid (HA) processing to individual hearing abilities. In a previous study, we

[*] Corresponding author: tneher@health.sdu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR),
Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau,
J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
ISBN: 978-87-990013-6-1.

screened almost 80 elderly hearing-impaired listeners with a large spread in the absolute across-ear difference in low-frequency (<2 kHz) pure-tone average hearing thresholds (ΔPTALF) in terms of the binaural contribution to speech-in-noise reception (Neher, 2017). To that end, we used the so-called binaural intelligibility level difference (BILD). The BILD is a measure of the improvement – or lack thereof – in speech-in-noise reception due to binaural processing (e.g., Kollmeier, 1996). Using virtual acoustics, we simulated a frontal target talker in the presence of a lateral speech-shaped noise masker and amplified the resultant stimuli according to the 'National Acoustic Laboratories–Revised Profound' (NAL-RP) fitting rule (Dillon, 2012). By taking the difference between the binaural and the monaural (i.e., better-ear) speech reception thresholds (SRTs), we then quantified the BILD, reflecting the change in signal-to-noise ratio (SNR) due to binaural interaction. Typically, normal-hearing listeners obtain BILDs of ~4 dB (e.g., Santurette and Dau, 2012).

In the current study, we tested a carefully selected subset of these listeners further. Using a computer simulation of a linked pair of behind-the-ear (BTE) HAs, we performed aided speech reception measurements with five directional processing conditions in three acoustic scenarios. The processing conditions differed in the trade-off between SNR improvement and binaural cue preservation. The acoustic scenarios differed primarily in terms of the noise characteristics (lateral speech maskers vs. spatially diffuse noise). Our aims were (1) to relate ΔPTALF and BILD to performance with the different directional processing schemes, and (2) to investigate if a simple binaural tone-in-noise detection measure can be used to predict the benefit from binaural cue preservation.

Below, we provide a summary of our methods and results. More detailed information can be found in (Neher *et al.*, 2017).

METHODS

Participants

Forty listeners aged 62-80 yr (mean: 73 yr) participated in the current study. Their pure-tone average hearing loss calculated across 0.5, 1, 2 and 4 kHz and left and right ears (PTA4) ranged from 35 to 69 dB HL (mean: 52 dB HL). The participants had either 'symmetric' ΔPTALF (N = 20; mean: 3 dB; range: 0-6 dB) or 'asymmetric' ΔPTALF (N = 20; mean: 23 dB; range: 15-39 dB). Furthermore, the two groups exhibited substantial and comparable spread in the BILD (ranges: 0.2 to 5.2 vs. –0.4 to 4.7 dB; means: 2.6 vs. 2.5 dB). To control for potentially confounding effects, we made sure that the two groups were matched in terms of age (means: 74 vs. 72 yr) and PTA4 (means: 52 vs. 53 dB HL).

In the study of Neher (2017), the 40 participants were characterised further using some psychoacoustic and cognitive tests. These included binaural tone-in-noise detection measurements (i.e., N_0S_0 and N_0S_π thresholds) at 0.5 and 1 kHz. Furthermore, they included a reading span test for the assessment of working memory capacity (Carroll *et al.*, 2015) and a 'distractibility' test for the assessment of selective attention

(Zimmermann and Fimm, 2012). Statistical analyses showed that the BILD was strongly correlated with the N_0S_π detection threshold at 500 Hz (Pearson's *r* correlation coefficient = –0.72, $p < 0.00001$) and that the two groups only differed in terms of ΔPTALF ($p < 0.00001$) and reading span ($p = 0.036$).

HA conditions

For simulating the different HA conditions, we used impulse response measurements made with a head-and-torso simulator equipped with two behind-the-ear (HA) dummies (Kayser *et al.*, 2009) together with the Master Hearing Aid research platform of Grimm *et al.* (2006). The directional processing conditions were all based on fixed, forward-facing microphone arrays, i.e., they were non-adaptive and steered towards 0° azimuth. The first (*pinna*) condition simulated two unilateral BTE devices with a modest degree of directivity above ~1 kHz. This resulted in a dichotic stimulus with binaural cues available across the entire frequency range. The second (*beamfull*) condition achieved maximal SNR improvement (~4.5 dB speech-weighted re. pinna) at the cost of binaural cue preservation. It resulted in a diotic stimulus across the entire frequency range. The third (*beam>0.8k*) and fourth (*beam<2k*) conditions were hybrid versions of the pinna and beamfull conditions. The beam>0.8k condition corresponded to the pinna condition below 0.8 kHz and to the beamfull condition above 0.8 kHz. The beam<2k condition corresponded to the pinna condition above 2 kHz and to the beamfull condition below 2 kHz. Thus, the beam>0.8k condition resulted in a dichotic stimulus in the low-frequency range and in a diotic stimulus in the mid- and high-frequency range. In contrast, the beam<2k condition resulted in a diotic stimulus in the low- and mid-frequency range and in a dichotic stimulus in the high-frequency range. Compared to the beamfull condition, the beam>0.8k and beam<2k conditions achieved less SNR improvement (~2 dB and ~3 dB speech-weighted re. pinna). The fifth (*beambetter*) condition was identical to the beamfull condition except that only the ear with the better speech-in-noise reception was stimulated (corresponding to bilateral contralateral routing of signals; BICROS). It therefore resulted in a monaural stimulus. Figure 1 shows polar patterns of the different directional processing conditions. Following the directional processing, we applied NAL-RP amplification to ensure adequate audibility.

Acoustic scenarios

We evaluated the different HA conditions in three acoustic scenarios. The scenarios comprised a frontal target talker uttering sentences from the Oldenburg sentence test (OLSA; Wagener *et al.*, 1999). As maskers, we used three types of signals: (1) a recording of another male speaker uttering OLSA sentences, (2) a modified version of the International Speech Test Signal (ISTS; Holube *et al.*, 2010), and (3) a recording made in a large cafeteria (T_{60} = 1.25 sec) during a busy lunch hour (Kayser *et al.*, 2009). The OLSA masker consisted of 10 sentences that were concatenated without any pauses. The fundamental frequency of the speaker uttering these sentences was very similar to that of the target speaker (~110 Hz). The ISTS masker used here was identical to the original ISTS except that its fundamental frequency was

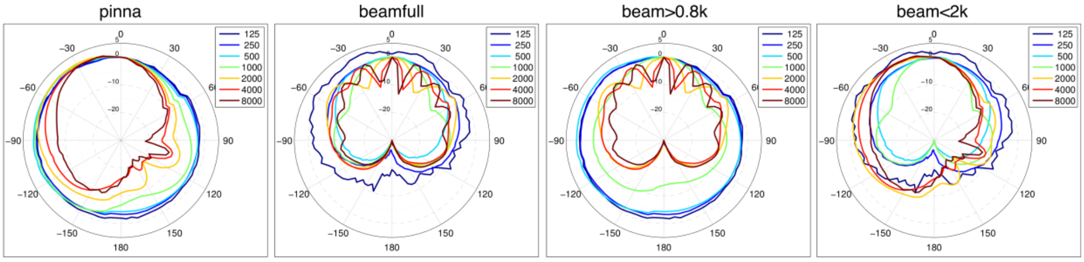


Fig. 1 (colour version online): Polar patterns of the pinna (left ear), beamfull (both ears), beam>0.8k (left ear), and beam<2k (left ear) settings calculated in octave bands with centre frequencies of 125, 250, 500, 1000, 2000, 4000 and 8000 Hz (see legend). The azimuth is in degrees and the gain in decibels.

lowered to match that of the target speaker. Thus, the main difference between the OLSA and ISTS maskers was that the latter was largely unintelligible. The OLSA and ISTS maskers were presented from ±60° azimuth. Below, we refer to the three stimulus conditions as the *olsa60*, *ists60* and *cafnois* scenarios.

For each combination of acoustic scenario and HA condition, we measured two SRTs (corresponding to 50%-correct speech intelligibility) per participant. A correlation analysis revealed that the test-retest reliability of these measurements was very good (all $r > 0.73$, all $p < 0.00001$).

RESULTS

Due to the bimodal distribution of the ΔPTALF data, we performed separate analyses of variance on the data from the symmetric and asymmetric groups. In each case, we included acoustic scenario and HA condition as within-subject factors and the BILD as a covariate. Furthermore, we initially also included age, PTA4, reading span and distractibility to control for potentially confounding effects due to these characteristics. Because age and distractibility did not contribute significantly to the models, we excluded them from all additional analyses.

For both groups, we found significant main effects of the BILD ($p < 0.001$) and acoustic scenario ($p < 0.00001$), a significant two-way interaction between HA condition and acoustic scenario ($p < 0.00001$) and a significant three-way interaction between the BILD, HA condition and acoustic scenario ($p < 0.016$). Follow-up analyses revealed (1) a strong negative association between the BILD and the SRT ($r < –0.76$), (2) better performance in the ists60 scenario than in the other two scenarios, (3) a very similar influence of the different HA conditions on performance in the olsa60 and ists60 scenarios but not in the cafnois scenario, (4) a differential influence of the BILD on performance with the different HA conditions in the osla60 and ists60 scenarios but not in the cafnois scenario, and (5) no performance benefits due to beambetter processing.

Symmetric group

Figure 2 shows scatter plots of the SRT and BILD data from the symmetric group for each acoustic scenario and HA condition together with regression lines. These plots indicate that, in situations with intelligible (olsa60) or unintelligible (ists60) speech maskers, participants with BILDs >2 dB profited from the preservation of low-frequency binaural cues (pinna and beam>0.8k). In contrast, for smaller BILDs and for spatially diffuse conditions (cafnois) in general, the maximal SNR improvement (beamfull) was beneficial. The plots also illustrate the negative association between the BILD and the SRT mentioned above.

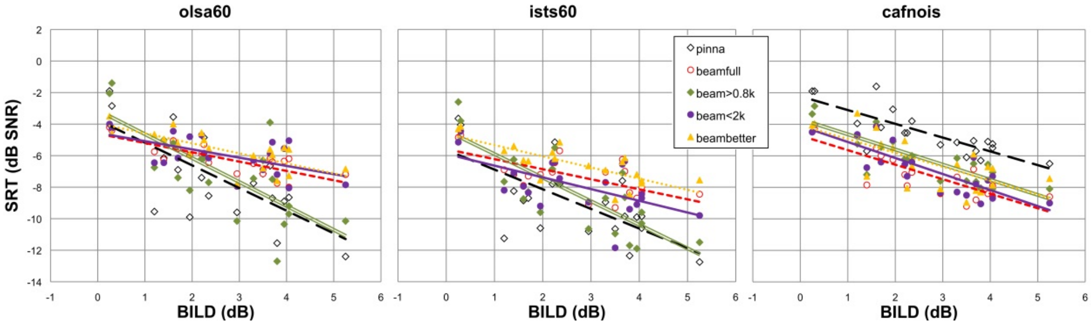


Fig. 2 (colour version online): Scatter plots of the BILD and SRT data for the *symmetric* group. Left: olsa60 scenario; Middle: ists60 scenario; Right: cafnois scenario. Least-squares regression lines corresponding to the pinna (long-dashed black line, unfilled black diamonds), beamfull (short-dashed red line, unfilled red circles), beam>0.8k (double green line, filled green diamonds), beam<2k (solid purple line, filled purple circles), and beambetter (dotted yellow line, filled yellow triangles) settings are also shown.

Asymmetric group

Figure 3 shows scatter plots of the SRT and BILD data from the asymmetric group for each acoustic scenario and HA condition together with least-squares regression lines. In general, these plots resemble those for the symmetric group (Fig. 2). For the asymmetric group, however, the benefit from low-frequency binaural cues (pinna and beam<0.8k) relative to more directionality (beamfull and beam<2k) occurred for participants with larger BILDs (>2.5 dB), leading to a reduction in the maximal benefit from binaural cue preservation (for a BILD of ~5 dB, ~2 dB for the asymmetric group vs. ~3 dB for the symmetric group).

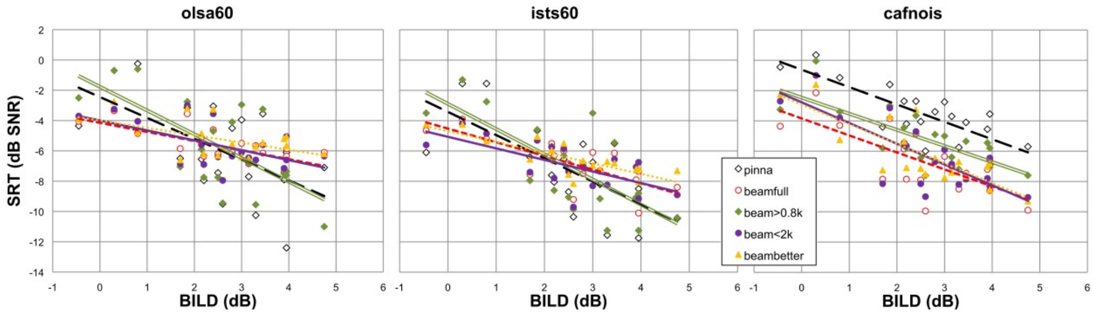


Fig. 3 (colour version online): Scatter plots of the BILD and SRT data for the *asymmetric* group. Left: olsa60 scenario; Middle: ists60 scenario; Right: cafnois scenario. Least-squares regression lines corresponding to the pinna (long-dashed black line, unfilled black diamonds), beamfull (short-dashed red line, unfilled red circles), beam>0.8k (double green line, filled green diamonds), beam<2k (solid purple line, filled purple circles), and beambetter (dotted yellow line, filled yellow triangles) settings are also shown.

Beambetter setting and abnormal BILDs

To test if HA users with clearly abnormal BILDs may benefit from the (rather extreme) beambetter setting, we analysed the data of a subset of participants with BILDs <1 dB (two 'symmetric' and three 'asymmetric' participants; mean BILD: 0.2 dB; range: –0.4 to 0.8 dB). Because some of the resultant datasets were not normally distributed, we used the Wilcoxon signed-rank test for this. Furthermore, we restricted our analysis to a comparison of the beambetter and beamfull settings. For none of the acoustic scenarios was there a significant difference between the two, nor was there one averaged across acoustic scenarios (all $p > 0.17$).

BILD vs. N_0S_π

As pointed out above, our participants had previously completed N_0S_π detection threshold measurements at 500 Hz, which were strongly correlated with the BILD data. To test if the BILD and N_0S_π measures can be used interchangeably to predict the effects of binaural hearing abilities on speech reception with bilateral directional processing, we repeated the analysis of the data from the symmetric group with the N_0S_π measure instead of the BILD included. There were significant effects of N_0S_π ($p < 0.001$), HA condition ($p < 0.021$), acoustic scenario ($p < 0.00001$), N_0S_π × HA condition ($p < 0.017$), HA condition × acoustic scenario ($p < 0.00001$) and N_0S_π × HA condition × acoustic scenario ($p < 0.009$). Thus, the results were very similar to those obtained with the BILD.

SUMMARY

In the current study, we investigated the influence of binaural hearing abilities, audiometric asymmetry <2 kHz and the acoustic scenario on aided speech reception with five directional processing schemes. The schemes, which were realized using virtual acoustics together with a computer simulation of a pair of completely occluding BTE devices, traded SNR improvement against binaural cue preservation below 800 Hz or above 2 kHz. In addition, they included a BICROS-like condition that combined maximal SNR improvement with better-ear stimulation. The participants were two groups of elderly individuals with symmetric or asymmetric hearing thresholds <2 kHz, large variation in the BILD, and no difference in age or PTA4. Our analyses revealed an influence of the BILD (or, alternatively, N_0S_π detection performance at 500 Hz) for intelligible (olsa60) and unintelligible (ists60) directional speech maskers from ±60° azimuth. Listeners with BILDs greater than 2-3 dB benefited more from low-frequency binaural cues than from greater directionality, whereas for smaller BILDs the opposite was true. Audiometric asymmetry reduced the influence of binaural hearing. Under spatially diffuse conditions (cafnois), performance was driven by SNR improvement, with the (maximally directional but diotic) beamfull setting giving the best results, irrespective of BILD and ΔPTALF status. The BICROS-like scheme did not result in any performance benefits, likely because only one of the participants tested here had a negative BILD (and thus a disbenefit from binaural interaction).

Together, these findings provide a valuable basis for adapting bilateral directional processing to the user and the acoustic scenario. Ongoing research is concerned with investigating their generalizability to clinical HA fittings.

ACKNOWLEDGEMENTS

This research was funded by the DFG Cluster of Excellence EXC 1077/1 “Hearing4all” and by Sonova AG, Switzerland.

REFERENCES

Carroll, R., Meis, M., Schulte, M., *et al.* (**2015**). “Development of a German reading span test with dual task design for application in cognitive hearing research,” Int. J. Audiol., **54**, 136-141. doi: 10.3109/14992027.2014.952458

Dillon, H. (**2012**). *Hearing Aids*, 2nd ed., Boomerang Press, Sydney, Australia.

Grimm, G., Herzke, T., Berg, D., and Hohmann, V. (**2006**). “The master hearing aid: A PC-based platform for algorithm development and evaluation,” Acta Acust. United Ac., **92**, 618-628.

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. doi: 10.3109/14992027.2010.506889.

Kayser, H., Ewert, S.D., Anemüller, J., *et al.* (**2009**). “Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses,” EURASIP J. Adv. Signal Process., **298605**, 1-10.

Kollmeier, B. (**1996**). “Computer-controlled speech audiometric techniques for the assessment of hearing loss and the evaluation of hearing aids,” In: Kollmeier, B. (Ed.), *Psychoacoustics, Speech and Hearing Aids*. World Scientific, Singapore, pp. 57-68.

Neher, T. (**2017**). “Characterizing the binaural contribution to speech-in-noise reception in elderly hearing-impaired listeners,” J. Acoust. Soc. Am., **141**, EL159-EL163, doi: 10.1121/1.4976327.

Neher, T., Wagener, K.C., and Latzel, M. (**2017**). “Speech reception with different bilateral directional processing schemes: Influence of binaural hearing, audiometric asymmetry, and acoustic scenario,” Hear. Res., **353**, 36-48. doi: 10.1016/j.heares.2017.07.014.

Santurette, S., and Dau, T. (**2012**). “Relating binaural pitch perception to the individual listener’s auditory profile,” J. Acoust. Soc. Am., **131**, 2968-2986, doi: 10.1121/1.3689554.

Wagener, K., Brand, T., and Kollmeier, B. (**1999**). “Development and evaluation of a sentence test for the German language. I-III: Design, optimization and evaluation of the Oldenburg sentence test,” Z. Audiol. (Audiol. Acoustics), **38**, 4-15, 44-56, 86-95.

Zimmermann, P., and Fimm, B. (**2012**). “Testbatterie zur Aufmerksamkeitsprüfung – Version Mobilität (Test battery for the assessment of attentional skills – Mobility version),” Psytest, Herzogenrath, Germany.

A bilateral hearing-aid algorithm that provides directional benefit while preserving situational awareness

TOBIAS PIECHOWIAK[1,*], ROB DE VRIES[2], CHANG MA[3], AND ANDREW DITTBERNER[3]

[1] *GN ReSound A/S, Ballerup, Denmark*

[2] *GN ReSound, Dept. R&D, Eindhoven, The Netherlands*

[3] *GN ReSound North America, Glenview, IL, USA*

A directional filter or beamformer is a classical approach to ensure speech intelligibility by suppressing distracting sounds from certain directions. However, they have some challenges on their own: (a) white-noise gain, (b) diminished benefit in reverberation, and (c) 'off-axis' audibility problems. In this study a new algorithm which is part of ReSound's Linx3D's Bilateral Directionality III is introduced. It is designed to provide good *situational awareness* (SA) while mainting *directional benefit* (DB). SA is maximized by combining the sensitivity of both left and right hearing aids to create a true binaural omnidirectional sensitivity pattern. DB is ensured by promoting the better-ear effect and thus allowing for better separation of sounds.

INTRODUCTION

The primary purpose of a hearing aid is to restore audibility of a target signal. The classical approach is to measure a pure-tone audiogram and apply frequency dependent gains on the microphone input signals. However, it often does not alleviate the problems hearing impaired have in noisy environments (Kochin, 2010). Directional filters (or beamforming filters) are one attempt to further address this problem by suppressing distracting sounds from certain (a-priori) known directions. They have some challenges on their own as, e.g.:

1. High white noise gain, i.e., creation of noise because of partially equalizing for inherent low-frequency roll-off;

2. Directional benefit dimishes quickly when reverberation is added (Ricketts, 2003); and

3. Directional filters impede 'off-axis' listening. Sounds from the side or rear are attenuated and might become inaudible creating problems when new sounds are introduced.

In this study we introduce a new (bilateral) algorithm, within ReSound's Linx3D Bilateral Directionality III, that is primarily targeting this last challenge, to promote

[*] Corresponding author: tpiechowiak@gnresound.com

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

better ‘off-axis’ listening while simultaneously not giving up on the directional benefit a beamformer provides for, e.g., increasing speech intelligbility. With other words, the new algorithm tries to combine the two complementary concepts *situational awareness* (SA) and *directional benefit* (DB). Basically, the new algorithm in itself constitutes a beamformer whose target is to provide a combined quasi-omnidirectional response across ears while simultaneously maintaining a large head-shadow (better-ear effect) that helps with source separation and speech intelligibility (Bronkhorst, 1988).

THE ALGORITHM

The algorithm processing scheme is depicted in Fig. 1.

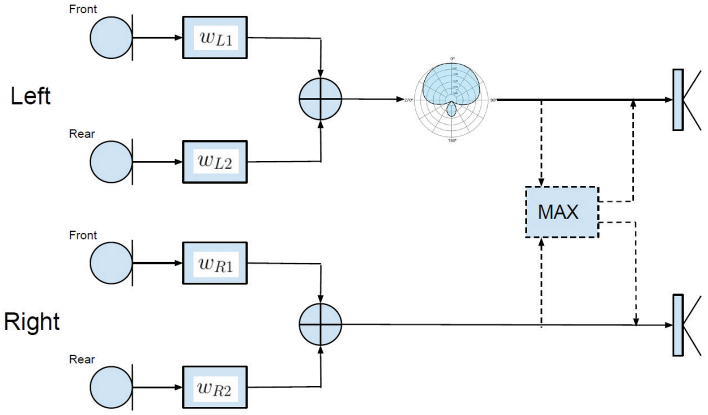


Fig. 1: Flowchart of the algorithm. The weights w_{Ri} are optimized in the way to give a quasi-omnidirectional response across ears. The weights on the left side w_{Li} are fixed to generate a hypercardioid at the left ear. Weights are optimized offline so the algorithm shown is not adaptive.

The $w_i(n)$ are fixed finite-impulse-response (FIR) filters associated with the i-th mircophone at the left ($w_{Li}(n)$) and right side ($w_{Ri}(n)$). The output of the right, respectively left side can be defined as

$$L(n,\Theta) = \sum_{i=1}^{2} w_{Li}(n) * h_{Li}(n,\Theta) * s(n) \qquad \text{(Eq. 1)}$$

$$R(n,\Theta) = \sum_{i=1}^{2} w_{Ri}(n) * h_{Ri}(n,\Theta) * s(n) \qquad \text{(Eq. 2)}$$

where $s(n)$ is the acoustic source and $h_i(n,\Theta)$ the hearing aid related impulse responses associated with the i-th microphone and azimuth angle Θ. Note, that the $w_{Li}(n)$ are not part of the optimization scheme. They are used to generate a hypercardiod on the left side as depcited in the figure and are fixed during optimization. The hypercardioid is part of providing directional benefit.

The algorithm itself is achieved by optimizing the weights $w_{Ri}(n)$ in a least-square sense:

$$\underset{w_{Ri}}{\operatorname{argmin}}\{Var[max(L(n,\Theta_k),R(n,\Theta_k))]\} \quad k \in \{1...K\} \qquad \text{(Eq. 3)}$$

where K is the number of sampled azimuth angles. In this study a resolution of 10° was used for that purpose. Minimizing the variance (*Var*) results in a quasi-omnidirectional response across ears aiming to provide high SA. Note, that the optimization is performed off-line, so no information is exchanged between devices during use of the algorithm. The resulting intensity plots for the left, respectively right side can be seen in Fig. 2.

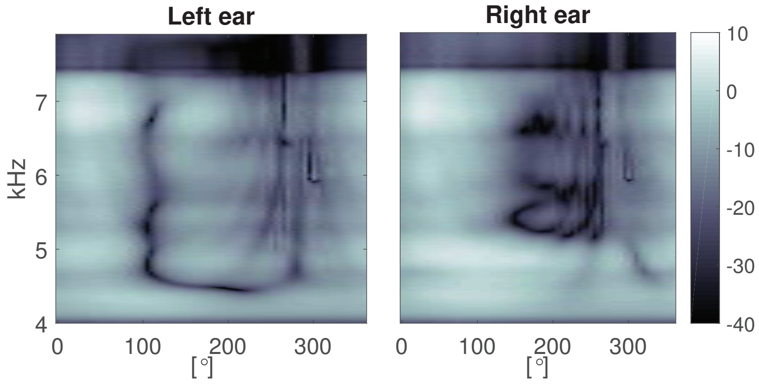


Fig. 2: Intensity plots for the optimized beampattern of the left and right side across azimuth and frequencies. Bright color illustrates high intensity. Responses are normalized to the maximum of the front microphone.

THE METRIC

The algorithm poses a fundamental question: What would be an appropriate metric for characaterizing the two-fold purpose of the algorithm? For example, the *Directivity Index* (DI) has usually been used to characterize the strength of the directional benefit of a beamformer that can be calculated on a manikin (Dittberner, 2007). DI in general correlates quite well with directional benefit perceived by subjects (Dittberner, 2007). However, this metric would not be able to account for SA since SA and DB are complementary concepts. Thus, the need for a reliable metric arises that considers both SA and DB simultaneously. For this purpose, a new metric is introduced. It consists of two indices, (a) the Situational Awareness Index (SAI) and (b) the Better-Ear Index (BEI).

Formally, SAI and BEI are defined as

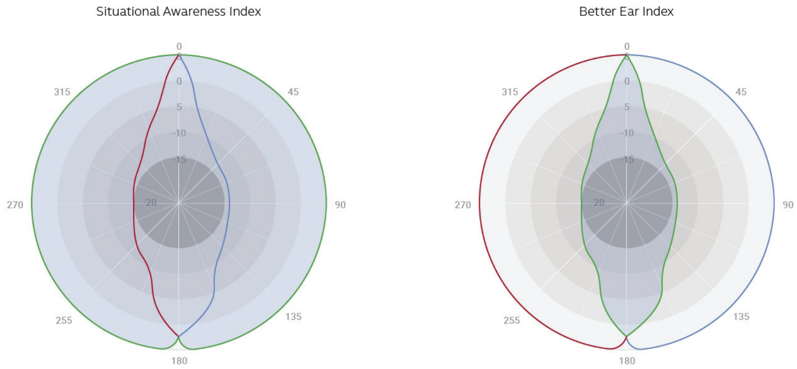


Fig. 3: Calculation of the metric. Left panel: Maximum power of across ears illustrates Situational Awareness Index (SAI). Right panel: Minimum power across ears (green line) illustrates Better-Ear Index. Grey shaded areas illustrate meaning of the indices.

$$SAI = 10 \cdot log_{10}\left(\frac{Std(max(P_L(\Theta_k), P_R(\Theta_k)))}{\overline{max(P_L(\Theta_k), P_R(\Theta_k)}}\right) \qquad \text{(Eq. 4)}$$

$$BEI = 10 \cdot log_{10}\left(\frac{min(P_L(\Theta_k), P_R(\Theta_k))}{\overline{min(P_L(\Theta_k), P_R(\Theta_k)}}\right) \quad k \in \{1...K\} \qquad \text{(Eq. 5)}$$

where P_L, respecticely P_R are the powers of the left, respectively right side at angles Θ_k, Std denotes standard deviation and $\overline{...}$ the mean. Note, that $P_L \propto \sum L(n,\Theta)^2$ and $P_R \propto \sum R(n,\Theta)^2$. The metric is defined as

$$Metric = BEI - SAI \qquad \text{(Eq. 6)}$$

In Fig. 3 the meaning of these indices is illustrated by the grey shaded areas. Note, that BEI resembles the definition of the classical directionality index (DI). However, while the perceptual effect of DI in its classical definition is based on the attenuation of sound from other than the frontal direction, providing directional benefit is achieved by the better-ear effect that. It describes a strategy of listening to a sound primarily through the ear at which the sound is strongest. It leads to a better separation of sources from different directions which helps with speech intelligibility. Additionally, the better-ear effect facilitates loudness summation by boosting frontal signals by 3 dB in contrast to sound coming from arbitrary other directions. Equation 6 now makes it possible to determine metric values for different algorithms characterizing how well they provide SA while at the same time maintaining DB. Figure 4 gives an example for the metric for an omnidirectional microphone on both ears and the new algorithm measured on a KEMAR manikin.

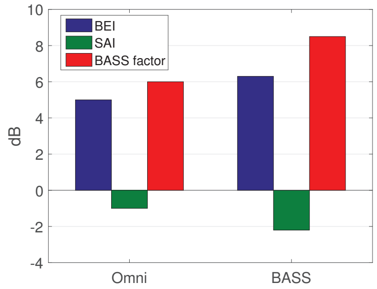


Fig. 4: SAI, BEI and the metric for an symmetric omnidirectional response on both ears and the new algorithm.

CORRELATION BETWEEN METRIC AND PERCEPTION

One question that arises from the previous sections is "what is the correlation between the metric and perception?" This section will try to address this question by measuring situational awareness and directional benefit for different metric values. In order to generate realistic sound environments and control the value of the metric, a combination of a virtual test environment with a room simulation software was applied in this study.

Room simulation software (MCRoomSim)

The room simulation software that was used in this study is MCRoomSim, a free-ware software tool (Wabnitz et al., 2010). MCRoomSim simulates both specular ('ray-tracing') and diffuse reflections in a rectangular 'shoebox' environment. It provides a MATLAB interface that allows for high level programming and set-up of simulation parameters. The output of the simulation is a matrix of room-impulse-responses from each source to each receiver channel which can be directly used in any audio application as was done in a speech-on-speech intelligibility test for this study.

Test setup

Figure 5 illustrates the setup for the measurement of situational awareness (a + b) and directional benefit (c):

For SA two distracting speech streams (red symbols) are presented from the frontal hemisphere while the target speech (green symbol) is presented either from the left or right side ('off-axis'). Intelligibility is measured for both situations independently and the obtained thresholds are averaged. In terms of our abovementioned example this would correspond to measure how sensitive one is to detecting sounds around you.

Values of metric and test environment

The polar patterns yielding different values of the metric which were applied in this study are shown in Fig. 6 . These patterns will be applied in MCRoomSim and were

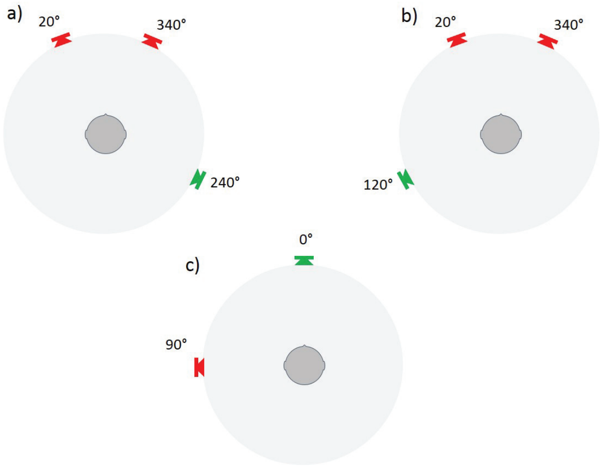


Fig. 5: Setup for measuring situational awareness (a + b) and directional benefit (c). Dark symbols denote distracting sound streams, light symbols target sound.

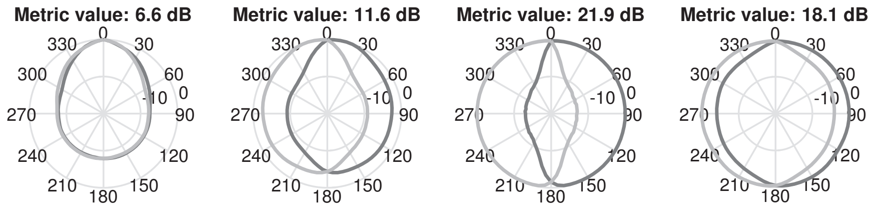


Fig. 6: Polar plots corresponding to increasing the values of the metric from left to right panel. Low values are characterized by a low situational awareness as well as small head shadow. Colors indicate sensitivity for the right (dark) and left ear (light).

applied as a direction-dependent hearing aid receiver gain. The simulated room had a low reverberation with an average broadband reverberation time of around 0.2 s. The speech reception thresholds were obtained with the help of a Danish HINT (Nielsen, 2014) implemented in MATLAB.

Subjects

Twelve normal-hearing subjects participated in the test.

RESULTS

Mean results for situational awareness and speech intelligibility are seen in the left and respectively right panel of Fig. 7. Lower speech reception thresholds (SRT) indicate a better performance. Dashed grey lines indicate trend lines. For situational awareness, the best linear fit to the data is given by $-0.53\frac{dB}{dB}\cdot x+8.02\,dB$ while for directional benefit the best fit is $-0.1\frac{dB}{dB} \cdot x - 8.34\,dB$. Threshold values for directional benefit are lower than for situational awareness. This is likely due to two effects: Thresholds for spatial separated sounds are lower for a single masker than for multiple masker sounds and the target receives a 3-dB boost due to addition of sounds from the frontal directions for all the investigated metric values.

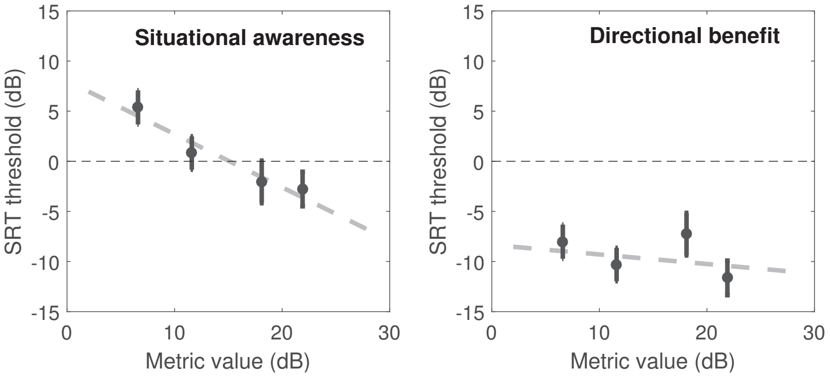


Fig. 7: Mean speech reception thresholds (SRT) for situational awareness and directional benefit. Dashed grey lines indicate trend lines.

The increase in directional benefit with larger metric values is not as large as in the case of situational awareness. However, there is one reason that the definition of the new metric makes sense: For this paradigm a higher metric value would result in less 'off-axis' attenuation and higher SRTs for a target from the front simply because less masking energy is attenuated. However, looking at the data this is clearly not the case. It indicates that the increase in masker energy with increasing metric values can indeed be compensated through the use of a larger head shadow and thus spatial separation that follows from an increasing BEI and thus higher metric values.

CONCLUSIONS

The new algorithm in Linx3D's Bilateral Directionality III tries to combine two aspects in a single hearing-aid microphone mode:

1. Situational awareness or ability to listen 'off-axis'

2. Speech intelligibility, the ability to understand speech coming from a certain direction in most cases the frontal direction

In order to be able to quantify such an approach an effective metric needed to be developed. The new defined metric fulfills this ne ed. The main purpose of this study was to correlate the (objective) metric values with perceptual data. In general, the most important finding in this study is that the new metric seems to be an appropriate tool for collectively quantifying SA and DB for hearing aid algorithms.

REFERENCES

Bronkhorst, A.W., 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.

Dittberner, A.B., and Bentler, R.A. (**2007**). “Predictive measures of directional benefit part 1: Estimating the directivity index on a manikin,” Ear Hearing, **28**, 26-45.

Kochkin, S. (**2010**). “Why my hearing aid is in the drawer: The consumer’s perspective,” Hearing Journal.

Nielsen, J.B., and Dau, T. (**2014**). “A Danish open set speech corpus for competing speech studies,” J. Acoust. Soc. Am., **135**, 407-420.

Ricketts, T.A., and Hornsby, B.W.Y.(**2003**). “Distance and reverberation effects on directional benefit,” Ear Hearing, **24**, 472-484.

Wabnitz, A., Epain, N., Craig, J., and van Schaik, A. (**2010**). “Room acoustic simulation for multichannel microphone arrays,” Proc. Int. Symp. Room Acoustics, 1-6.

Comparison of objective and subjective measures of cochlear compression in normal-hearing and hearing-impaired listeners

KONSTANTINOS ANYFANTAKIS, EWEN N. MACDONALD, BASTIAN EPP, AND MICHAL FERECZKOWSKI*

Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

Among several behavioural methods for estimating the basilar membrane input/output function, the temporal masking curve is the most popular. Distortion product otoacoustic emissions provide an objective measure for estimating cochlear compression. However, estimates from both methods have been poorly correlated in previous studies. We hypothesise that this could be due to the interplay between generator and reflection components in the recorded otoacoustic emissions. Here, compression estimates obtained with the two methods were compared at three audiometric frequencies (1, 2, and 4 kHz) for 10 normal-hearing and 6 hearing-impaired listeners. Distortion-product otoacoustic emissions were evoked using continuously-swept tones, to separate the generator component and investigate the corresponding compressive characteristic. For hearing imapired listeners, the estimates from the two methods were highly correlated.

INTRODUCTION

While it is not possible to directly measure the basilar membrane input-output (BMI/O) characteristic in humans, several indirect methods have been proposed. They can be classified into psychophysical and physiological.

Currently, the temporal masking curve paradigm (TMC, Nelson *et al.*, 2001) is the most widely used behavioural method for estimating BMI/O. However, the validity of the method and its several assumptions have been questioned. For instance, Wojtczak and Oxenham (2010) suggested that BM compression may be overestimated in TMC experiments, due to slower recovery from forward-masking for an off-frequency masker than for an on-frequency masker.

While distortion-product otoacoustic emissions (DPOAEs) may be difficult to obtain for hearing-impaired (HI) listeners, their presence is an indicator of active outer hair cells (OHC) and BM compression is believed to depend on OHC activity. Specifically, DPOAEs arise in the presence of two tonal signals (with frequencies $f1 < f2$) and the strength of the $2f1 - f2$ DP component is assumed to reflect the strength of the nonlinearity close to the $f2$ characteristic place on the BM. As the levels of the two

*Corresponding author: mfer@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

primaries increase, so does the DPOAE response and a DPI/O characteristic can be derived, as a function of the $f2$-level ($L2$). Several level rules ($L1$ as a function of $L2$) have been proposed to maximize the DPOAE, such as the scissors rule (DP-SC, Kummer *et al.*,1998) or equal-level rule (DP-EQ). The response-level maximization is desired in order to improve the SNR of DP recordings. However, it is not clear whether any of these rules guarantees maximum DP response at all $L2$ levels, for individual listeners.

Investigations of DPI/Os are often complicated by distinctive fine structure in the recorded DPOAE spectrum. The fine structure arises due to interference between the generator and reflection DP components. Since the reflection component does not directly reflect the state of the OHCs in the generator region (Abdala and Kalluri, 2017), the isolated generator component is a more accurate measure of the DPI/O at the $f2$ place.

A comparison of BMI/Os estimated with TMCs and DPOAEs was made for normal hearing (NH) listeners (Johannesen and Lopez-Poveda, 2008). Correlation between the corresponding compression exponent estimates was found at 4 kHz, but not at other frequencies. So far, no correlation has been found in hearing-impaired (HI) listeners. If the two methods gave correlated results, this would support both methods. Therefore, the main aim of the study was to reassess the correlation between the compression ratios (CR) of the BMI/O functions inferred from behavioural and objective methods for NH and HI listeners, taking into account recent developments in both physiological and psychophysical procedures. Specifically, a source-unmixing technique (Long *et al.*, 2008) was employed here. Additionally, forward pressure level (FPL, Scheperle *et al.*, 2008) calibration was performed to reduce the influence of ear-canal acoustics on the DPI/O input. Moreover, to assure the testing of a wide BMI/O dynamic range, TMCs were obtained using the Grid method (Fereczkowski *et al.*, 2016).

METHOD

Listeners

Single ears from ten NH (all thresholds $\leq$ 20 dB HL, 125-8000 Hz) and six HI listeners with sensorineural hearing loss participated in the experiments. The audiometric thresholds of the HI listeners varied between 25 and 70 dB HL at the tested frequencies (1, 2, and 4 kHz).

Measurement of DPOAEs

An Etymotic Research ER-10X probe was used for collection of the DPOAE recordings. DP-primaries were continuously swept tones with a frequency ratio of 1.22 and the sweep rate was set to 2 s/octave, as in Long *et al.* (2008). Since measurements were made for three discrete frequencies, the following sweep frequency ranges of the second primary were chosen: 0.75-1.5 kHz, 1.5-3 kHz, and

3-6 kHz. The ranges were selected to place the target frequencies near the temporal centre of the sweep to avoid edge effects.

Four levels of the second primary ($L2$) were used (35, 50, 65, and 80 dB SPL), to span the compressive range of the BMI/O for NH listeners (Neely *et al.*, 2003). Two primary-level rules were used: (1) the scissors rule (DP-SC), where $L1 = 0.4 * L2 + 39$ for $L2$ below 65 dB SPL and $L1 = L2$ at and above 65 dB SPL; and (2) the equal-level rule (DP-EQ), where $L1 = L2$ for all L2 values. When L2 was at or above 65 dB SPL, the two rules resulted in the same L1. The primaries were calibrated in situ approximately once per minute via the FPL, in order to control the level of the stimuli at the eardrum.

In each of the six tested conditions (three target frequencies and two level-rules) 108 recordings were performed per L2 level. The SNR acceptance criterion was set at 5 dB. The least-squares-fit procedure was used to isolate the DP-generator component and thus reduce the fine structure in the DP spectrum (Long *et al.*, 2008). If the generator-component response levels could be estimated for at least two $L2$ values, the correspondiing CR was estimated as an inverse of the regression slope.

Temporal masking curves (TMC)

The TMC method is based on forward masking, where the listener's task is to detect a target tone following a masker tone. Pure tones were used with a duration of 200 ms (masker) and 16 ms (target). All tones were gated with 8 ms raised-cosine ramps, hence the target had no steady-state portion. The target frequencies were the same as in the DPOAE experiment (1, 2, and 4 kHz). Four conditions were used. In three on-frequency conditions, the masker frequency was same as the target frequency. The fourth condition was the off-frequency condition, where a 2.2-kHz masker and 4-kHz target served to obtain a single linear reference for all on-frequency conditions. The single-reference approach is similar to that of Johannesen and Lopez-Poveda (2008) and is based on the assumption of frequency-independence of post-cochlear decay. When elevated thresholds and the maximum level limitation rendered the 4 kHz off-frequency TMC unobtainable, a 2-kHz off-frequency TMC was collected instead (with the masker frequency set to 1.1 kHz). The on-frequency thresholds were taken as BMI/O input estimates and the off-frequency thresholds (obtained for corresponding masker-target time gaps) served as output-level estimates (Nelson *et al.*, 2001). To aid comparability with the DP-based fits, only the TMC-BM I/O points within the input range of 35-80 dB SPL were considered for a regression fit. CR estimates were obtained as in the DPOAE case.

The Grid method (Fereczkowski *et al.*, 2016), which adaptively varies masker-target gap and masker level in each experimental run was used to estimate the masked thresholds of a 12 dB SL target as a function of the time gap. A 3-alternative forced-choice paradigm with a 1-up 2-down step-rule variant of the Grid method was employed. This method was used to enable testing a wide range of masker-target gaps and thus maximize the tested range of the estimated masked thresholds in each condition (Fereczkowski *et al.*, 2017). The set of testable gaps was defined as 10-250 ms with a 5-ms step. The corresponding set of testable masker-levels was -10 to 95 dB SPL for NH listeners and up to 100 dB SPL for HI listeners. The step size was 3 dB. The maximum level was reduced if a listener reported discomfort due to excessive loudness. At least two hours of training were administered to each listener. Six test runs were performed per test condition, to reduce the variability of the threshold estimates (Rosengard *et al.*, 2005).

RESULTS

Figure 1 illustrates the BMI/O estimates obtained for a representative NH listener (top panels) and HI listener (bottom panels). Each panel-column presents data for a single target frequency (1, 2, and 4 kHz from left to right). For the NH listener (top three panels), the slopes of the fitted lines are comparable between methods, particularly between the TMC and the DP-SC paradigms at 2 and 4 kHz. CE estimates from DP-EQ were usually higher than both TMC and DP-SC estimates (e.g., at 1 and 2 kHz). In some cases (1 kHz), the three methods returned estimates that did not show any clear correspondence. As shown in the bottom three panels of Fig. 1, several of the DPOAE data points failed to reach the 5 dB SNR criterion, particularly for frequencies above 1 kHz. This limited the number of compression slope estimates obtained for the HI listeners. Since the measured DP-EQ responses were generally lower than those from the SC paradigm, the 5-dB criterion was met less often in the EQ paradigm. Out of six HI listeners, only two returned more than 1 DP response at 2 kHz (two cases per paradigm), and just one at 4 kHz (DP-SC paradigm only).

The left panel of Fig. 2 presents scatterplots of CR estimates from the NH (top subpanels) and HI (bottom subpanels) listeners. The two left subpanels compare the TMC-based CRs (abscissa) and DP-SC inferred CRs (ordinate). The two right subpanels show the corresponding comparison between the TMC-based CRs and those from the DP-EQ paradigm. The data is aggregated across frequencies, due to the low number of DP-CR estimates at frequencies above 1 kHz obtained for HI listeners. For NH listeners, the CR estimates from both objective methods were not normally distributed. A Friedman's test showed a significant difference between the CR estimates obtained from the behavioral and objective methods [$\chi^2(2) = 35.4$, $p < 0.001$]. A post-hoc Bonferroni-corrected Yuen's paired-sample test showed that the TMC CR estimates were significantly higher than the corresponding DP-EQ estimates (trimmed mean difference of 2.06, $p < 0.001$) and that there was a trend towards TMC-CR estimates being higher than the corresponding DP-SC estimates (trimmed mean difference of 0.68, $p < 0.034$). The DP-EQ estimates were also significantly lower than the DP-SC estimates [$t(17) = 8.7, p < 0.001$] and the trimmed-mean difference was 1.38. Spearman's rank correlation coefficients between TMC- and DP-based estimates were low and insignificant ($\rho = -0.1$, $p < 0.57$ for DP-SC method and $\rho = 0.26$, $p < 0.18$ for DP-EQ method). The Spearman's correlation coefficient between the two DP methods was low (0.27) and insignificant ($p < 0.153$).

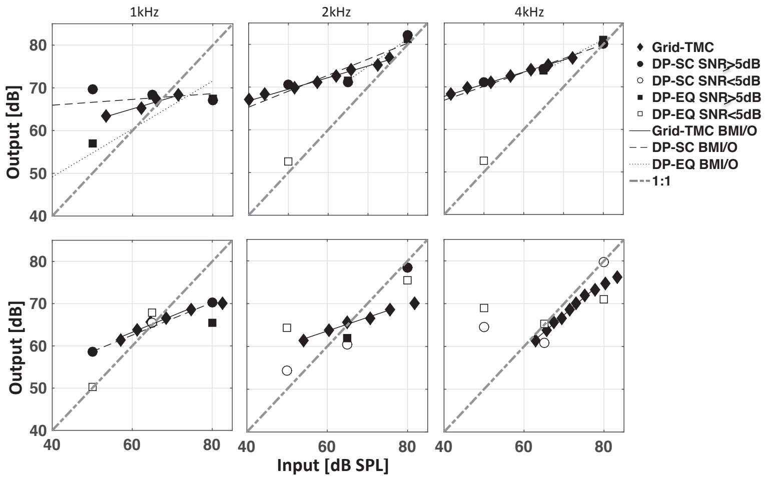


Fig. 1: BM I/O estimates from a representative NH listener (top panels) and a representative HI listener (bottom panels). Diamonds, circles and squares represent data points inferred from TMC, DP-SC and DP-EQ paradigms, respectively. Open symbols correspond to DP responses that did not meet the 5 dB SNR criterion. The solid circles and squares were fitted with straight lines, to estimate the CR of the corresponding DPI/O function. The dashed and dotted lines show the fits to the DP-SC and EQ paradigm data, respectively. The solid line represents the linear fit to the TMC-based estimates. The dash-dot line represents the linear reference (1 dB/dB). To aid visual comparability, an offset was added to each DPI/O curve, such that it coincides with the corresponding TMC-based I/O curve at the 75 dB input level.

For HI listeners, CR estimates from all three methods were normally distributed. The DP-SC CR estimates were significantly correlated with those from the TMC method (Pearson's $r = 0.77, n = 8, p < 0.026$) and the TMC CR estimates were on average lower (0.41), but the difference was not significant [$t(7) = 1.49, p < 0.18$]. The Pearson's correlation coefficient between the DP-EQ and TMC CR estimates was 0.8, i.e., comparable to the DP-SC case, but it did not reach significance ($n = 6, p < 0.057$). The average difference between the DP-EQ and TMC CR estimates was low (0.06) and not statistically significant [$t(5) = 0.24, p < 82$].

DISCUSSION

Out of the two physiological CR estimates, the DP-SC showed a better correspondence with the TMC-based estimate. First, the average difference between the DP-SC and TMC CR estimats was insignificant in NH and Hi listeners. Second, both measures

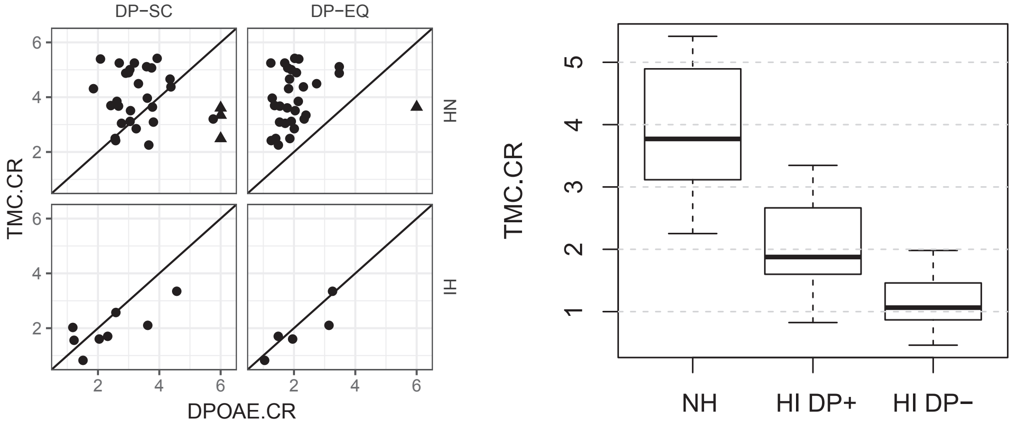


Fig. 2: Left: Scatterplots between the TMC and DP-SC (top panels) and DP-EQ (bottom panels) inferred CEs for the three tested frequencies. Right: DPOAE presence as indicator of BM compression. Each boxplot represents the TMC-inferred CRs from NH listeners and HI listeners with and without DPOAEs measured above the SNR criterion (see Discussion).

were strongly and significantly correlated in HI listeners. The lack of correlation between NH-CR estimates from the two methods is expected, under the assumption that the sporead in NH listeners data is an effect of measurement noise. In case of HI listeners, the dynamic range of the obtained estimates was larger than in NH listeners, hence the effect of measurement noise was smaller. To test this assumption, the between-method variability of the estimates was tested in NH and HI listeners by comparing geometric standard deviations (GSD) of the ratios of corresponding CR estimates obtained from the two methods. The GSD was 2.10 in NH and 1.48 in HI listeners. The NH value is inflated by three individual DP-SC CR estimates above 8, i.e., 2 times higher than the average NH value of 4 found in literature. Excluding these three values from the analysis returns a NH GSD of 1.43. This suggests that the between-method variability in NH listeners is comparable to or even higher than that in HI listeners, supporting the tested assumption. Thus, the good agreement between DP-SC and TMC results in HI listeners suggest that both methods estimate the same quality of the auditory pathway. Since DPOAEs are assumed to be generated by the cochlear nonlinearity and the generator component is assumed to reflect the state of OHCs near the $f2$ characteristic place, the observed correlations provide evidence that the TMC method is estimating BM compression. However, this conclusion is based on just eight data-points from HI listeners where CR could be estimated from DP-SC recordings.

In contrast, the CR estimates from the DP-EQ method were on average lower than the TMC and DP-SC based estimates in NH listeners, and no significant correlation was found between the DP-EQ and TMC estimates. However, the average difference between DP-EQ and TMC CR estimates was low and insignificant in HI listener group. The lower CR estimates from the DP-EQ method are a consequence of lower DP response levels elicited by this method, compared to those elicited by the DP-SC method for $f2$ levels below 65 dB SPL. In some cases the difference in responses levels exceeded 15 dB (e.g., top-right panel of Fig. 1). This suggests that the equal-level rule is less effective in eliciting BM response than the DP-SC rule, for input levels lower than 65 dB SPL, at least for NH listeners. Moreover, the DP-EQ method returned fewer CR estimates than the DP-SC method, in HI listeners at 2 and 4kHz.

DPOAE presence as indicator of compression

If an HI listener does not have functioning OHCs at some frequencies, then no measurable DP response should be obtained, regardless of the level rule. Thus, it can be hypothesized that such HI listeners will show lower behavioural CR estimates than those HI listeners with measurable DPOAEs. To test this hypothesis, a comparison was made between the TMC inferred CRs from cases with no DP data points above the SNR criterion and from cases with at least one data point above the SNR criterion. The right panel of Fig. 2 illustrates this comparison. The three boxplots show TMC-CRs from three groups of listeners: NH and HI with and without measurable DP responses. The median TMC-CR for NH listeners was 3.77. The median value of the TMC inferred CRs of the cases with and without DP responses were 1.88 and 1.06. A linear mixed-effect model was fitted to the data. The fixed effects selected for the model were hearing threshold, tested frequency, DP-response presence and the interaction of DP presence and the hearing threshold. The subject was selected as the random effect. According to the model, the DP presence was the only significant fixed predictor of the CR ($p < 0.001$) in the HI groups, which also means that there was a significant difference betwen the two HI groups. Moreover, since the median CR of the no-DP group was close to 1, it can be hypothesized that the lack of measurable DPOAEs indicates a linear BMI/O.

CONCLUSION

BMI/O estimates were inferred from a behavioural method (TMC) and two physiological paradigms (DPOAEs with scissors and equal-level rules) in NH and HI listeners. While the DP-EQ method seems not to elicit maximum response from the BM, the CR estimates from the DP-SC method were comparable to those from the TMC method, particularly in HI listeners, where no significant bias and a significant correlation was found. However, this finding is based on few data points, since physiological CR estimate was obtained in 8 out of 18 HI cases. The median TMC CR estimates in HI listeners with and without measured DP responses were 1.88 and 1.06, respectively, and the difference between the two groups was significant. Altogether, these results suggest that both the DP-SC and the TMC method estimate peripheral compression.

REFERENCES

Abdala, C., and Kalluri, R. (**2017**). “Towards a joint reflection-distortion otoacoustic emission profile: Results in normal and impaired ears,” J. Acust. Soc. Am., **142**, 812-824. doi: 10.1121/1.4996859

Fereczkowski, M., Dau, T., and MacDonald, E.N. (**2016**). “Grid-a fast threshold tracking procedure,” 63rd Open Seminar on Acoustics, pp. 545-553.

Fereczkowski, M., Jepsen, M.L., Dau, T., and MacDonald, E.N. (**2017**). “Investigating time-efficiency of forward masking paradigms for estimating basilar membrane input-output characteristics,” PloS One, **12**, e0174776. doi: 10.1371/journal.pone.0174776

Johannesen, P.T., and Lopez-Poveda, E.A. (**2008**). “Cochlear nonlinearity in normal-hearing subjects as inferred psychophysically and from distortion-product otoacoustic emissions,” J. Acoust. Soc. Am., **124**, 2149-2163. doi: 10.1121/1.2968692

Kummer, P., Janssen, T., and Arnold, W. (**1998**). “The level and growth behaviour of the 2f1-f2 distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss,” J. Acoust. Soc. Am., **103**, 3431-3444.

Long, G.R., Talmadge, C.L., and Lee, J. (**2008**). “Measuring distortion product otoacoustic emissions using continuously sweeping primaries,” J. Acoust. Soc. Am., **124**, 1613-1626. doi: 10.1121/1.2949505

Neely, S.T., Gorga, M.P., and Dorn, P.A. (**2003**). “Cochlear compression estimates from measurements of distortion-product otoacoustic emissions,” J. Acoust. Soc. Am., **114**, 1499-1507. doi:10.1121/1.1604122

Nelson, D.A., Schroder A.C., and Wojtczak M. (**2001**). “A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners.” J. Acoust. Soc. Am., **110**, 2045-2064. doi: 10.1121/1.1404439

Rosengard, P.S., Oxenham, A.J., and Braida, L.D. (**2005**). “Comparing different estimates of cochlear compression in listeners with normal and impaired hearing,” J. Acoust. Soc. Am., **117**, 3028-3041. doi: 10.1121/1.1883367

Scheperle, R.A., Neely, S.T., Kopun, J.G., and Gorga, M.P. (2008), “Influence of in situ, sound-level calibration on distortion-product otoacoustic emission variability,” J. Acoust. Soc. Am., **124**, 288-300. doi: 10.1121/1.2931953

Wojtczak, M., and Oxenham, A.J. (**2010**). “Recovery from on-and off-frequency forward masking in listeners with normal and impaired hearing,” J. Acoust. Soc. Am., **128**, 247-256. doi: 10.1121/1.3436566

Effect of musical training on fundamental frequency discrimination for older normal-hearing and hearing-impaired listeners

FEDERICA BIANCHI[1,*], TORSTEN DAU[1], AND SÉBASTIEN SANTURETTE[1,2]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Department of Otorhinolaryngology, Head and Neck Surgery & Audiology, Rigshospitalet, Copenhagen, Denmark*

Hearing-impaired (HI) listeners, as well as elderly listeners, typically have a reduced ability to discriminate the fundamental frequency (F_0) of complex tones compared to young normal-hearing (NH) listeners. Several studies have shown that musical training, on the other hand, leads to improved F_0-discrimination performance for NH listeners. It is unclear whether a comparable effect of musical training occurs for listeners whose sensory encoding of F_0 is degraded. To address this question, F_0 discrimination was investigated for three groups of listeners (14 young NH, 9 older NH and 10 HI listeners), each including musicians and non-musicians, using complex tones that differed in harmonic content. Musical training significantly improved F_0 discrimination for all groups of listeners, especially for complex tones containing low-numbered harmonics. In a second experiment, the sensitivity to temporal fine structure cues (TFS) was estimated in the same listeners. Although TFS cues were degraded for the two older groups of listeners, musicians showed better performance than non-musicians. Additionally, a significant correlation was obtained between F_0-discrimination performance and sensitivity to TFS cues for complex tones with low and intermediate harmonic numbers. These findings suggest that musical training may enhance both sensory encoding of TFS cues and F_0 discrimination in young and older listeners with or without hearing loss.

INTRODUCTION

The effects of musical training on fundamental frequency (F_0) discrimination have been largely investigated for young normal-hearing (NH) listeners. Behavioral studies have shown that young NH musicians perform two to six times better than non-musicians in complex-tone F_0 discrimination (Spiegel and Watson, 1984; Micheyl *et al.*, 2006; Bianchi *et al.*, 2016). However, little is known about the effects of musical training for older and hearing-impaired (HI) listeners. The aim of this study was to assess whether older and HI listeners show a benefit of musical training and to clarify the extent to which the degradation of peripheral cues is a limiting factor.

[*] Corresponding author: fbia@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.

ISBN: 978-87-990013-6-1.

The ability to discriminate F_0 changes is assumed to be partly limited by the frequency resolution of the peripheral auditory system. The harmonic overtones of a complex tone are considered to be resolved when they are processed within distinct auditory filters (up to the 8th harmonic for NH listeners; Plomp (1964)), and unresolved when neighbouring harmonics interact within the same filter (above the 13th harmonic). It has been shown that HI listeners with sensorineural hearing loss (SNHL) have a reduced ability to discriminate the F_0 of complex tones with resolved harmonics relative to young NH listeners (Moore and Peters, 1992; Bernstein and Oxenham, 2006). This perceptual deficit may be ascribed to a variety of factors, such as reduced frequency selectivity (Bernstein and Oxenham, 2006), degraded temporal fine structure processing (TFS; Hopkins and Moore, 2007) and decreased neural synchrony. Older listeners also show reduced F_0 discrimination (Moore and Peters, 1992), possibly due to the degradation of TFS cues, despite normal audiometric thresholds and filter bandwidths (Hopkins and Moore, 2011).

In this study, two experiments were performed using three groups of listeners, young NH (YNH), older near-NH (ONH) and older HI, each including musicians and non-musicians. In the first experiment, F_0-discrimination performance was investigated using complex tones that differed in harmonic content to clarify how the effect of musical training varies when frequency selectivity and TFS sensitivity are degraded. In the second experiment, the ability to use TFS cues was assessed for the three groups and compared with the outcomes of the first experiment.

METHOD

Listeners

Fourteen YNH listeners (7 musicians, 7 non-musicians; mean age 25 ± 4 years), nine ONH listeners (3 musicians, 6 non-musicians; mean age 62 ± 5 years) and ten HI listeners (5 musicians, 5 non-musicians; mean age 68 ± 6 years) participated in this study. All YNH listeners had hearing thresholds lower or equal to 20 dB hearing level (HL) between 125 Hz and 8 kHz. The ONH listeners had hearing thresholds lower than or equal to 25 dB HL up to 4 kHz. The HI listeners had hearing thresholds up to 70 dB HL up to 4 kHz. Musicians had at least eight years of formal music education and non-musicians less than 3 years. One non-musician underwent musical training for 6 years, but stopped 40 years before his participation in this study.

Experiment I: F_0 discrimination

A three-alternative forced choice (3-AFC) paradigm was used in combination with a weighted up-down method to estimate 75% correct performance. In each trial, two intervals contained a reference complex tone with a fixed F_0 (125 Hz) and one interval contained the target complex tone with a higher F_0. The task was to select the interval containing the tone with the highest pitch. The difference in F_0 between the reference and the target, ΔF_0, was initially set to 20% and was decreased after each correct response and increased after each incorrect response. The threshold for each condition was measured four times. The first repetition was considered as training and the last three were used to calculate the final F_0-discrimination threshold (F_0DL).

Five conditions were tested: a resolved condition (RES, harmonics: 3-9), an intermediate condition (INT, harmonics: 10-16), two unresolved conditions (UN1, harmonics: 17-23; UN2, harmonics: 17-36) and a broadband condition (ALL, harmonics: 3-36). To avoid spectral edges as a discrimination cue, the lowest harmonic number was roved within each trial, such that the three complex tones had lowest harmonic numbers of $N-1$, N and N+1 in a random order, where N was the lowest nominal harmonic number in each condition (Bernstein and Oxenham, 2003).

All signals were 300-ms complex tones embedded in broadband threshold equalizing noise (TEN). The complex tones were created by summing harmonic components either in sine, Schroeder positive or Schroeder negative phase (Schr + or -) to vary the envelope peakiness. For the NH listeners, the TEN level was set to 55 dB SPL per equivalent rectangular bandwidth (ERB_N). For the HI listeners, the level of the TEN per ERB_N was set to the maximum hearing threshold up to 4 kHz. Each harmonic of each complex tone was set at 12.5 dB sensation level (SL) re the threshold in the TEN.

Experiment II: IPD detection

To obtain an estimate of interaural phase sensitivity, the highest frequency at which an interaural phase difference (IPD) of 180°could be detected was measured using a 2-AFC paradigm with a one-up two-down tracking rule (71% correct performance). For each trial, the reference interval contained four diotic pure tones ("AAAA", IPD = 0°), each 400 ms in duration with a 100-ms inter-stimulus interval.The target interval contained two diotic and two dichotic tones (IPD = 180°), presented in a interleaved manner ("ABAB"). The interval between reference and target was of 333 ms. The task was to select the interval containing the tones that were perceived as shifting location inside the head. The starting frequency was 500 Hz. The tones were presented at 35 dB SL. The experiment was carried out three times, and the final threshold was calculated as the mean of three repetitions. Prior to carrying out the IPD experiment, the listeners had a short familiarization session (2 minutes) with a similar task, where an interaural level difference was introduced in the dichotic conditions instead of an IPD.

RESULTS

Experiment I: F_0 discrimination

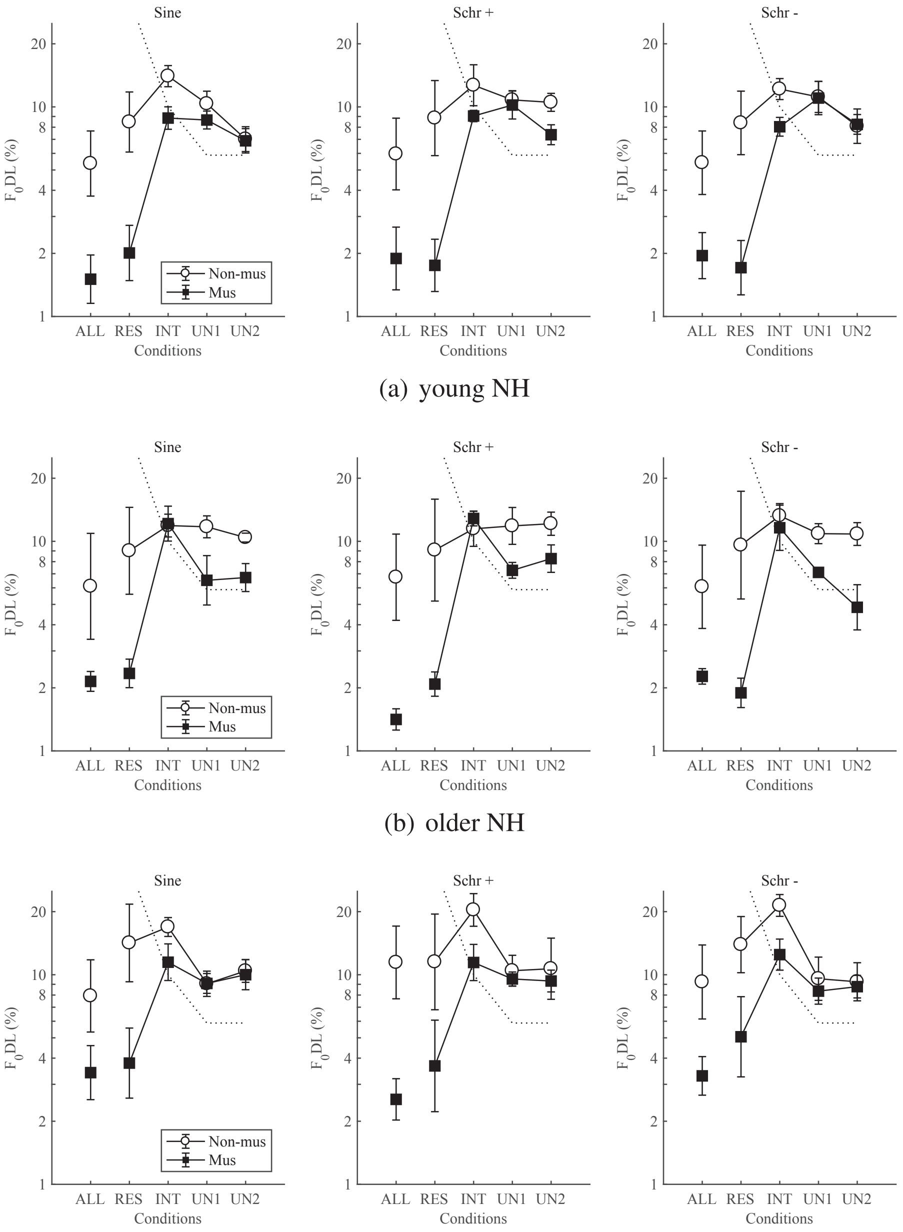
The mean F_0DLs for the three groups of listeners are presented in Fig. 1. Performance was most accurate (i.e., lowest thresholds) for the ALL and RES conditions and worsened for the INT and UN conditions. Performance was worse for the ONH and HI listeners for the ALL, RES and INT conditions than for the YNH listeners. However, the effect of musical training was similar across the three groups, with significantly lower thresholds for musicians in the ALL and RES conditions. An analysis of variance (ANOVA) with factors condition, musicianship, group, and phase gave significant effects of condition [F(4,457) = 54.8; $p < 0.001$), musicianship [F(1,457) = 127.2; $p < 0.0001$], and group [F(2,457) = 12.4; $p < 0.001$], a significant interaction between musicianship and condition [F(4,457) = 20.4; $p < 0.001$], and a marginally significant interaction between group and condition [F(8,457) = 1.96; p = 0.050]. Phase was not significant, nor the interaction between musicianship and group.

The dashed line in Fig. 1 shows the thresholds (66.7% correct) predicted if performance had solely been based on spectral edge cues. Although 66.7% is lower than the tracked 75% correct performance, it is possible that thresholds significantly above the dashed line were based on spectral edge cues, rather than F_0s cues (Bernstein and Oxenham, 2003). Since most thresholds in the UN conditions were significantly above the dashed line, it cannot be excluded that for this condition the listeners used spectral edges as a cue, rather than F_0 cues.

Experiment II: IPD detection

Figure 2 depicts the highest frequency (f_{max}) at which an IPD was detected for each listener group. YNH musicians were sensitive to the IPD shift, on average, up to 1281 Hz, while YNH non-musicians were sensitive up to 1116 Hz. Sensitivity to IPD decreased for the ONH listeners (musicians: 1022 Hz; non-musicians: 761 Hz), and for the HI listeners (musicians: 993 Hz; non-musicians: 820 Hz). An ANOVA with factors group and musicianship showed a significant effect of both factors [group: F(2,26) = 8.09, p = 0.002; musicianship: F(1,26) = 6.87, p = 0.014]. The interaction was not significant. Although there was an overall trend for musicians to be sensitive to IPD up to higher frequencies, posthoc *t*-tests revealed that the effect of musicianship was not significant within each group. Additionally, the group difference was mostly driven by age (the thresholds for the ONH and HI groups were not significantly different).

Spearman correlations were calculated between the IPD f_{max} thresholds and the F_0-discrimination performance (Experiment I). A significant correlation was found for the ALL condition ($r = -0.48$; p = 0.007), RES condition ($r = -0.48$; p = 0.006), and INT condition ($r = -0.36$; p = 0.043) but not for the UN conditions (Fig. 3). This finding suggests that TFS cues may play a role for F_0 discrimination of complex tones containing low and intermediate numbered harmonics (Moore and Moore, 2003). No significant correlations were obtained for the musicians alone (N = 15), suggesting that the degradation of TFS cues with age did not affect the musicians' F_0-discrimination performance.



(a) young NH

(b) older NH

(c) older HI

Fig. 1: Mean F$_0$DLs and standard errors for the three groups of listeners (N = 33): a) young NH listeners; b) older NH listeners; c) older HI listeners. Musicians are depicted with filled squares and non-musicians with open circles. Left panels: sine phase condition; Middle panels: Schroeder positive; Right panels: Schroeder negative. The dashed line depicts the predicted thresholds (66.7% correct) if the listener used only spectral edge cues.

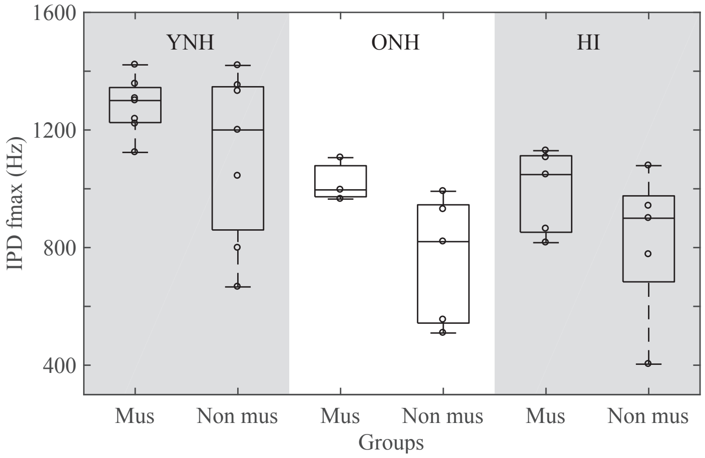


Fig. 2: Highest frequency at which an IPD was detected. The median for each group of listeners is depicted together with the 25th and 75th percentiles. The individual results are depicted by the open circles (N = 32 listeners; one ONH listener non-musician could not perform the task).

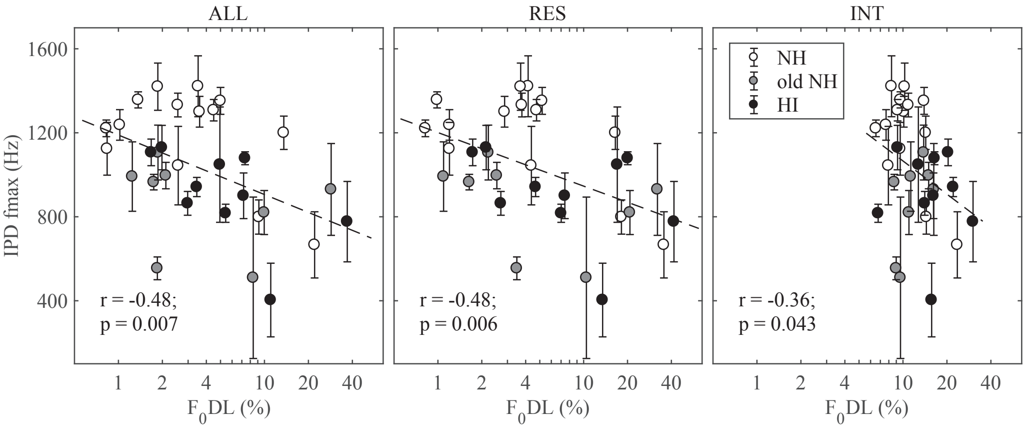


Fig. 3: Scatter plot and Spearman correlations between the IPD f_{max} thresholds and the F_0DLs averaged across phase conditions (N = 32 listeners), for the ALL (left panel), RES (middle panel), and INT (right panel) conditions.

DISCUSSION

The aim of this study was to clarify whether listeners with degraded processing of F_0 cues would show a benefit of musical training for F_0 discrimination, comparable to that observed for YNH listeners. Experiment I showed a similar benefit of musicians for the three groups of listeners (confirmed by the absence of a significant interaction of group and musicianship), with the largest benefit observed in the ALL and RES conditions. The fact that the benefit of musicians was larger in these two conditions may be ascribed either to a training-dependent effect that may be more salient for complex tones containing lower-numbered harmonics (Bianchi *et al.*, 2017) or to the random changes in the lowest harmonic number which may be more distracting for non-musicians. In favor of this last hypothesis, the F_0DLs were significantly lower (better) for the ALL than for the RES condition for non-musicians (posthoc *t*-test, $p < 0.001$), but not for musicians. This may be due either to the contribution of high-numbered harmonics in the ALL condition for non-musicians or to a reduced distraction in the ALL condition from the spectral upper-edge pitch.

The F_0DLs obtained in this study for YNH listeners in the RES condition were, on average, 1.8% for musicians and 8.5% for non-musicians. These discrimination thresholds are much higher than the F_0DLs obtained in previous studies for resolved complex tones presented at similar sensation levels as in the current study (Oxenham *et al.*, 2009; Bianchi *et al.*, 2016). This difference may be ascribed to the distracting effect of the randomization of the lowest harmonic number. Since the lowest harmonic number could differ by $\pm$ 1 across intervals, spectral edge pitch was a strong distracting cue especially for the RES condition. In this condition, spectral edge cues helped in the discrimination task only when the lowest harmonic number of the target was higher than both references (i.e., in one out of three cases, hence at chance level). In the remaining cases, the spectral edge cue was a disrupting cue in the F_0-discrimination task, leading to higher thresholds.

Although one additional aspect of this study was to investigate the effect of musical training for different harmonic phases, there was no significant difference in thresholds between sine and Schroeder phase, in contrast to the results of Houtsma and Smurzynski (1990). A possible explanation may be that we used a noise level high enough to mask distortion products, in combination with a low sensation level, which has been shown to lead to higher F_0DLs (Oxenham *et al.*, 2009). When the F_0DLs are high, spectral edge pitch may help in the discrimination task for the INT and UN conditions. This may explain the very small (or absent) benefit of musicians for the INT and UN conditions, as well as the absence of significant phase effects for the UN conditions (Oxenham *et al.*, 2009).

Overall, the findings of this study suggest that the F_0-discrimination performance of all listeners depends on sensitivity to TFS cues for complex tones containing low and intermediate numbered harmonics (Moore and Moore, 2003). Although limited by age, TFS cues were generally enhanced in musicians. This may be explained by enhanced neural synchrony in the brainstem of musicians (Parbery-Clark *et al.*, 2012), which could increase the sensitivity to small time differences and lead to a more accurate representation of pitch (Bianchi *et al.*, 2017). These findings suggest that although the sensory encoding of pitch cues was degraded in older and HI listeners, a benefit of musical training comparable to that of YNH listeners was still present and could account for improved encoding of TFS cues and F_0 discrimination. Hence, music-training paradigms in older listeners may be considered as a tool to improve auditory perceptual skills, although the effects may be different if musical training is applied later in life after hearing loss onset.

REFERENCES

Bernstein, J.G.W., 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. doi: 10.1121/1.1572146

Bernstein, J.G.W., and Oxenham, A.J. (**2006**). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," J. Acoust. Soc. Am., **120**, 3929-3945. doi: 10.1121/1.2372452

Bianchi, F., Santurette, S., Wendt, D., and Dau, T. (**2016a**). "Pitch Discrimination in Musicians and Non-Musicians: Effects of Harmonic Resolvability and Processing Effort," J. Assoc. Res. Otolaryngol., **17**, 69-79. doi: 10.1007/s10162-015-0548-2

Bianchi, F., Hjortkjær, J., Santurette, S., Zatorre, R.J., Siebner, H.R., and Dau, T. (**2017**). "Subcortical and cortical correlates of pitch discrimination: Evidence for two levels of neuroplasticity in musicians," Neuroimage. doi 10.1016/j.neuroimage.2017.07.057

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. doi: 10.1121/1.2749457

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. doi: 10.1121/1.3585848

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. doi: 10.1121/1.399297

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A.J. (**2006**). "Influence of musical and psychoacoustical training on pitch discrimination," Hear. Res., **219**, 36-47. doi: 10.1016/j.heares.2006.05.004

Moore, B.C.J., and Peters, R.W. (**1992**). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," J. Acoust. Soc. Am., **91**, 2881-2893.

Moore, B.C.J., and Moore, G.A. (**2003**). "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. doi: 10.1016/S0378-5955(03)00191-6

Oxenham, A.J., Micheyl, C., and Kleebler, M.V. (**2009**). "Can temporal fine structure represent the fundamental frequency of unresolved harmonics?" J. Acoust. Soc. Am., **125**, 2189-2199.

Parbery-Clark, A., Anderson, S., Hittner, E., and Kraus, N. (**2012**). "Musical experience offsets age-related delays in neural timing," Neurobiol. Aging, **33**, 1483:e1-e4. doi: 10.1016/j.neurobiolaging.2011.12.015

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

Spiegel, M.F., and Watson, C.S. (**1984**). "Performance on frequency discrimination tasks by musicians and non-musicians," J. Acoust. Soc. Am., **76**, 1690-1695.

Spectral and binaural loudness summation in bilateral hearing aid fitting

MAARTEN VAN BEURDEN[1,2], MONIQUE BOYMANS[1,2], MIRJAM VAN GELEUKEN[1], DIRK OETTING[3], AND WOUTER A. DRESCHLER[1,*]

[1] *Academic Medical Center, Department of Clinical & Experimental Audiology, Amsterdam, The Netherlands*

[2] *Libra Rehabilitation & Audiology, Eindhoven, The Netherlands*

[3] *HörTech gGmbH and Cluster of Excellence Hearing4all, Oldenburg, Germany*

Aversiveness of loud sounds is a frequent complaint by hearing-aid users, especially when fitted bilaterally. This study investigates whether loudness summation can be held responsible for this finding. Two aspects of loudness summation should be taken into account: spectral loudness summation for broadband signals and binaural loudness summation for signals that are presented binaurally. In this study different aspects were investigated: (1) the effect of different symmetrical hearing losses according to the classification of Bisgaard *et al.* (2010): N2, N3, N4, S2, and S3, and (2) the effect of spectral shape of broadband signals, by using high frequency noise and low frequency noise. For the measurements we used a well-standardized technique "Adaptive Categorical Loudness Scaling" (ACALOS). Also loudness matching was applied as a potentially clinical technique to get information about the individual loudness perception. Results show large individual differences in binaural loudness perception especially for broadband stimuli.

INTRODUCTION

Nowadays, the majority of listeners with hearing loss are fitted bilaterally. The use of two hearing aids increased over the last decades and reached values of about 75% in the US (Kochkin, 2009) and about 70% in Europe (see www.ehima.com). Bilaterally fitted hearing aids have been shown to improve speech intelligibility both in quiet and in noise and to improve localization (Boymans *et al.*, 2008; 2009). However, with respect to aversiveness of loud sounds bilateral fittings typically have poorer scores than unilateral fittings (Boymans *et al.*, 2009). Loudness complaints remain a major reason for revisiting the hearing aid dispenser (Jenstad *et al.*, 2003) and averseness of loud sounds is one of the main reasons to be dissatisfied with a hearing aid fitting (Hickson *et al.*, 2010).

It is generally accepted that hearing aid rehabilitation involves successive steps, starting with a first-fit based on a prescriptive formula, followed by individual fine tuning based on subjective responses and/or technical measurements using in-situ responses.

[*] Corresponding author: w.a.dreschler@amc.uva.nl

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

Over the years a number of prescriptive formulas have been developed. The linear prescriptive formulas (e.g., NAL-R) have been replaced by non-linear prescriptions as NAL-NL2 (Dillon, 2012), taking into account that the amount of gain required is not only frequency dependent, but also level dependent.

Nonlinear fitting formulas show some relationship with the loudness growth at different frequencies. The level of detail of knowledge about loudness perception required for an effective first-fit setting is still in debate. But the dynamic range as the frequency-dependent range between the individual thresholds and the levels of uncomfortable loudness is generally accepted and applied in different forms in nonlinear prescriptive formulas.

Due to the fact that the hearing loss is often strongly frequency-dependent, loudness growth is usually measured with narrow-band signals. Loudness curves measured in individual hearing-impaired subjects can be compared with loudness curves of normal-hearing listeners and thus transferred into level-dependent gain prescriptions for hearing aid amplification settings to normalize loudness (Herzke and Hohmann, 2005).

However, in this approach, two aspects of loudness perception are not taken into account: spectral loudness summation (in case of the presentation of broadband signals instead of narrow-band signals) and binaural loudness summation (in case of bilateral presentation instead of unilateral). This includes also the binaural loudness perception of broadband signals which can be referred to as binaural spectral loudness summation. This combined effect has to be considered because typically two hearing aids are worn and they will typically process broadband signals as speech or environmental sounds.

These three types of loudness summation may require individual corrections. Recent data of hearing-impaired listeners (Oetting *et al.*, 2016) showed large individual differences in spectral loudness summation and binaural loudness summation after careful narrowband loudness normalization. Some of the listeners showed loudness perception for binaural broadband signals that was fully in agreement with normal-hearing reference data whereas others showed a higher-than-normal loudness sensitivity of up to 30 dB SPL for the binaurally presented broad-band signals. Given the magnitude of the inter-individual differences found, it can be assumed that these findings are relevant for loudness adjustments during bilateral hearing aid fittings.

In this study we measured spectral and binaural loudness summation as well as the combination, binaural spectral loudness summation. Listeners with different audiometric shapes were tested to investigate if the shape of the audiogram could explain the individual differences.

METHODS

Subjects

The inclusion criteria were: Age above 18 years; Native speaker of Dutch.

From the clinical files we selected subjects with mild to moderate symmetrical hearing losses (differences between both ears at 0.5, 1, 2 and 4 kHz < 10dB) and their pure-tone audiograms were classified according to Bisgaard *et al.* (2010). Twelve women and 10 men participated with an average age of 70 years. The classifications of the audiograms of the 44 ears included can be seen in Fig. 1.

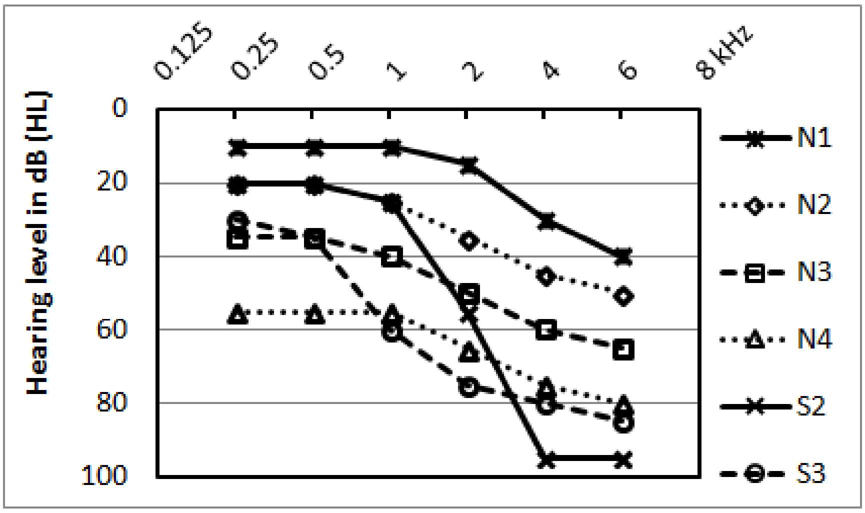


Fig. 1: Standard audiograms according to Bisgaard *et al.* (2010). There are 1, 11, 18, and 6 ears in the categories N1 to N4, and 6 and 2 ears in categories S2 and S3, respectively.

Equipment

All measurements were conducted in a sound-insulated booth in two sessions of about 2 hours each. Pure-tone audiograms (air and bone conduction) were measured with DECOS audiometers, using TDH39 headphones. Sennheiser HDA 200 headphones were used for the loudness scaling and the loudness matching. Both procedures were conducted using the framework for psychoacoustic experiments (Ewert, 2013). Signals were presented using a RME Fireface UC at 44.1 kHz. Headphones were calibrated using a Brüel & Kjær artificial ear type 4153, a 0.5-inch microphone type 4134, a microphone preamplifier type 2669, and a measuring amplifier type 2610. Headphones were free-field equalized according to ISO 389 (2004) and levels are expressed as the equivalent free-field level in dB SPL(FF).

Stimuli

All stimuli were 1-s noises with 50-ms rise and fall ramps. For the narrow-band signals one-third octave low-noise noises (LNN; Kohlrausch *et al.*, 1997) were used. The narrow-band stimuli had center frequencies of 250, 500, 1000, 2000, 4000, and 6000 Hz. The stimuli to assess loudness summation effects consisted of uniformly exciting noise (UEN, Fastl and Zwicker, 2007) with bandwidths of 1, 5, and 17 Barks, referred to as UEN1 (bandwidth: 210 Hz), UEN5 (1080 Hz) and UEN17 (5100 Hz), respectively. The UEN noises were centered at 10.5 Bark (1370 Hz).

In addition to the UEN noises a speech shaped noise referred to as IFnoise (international female noise) was included in the test battery. The IFnoise was generated to match the spectral shape as the long-term average speech spectrum for females (Byrne *et al.*, 1994).

Loudness Scaling Procedure

After the inclusion criteria were checked, categorical loudness scaling using ACALOS was performed to measure the individual loudness perception. During the ACALOS procedure listeners had to rate the perceived loudness on an 11-point scale from “not heard” to “too loud”, which were transformed into numerical values in “Categorical Units” (CU) from 0 to 50. Stimuli were presented in a pseudo-random order with levels between −10 and 105 dB HL. A monotonically increasing loudness function was fitted to the responses for each of the ACALOS measurements using the BTUX fitting method (Oetting *et al.*, 2014). The model function consists of two linear parts with independent slopes m_{low} and m_{high} with a smooth transition range (see Brand and Hohmann, 2002).

Before loudness summation was determined for the broadband signals the UEN and IFnoise noises were corrected for each hearing impaired subject individually aiming to restore the loudness of the narrow-band signals to that of the average normal hearing listener (narrow-band loudness normalization). The required gain (Fig. 2) was defined as the difference in level for each loudness category between the individual loudness functions of the narrow-band signals and the average normal hearing loudness function. To quantify the level of correction, for each narrow-band signal the compression ratio (CR) was calculated defined as the ratio between input and output level at 40 and 80 dB input level according to:

$$CR = \frac{\Delta in}{\Delta out} = \frac{\Delta in}{\Delta in - \Delta gain} = \frac{80-40}{80-40-(G40-G80)} = \frac{40}{40-\Delta gain} \qquad Eq.(1)$$

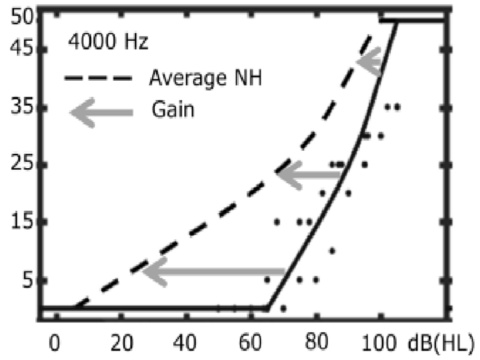


Fig. 2: Gain correction at 4000 Hz to the normal-hearing reference.

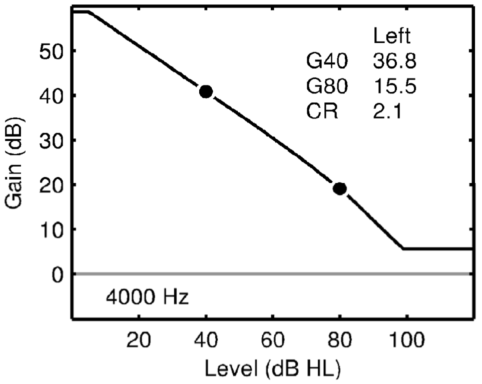


Fig. 3: Compression ratio of 2.1, calculated for the 4 kHz signal of Fig. 2.

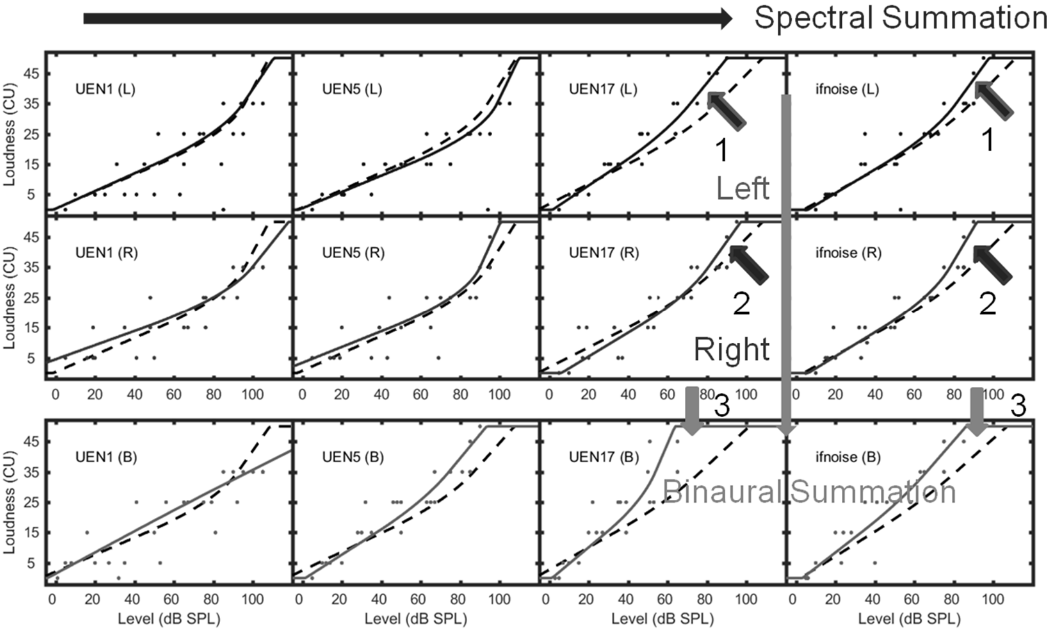


Fig. 4: Results for a hearing-impaired listener for spectral loudness summation for signals with increasing bandwidth (from left to right) and binaural loudness summation: from unilateral (upper rows) to bilateral (bottom row) presentation.

RESULTS

The narrow-band loudness normalization fitting method showed decreasing gains with increasing presentation level (Fig. 3). The results show that this fitting was able to restore normal loudness perception of narrow-band signals (UEN1, left panels in Fig. 4). However, normal loudness perception for narrow-band signals is no guarantee for normal loudness perception for broadband and binaurally presented signals, in fact, huge inter-individual variability was found in these conditions. Examples of such differences are shown in Fig. 4. In Fig. 4 spectral loudness summation (with increasing bandwidth from left to right) is shown to be higher than normal at high levels for both UEN17 and IFnoise at both ears (see arrows 1 and 2). In the same subject binaural spectral loudness summation (lower panel) for these same stimuli is even higher (arrows 3).

Figure 5a shows individual data per ear for the differences in spectral loudness summation at 35 CU (calculated as the level differences of the average level of UEN17 and IF noises relative to the average level of UEN1 and UEN5). Even within a Bisgaard classification large inter-individual differences in loudness summation were found. The differences between hearing loss configurations suggest a trend for more spectral loudness summation for hearing impaired subjects with increasing hearing loss, especially at configuration N4. Figure 5b shows the binaural loudness

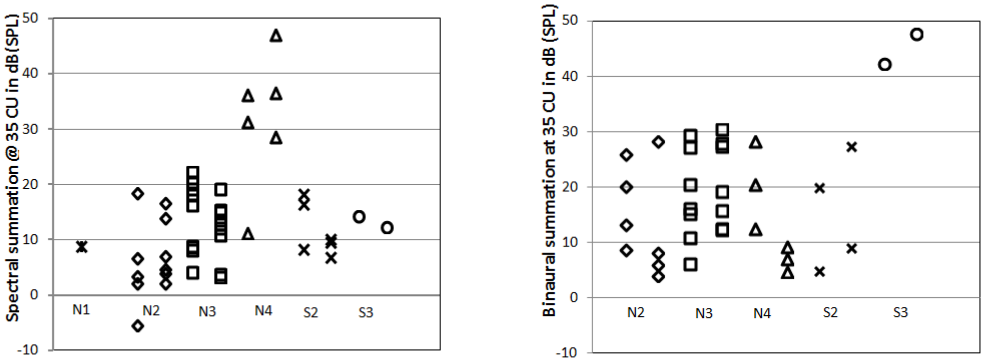


Fig. 5: Individual spectral loudness summation (5a) and binaural loudness summation (5b) per ear. Left hand side: left ear. Right hand side: right ear.

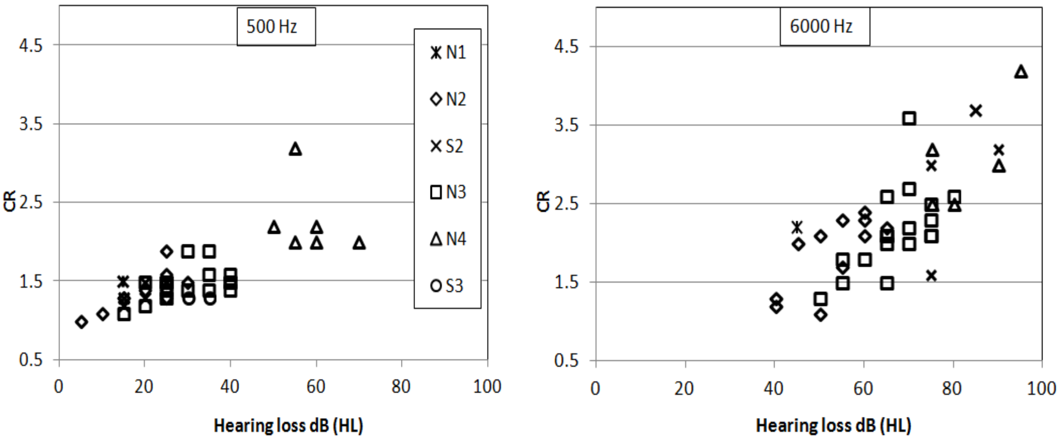


Fig. 6: Compression ratios for all subjects at 500 Hz (left panel) and at 6000 Hz (right panel) for different hearing loss categories.

summation effect (calculated as the level differences of the average binaural UEN 17 and IF noises relative to the average level of the monaural UEN17 and IF noises for both ears individually) as a function of hearing loss configuration. For binaural loudness summation the group data are more uniform across different audiogram configurations. However, we also found an extreme high level of binaural loudness summation for the single subject with hearing loss configuration S3.

Figure 6 shows the calculated CRs per ear for different hearing loss categories for narrow-band signals at 500 Hz and 6000 Hz. At both frequencies CRs tend to increase with increasing hearing losses. The CRs at 6000 Hz are higher than at 500 Hz, not only due to the fact that the hearing losses in the higher frequencies are on average higher, but also due to the fact that the increase of CR with increasing hearing loss tends to be stronger at 6000 Hz than at 500 Hz. CRs may therefore be used to characterize the amount of loudness compensation used in a certain narrow-band region.

DISCUSSION AND CONCLUSIONS

Although narrow-band loudness normalization has proven to give good individual results for narrowband signals, the current results show that this does not guarantee loudness normalization for broadband stimuli or stimuli presented binaurally. Furthermore, the large individual variability of spectral and binaural loudness summation could not be predicted from the hearing loss configuration. The only observed trends were higher spectral loudness summation in listeners with N4 audiograms and higher binaural loudness summation in a subject classified as S3.

In further analysis the compression ratios may be useful to investigate if the amount of loudness summation can be predicted by the amount of loudness compensation (applied gain for broadband signals).

The high individual variability in loudness perception for binaurally presented broadband signals can be one of the causes of aversiveness for loud sounds of bilateral hearing aid users. The individual differences are that large that they should be taken into account during the hearing aid fitting procedure. Currently, the most common fitting rules only utilize average gain corrections for bilateral fittings that are identical for all hearing-impaired subjects. NAL-NL1 and NAL-NL2 utilize a bilateral compensation (reduction in gain) with respect to a unilateral fitting of 3 and 2 dB, respectively, for input levels below 40 dB increasing to 6 and 8 dB, respectively at 90 dB SPL and above. (Byrne *et al.*, 2001; Keidser *et al.*, 2012). Our results show bilateral summation effects above 20 dB at high levels.

Therefore, there is a need to adjust fitting rules for bilaterally fitted hearing aids to take the large individual differences in loudness summation into account. A loudness-based approach based on individual measurements will require extra tests and thus requires extra time that is usually scarce. For this purpose we compared loudness matching with loudness scaling to find out if the first method is applicable in clinical practice and can be used as an alternative with reduced testing time. Preliminary results indicate that loudness matching could be suitable. A typical loudness scaling condition takes about 2 minutes. A single comparison between conditions therefore takes about 4 minutes. The loudness matching procedure compares 15 conditions in about 10 minutes.

More important is that the loudness matching produces about equivalent results to the loudness scaling data. That is, in one and the same subject, the amount of loudness difference between two stimuli (narrow-band vs. broad-band and monaural vs. binaural) at 35 CU is in qualitative agreement in both procedures.

REFERENCES

Bisgaard, N., Vlaming, M.S., and Dahlquist, M. (**2010**). "Standard audiograms for the IEC 60118-15 measurement procedure," Trends Amplif., **14**, 113-120.

Boymans, M., Goverts, S.T., Kramer, S.E., Festen, J.M., and Dreschler, W.A. (**2008**). "A prospective multi-centre study of the benefits of bilateral hearing aids," Ear Hearing, **29**, 930-941.

Boymans, M., Goverts, S.T., Kramer, S.E., Festen, J.M., and Dreschler, W.A. (**2009**). "Candidacy for bilateral hearing aids: a retrospective multicenter study," J. Speech Lang. Hear. Res., **52**, 130-140.

Brand, T., and Hohmann, V. (**2002**). "An adaptive procedure for categorical loudness scaling," J. Acoust. Soc. Am., **112**, 1597-1604.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., *et al.* (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am., **96**, 2108-2120.

Byrne, D. Dillon, H., Ching, T., Katsch, R., and Keidser, G. (**2001**). "NAL-NL1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures.," J. Am. Acad. Audiol., **12**, 37-51.

Dillon, H. (**2012**). *Hearing Aids Second Edition.* Sydney, Australia: Boomerang Press, pp. 226.

Ewert, S.D. (**2013**). "AFC – a modular framework for running psychoacoustic experiments and computational perception models," 39 Tagung der Deutschen Arbeitsgemeinschaft für Akustik (DAGA).

Fastl, H., and Zwicker, E. (**2007**). In: *Psychoacoustics: Facts and Models, third ed.* Springer, Berlin.

Herzke, T., and Hohmann, V. (**2005**). "Effects of instantaneous multiband dynamic compression on speech intelligibility," EURASIP J. App. Sig. P., **18**, 3034-3043.

Hickson, L., Clutterbuck, S., and Khan, A. (**2010**). "Factors associated with hearing aid fitting outcomes on the IOI-HA," Int. J. Audiol., **49**, 586-595.

ISO 389-8 (**2004**). *Acoustics - Reference zero for the calibration of audiometric equipment.*

Keidser, G., Dillon, H., Carter, L., and O'Brien, A. (**2012**). "NAL-NL2 empirical adjustments," Trends Amplif., **16**, 211-223.

Kochkin, S. (**2009**). "MarkeTrak VIII: 25-year trends in the hearing health market," Hearing Review, **16**, 12-31.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A.J., and Püschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acta Acust. United Ac., **83**, 659-669.

Jenstad, L.M., Van Tasell, D.J., and Ewert, C. (**2003**). "Hearing aid troubleshooting based on patients' descriptions," J. Am. Acad. Audiol., **14**, 347-360.

Oetting, D., Brand, T., and Ewert, S.D. (**2014**). "Optimized loudness-function estimation for categorical loudness scaling data," Hear. Res., **316**, 16-27.

Oetting, D., Hohmann, V., Appell, J.-E., Kollmeier, B., and Ewert, S.D. (**2016**). "Spectral and binaural loudness summation for hearing-impaired listeners," Hear. Res., **335**, 179-192.

Auditory disabilities, individual fitting targets, and the compensation power of hearing aids

SIMON LANSBERGEN[1], INGE DE RONDE-BRONS[1], MONIQUE BOYMANS[1], WIM SOEDE[2], AND WOUTER DRESCHLER[1,*]

[1] *Clinical & Experimental Audiology AMC, Amsterdam, The Netherlands*

[2] *Audiology, LUMC, Leiden, The Netherlands*

There is lack of a systematic approach how to select an adequate hearing aid and how to evaluate its efficacy towards the personal needs of rehabilitation. The goal of this study was to examine the applicability and added value of two widely used self-reporting questionnaires (COSI and AVAB) in relation to the evaluation of hearing aid fitting. We analysed responses from 740 subjects who filled in the questionnaires pre and post hearing aid fitting. Results show a moderate to strong correspondence between COSI scores for overall degree of change and final ability. Most COSI responses are at or near the maximum possible score and show slight differences in overall scores considering the effect of hearing aid experience or hearing loss. AVAB results reveal a more refined evaluation of the hearing aid fitting. Combining the advantages of both methods results in a profound evaluation of hearing aid rehabilitation. Our results suggest that both methods should be used complementary, rather than separately.

INTRODUCTION

A patient's personal experience and judgement are known to be essential factors in the rehabilitation with hearing aids. Self-reporting questionnaires are by design very suitable methods to collect and assess such information. The Amsterdam Inventory for Auditory Disability and Handicap (AIADH), developed by Kramer *et al.* (1995) is an example of a questionnaire to assess hearing disabilities in daily life with a high reliability and validity (Meijer *et al.*, 2003). In this study we used a slightly adapted version of the AIADH, called AVAB (in Dutch: Amsterdam Questionnaire for Auditory Disabilities), resulting into a six dimensional profile: detection of sounds, speech in quiet, speech in noise, auditory localization, sound discrimination, and noise tolerance (Dreschler and de Ronde-Brons, 2016). Not only could the characteristics of the AVAB be advantageous in selecting and fitting a hearing aid according to the specific needs of a patient, it might also be an adequate tool for evaluating the benefit of a hearing aid with respect to different aspects of auditory functioning (see also Fuente *et al.*, 2012).

*Corresponding author: w.a.dreschler@amc.uva.nl

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

AVAB is a questionnaire limited to a fixed list of general listening conditions, which are not necessarily applicable for each patient. This could be considered an important drawback. Alternatively, Dillon *et al*. (1997) introduced the Client Oriented Scale of Improvement (COSI) for the evaluation of hearing aids, which makes use of personally defined targets for rehabilitation. This makes COSI very useful for individual patients, but complicates the comparison of needs or benefits for groups of patients. To overcome the problem of low comparability between individual targets, Dillon *et al*. (1997) proposed to categorize each target into a total of sixteen pre-designated categories. Zelski (2000) showed a high level of inter-observer agreement in assigning COSI targets to those categories, but concluded that the amount of categories could be reduced. It has been shown by Dreschler and de Ronde-Brons (2016) that individual COSI targets can be categorized to match the same six dimensions as the AVAB auditory disability profile. This opens the possibility to compare individual hearing disabilities (AVAB) and individual compensation targets (COSI) along the same dimensions and to combine AVAB and COSI results for each individual.

The goal of this study was to examine the applicability and added value of the combined use of AVAB and COSI in relation to the evaluation of a hearing aid fitting. The analyses primarily address the correspondence between the AVAB and COSI results, and the effects of hearing loss and level of experience on these results.

METHODS

Over a period of 10 months data were collected from various hearing aid dispensers that took part in a study which explored the advantages of self-report questionnaires in the hearing aid rehabilitation process. Auditory disability, before and after the hearing aid fitting, was assessed by the AVAB method. In addition, the COSI method was implemented to define individual targets and to measure the degree of change due to the hearing aid fit and the final ability afterwards with respect to the individual targets.

Prior to the hearing aid selection and fitting process, pre-AVAB questionnaires were administered to the subjects, followed by a question to describe a maximum of 5 situations in which they experience hearing difficulties. These situations formed the basis for formulating the COSI in dialogue with the hearing aid dispenser. The dispenser assigned matching AVAB dimensions to each COSI target (multiple dimensions per target were possible). Additionally, pure tone audiometry and speech audiometry were deemed mandatory aspects for the selection of a new hearing aid. Once fitted and after a trial period COSI targets were evaluated resulting in scores for degree of change and final ability for each individual target, again in dialogue with the hearing-aid dispenser. Furthermore a post-AVAB questionnaire was administered. Speech intelligibility in quiet, with and without the fitted hearing aid, was assessed as part of the final assessment of the benefit of the fitting. The fitting, trial and evaluation process were similar for first time users and experienced users.

Subjects

A representative sample of both new hearing aid users and experienced hearing aid users who needed replacement of their hearing aid were included from 64 hearing aid dispensers in the Netherlands. Subjects participated voluntarily and were included when agreeing upon usage of their data as they fully completed the hearing aid fitting process including the purchase of the hearing aid. Subjects with a CROS or biCROS-fitting were excluded.

RESULTS

Data from 1223 subjects were collected, but data from 483 subjects were incomplete. A number of 740 subjects fulfilled the criteria of inclusion and their data were used for further analysis. The median of the trial period after the hearing aid fitting was 47 days. Of the total group 58% was male and 42% female and about half (54%) of them were first time hearing aid users. The median age of the total group was 72 years. Pure tone threshold averages for 0.5, 1, 2 and 4 kHz were calculated for the better ear (PTAB) of all subjects, which showed a median hearing loss of 44 dB HL. Pure tone frequencies that exceeded the maximum output of the audiometer were denoted 125 dB HL. The median difference between PTAB and the pure tone average at the other ear was 5 dB HL, indicating that by far most of the subjects had a symmetrical hearing loss. As a consequence, 90% of the fittings were bilateral. COSI responses showed that on average 3.8 fitting targets were formulated per individual. These fitting targets were attributed to AVAB dimensions by the hearing aid dispenser, who indicated on average 1.6 matching dimensions per fitting target.

The overall scores for AVAB and COSI

Overall AVAB and COSI scores (i.e., the mean of the scores per dimension for each subject, not the mean of all individual items) were analysed by examining the cumulative distributions of the reported scores. These cumulative plots show the percentage of subjects whose COSI or AVAB score had a value less than or equal to the score indicated on the x-axis. Figure 1a shows that both methods reveal large benefits of hearing aid fitting. In the COSI results there is a strong visual correspondence between the distribution of COSI scores for overall degree of change and final ability. These results are in line with the findings previously described by Dillon *et al.* (1999), which have also been plotted in Fig. 1a. The extent of similarity in our data is emphasized by a moderate to strong correlation (Spearman's rho = 0.69), confirming the close relation between the two reported scores, not only on a group level but also on an individual basis.

Further analyses reveal ceiling effects, most pronounced in the COSI scores (see Fig. 1b). In fact, over 87% of all subjects reported mean scores equal or greater than 4, and 32% even reported the maximum score on all given targets.

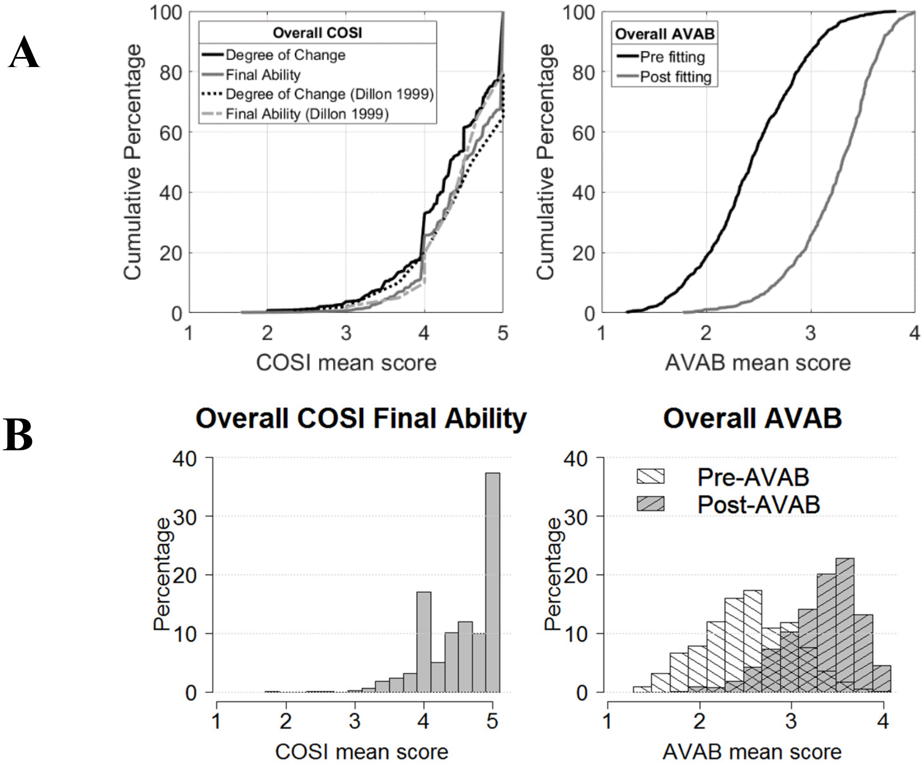


Fig. 1: (A) Cumulative distributions of overall mean COSI results (left), degree of change (black) and final ability (grey), dotted and striped lines show results found by Dillon *et al.* (1999). AVAB results (right) show overall mean pre-fitting results (black) and post-fitting results (grey). (B) Histogram of overall COSI Final Ability scores (left), and pre- and post- AVAB scores (right).

Effects of hearing aid experience

Effects of hearing aid experience of overall COSI final ability and overall pre- and post-fitting AVAB results are shown in Fig. 2. The responses of both COSI and AVAB were divided between first time users (54%) and experienced users (46%). COSI shows a slight difference between first time users and experienced users, while both pre- and post-fitting AVAB results show apparent differences.

Effects of degree of hearing loss

To analyse the effects of hearing loss on overall COSI and AVAB results, subgroups were composed based on pure tone average of the better ear (PTAB): ≤ 35 dB HL, 36-45 dB HL, 46-55 dB HL, and > 55 dB HL. The cumulative distributions (Fig. 3) of the COSI results show a strong visual correspondence between subgroups, except for the group comprised of the largest hearing losses. In contrast, overall AVAB results pre- and post-fitting are well distinguishable and show higher average scores for subjects with less severe hearing losses at the better ear.

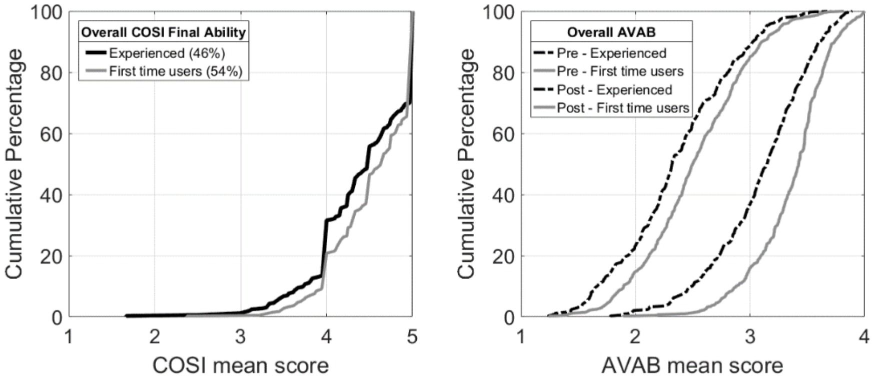


Fig. 2: overall COSI final ability (left) and pre/post-AVAB (right) cumulative distributions for first time users and experienced users.

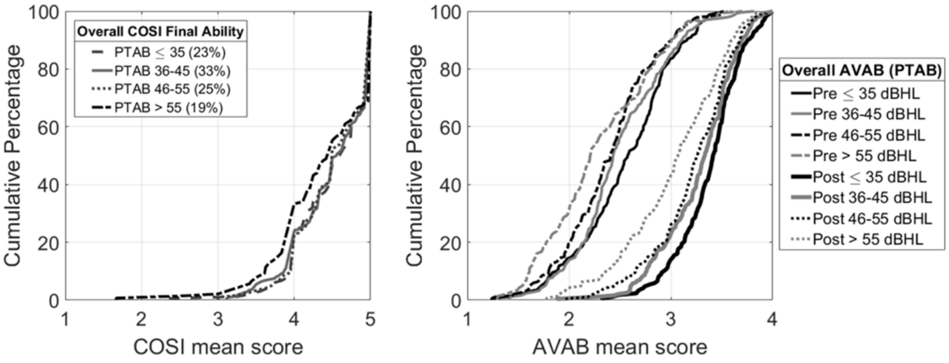


Fig. 3: Overall COSI final ability (left) and pre/post-AVAB (right) cumulative distributions for different groups of PTAB.

Effects for different dimensions of auditory functioning

Individual COSI targets can be summarized and categorized according to the six auditory disability dimensions resulting from the AVAB questionnaire. This results in specific distributions among the six dimensions. The largest contribution of matched COSI targets was to the dimension *speech in noise* (98% of the subjects had at least one target for this dimension), followed by *speech in quiet* (75%), *discrimination* (41%), *detection* (37%), *localization* (37%), and lastly *noise tolerance* (23%). A total of 2844 COSI fitting targets was formulated.

Figure 4 shows boxplots for pre- and post-fitting AVAB scores and COSI final ability scores in all six dimensions. It should be noted that these results comprise different numbers of responses between COSI and AVAB per dimension, which prevent a direct comparison. The median post-fitting AVAB scores were found to be higher relative to pre-fitting AVAB scores in all six dimensions. More specifically, pre- and post-fitting AVAB scores shows clear differences in the degree of benefit among each of the six dimensions. The *speech in noise* dimension showed the largest difference between pre- and post-fitting AVAB score, the smallest effect is denoted by the *tolerance* dimension. Differences between dimensions were less pronounced in the average COSI final ability results.

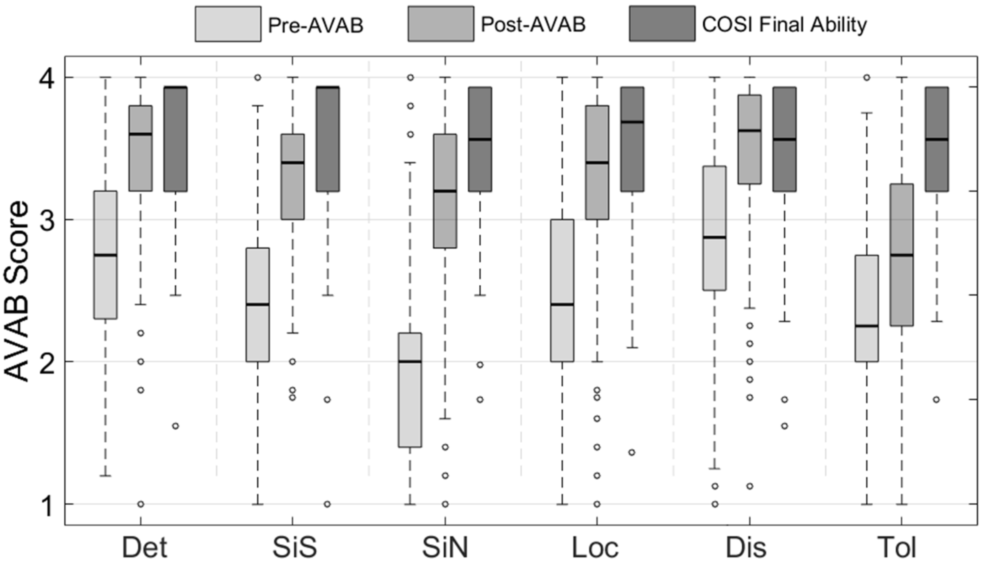


Fig. 4: Boxplots of COSI and pre/post AVAB scores per auditory disability dimension: Det=Detection; SiS=Speech in silence; SiN=Speech in noise; Loc=Localization; Dis=Discrimination; Tol=Noise tolerance.

DISCUSSION

Our study focused on the combination of two self-report questionnaires (AVAB and COSI) for the selection and evaluation of hearing aids. In a representative population of hearing aid users, both AVAB and COSI show a beneficial effect of fitting new hearing aids for six dimensions of auditory functioning. AVAB scores show more differentiation than COSI scores between user types (first time user or experienced user), degrees of hearing loss, and between the six dimensions.

The current study indicates that the two outcome measures resulting from COSI (degree of change and final ability scores) are closely related. Both measures show similar overall cumulative distributions, as well as a moderate to strong correlation between individual scores. These results match those found by Dillon *et al.* (1999) and suggest that there is no apparent distinction between the two measures. Therefore, it could be argued that merely evaluating final ability could be sufficient to assess individual COSI targets. COSI scores had a skewed distribution, with a tendency towards maximum scores. A possible explanation for the ceiling effect in the COSI scores might be a biased judgement by the audiologist/dispenser. On the other hand, Dillon *et al.* (1999) reported very similar results concerning the observed ceiling on the COSI results and argue that there may be a tendency for individuals to exaggerate their level of satisfaction. COSI targets are central to the rehabilitation process and efforts will be made to achieve maximum results on each of these targets, which implies considerable attention from the dispenser for the subject's COSI targets. This is not necessarily the case for conventional questionnaires such as AVAB of which not all items are equally relevant or even applicable to the subjects rehabilitation needs. In other words, greater attention to the COSI targets might contribute to the ceiling effect in final ability scores.

Although AVAB post scores also show a skewed distribution (Fig. 1), AVAB scores vary more between subjects than COSI scores. As a result, AVAB scores show differences between groups of user types (first time or experienced users) and degrees of hearing loss, which were not matched by COSI scores (Figs. 2 and 3). Furthermore, AVAB scores differ more between the six dimensions than COSI scores (Fig. 4). One reason for this might be that within the AVAB questionnaire all subjects had answered questions about all six dimensions, whereas COSI included only a limited range of situations. Assignment of these situations to the six dimensions is subjective and might differ between dispensers, although previous results show high inter-observer agreement. Also, multiple dimensions could be assigned to one target, resulting in the same score for different dimensions for one COSI target. This reduces the ability to discriminate between dimensions in final COSI scores.

Due to the low variability in scores, COSI in its current form appears to have limited value for evaluating effects of hearing aid fitting between different groups of users. AVAB, on the other hand, seems to be a useful outcome measure for such analyses. However, for counseling purposes COSI forms a useful addition to the AVAB questionnaire in that it provides concrete targets for an individual hearing aid fitting. Assigning the COSI targets to the six AVAB dimensions, supports the interpretation and weighting of the AVAB results for an individual, and the translation into important hearing aid functions and settings. On the other hand, the AVAB has added value in combination with the COSI in that it always provides results for all six defined dimensions of auditory functioning and therefore provides a broader overview of the fitting results. In addition, by first completing the AVAB questionnaire pre fitting, subjects are encouraged to think about their hearing ability in a broad range of situations before they define their individual need for rehabilitation.

In conclusion, both COSI and AVAB are very suitable in the evaluation of hearing aid rehabilitation, each method having specific strengths and weaknesses. AVAB contributes to the formulation of individual needs of rehabilitation used by COSI and provides detailed information on pre- and post-fitting evaluation. COSI is a very strong tool for the assessment of individual rehabilitation needs but is less sensitive for comparison between groups due to responses at or near the top of the response scale. AVAB on the other hand seems to be a useful tool for such comparisons and provides a broader insight in the auditory functioning of individuals. These differences between COSI and AVAB point to the fact that both methods should be used complementary, rather than separately.

ACKNOWLEDGEMENTS

Data collection was organized by the PACT Foundation, a network of co-operating Audiological Centers in the Netherlands. The authors like to thank Bert van Zanten (Academic Medical Center Utrecht) and André Goedegebure (Erasmus Medical Center in Rotterdam) for their efforts in data collection.

REFERENCES

Dillon, H., James, A., and Ginis, J. (**1997**). "Client Oriented Scale of Improvement (COSI) and its relationship to several other measures of benefit and satisfaction provided by hearing aids," J. Am. Acad. Audiol., **8**, 27-43.

Dillon, H., Birtles, G., and Lovegrove, R. (**1999**). "Measuring the outcome of a national rehabilitation program: Normative data for the Client Orientated Scale of Improvement (COSI) and the Hearing Aid User's Questionnaire (HAUQ)," J. Am. Acad. Audiol., **10**, 67-79.

Dreschler, W.A., de Ronde-Brons, I. (**2016**). "A profiling system for the assessment of individual needs for rehabilitation with hearing aids," Trends Hear., **20**, doi: 10.1177/2331216516673639

Fuente, A., McPherson, B., Kramer, S.E., Hormazábal, X., and Hickson, L. (**2012**). "Adaptation of the Amsterdam Inventory for Auditory Disability and Handicap into Spanish," Disabil. Rehabil., **34**, 2076-2084. doi: 10.3109/09638288.2012.671884

Kramer, S.E., Kapteyn, T.S., Festen, J.M., and Tobi, H. (**1995**). "Factors in subjective hearing disability," Int. J. Audiol., **34**, 311-320. doi: 10.3109/00206099509071921

Meijer, A.G.W., Wit, H.P., Tenvergert, E.M., Albers, F.W.J., and Kobold, J.P.M. **(2003)**. "Reliability and validity of the (modified) Amsterdam Inventory for Auditory Disability and Handicap," Int. J. Audiol., **42**, 220-226. doi: 10.3109/14992020309101317

Zelski, R.F. (**2000**). "Use of the Client Oriented Scale of Improvement as a clinical outcome measure in the Veterans Affairs national hearing aid program," Graduate dissertation.

Clinical measures for investigating hidden hearing loss

PERNILLE HOLTEGAARD[1], JOSEFINE J. JENSEN[2], AND BASTIAN EPP[1]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Department of Nordic Studies and Linguistics, Copenhagen University, Copenhagen, Denmark*

The present study compared clinical measures of auditory function in two listener groups prone to hidden hearing loss relative to a control group: a) listeners with tinnitus, and b) listeners with a history of noise-exposure. Auditory brainstem response (ABR) wave I, III and V were measured in response to a 4-kHz tone burst to quantify the level-growth of wave I and the amplitude difference between waves I-III and I-V. In addition, speech-in-noise performance using “Dantale I” and the Danish hearing in noise test (HINT) were assessed. The ABR wave-I level growth showed no difference between the tinnitus-, noise-exposed- and control group. The listeners with tinnitus had, however, significantly larger wave I-III differences indicating a gain at brainstem level. While the ABR results support that the wave I-III difference can be used as a physiological indicator of tinnitus, none of the applied audiological methods show signs of a noise-induced hidden hearing loss in the tested listener groups.

INTRODUCTION

It has been a common assumption that temporary threshold shifts (TTS) following a noise-exposure were not hazardous as the observed pathology and shift in threshold were, as the name suggests, temporary. It was further assumed that hazardous sound levels, causing permanent deficits, primarily targets and damages outer hair cells (OHC) (Puel *et al.*,1988; Lawner *et al.*, 1997). Such damage causes reduced sensitivity to soft sounds which can be assessed with standard pure-tone audiometry. Recent animal studies suggest, however, that 40 dB TTS cause immediate permanent damage of the inner hair cell (IHC) synaptic ribbons and afferent type I nerve fibres prior to any involvement of the OHCs (Kujawa and Liberman, 2009). This damage was reflected in the ABR as significantly reduced amplitude of wave I in response to supra-threshold level stimuli, while thresholds measured using ABR wave V normalised. Following this acute synaptic damage, a slowly progressive loss of cell bodies in the spiral ganglion was observed (Kujawa and Liberman, 2009). This suggests that noise-exposure causing TTS can cause immediate synaptic damage and progressive nerve damage (i.e., noise-induced neural degeneration, NIND) without affecting threshold sensitivity to pure-tones. One explanation for the restoration of

*Corresponding author: perholt@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.

ISBN: 978-87-990013-6-1.

thresholds could be the finding that high spontaneous rate fibres (HSRFs) were largely unaffected by noise exposure, whereas the synaptic damage predominantly affected low spontaneous rate fibres (LSRFs) (Furman *et al.*, 2013). LSRFs have higher thresholds and have been suggested to be responsible for the coding of mid- to high-intensity stimuli (Liberman, 1978; Taberner and Liberman, 2005). In addition to coding supra-threshold stimuli, it has also been suggested that LSRFs are critical for processing of auditory stimuli in the presence of high-level background noise (Costalupes *et al.*, 1984). The consequences of NIND are therefore assumed to not be reflected in the audiogram, and have been given the term "hidden hearing loss" (Schaette and McAlpine, 2011). Thus, if acoustic overexposure also causes NIND of the LSRFs in humans, this may help to explain auditory disorders defined as difficulties processing speech in challenging listening environments, despite normal pure-tone thresholds (Zhao and Stephens, 1996). NIND has also been suggested to be a potential contributor to tinnitus in the absence hearing loss (Schaette and McAlpine, 2011).

So far physiological evidence of NIND has been shown for both mice (Kujawa and Liberman, 2009) and guinea pigs (Furman *et al.,* 2013). In humans, it is not possible to expose listeners to noise in order to investigate its consequences on physiology and auditory perception. Therefore, efforts have been made to investigate deficits in listener groups with a history of noise-exposure using behavioural and physiological measures. The reported results are, however, inconclusive. Significantly lower ABR wave I amplitudes in response to high-level stimuli have been found in listener groups with a self-reported higher exposure history compared to a control group with less reported exposure history (Bramhall *et al.,* 2016; Liberman *et al.,* 2016). Liberman *et al.* (2016) also reported significantly poorer performance on speech recognition in noise in the high-exposure group compared to the control group. In addition, a relationship between ABR wave I amplitude and self-reported noise exposure for female listeners, but not male listeners, has been reported (Stamper and Johnson, 2015). The findings of these studies support the assumption that NIND also exists in human listeners. A large study performed with 129 normal-hearing listeners found, however, no correlation between ABR wave I amplitude and history of noise-exposure, or any correlation between behavioural performance and noise-exposure (Prendergast *et al.*, 2016). Hence, it has still to be revealed if NIND occurs in human listeners, and if NIND can explain auditory deficits such as tinnitus or impaired speech recognition in noise, in the presence of normal threshold sensitivity.

The present study investigates if listeners prone to hidden hearing loss will: a) show a shallower slope in the level-growth of wave I, b) have larger amplitude gap between waves I-III and I-V, c) show poorer speech-performance in noise than the control group, and d) if level-growth is correlated with speech-in-noise performance.

METHOD

Listeners: Two test groups and a corresponding control group were included in the study: a) listeners with tinnitus who reported chronic tinnitus for a minimum of one year (tinnitus group; n = 7, mean age 26.8 ± 1.9 years), and b) listeners with a self-reported history of over-exposure working in loud sound environments (professional musicians) for at least 5 hours a day, 5 days a week for at least 1 year (exposure group; n = 9, mean age 25.6 ± 4.1 years). A loud sound environment was defined as an environment in which one would need to raise his or her voice in order to communicate. All listeners across the groups were young normal hearing (pure-tone thresholds ≤ 15 dB HL between 0.25 – 8 kHz) listeners between the ages of 18-35 years. The control group consisted of 9 listeners (mean age = 25.11 years ± 4.7). All participants provided informed consent and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

ABR: ABRs were recorded using the Interacoustics Eclipse ABR system (EP15/EP25). Disposable non-invasive inverting electrodes were attached to the mastoids, a non-inverting electrode was placed on the middle of the forehead just below the hairline, and the ground electrode was placed below the non-inverting electrode. An impedance of < 3 kΩ was ensured before initiating the measurement. Listeners were instructed to relax and preferably sleep while lying in an electrically shielded sound proof booth. 4-kHz tone burst stimuli of 1.25 ms, using Blackman window, were presented monaurally using ER2 insert earphones at peak-equivalent sound pressure levels (peSPL) of 97, 102 and 107 dB peSPL. The stimuli were presented with alternating polarity at a rate of 11.1/s and the ABR waveforms were recorded from −5 to 11 ms. Each measurement continued until a residual noise level of ≤ 30 nV was obtained or a maximum of 4000 sweeps were recorded. ABR wave I-V peak-to-trough amplitudes were selected manually.

Discrimination score (DS): DS was measured using the "Dantale I" material comprising wordlists consisting of 25 single monosyllabic words in speech-shaped noise (Elberling *et al.*, 1989). For each ear 3 lists were presented at 3 different SNR levels (10, 5 and 0 dB SNR). The speech level was kept at a constant level of 70 dB, while the noise was started at 60 dB and increased in steps of 5 dB for each list. Scoring was kept in percentage correctly repeated words of the 25 presented words.

Hearing in noise test (HINT): The Danish HINT sentences (Nielsen and Dau, 2009) were presented monaurally in speech shaped background noise of 70 dB SPL. The stimuli were generated in MATLAB (The Mathworks, MA, USA) and presented over headphones (Sennheiser HDA200). Lists of 20 sentences were presented, and the listener was asked to repeat the sentences after best ability. The speech level was adaptively adjusted dependent on the response of the listener. The test was always started with an SNR of 0, with an initial speech level of 70 dB SPL. Before each test, a list of 20 training sentences was completed to familiarize the listener with the test and to eliminate training effects.

Speech recognition in noise was measured for the control and exposure group. However, not all listeners completed the speech task. For DS all the exposure group listeners, but only 5 out of 9 control listeners completed the task (n = 14). For HINT data was measured from 8 of the listeners in the exposure group, but only 2 in the control group (n = 10).

Statistical evaluation: Statistics were calculated using Mann-Whitney U, and linear regression was calculated using the statistical software R (R Core Team, Vienna, Austria).

RESULTS

Figure 1 depicts level-growth (amplitude of ABR wave I as a function of level increase) for the groups for right (left panel) and left (right panel) ear separately. The average level-growth from 97 to 107 dB peSPL for the right ear was 0.108 µV for the control group, 0.104 µV for the exposure group and 0.092 for the tinnitus group. Statistical analysis showed no significant differences of level-growth across listener groups. Significantly lower level-growth was, however observed for the left ear of the exposure group between 102-107 dB peSPL (U = 12, p < 0.01, one-tailed) compared to the control group. The tinnitus group had significantly steeper level-growth from 97-102 dB peSPL (U = 17, p < 0.05, two-tailed) compared to the control group.

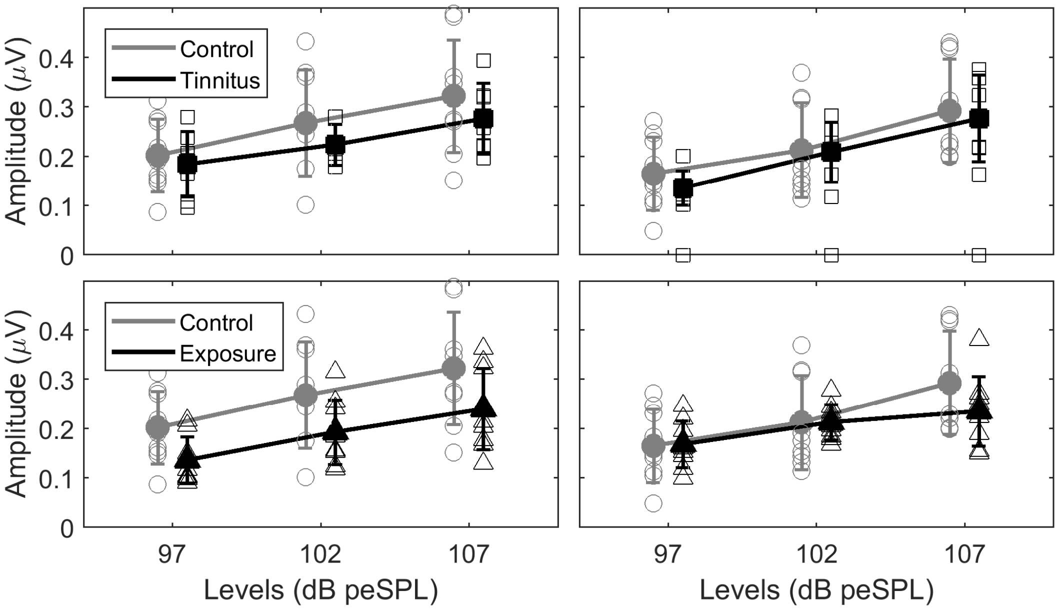


Fig. 1: Wave I amplitude measured at 97, 102 and 107 dB peSPL. Results from the right ear are shown in the left panel and results from left ear are shown in the right panel. The upper panels show results from the control (grey circles) and tinnitus group (black squares). The bottom panels depict the exposure (black triangles) and the control group (grey circles). Mean values across groups are shown as solid markers and the individual data are shown as open markers.

Figure 2 shows the results of the discrimination score (DS) as a function of the level growth. No significant difference was observed between the DS of the two groups for the right ear values ($U = 33$, $p > 0.05$), but a significant difference was observed for the left ear ($U = 20.5$, $p < 0.05$, one-tailed). The Pearson correlation coefficient also showed a significant positive correlation between level-growth and DS with 0 dB SNR (the most challenging SNR) on the left ear ($r = 0.53$, $p < 0.025$, one-tailed). However, there was no correlation between these two variables for the right ear ($r = 0.25$, $p > 0.05$). Measures of the Danish HINT did not show significant differences across groups, and no significant correlations between HINT and level-growth were found on either right ($r = 0.25$, $p < 0.05$, one-tailed) or left ($r = 0.15$, $p < 0.05$, one-tailed) ear of the listeners.

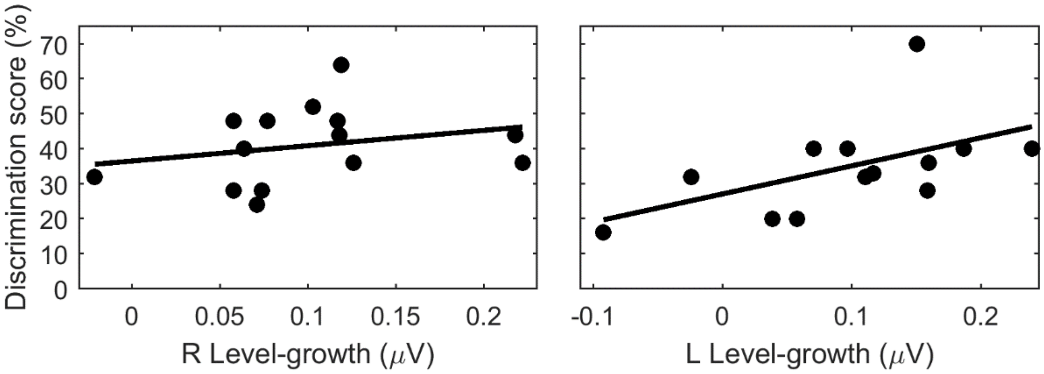


Fig. 2: DS measured with an SNR of 0 dB SNR as a function of ABR wave I level-growth for right ear (left graph) and left ear (right graph).

Figure 3 shows the difference in amplitude between ABR waves I-III between the control and tinnitus group (two upper panels) and the control and exposure group (bottom two panels) measured at a level of 107 dB peSPL. Significantly larger amplitude difference between waves I-III was observed for the listener group with tinnitus compared to the control group for both right ($U = 15$, $p < 0.025$, one-tailed) and left ear ($U = 12$, $p < 0.01$, one-tailed). This significant difference was not observed between the control and exposure group either for the right ($U = 35$, $p > 0.05$) or left ear ($U = 32.5$, $p > 0.05$).

No significant difference between ABR waves I-V was observed between the exposure and control group for right ($U = 33$, $p > 0.05$) or left ear ($U = 37$, $p > 0.05$) nor for the tinnitus and control group right ($U = 28.5$, $p > 0.05$) or left ear ($U = 22$, $p > 0.05$, one-tailed).

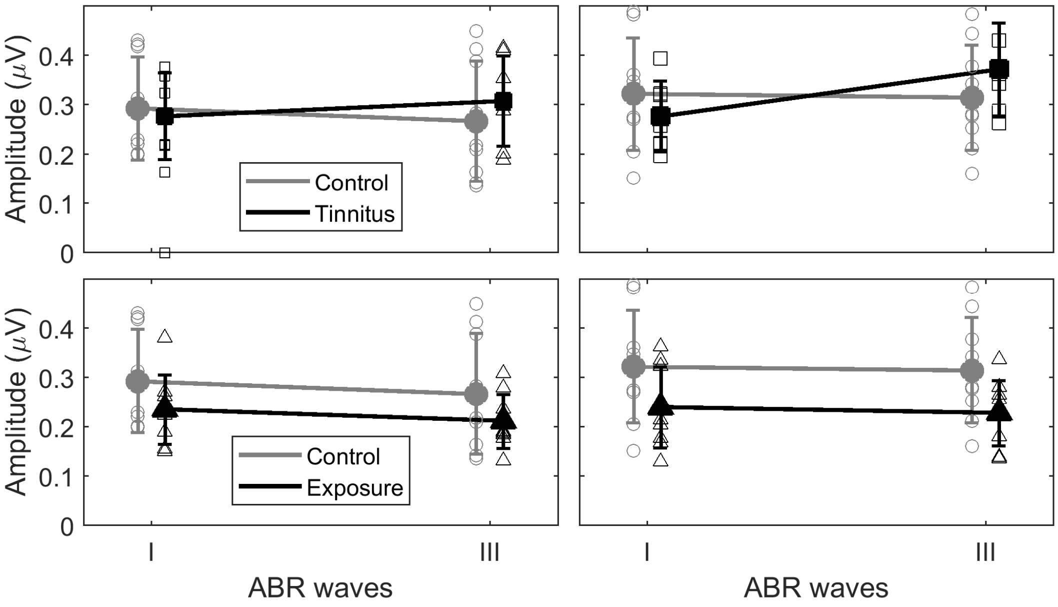


Fig. 3: Amplitude difference between ABR waves I-III for right (left panels) and left ear (right panels). The upper graphs show results from the control (grey circles) and tinnitus group (black squares), while the bottom graphs depict the exposure (black triangles) and the control group (grey circles). Mean values across groups are shown as solid markers and the individual data are shown as open markers.

DISCUSSION

The hypothesis of the current study was that if NIND occurs in human listeners (Kujawa and Liberman, 2009), ABR wave I level-growth will be significantly lower for a group with a history of working in higher-level sound-exposure or with chronic tinnitus with a normal audiogram. This hypothesis was not supported by the data.

Significantly lower level-growth was only found for the exposure group on the left ear between 97 and 102 dB peSPL ($U = 12$, $p < 0.01$, one-tailed). Despite of this a significant correlation ($r = 0.53$, $p < 0.025$, one-tailed) between level-growth and DS was observed for the left ear of listeners in the control and exposure group. Left ear DS was in fact also significantly poorer in the exposure group compared to the control group ($U = 20.5$, $p < 0.05$, one-tailed). These data and the fact that DS was not correlated with age ($r = 0.03$, $p > 0.05$) could thus support a potential relationship between the pathology of noise-induced synaptic and neural degeneration and auditory disorders despite normal audiogram. However, level-growth was also significantly correlated with age for both right ($r = 0.45$, $p < 0.05$, one-tailed) and left ear ($r = 0.49$, $p < 0.025$, one-tailed). Multiple regression was performed to take the age variable into account. This reinforced the significant correlation between DS and level-growth ($r = 0.61$, $p < 0.025$, one-tailed). For the right ear this relationship can be rejected as level-growth and DS did not correlate significantly ($r = 0.25$, $p > 0.05$). The lack of significant difference on the HINT scores of the groups and correlation with level-growth suggests that NIND does not occur in humans. However, since the HINT material consists of natural sentences, it cannot be ruled out that cognitive abilities could have affected the results.

The hypothesis that the two groups assumed prone to NIND show larger amplitude differences between waves I-III was not confirmed for the exposure group relative to the control group. A significant difference between the tinnitus and control group was however confirmed with significantly larger amplitude difference between waves I-III for both right ($U = 15$, $p < 0.025$, one-tailed) and left ear ($U = 12$, $p < 0.01$, one-tailed) for the tinnitus group compared to the control group. This is in agreement with previous literature suggesting a relationship between the diagnosis of tinnitus, despite normal threshold sensitivity, and neural gain at brainstem level (Hickox and Liberman, 2014; Knipper *et al.*, 2013; Schaette and McAlpine, 2011).

Despite the possibility that NIND is absent in the tested listeners, one might speculate that the results of the present study can be explained by individual susceptibility to noise and developing hearing impairment or NIND. In such a case, the grouping variable 'noise exposure' might reduce the significance of the results.

CONCLUSION

The current study did not find evidence of reduced ABR wave I level-growth in groups assumed prone to NIND using clinical measures. The findings of this study cannot confirm the presence of NIND in human listeners. The significant correlation between level-growth and DS on the left ear, however support the assumption that poorer speech recognition in noise, despite normal audiometric thresholds, can be attributed to loss of LSRFs. Furthermore, the findings of increased amplitude difference between waves I-III in the listeners with tinnitus support evidence of a relation between hyperactivity in the brainstem and tinnitus. Overall, the results of the clinical measures used in the current study either suggest that NIND is not present in the tested listeners or that these measures are not sensitive enough to reveal a clear connection between noise exposure and NIND in human listeners.

ACKNOWLEDGEMENTS

This work is supported by the Oticon Foundation.

REFERENCES

Bramhall, N.F., Konrad-Martin, D., McMillan, G.P., and Griest, S.E. (**2017**). "Auditory brainstem response altered in humans with noise exposure despite normal outer hair cell function," Ear Hearing, **38**, E1-E12. doi: 10.1097/AUD.0000000000000370.

Costalupes, J.A., Young, E.D., and Gibson, D.J. (**1984**). "Effects of continuous noise backgrounds on rate response auditory nerve fibers in cat," J. Neurophysiol., **51**, 1326-1344.

Elberling, C., Ludvigsen, C., and Lyregaard, P.E. (**1989**). “DANTALE – a new Danish speech material,” Scand. Audiol., **18**, 169-175. doi:10.3109/01050398909070742.

Furman, A.C., Kujawa, S.G., and Liberman M.C. (**2013**). “Noise induced cochlear neuropathy is selective for fibers with low spontaneous rates,” J. Neurophysiol., **110**, 577-586.

Hickox, A.E., and Liberman, M.C. (**2014**). “Is noise-induced cochlear neuropathy key to the generation of hyperacusis or tinnitus?” J. Neurophysiol., **111**, 552-564.

Knipper, M., Dijk, P.V., Nunes, I., Rüttiger, L., and Zimmermann, U. (**2013**). “Advances in the neurobiology of hearing disorder: Recent developments regarding the basis of tinnitus and hyperacusis,” Prog. Neurobiol., **111**, 17-33.

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.

Lawner, B.E., Harding, G.W., and Bohne, B.A. (**1997**). “Time course of nerve-fiber regeneration in the noise damaged mammalian cochlea,” Int. J. Dev. Neurosci., **15**, 601-617.

Liberman, M.C. (**1978**). “Auditory-nerve response from cats raised in a low-noise chamber,” J. Acoust. Soc. Am., **63**, 442-455.

Liberman, M.C., Epstein, M.J., Cleveland, S.S., Wang, H., and Maison, S.F. (**2016**). “Toward a differential diagnosis of hidden hearing loss in humans,” PLoS ONE, **11**, e0162726. doi: 10.1371/journal.pone.0162726.

Nielsen, J.B., and Dau, T. (**2009**). “Development of a Danish speech intelligibility test,” Int. J. Audiol., **48**, 729-741. doi: 10.1080/14992020903019312

Prendergast, G., Guest, H., Munro, K.J., Kluk, K., Leger, A., Hall, D.A., Heinz, G.A., and Plack, C.J. (**2017**). “Effects of noise exposure on young adults with normal audiograms I: Electrophysiology,” Hear. Res., **344**, 68-81. doi: 10.1016/j.heares.2016.10.028

Puel, J.L., Bobbin, R.P., and Fallon, M. (**1988**). “The active process is affected first by intense sound exposure,” Hear. Res., **37**, 53-64.

Schaette, R., and McAlpine, D. **(2011**). “Tinnitus with a normal audiogram: Physiological evidence for hidden hearing loss and computational model,” Eur. J. Neurosci., **23**, 3124-3138.

Stamper, G.C., and Johnson, T.A. (**2015**). “Letter to the Editor: Examination of potential sex influences in Stamper, G.C & Johnson, T.A. (2015). Auditory function in normal-hearing, noise-exposed human ears, Ear Hearing, 36, 172-184,” Ear Hearing, **36**, 738-740. doi: 10.1097/AUD.0000000000000228

Taberner, A.M., and Liberman, M.C. (**2005**). “Response properties of single auditory nerve fibers in the mouse,” J. Neurophysiol., **93**, 557-569.

Zhao, F., and Stephens, D. (**1996**). “Hearing complaints of patients with King-Kopetzky Syndrome (obscure auditory dysfunction),” Br. J. Audiol., **30**, 397-402.

The scale illusion detection task: Objective assessment of binaural fusion in normal-hearing listeners

NICLAS A. JANßEN[1,*], LARS BRAMSLØW[2], SØREN RIIS[3], AND JEREMY MAROZEAU[1]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Eriksholm Research Centre, Snekkersten, Denmark*

[3] *Oticon Medical, Smørum, Denmark*

The normal auditory system can fuse sounds from both ears into a single sound object (binaural fusion). This ability can be assessed subjectively by asking whether listeners perceive one or two sounds or by the scale illusion percept described by Deutsch (1975). The aim of the current study is to develop an objective task to measure binaural fusion. Twelve normal-hearing participants had to detect one deviant note within a stream composed of a repeating melody while simultaneously being presented with another stream of randomized notes. The experiment included 3 conditions. First, in a monaural condition both streams were presented to the same ear. Then, in a binaural condition every second note from each of the two streams was presented to the other ear. Finally, in a binaural control condition, the timbre of all the notes presented to one ear was altered severely, to prevent binaural fusion. The expected result was a better detection of deviant notes for listeners that are able to fuse streams across the two ears. Each condition had 24 repetitions. In the binaural and monaural conditions, average performance was about 80% correct, while the control condition showed a significantly lower performance of about 50%. Thus, this type of experiment can be used to test objectively if fusion takes place. It lays the foundation for further studies with bilateral and bimodal cochlear implant listeners.

INTRODUCTION

While listening with both ears, humans fuse sounds binaurally into a common percept. Consequently, the voice of one speaker is perceived as one sound object, rather than two separate voices and information from both ears can be utilized to localize sounds or achieve superior speech perception in noise (cf. Middlebrooks *et al.*, 2017).

This ability to fuse binaurally presented sounds into one percept has been demonstrated elegantly by the scale illusion percept described by Deutsch (1975). It is based on a complex stimulation pattern, which consists of two melodies at two different frequency ranges (one high and one low). Both melodies go up and down in frequency over eight notes (see Fig. 1). These are presented in such a way that every

*Corresponding author: najan@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

second note from each stream is played to the other ear. Yet, listeners most often perceive the two ordered melodies, each of them lateralized to one side.

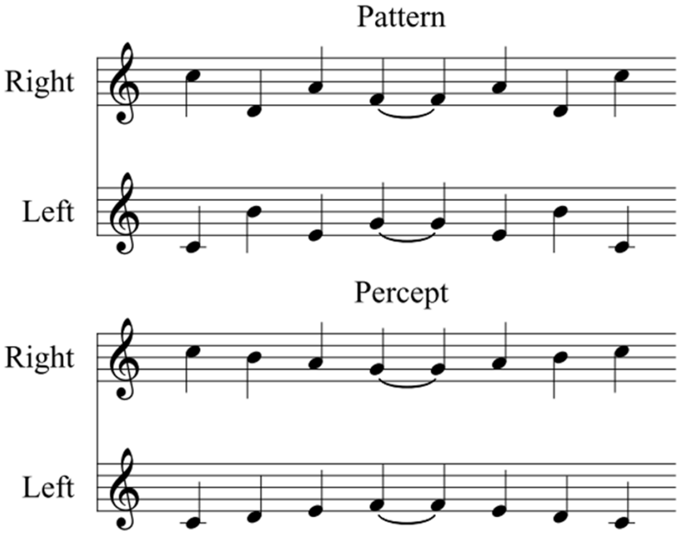


Fig. 1: Melodic pattern and percept for the original scale illusion (Deutsch, 1975).

These results clearly illustrate the ability to fuse sounds binaurally, as there is no other explanation for the reported percepts. However, the scale illusion experiment was based on a subjective task. Participants had to describe and draw their percepts. Such descriptions are open to interpretation by the experimenter and carry the risk of a potential biases. The aim of the current study is to find a fast and objective measure for binaural fusion by using a forced-choice detection paradigm.

PARTICIPANTS

The new experiment has been evaluated with 12 normal-hearing listeners. They have been recruited from the students and staff at the Technical University of Denmark. Their age ranges from 24 to 31 years with a mean of about 26.95 and a standard deviation of about 1.73 years. Of the participants, 25.0% were female, 58.3% reported musical experience such as regular singing or playing an instrument and 75.0% were right-handed.

All participants provided informed consent and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

STIMULI

The experiment was based on the scale illusion percept, featuring notes from the same set of frequencies. Unless denoted otherwise, these notes were presented as pure tones using the corresponding frequencies from the western tuning of the twelve-tone equal temperament scale. The duration of the notes was 250 ms, as in the original study. Further, half Hann-window ramps of 10 ms were applied to onset and offset of the notes to prevent spectral splatter. All the stimuli were loudness balanced using an adjustment procedure. This balancing has been performed twice for all frequencies used in the experiment, based on an external reference sound with preferred most comfortable loudness level chosen by the participant.

Stimuli were presented in a double-walled sound isolated listening booth via Sennheiser HDA-200 headphones.

METHOD

Like in the original scale illusion experiment by Deutsch (1975), our new experiment used two melodies. We will refer to these two as the target stream and the distractor stream (or melody). The target stream was meant to be followed by the listener. It was chosen to be identical to one of the eight-note melodies in the scale illusion, either of the higher or lower frequency range. The distractor stream then consists of eight notes with frequencies chosen randomly from the other frequency range (i.e., if the target stream was of the higher frequency range, the distractor was of the lower range and vice versa). Both streams together were presented in the same way as in the original illusion, i.e., every second note from each stream is presented to the other ear in a pattern symmetrical to the middle of the eight notes sequence (see Fig. 2).

If a listener could fuse, he or she would be able to follow the target melody stream described above, while perceiving the distractor stream on the other side. If the listener was not able to fuse, he or she would perceive random melodies on both sides.

The task of the experiment was to detect a deviant note introduced into the target stream. A listener who fuses the binaural input into one stream was expected to be able to detect this deviant easily, whereas somebody who does not fuse the binaural input will fail to detect the change due to the random input in both ears.

First in each trial, the participant was presented with the target stream alone twice, to signal which stream to listen for. After that, target and distractor streams were presented six times. Of these six repetitions, the first three serve the purpose to allow for a build-up of streaming, as it has been reported that streamed percepts arise gradually over several seconds (Bregman, 1994, cf. Fig. 3). One of the last three repetitions contains the deviant note, as depicted exemplarily in interval B of Fig. 2. The currently presented interval was indicated on the graphical user interface. This arrangement thus represented a 3-alternative forced-choice paradigm. This prevents a bias in answers, compared to a yes/no task, since participants know that the deviant has occurred in one of the intervals (Wickens, 2002).

The deviant could occur at any of the inner 6 notes within a repetition of the melody, but not the first or last position, to prevent confusion of the interval. Its occurrence was counterbalanced with respect to the side of presentation, interval, higher or lower melody as the target stream and the order of trials has been randomized per participant to avoid order effects.

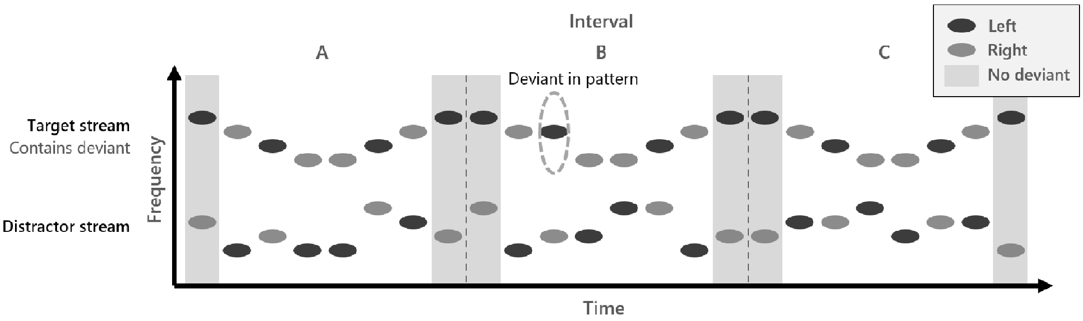


Fig. 2: Pattern of the scale illusion detection task, based on Deutsch (1975). A random deviant occurs in the ordered stream. The other stream consists of random sounds. There is always one higher and one lower stream and half of each stream is presented to the other ear.

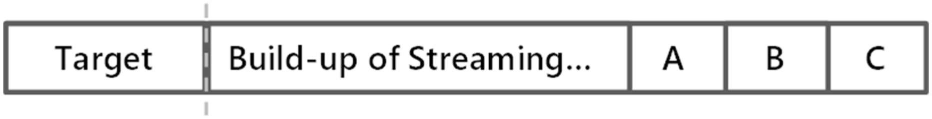


Fig. 3: Structure of a single trial. The Target consists of two repetitions of the target stream alone (either the low or the high stream of notes). After that, the randomized stream joins it. The streaming can build up over 3 iterations of the basic sequence, before finally, the deviant occurs in one of the three last repetitions, A, B, or C.

Before conducting the test, participants underwent a training session. This training session was almost equal to the experiment itself, with the difference that the participants were given feedback whether they answered correctly. It was furthermore possible to repeat the stimuli.

The experiment featured further three conditions: a test condition for binaural fusion and two control conditions to verify that the task was indeed performed by fusing the target melody from the binaural input. Each of the three conditions was repeated 24 times.

Binaural test condition

In the binaural test condition, both the target and the distractor stream were presented in the same fashion as the original scale illusion, i.e., every second note from each stream was presented to the other ear. This condition thus requires binaural fusion to perceive the two streams.

Monaural control condition

In the monaural control condition, the target and the distractor streams were presented to the same ear. In this condition, listeners should be able to segregate the two streams based on their frequency range, even if they could not perform this segregation binaurally.

Binaural control condition

In the binaural control condition, target and distractor streams were again presented binaurally, as in the binaural test condition – but binaural fusion was prevented by an altered timbre for all sounds in one ear.

In addition to pitch, also timbre represents a grouping cue (cf. Bregman, 1994; Deutsch, 1999). Here, it was altered severely by manipulating the temporal envelope and harmonics: Compared to the normal presentation, the duration of the notes itself have been shortened to 200 ms, while keeping the overall interval at 250 ms. Additionally, the half Hann-window ramps have been set to a duration of 50 ms. This corresponds to changes in the attack, release and sustain of the notes. Furthermore, harmonics of the pure tone at 2, 3, 4, and 5 multiples of the fundamental frequency have been added.

These changes were therefore expected to lead to a breakdown of performance that reflects the effect of binaural fusion when compared to the binaural test condition.

ANALYSIS

A binomial distribution underlies this task, since the answer is either correct or false. Therefore, the significance levels for $p \approx 0.333$ are given by: $\Pr(X \geq 13, n=24) \approx 0.0284$ (* ; ≥ 54.17 %), $\Pr(X \geq 15, n=24) \approx 0.00323$ (** ; ≥ 62.50 %) and $\Pr(X \geq 16, n=24) \approx 0.00860$ (*** ; ≥ 66.67 %).

RESULTS

The results are presented in Fig. 4 with the detection performance plotted as percent correct for the three conditions, binaural, binaural control and monaural. The average performance for the binaural and monaural conditions lies at about 80% correct (***), while the binaural control with altered timbre shows an average performance slightly above 50 % and still not significantly above chance performance.

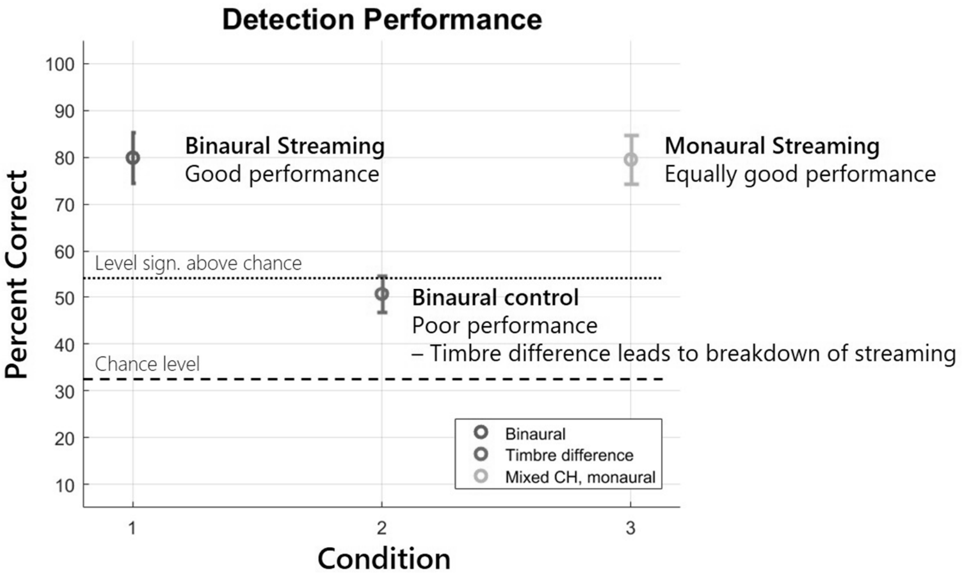


Fig. 4: Results of the scale illusion detection experiment in percent correct for the three conditions: **C1:** binaural streaming, **C2:** binaural control and **C3:** monaural streaming. The bars give the range for the standard error, the dashed line the chance level at 33.33 % and the dotted line indicates the performance level significantly above chance, 54.17 % (cf. Analysis).

CONCLUSIONS

This study describes a task that proves binaural streaming and fusion ability of normal hearing participants. It is based on two melody streams that form the scale illusion. The participants showed that they can build a fused stream out of the binaural input, follow this stream and successfully detect the deviant note embedded into it. Further, in this task their monaural performance equals their binaural performance. The participants therefore do not have more difficulty following the streams binaurally. Additionally, the results of the binaural control condition demonstrate that a severe difference in timbre across ears leads to a breakdown of performance. When the two ears have such a severely different timbre, the components of the streams can no longer be identified as belonging together and the binaural information is no longer fused into a segregated object. Thus, participants can no longer follow the melody stream and identify the deviant note. Besides, this breakdown in performance shows that single-ear listening is insufficient to score well in the binaural condition, validating the experiment’s design.

Variants of this task can potentially be used to test fusion ability of cochlear implant users (both bilateral and bimodal), where it is unclear whether binaural fusion takes place. This will be investigated in further studies.

ACKNOWLEDGEMENTS

The authors would like to thank everybody who participated in the experiment for their help. We could not have done it without you!

REFERENCES

Bregman, A.S. (**1994**). *Auditory Scene Analysis. The Perceptual Organization of Sound*. MIT Press, ISBN: 9780262521956.

Deutsch, D. (**1975**). "Two-channel listening to musical scales," J. Acoust. Soc. Am., **57**, 1156-1160. doi: 10.1121/1.380573

Deutsch, D. (**1999**). "Grouping mechanisms in music," in *The Psychology of Music*. Academic Press, ISBN: 9780123814609, pp. 299-348.

Middlebrooks, J.C., Simon, J.Z., Popper, A.N., and Fay, R.R. (**2017**). *The Auditory System at the Cocktail Party*. Springer Handbook of Auditory Research, **60**, doi: 10.1007/978-3-319-51662-2

Wickens, T.D. (**2002**). *Elementary Signal Detection Theory*. Oxford University Press, Oxford, ISBN: 978-0195092509, pp. 3-37, 93-110, 237-241.

Verbal attribute magnitude estimates of pulse trains across electrode places and stimulation rates in cochlear implant listeners

WIEBKE LAMPING[1,*], SÉBASTIEN SANTURETTE[1,2], AND JEREMY MAROZEAU[1]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Department of Otorhinolaryngology, Head and Neck Surgery & Audiology, Rigshospitalet, Copenhagen, Denmark*

For cochlear implant users, temporal and place cue are assumed to vary along two orthogonal perceptual dimensions linked to pitch height and timbre. Here, the effect of electrode place, pulse rate, and amplitude modulation frequency on those perceptual dimensions was investigated. Combinations of different electrode places with differing pulse rates or modulation frequencies were presented to the participants while they were asked to rate pitch height and sound quality using multiple verbal attributes. The results indicate that temporal and place cues induce two perceptual dimensions that can be both linked to pitch and timbre.

INTRODUCTION

Pitch is one of the primary auditory sensations and plays an important role when defining and differentiating our acoustic environment. Although human listeners perform remarkably well in discriminating and ranking pitch, this task remains difficult for hearing-impaired listeners. Especially for cochlear implant (CI) users the perception of musical and voice pitch has been shown to be problematic (cf. McDermott, 2004, for a review). For normal-hearing listeners, pitch has been suggested to have multiple dimensions such as pitch height or pitch chroma. However, when dealing with the different cues that can induce a pitch-like sensation in CI users, there seems to be a lack of definition (Oxenham, 2008).

The implant can provide three different types of potential pitch cues that can be manipulated independently and that have been assumed to elicit a change in pitch height: (i) Place cues are provided by a change in place of electrode; (ii) Rate cues are associated with the pulse rate in pulses per second (pps); and (iii) Modulation cues can be provided by imposing an amplitude modulation on a sufficiently high carrier pulse train (e.g., Tong *et al.*, 1983; Shannon, 1983; McKay and Carlyon, 1999). Further, pitch can be either increased or decreased when both rate and place cues are varied in complementary or contradictory directions, respectively (Zeng, 2002). However, pitch perception remains poor in most CI users: Rate pitch deteriorates above a specific

[*] Corresponding author: wila@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

"upper limit" (generally 300 to 500 pps) and place cues are limited by the number of implantable electrodes, current spread, and shallow insertion depths.

Even though the sensations induced by rate cue and place cue can be ranked from low to high, they have been shown to be independent as they vary along two orthogonal perceptual dimensions (Tong *et al.*, 1983). It has been hypothesised that the dimension connected to rate may be linked to pitch height whereas the dimension connected to place may rather be linked to timbre (McDermott and McKay, 1997; McKay *et al.*, 2000). Particularly brightness, a timbre attribute associated with the spectral centroid of a sound in normal-hearing listeners, has been assumed to correlate with electrode place. Fearn and Wolfe (2000) tried to determine perceptual features other than pitch by assessing the sound quality of regular pulse trains while varying place and rate parameters. They let six CI recipients scale the pitch and sound quality for stimuli from 100 to 1000 pps presented on apical, middle, and basal bipolar electrode pairs. Results showed that low pulse rates presented on the basal electrodes were rated with the poorest sound quality and participants reported that these stimuli were rather perceived like buzzing sounds. In a similar study, Landsberger *et al.* (2016) also found ratings of the attribute "clean" to be low for low-rate stimuli presented at basal cochlear locations. Still, "cleanness" remained high when low-rate pulse trains were presented at apical locations, suggesting better sound quality when temporal code is provided apically.

It remains unclear which specific sound sensations can be linked to the physical parameters of pulse rate, place of electrode, and modulation frequency. The aim of the present study was to investigate the effect of these parameters on the perceptual dimensions associated with pitch and timbre by using verbal attributes, and to assess whether they induce independent dimensions. It was also assessed whether both changes in pulse rate and modulation frequency led to a similar patterns of results for the same timbre attributes.

METHODS

Participants

Five adult native Danish speaking participants with Nucleus devices were tested at the Technical University of Denmark (DTU). Specific participant demographics are presented in Table 1. All participants provided written informed consent and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391). All tested electrodes were present in the participant's clinical map.

Stimuli

Stimuli consisted of single electrode, cathodic-first biphasic pulse trains. All stimuli were presented with a pulse duration of 25 μs, an interphase gap of 8 μs, and in monopolar mode. Two different sets of stimuli were generated. The first set was

Participant	Age in years	Years of implant use	Age of onset hearing loss
C1	73	13	20
C2	19	14	Birth
C3	45	2	25
C4	64	15	13
C5	43	5	Birth

Table 1: Details of the five CI users who participated in the experiment.

created by all possible combinations of electrode numbers 22, 18, 14, and 10 and pulse rates of 80, 150, 300, 600, and 1200 pps. The second stimulus set was composed of amplitude-modulated pulse trains with modulation frequencies of 80, 150, 200, 300, and 400 Hz imposed on a constant carrier of 1200 pps, presented via the same electrodes as in set 1. The amplitude of each stimulus was adjusted for presentation at a comfortable and equally loud level, as described in the following.

Procedure

The loudness growth for all stimuli was estimated before loudness balancing. On a single electrode, a stimulus was played initially below threshold and then gradually increased in 0.88-dB steps from threshold to upper comfort level. The 10-point loudness scale from Advanced Bionics was used to let the participants indicate the loudness level of each stimulus presentation. For loudness balancing, the reference stimulus was a 300-pps pulse train on electrode 18. This stimulus was first adjusted to have the most comfortable level. Thereafter, pulse trains of the same rate but differing in electrode number were balanced to this reference. For this, two stimuli were presented with a duration of 500 ms and with a 500-ms interstimulus interval at amplitudes corresponding to what had been previously described as the most comfortable loudness in the first interval, and a lower loudness level in the second. After participants adjusted the loudness of the test stimulus to be the same of the reference, both reference and test stimulus were swapped and the test stimulus was presented at the previously determined comfort level in the first interval while the reference was balanced to it. The adjusted level was calculated by averaging the current difference in the logarithmic domain. Once the 300-pps pulse trains were set to equal loudness, the 80-pps, 150-pps, 600-pps, and 1200-pps pulse trains were each balanced to the 300-pps pulse train on the same electrode.

After loudness balancing, participants were familiarised to the range of stimuli and definitions and descriptions for all attributes were provided in Danish, taken from the DELTA lexicon of sound-describing words (Pedersen, 2008). The listeners were then presented with one randomly selected single electrode pulse train with a duration of 2 s and asked to rate “pitch height”, as well as sound quality, using multiple verbal attributes, i.e., “calm”, “loud”, “clean”, “complex”, “bright”, “lively”, “rough”, “boomy”, and “humming” which were translated into Danish (i.e., “høj”, “rolig”, "kraftig", "ren", "kompleks", "lys", "livlig", "ru", "dybtoneresonant", "summende"). Responses were collected on continuous verbal attribute magnitude estimate scales ranging from 0 to 100, with 100 translating into a full agreement between the attribute and the sound specific sensation and 0 to the opposite. All attributes were displayed at the screen at the same time and in random order. In a single trial, participants could click on a "play" button to be presented with the stimulus and were encouraged to repeat the sounds as often as necessary. The procedure was repeated until 3 measurements were collected for each stimulus with each of the ten descriptors for all participants.

To reduce variability and investigate the relationship between the physical parameters and pitch height, brightness, and roughness further, five more repetitions were conducted for these attributes with the same participants.

RESULTS

Results in Fig. 1 show the principal component analysis (PCA) for scalings of all 10 attributes with variables plot (left), scores plot (right) for stimulus set 1 (top) and stimulus set 2 (bottom). The number of dimensions kept in the results was estimated by using the generalised cross-validation approximation method. The data are scaled to unit variance.

The scores plot for stimulus set 1 seen in Fig. 1 (upper right) shows that the first two principal components can account for around 80% of the variance. For stimulus set 2 (Fig. 1, bottom right), approximately 70% of the variance can be explained by components 1 and 2.

The variables plot for both stimulus sets (Fig. 1, left panels) shows that many of the chosen attributes covary. However, the majority of the attributes lies orthogonal to the attribute pitch height, e.g., roughness, complexity, cleanness, calmness, etc. As all attributes were supposed to be equated in loudness, the attribute loud is only showing weak correlation with low-rate pulse trains on apical electrodes. Brightness, which has previously been associated with place of excitation, does neither show the same ratings as pitch height, nor is orthogonal to it.

Results from the repeated measurements on pitch height, brightness, and roughness are shown in Fig. 2 and Fig. 3. The results are analysed by means of a mixed model with two within-listener factors, pulse rate and electrode place, and the random effect participant.

Scalings for pitch height, as seen in Fig. 2 (left), were in agreement with previous findings showing a significant dependency of pitch on electrode place [$F(3,4) = 31.24, p < 0.005$] and pulse rate [$F(4,4) = 33.80, p < 0.005$] (e.g., Fearn and Wolfe, 2000; Landsberger *et al.*, 2016), while showing no significant interaction effect. For roughness (Fig. 2, middle), participant was a significant random effect ($p < 0.005$) too and pulse rate was a significant main factor [$F(4,4) = 34.77, p < 0.005$]. However, post-hoc paired t-tests with Bonferroni adjustments indicated no significant difference

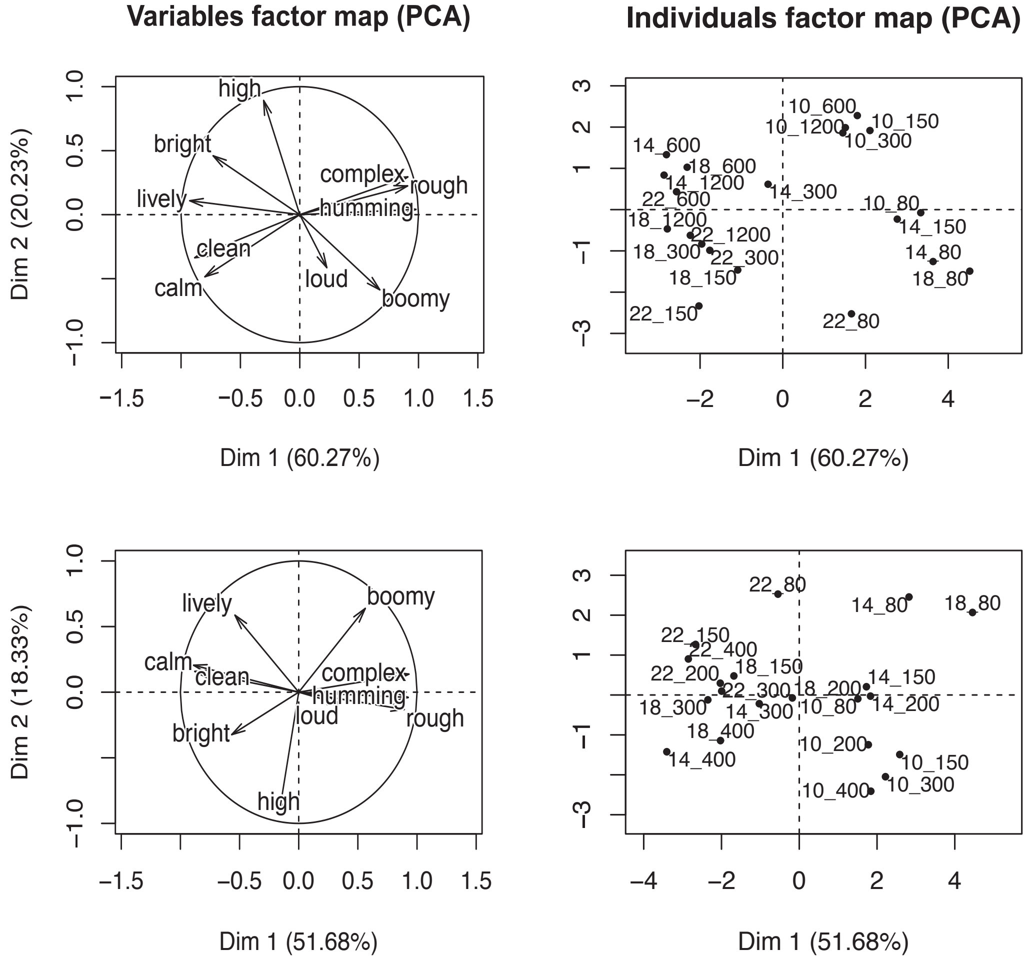


Fig. 1: Variable (left) and scores plot (right) of a principal component analysis for the ten attributes used in the experiment. Top and bottom show results for stimulus set 1 and 2, respectively. The scores plot shows electrode numbers followed by pulse rate.

for rates above 600 pps. Electrode place showed a non-significant tendency of low-rate pulse trains being rated as less rough when presented at apical cochlear locations than at basal locations [$F(3,4) = 1.06, p = 0.37$]. For brightness (Fig. 2, right), pulse rate [$F(4,4) = 36.19, p < 0.005$] and electrode place [$F(4,4) = 12.33, p < 0.005$] were significant main effects. Interestingly, brightness was the only attribute for which there was a significant interaction between the two main factors [$F(4,4) = 3.8, p < 0.05$].

Figure 3 shows scalings for modulated pulse trains (stimulus set 2). For pitch height, rate [$F(4,4) = 36.28, p < 0.005$], electrode place [$F(3,4) = 15.33, p < 0.005$], and participant ($p < 0.05$) were significant. For roughness, the effects of rate [$F(4,4) = 33.42, p < 0.005$] and electrode [$F(3,4) = 11.20, p < 0.005$] were significant. For

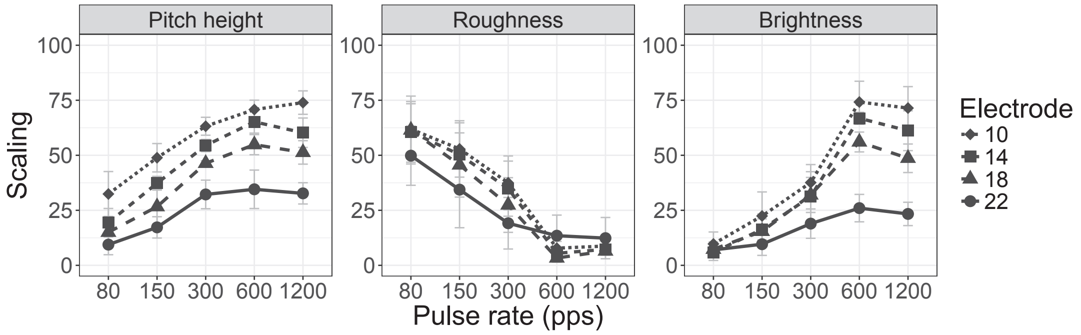


Fig. 2: Average of scaled values for all participants for pitch height (left), roughness (middle), and brightness (right) for unmodulated pulse trains. Electrode 22 is the most apical electrode in the Cochlear device. Error bars depict the standard error.

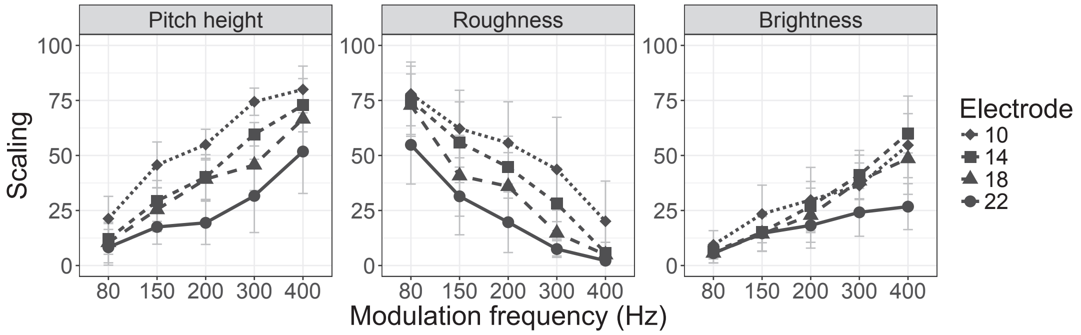


Fig. 3: Average of scaled values for all participants for pitch height (left), roughness (middle), and brightness (right) for modulated pulse trains on a constant carrier rate of 1200 pps. Error bars depict the standard error.

brightness, rate [$F(4,4) = 25.78, p < 0.005$], electrode [$F(3,4) = 4.44, p < 0.05$], and participant ($p < 0.005$) were significant effects as well. However, for the brightness attribute only scalings on electrode 22 differed significantly from those for other electrodes. No significant interaction effect was found for any attribute using stimulus set 2.

Stimulus set 1 and 2 showed very similar results: There was no significant difference in scalings between the results of these two sets for frequencies of 80, 150, and 300 Hz, and pulse rates of 80, 150, and 300 pps. The only attribute for which a significant difference between stimulus sets emerge, was roughness [$F(1,4) = 7.63, p < 0.05$].

DISCUSSION

The results of the PCA showed that most of the variance in the data set could be explained by the first two principle components. Further, first and second principal components seemed to be related to the pulse rate and electrode place, respectively. The majority of attributes lay orthogonally to pitch height, e.g., roughness or cleanness, which may be connected to sound quality or pleasantness. The lack of correlation between pitch height and roughness suggests that different rate and place combinations may induce similar pitch-like sensations but that their sound qualities might differ substantially. Scalings for brightness lay in-between these two dimensions, suggesting a combined effect along the first two principal components.

Scalings for pitch height were consistent with previous literature (e.g., Fearn and Wolfe, 2000) as they show the expected changes with pulse rate and electrode place. Roughness, as a possible indicator for sound quality, showed less dependency on electrode place than in previous results, (e.g., "cleanness", Landsberger *et al.*, 2016), despite a non-significant trend in the scalings. The smaller number of participants in the present study compared to Landsberger *et al.* (2016) may explain the lack of significance. Further, this trend was significant for the amplitude modulated pulse trains in stimulus set 2. Lower scalings for roughness on low-rate apical pulse trains may be linked to the idea of a better place-rate match for this type of stimulation (Oxenham *et al.*, 2004). Apart from better sound quality at apical cochlear regions, other studies, such as Macherey *et al.* (2011) and Stahl *et al.* (2016), also suggested that temporal processing could be improved when provided apically. They found a significantly higher upper limit and lower rate discrimination thresholds at the apex relative to more basal cochlear locations. These results suggest that temporal coding, i.e., rate pitch, is likely be conveyed more pleasantly but also more adequately when provided apically. Finally, scalings for most attributes did not reveal a significant interaction effect, as shown before (see McKay *et al.*, 2000). However, it is interesting to note that this is not the case for the attribute brightness. It seems that differences in brightness scalings emerged only for high rates where a change in the temporal code no longer evokes a change in perceived pitch and only place of excitation cues are available.

Similar results were obtained for stimulus set 1 and 2. This may indicate similarities in sound quality and seems consistent with measures of temporal acuity in CI listeners. Kong *et al.* (2009) showed that rate discrimination thresholds have similar patterns for both modulated and unmodulated pulse trains, indicating a similar pitch salience for these stimuli.

CONCLUSION

The statistical analysis revealed no significant interaction effect between temporal and place cues, apart for scalings for the attribute brightness. This may suggest that the two cues are not totally independent, at least when scaling this particular attribute. A comparison between scalings for modulated and unmodulated pulse trains only showed a significant difference between the two sets for the attribute roughness. Results suggest that neither pitch nor timbre exclusively covary with electrode place, pulse rate, or modulation frequency.

REFERENCES

Fearn, R., and Wolfe, J. (**2000**). "Relative importance of rate and place: Experiments using pitch scaling techniques with cochlear implants recipients," Ann. Oto. Rhinol. Laryngol. Suppl., **185**, 51-53. doi: 10.1177/0003489400109S1221

Kong, Y.Y., Deeks, J.M., Axon, P.R., and Carlyon, R.P. (**2009**). "Limits of temporal pitch in cochlear implants," J. Acoust. Soc. Am., **125**, 1649-1657. doi: 10.1121/1.3068457

Landsberger, D.M., Vermeire, K., Claes, A., *et al.* (**2016**). "Qualities of single electrode stimulation as a function of rate and place of stimulation with a cochlear implant," Ear Hearing, **37**, 149-159. doi: 10.1097/AUD.0000000000000250

Macherey, O., Deeks, J.M., and Carlyon, R.P. (**2011**). "Extending the limits of place and temporal pitch perception in cochlear implant users," J. Assoc. Res. Otolaryngol., **12**, 233-251. doi: 10.1007/s10162-010-0248-x

McDermott, H.J., and McKay, C.M. (**1997**). "Musical pitch perception with electrical stimulation of the cochlea," J. Acoust. Soc. Am., **101**, 1622-1631. doi: 10.1121/1.418177

McDermott, H.J. (**2004**). "Music perception with cochlear implants: A review," Trends Amplif., **8**, 49-82. doi: 10.1177/108471380400800203

McKay, C.M. and Carlyon, R.P. (**1999**). "Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains," J. Acoust. Soc. Am., **107**, 347-357. doi: 10.1121/1.424553

McKay, C.M., McDermott, H.J., and Carlyon, R.P. (**2000**). "Place and temporal cues in pitch perception: Are they truly independent?," Acoust. Res. Lett. Onl., **1**, 25-30. doi: 10.1121/1.1318742

Oxenham, A.J., Bernstein, J.G.W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. USA, **101**, 1421-1425. doi: 10.1073/pnas.0306958101

Oxenham, A.J. (**2008**). "Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants," Ann. Trends Amplif., **12**, 316-331. doi: 10.1177/1084713808325881

Pedersen, T.H. (**2008**). "The Semantic Space of Sounds – Lexicon of Sound-Describing Words," DELTA Technical note.

Shannon, R.V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res., **11**, 157-189. doi: 10.1016/0378-5955(83)90077-1

Stahl, P., Macherey, O., Meunier, S, *et al.* (**2016**). "Rate discrimination at low pulse rates in normal-hearing and cochlear implant listeners: Influence of intracochlear stimulation site," J. Acoust. Soc. Am., **139**, 1578-1591. doi: 10.1121/1.4944564

Tong, Y.C., Blamey, P.J., and Dowell, R.C. (**1983**). "Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant," J. Acoust. Soc. Am., **74**, 73-80. doi: 10.1121/1.389620

Zeng, F.G. (**2002**). "Temporal pitch in electric hearing," Hear. Res., **174**, 101-106. doi: 10.1016/S0378-5955(02)00644-5

On the cost of introducing speech-like properties to a stimulus for auditory steady-state response measurements

SØREN LAUGESEN[1,*], JULIA E. RIECK[1,2], CLAUS ELBERLING[3], TORSTEN DAU[4], AND JAMES M. HARTE[1]

[1] *Interacoustics Research Unit, Kgs. Lyngby, Denmark*

[2] *Faculty of Sciences, Free University, Amsterdam, The Netherlands*

[3] *Virum, Denmark*

[4] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

For the purpose of objectively validating hearing-aid fittings in pre-lingual infants, auditory steady-state response (ASSR) measurements are investigated. This paper examines the cost of introducing speech-like features into the ASSR stimulus, which is done to ensure that the hearing aid processes the stimulus as if it were real speech. The main findings were a reduction in ASSR amplitude of 4 dB and an increase in detection time by a factor of 1.6, while detection rates were unaffected given sufficient recording time.

INTRODUCTION

The success of new-born hearing-screening has led to very young infants being fitted with hearing aids. Standard tools for validation of these fittings (aided audiometry, questionnaires, etc.) are either impossible or highly unreliable in very young infants. Therefore, objective methods based on electrophysiology are investigated (e.g., Punch *et al.*, 2016). Here, an approach using the auditory steady-state response (ASSR) is considered (Picton *et al.*, 1998).

Aided-ASSR measurements are associated with several challenges. As the ASSR stimulus passes through a hearing aid, it must be ensured that correct gain is applied and that the correct signal processing features relevant for speech are selected. This can be achieved by manipulating the settings of the hearing aid, e.g., by turning off problematic helping systems, such as noise reduction or directionality (Billings *et al.*, 2007; Carter *et al.*, 2013; Easwar *et al.*, 2015). However, this weakens the ecological argument that the hearing aid is in a normal mode of operation. An alternative approach is to construct an ASSR stimulus with sufficiently speech-like properties that the hearing aid automatically classifies the stimulus as speech. The benefit in terms of strengthening the counselling of the infant's parents suggests the latter approach, because in that case the hearing aid can be fitted to the infant and tested in the exact same setting as it will be used in daily life. In addition, a speech-like stimulus will corroborate the relevance of the measurement for both parents and clinicians.

[*] Corresponding author: slau@iru.interacoustics.com

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

For this purpose, narrow-band (NB) CE Chirps® (Elberling and Don, 2010) were modified to have speech-like properties, and it was verified that this stimulus indeed is classified as speech by the hearing aids tested, by observation in the fitting software. However, since the NB-CE chirps were designed for optimal efficiency, it is expected that any change, such as adding speech-like features to the chirps, will come at a cost of reduced ASSR amplitudes and increased detection times. This was investigated in the present experiment. In addition, the observed changes to the measured ASSR amplitudes and response times are compared to an objective characterisation of the speech-modified versus the standard chirps, based on modulation power.

METHOD AND MATERIAL

In order to isolate the effects of the stimulus modifications, the experiment was carried out with young adult normal-hearing test subjects ($N = 10$) and stimulation provided through insert phones. Individual real-ear measurements (REM) were performed in terms of the real-ear unaided gain (REUG) and these results were used to shape the stimuli to mimic the free-field stimulation eventually needed.

The NB-CE Chirps® consist of four one-octave-wide chirps centred at 500, 1000, 2000, and 4000 Hz, and the speech modifications (patent pending) were derived from the International Speech Test Signal (ISTS, Holube *et al.*, 2010). The root-mean-square levels of the individual octave-band chirps of both stimuli were scaled to match the one-octave band levels of the ISTS at its nominal broad-band level of 65 dB SPL, see Fig. 1. The chirp-band-specific repetition rates were 90.8, 94.7, 102.5, and 96.7 Hz, respectively, and each chirp-train was designed with alternating polarity.

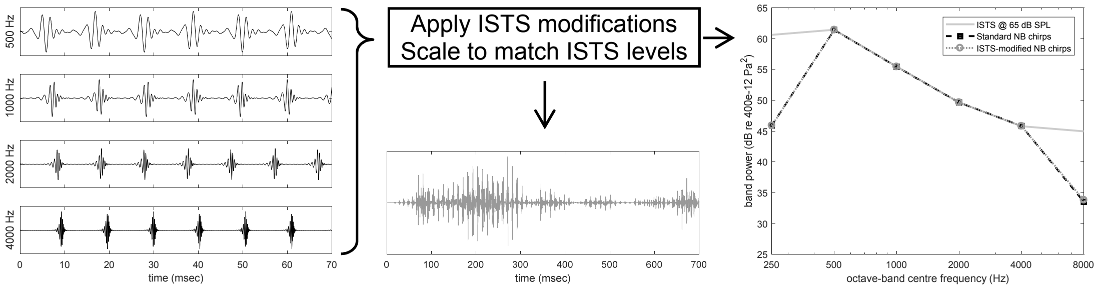


Fig. 1: Sketch of the construction of the stimuli. Left: individual one-octave-band chirps. Middle: resulting waveform of the ISTS-modified chirps. Right: resulting one-octave band spectra compared to that of the ISTS.

ASSR recordings were made using standard clinical 4-electrode montages (high forehead ground, ipsi- and contra-lateral mastoids active, and cheek reference), with 15 minutes of recording time per condition, in one ear at a time. The test subjects were lying comfortably on a bed in a darkened and sound-treated room. Test and re-test recordings were made for all conditions. The Interacoustics Eclipse unit was used as a front-end, with line-level signals routed to an RME Fireface UC soundcard that also delivered the stimuli to the Etymotic Research ER-1 insert phones. Recording and playback were handled by custom Matlab software. The sampling frequency was 32 kHz and the analysis block length was 65536 samples, corresponding to 2.048 sec.

The recordings were analysed with a 40 µV artefact rejection threshold, weighted averaging (John *et al.*, 2001), and simple F-test detection (Dobie and Wilson, 1996) for each of the first 6 response harmonics separately. (A multi-harmonic detector for the ISTS-modified stimulus has yet to be developed.) The outcomes considered were:

- Noise levels, estimated across 20 frequency bins distributed evenly around each response harmonic, excluding frequency bins close to harmonics of 50-Hz line noise, GSM interference, etc.
- Noise-corrected ASSR amplitudes (Dobie and Wilson, 1996). The estimated noise power was subtracted from the response-bin power to yield the noise-corrected power used to compute the ASSR amplitude, which was converted to dB to allow analysis of the relative changes in response with stimulus type.
- Detection times, evaluated as the first time a response was detected with a 5% criterion in successive weighted averages, ignoring Bonferroni correction.

The outcomes were finally analysed with a mixed-model ANOVA with *Test ear* as a random effect and *Stim freq*, *Stim type*, and *Harmonic* as fixed effects.

EXPERIMENTAL RESULTS

Figure 2 shows the number of successful detections in terms of percentages (detection rates). The upper panel shows results for the first 6 harmonics individually, whereas the lower panel shows the detection rates accumulated across harmonics, meaning that an ASSR is detected in either of the first *n* harmonics, *n* = 1, …, 6.

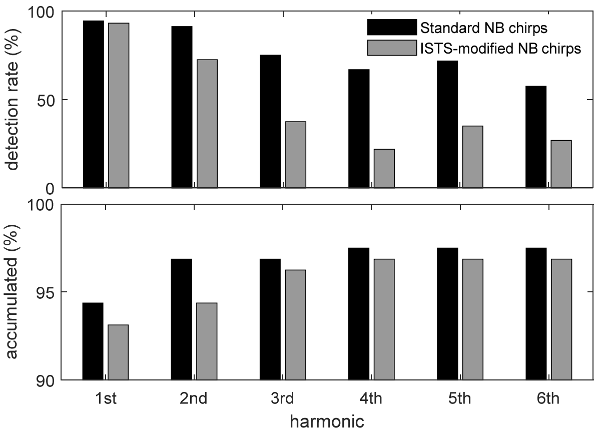


Fig. 2: Detection rates across *Test ear* and *Stim freq* for each *Harmonic* separately (top) and accumulated (bottom), for each *Stim type*.

Considering only the dominant 1st harmonic, the detection rates are very similar for the two types of stimuli. Individually, the higher harmonics provide fewer detections for the ISTS-modified than for the standard stimulus, but nevertheless the accumulated percentages are also very similar.

Figure 3 displays ASSR magnitudes (top panels) and detection times (lower panels) for each harmonic, stimulus band centre frequency, and stimulus type. Both three-way interactions in the statistical models (*Harmonic* by *Stim freq* by *Stim type*) were statistically significant. There are several interesting observations to make from Fig. 3. With increasing harmonic number, the ASSR magnitudes are generally reduced while detection times are increased, as expected. It should be noted that, particularly for the ISTS-modified stimulus, the number of detected responses decreases towards the higher harmonics, which implies that the estimated mean values

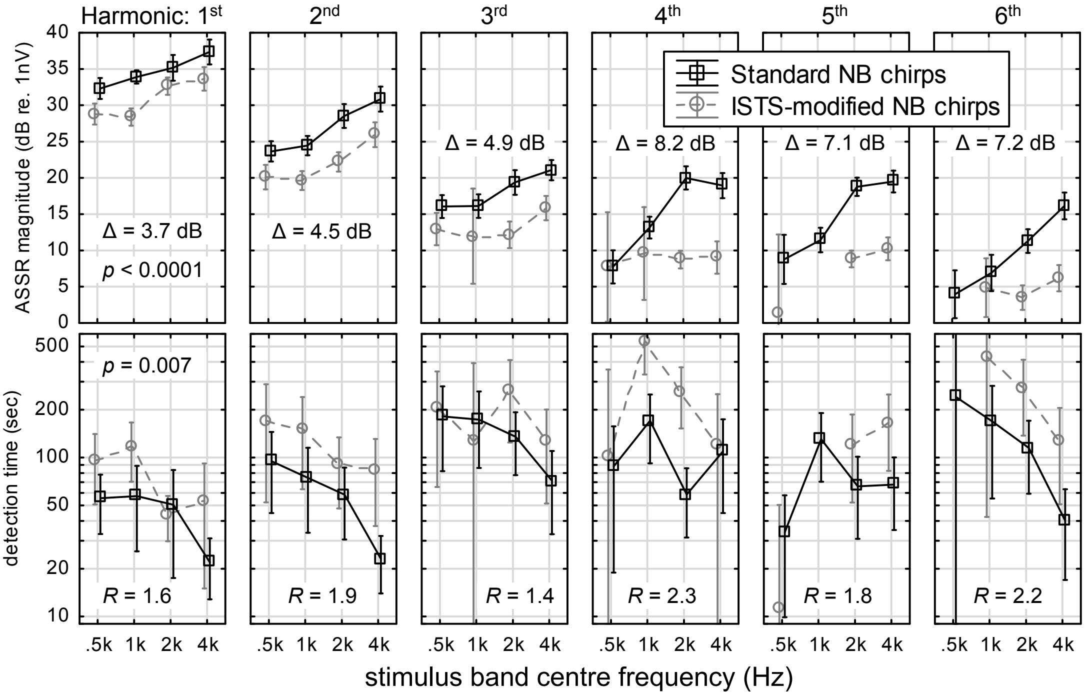


Fig. 3: Noise-corrected ASSR magnitudes (top) and detection times (bottom, logarithmic scale) for each *Harmonic*, *Stim freq*, and *Stim type*, averaged across all ears in which detections were obtained. Error bars indicate 95% confidence intervals. Mean *Stim type* magnitude differences, Δ, mean response-time ratios, *R*, and interaction *p*-values are indicated.

and their patterns are less reliable. Considering the difference between the two stimulus types, the ISTS-modified stimulus produces lower magnitudes and longer detection times than the standard stimulus, which again was expected. It is noteworthy that the patterns of magnitude versus stimulation frequency seem stable up to the 3rd harmonic in terms of a constant difference between the stimuli, while greater mean differences (the inserted Δ-values) and differences in slopes between stimuli can be observed at the higher harmonics. The detection-time data are more variable, as indicated by the wider error bars, but it is striking that the relative increase in detection time between stimuli (the inserted *R*-values) is almost constant across harmonics. This is in contrast to the observed increase in the Δ-values for the ASSR magnitude.

Figure 4 shows the estimated noise levels across conditions where detection was obtained. These results show a reduction in noise level with increasing harmonic number, which was expected since biological noise typically has a 1/*f*-shaped spectrum towards low frequencies (Pritchard, 1992). Note that the plotted *Harmonic* by *Stim type* interaction just fails to reach significance ($p = 0.06$).

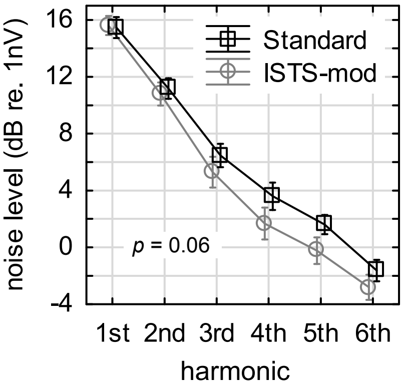


Fig. 4: Mean noise levels across *Test ear* and *Stim freq*.

STIMULUS ANALYSIS

To characterise the stimulus waveforms, an un-normalised modulation spectrum analysis was applied. This analysis was introduced under the assumption that the ASSR is driven by a non-linear representation of the stimulus (for example, after rectification), here modelled in terms of envelope power. This approach requires that the stimuli under comparison are scaled to the same level, in this case the nominal SPL of 65 dB. The two stimuli (both consisting of all four NB chirps) were first passed through gammatone filters (Johannesma, 1972) corresponding to the stimulus frequency band of interest, to mimic the frequency specificity of the auditory system. The results are displayed in Fig. 5 for the two stimuli and two representative stimulus-band centre frequencies: 500 and 2000 Hz. For the standard stimulus (left- hand panels), the modulation power is almost entirely represented at the response harmonics, i.e., the repetition rate of the respective stimulus band and its harmonics. There are smaller modulation power components present at frequencies not belonging to any stimulus repetition rate; these occur because of interactions among the four different repetition rates that are present at the same time in the stimulus. For the ISTS-

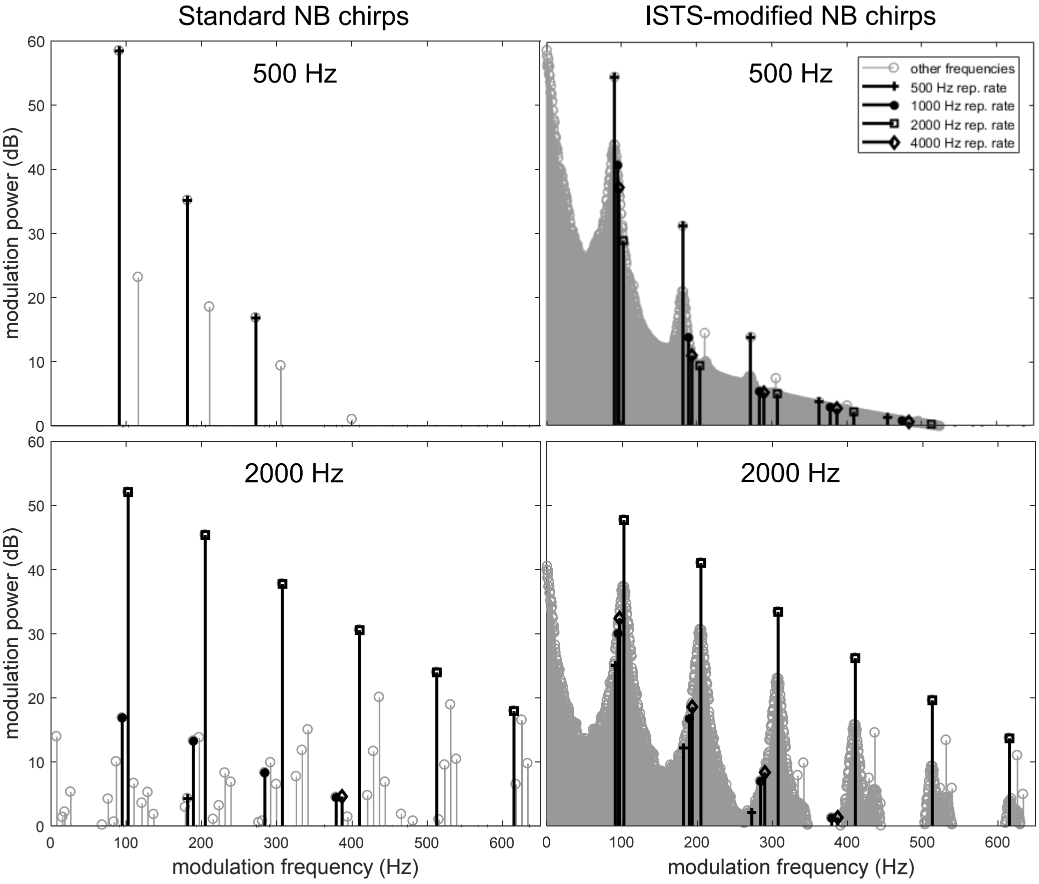


Fig. 5: Un-normalised modulation power spectra of the two stimuli, computed with gammatone filtering at 500 Hz (top) and 2000 Hz (bottom). The stimulus harmonics for each stimulus frequency band are highlighted.

modified stimulus, the modulation power is clearly smeared out around the respective repetition rate and its harmonics. This is a consequence of the additional temporal envelope fluctuations imposed from the ISTS. The modulation power at the response harmonics still stand out in the right-hand panels of Fig. 5, but their magnitudes are reduced compared to the standard stimulus. Table 1 lists the change in modulation power for all four stimulation bands and the first three harmonics. These results show that the estimated change in modulation power is similar across the first response harmonics and among the stimulus bands. The mean reduction is $\Delta_{mod.power}$ = 4.5 dB.

Response harmonic	Analysis band			
	500 Hz	1 kHz	2 kHz	4 kHz
1st	4.0	5.4	4.3	4.9
2nd	3.9	5.3	4.4	4.9
3rd	3.1	5.1	4.4	4.9

Table 1: Reduction in modulation power (in dB) due to the ISTS-modifications to the stimulus.

DISCUSSION

First, the relation between stimulus modulation power and observed ASSR magnitude is considered. For reference, two examples of similar data (‘physiological input/output curves’) are shown in Fig. 6. Both sets of results indicate a slope of about s = 0.8 between dB measures of modulation power and response magnitude. Applying this to the dominant 1st-harmonic data from the present experiment yields

$$\widehat{\Delta_{ASSR}} = \Delta_{mod.power} \times s = 4.5 \text{ dB} \times 0.8 \text{ dB/dB} \approx 3.6 \text{ dB},$$

which agrees very well with the observed Δ_{ASSR} = 3.7 dB from Fig. 3 (top-left panel). In addition, the modulation power analysis reproduces the trend that ASSR magnitude drops more rapidly across harmonics for the lower stimulation frequencies than for the higher (Fig. 3, top row), at least considering the standard NB chirps. This agrees with the successively fewer stimulus line-spectrum components present within an auditory (or gammatone) filter towards lower stimulus band centre frequencies. For the ISTS-modified chirps, the higher-order harmonic responses appear to be limited by the noise floor, which probably conceals the aforementioned effect.

Secondly, the cost of introducing the ISTS-modifications to the NB-CE chirps is considered. By comparing the changes to ASSR magnitude and detection time between the two stimuli, it is seen that the observed reduction in ASSR magnitudes is out of proportion with the increase in detection time, particularly at the higher harmonics. For example, at the 1st harmonic, Δ_{ASSR} = 3.7 dB suggests an increase in detection time by a factor of R = 2.4, where R = 1.6 was observed; at the 6th harmonic Δ_{ASSR} = 7.2 dB suggests R = 5.3, with R = 2.2 observed. The (albeit non-significant) difference in the noise-level patterns (Fig. 4) may hint at lower noise levels for the ISTS-modified relative to the standard NB chirps towards the higher harmonics, which would partly offset the effect of lower ASSR magnitudes on detection time. In addition, note that the detection times were determined from successively averaged un-weighted spectra, whereas the ASSR magnitudes were determined from weighted-

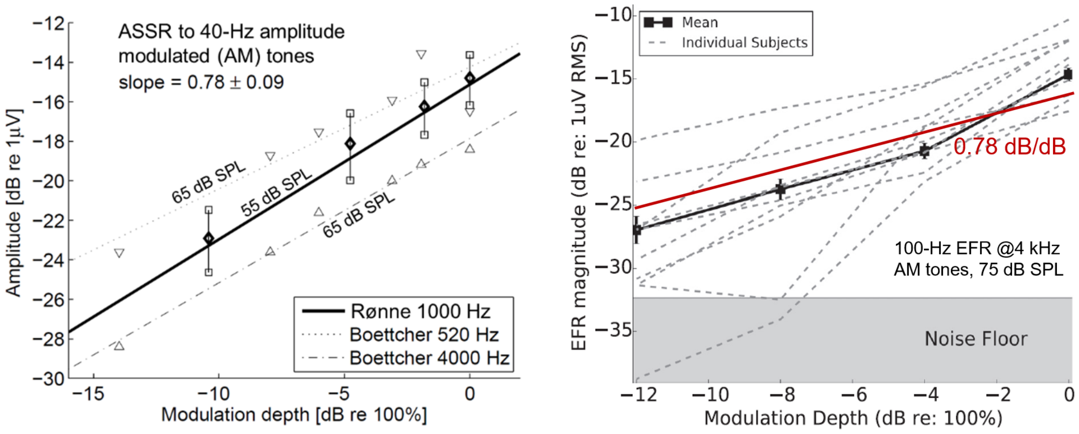


Fig. 6: Two examples of physiological input/output curves. Left: 40-Hz ASSR measurements (Rønne, 2012; Boettcher *et al.*, 2001). Right: 100-Hz envelope-following response (EFR) measurements (Bharadwaj *et al.*, 2015). The regression line from the left panel is superimposed on the right panel.

average spectra across all accepted blocks of the full 15-minute recordings.

Finally, it is encouraging to observe very similar detection rates for the two stimuli (Fig. 2). The reduced response contribution from successively higher harmonics seen for the ISTS-modified versus the standard NB chirps, observed in Fig. 2, top panel, indicates that a multi-harmonic detector, e.g. the q-sample detector (Cebulla *et al.*, 2006) may provide less benefit for the ISTS-modified stimulus compared to what has been found for standard stimuli. On the other hand, the accumulated detection rates in Fig. 2, bottom row, show bigger improvement from including more harmonics for the ISTS-modified than for the standard stimulus.

CONCLUSIONS

The consequences of adding speech-like properties to the NB-CE Chirps® for ASSR recordings – for the purpose of hearing-aid validation in infants – were investigated in young adult normal-hearing test subjects. The main findings were:

- Detection rates were very similar for the speech-modified and the standard stimuli.
- ASSR magnitude decreased by about 4 dB (for the dominant 1st response harmonic).
- Detection times increased relatively less, by a factor of 1.6.

The reduced response magnitude and increased detection time seem acceptable, given the potential for allowing aided ASSR recordings to be carried out with hearing aids in their daily-life mode of operation.

The un-normalised modulation power spectrum including pre-processing through gammatone filters appears to be a useful tool for characterising the efficacy of complex stimuli for ASSR measurements.

Future work will extend the investigations to infants fitted with hearing aids.

REFERENCES

Bharadwaj, H.M., Masud, S., Mehraei, G., Verhulst, S., and Shinn-Cunningham, B. G. (**2015**). "Individual differences reveal correlates of hidden hearing deficits," J. Neurosci., **35**, 2161-2172.

Billings, C.J., Tremblay, K.L., Souza, P.E., and Binns, M.A. (**2007**). "Effects of hearing aid amplification and stimulus intensity on cortical auditory evoked potentials," Audiol. Neurootol., **12**, 234-246.

Boettcher, F.A., Poth, E.A., Mills, J.H., and Dubno, J.R. (**2001**). "The amplitude-modulation following response in young and aged human subjects," Hear. Res., **153**, 32-42.

Carter, L., Dillon, H., Seymour, J., Seeto, M., and Van Dun, B. (**2013**). "Cortical auditory-evoked potentials (CAEPs) in adults in response to filtered speech stimuli," J. Am. Acad. Audiol., **24**, 807-822.

Cebulla, M., Stürzebecher, E., and Elberling, C. (**2006**). "Objective detection of auditory steady-state responses: comparison of one-sample and q-sample tests," J. Am. Acad. Audiol., **17**, 93-103.

Dobie, R.A. and Wilson, M.J. (**1996**). "A comparison of t test, F test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids," J. Acoust. Soc. Am., **100**, 2236-2246.

Easwar, V., Purcell, D.W., Aiken, S.J., Parsa, V., and Scollie, S.D. (**2015**). "Evaluation of speech-evoked envelope following responses as an objective aided outcome measure: effect of stimulus level, bandwidth, and amplification in adults with hearing loss," Ear Hearing, **36**, 635-652.

Elberling, C., and Don, M. (**2010**). "A direct approach for the design of chirp stimuli used for the recording of auditory brainstem responses," J. Acoust. Soc. Am., **128**, 2955-2964.

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.

Johannesma, P.I. (**1972**). "The pre-response stimulus ensemble of neurons in the cochlear nucleus," Symposium on Hearing Theory, Eindhoven, Holland, 58-69.

John, M.S., Dimitrijevic, A., and Picton, T.W. (**2001**). "Weighted averaging of steady-state responses," Clin. Neurophysiol., **112**, 555-562.

Picton, T.W., Durieux-Smith, A., Champagne, S.C., Whittingham, J., Moran, L.M., Giguère, C., and Beauregard, Y. (**1998**). "Objective evaluation of aided thresholds using auditory steady-state responses," J. Am. Acad. Audiol., **9**, 315-331.

Pritchard, W.S. (**1992**). "The brain in fractal time: 1/f-like power spectrum scaling of the human electroencephalogram," Int. J. Neurosci., **66**, 119-129.

Punch, S., Van Dun, B., King, A., Carter, L., and Pearce, W. (**2016**). "Clinical experience of using cortical auditory evoked potentials in the treatment of infant hearing loss in Australia," Semin. Hear., **37**, 36-52.

Rønne, F.M. (**2012**). "Modeling auditory evoked potentials to complex stimuli," PhD thesis, Technical University of Denmark.

The relationship between stream segregation of complex tones and frequency selectivity

SARA M. K. MADSEN[1,*], TORSTEN DAU[1], AND BRIAN C. J. MOORE[2]

[1] *Hearing Systems group, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Department of Psychology, University of Cambridge, Cambridge, UK*

The discrimination of changes in fundamental frequency (F0) is better for complex tones with low than with high harmonics, perhaps because the low harmonics are spectrally resolved. The reduced frequency selectivity of hearing-impaired (HI) participants may lead to poorer resolution of low and medium harmonics. This may adversely affect F0 discrimination and in turn reduce the extent of perceptual segregation (streaming) of a rapid sequence of complex tones. We assessed how the streaming of complex tones is affected by harmonic rank and whether HI listeners are less able to segregate tones with low and medium harmonics than near normal-hearing (NH) participants. Subjective streaming was assessed for complex tones that were bandpass filtered between 2 and 4 kHz. Harmonic rank was varied by changing the baseline F0 (with differences in F0 from 5 to 11 semitones). Auditory filter shapes were estimated from notched-noise masking using a 2-kHz signal. The auditory filters were wider for the HI than for the NH participants. Streaming decreased with increasing harmonic rank but was similar for the two groups. Streaming scores were not correlated with auditory filter bandwidths. The results suggest that the effects of harmonic rank on streaming cannot be explained in terms of resolvability.

INTRODUCTION

The extent to which a rapid sequence of tones is perceived as one stream or two streams depends on the perceptual difference between successive tones, produced, for example, by differences in frequency or level; the greater the perceptual difference, the greater is the extent of stream segregation (Moore and Gockel, 2002). When successive complex tones differ in fundamental frequency (F0), the degree of segregation may therefore depend on the salience of the F0. Complex tones with low harmonics have higher pitch salience than tones with only high harmonics and, correspondingly, F0 discrimination limens increase (performance worsens) when the rank (also called the harmonic number) of the lowest harmonic is increased above about eight (Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2006a). Both of these effects may be related to the fact that lower harmonics are better resolved

*Corresponding author: samkma@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

than higher harmonics (Bernstein and Oxenham, 2006b). It is therefore possible that, for a given difference in F0, ΔF0, stream segregation will be greater for complex tones with low resolved harmonics than for tones with only high unresolved harmonics.
It has been shown that participants can segregate streams of complex tones with only high unresolved harmonics (Vliegen and Oxenham, 1999; Vliegen *et al.*, 1999; Grimault *et al.*, 2000; Grimault *et al.*, 2001; Madsen *et al.*, 2015). However, whereas one study found that stream segregation was similar for pure tones, complex tones with low harmonics and complex tones with only high harmonics (Vliegen and Oxenham, 1999), two other studies found a significant decrease in perceived segregation with increasing harmonic rank (Grimault *et al.*, 2000; Madsen *et al.*, 2015). However, both of these studies tested only young NH participants and the results do not make it possible to determine whether the effect of harmonic rank on stream segregation was solely an effect of resolvability; harmonic rank itself may play a role, as has been argued to be the case for the F0 discrimination of complex tones (Bernstein and Oxenham, 2003).

Here, we measured stream segregation and estimated auditory filter bandwidths using notched-noise measurements for older age-matched near-NH and HI-impaired participants. The goals were (1) to establish how the sequential stream segregation of complex tones was affected by harmonic rank and (2) to determine if the pattern of the results could be explained by the resolvability of the harmonics in the tones.

PARTICIPANTS

Eight near-NH (three male) and 13 HI (seven male) participants were tested. The NH participants were 52-78 years old (mean = 63 years, SD = 9 years) and the HI participants were 49-80 years old (mean = 67 years, SD = 9 years). The pure-tone audiometric threshold averaged across 2, 3, and 4 kHz (PTA) was required to be ≤25 dB HL for the NH participants and was between 26 and 55 dB HL for the HI participants. Audiometric thresholds for the test ear (the ear with the lower PTA) of each participant are shown in Fig. 1.

STREAM-SEGREGATION EXPERIMENT

Method

The stimuli for the stream-segregation experiment consisted of sequences of ABA_ triplets where A and B are different tones and "_" represents a pause. These sequences are perceived as having a galloping rhythm when the A and B tones are similar to each other (integration; **Fig. 2**A, upper panel) and as being two streams – with one faster than the other – if the A and B tones are more different from each other (segregation; Fig. 2A, lower panel). The ABA_ sequences had an overall duration of about 8 s. Each tone had a duration of 90 ms, and tones within a triplet were separated by a 20-ms pause. Consecutive ABA_ triplets were separated by 110-ms pauses. The tones were gated on and off with 20-ms raised-cosine ramps.

All tones contained multiple harmonics, but they were bandpass filtered between 2 and 4 kHz (Fig. 2B) to limit the audible range of the harmonics. The harmonic rank

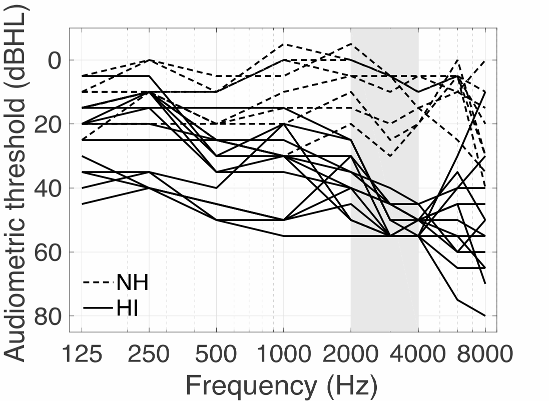


Fig. 1: Audiograms of the test ears of each NH and HI participant.

was varied by varying the F0. Four A-tone F0s were used, and the B-tone F0 was 5, 7, or 11 semitones (ST) higher than the A-tone F0. The F0s of the A- and B-tones were fixed within each trial. The overall level of each tone was 80 dB SPL, and the level of the threshold-equalising noise (TEN) used to mask combination tones and to limit the audibility of stimulus components falling in the filter skirts was 55 dB SPL/ERBn, where ERBn is the average value of the equivalent rectangular bandwidth of the auditory filter for NH listeners (Glasberg and Moore, 1990). The experiment aimed to assess the proportion of time that the streams were perceived as segregated when actively trying to segregate them. The participants were therefore asked to try to hear the sequence as segregated and to press one key when they heard one stream and a different key when they heard two streams.

Results

Participants varied markedly in the time that they took to first press the two-streams key and they rarely made more than one key press within a trial. Hence, the results are presented as the proportion of trials for which the two-streams key was the last key pressed. For brevity, we refer to this as "proportion segregated". This measure is similar to the measure used in other studies (Vliegen and Oxenham, 1999; Grimault *et al.*, 2000; Grimault *et al.*, 2001). Increasing the F0 difference between the A and B tones increased the proportion segregated (Fig. 3). Also, generally, the proportion segregated increased with increasing F0 (i.e., with decreasing harmonic rank) as hypothesized. However, the results were similar for the NH and HI participants.

The streaming scores were averaged across ΔF0 values for each A-tone F0 and were analysed with a mixed linear model with harmonic rank and participant group (NH or HI) as fixed factors and participants as a random factor. There was a significant effect of harmonic rank [$F(3,60) = 6.46$, $p < 0.001$] but no effect of participant group [$F(1,19) = 0.11$, $p = 0.74$] and no interaction [$F(3,57) = 1.21$, $p = 0.31$].

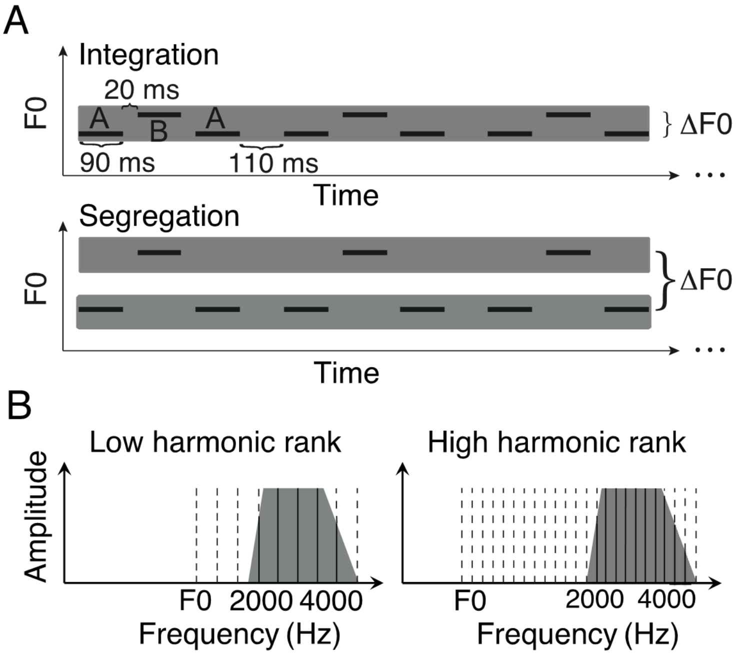


Fig. 2: Illustration of the stimuli used in the stream-segregation experiment. A: Illustration of the ABA_ tone sequences. Upper panel: When A and B were sufficiently similar to each other, all tones were perceived as being in one stream with a galloping rhythm (integration). Lower panel: When A and B were more different from each other, they were perceived as two separate streams. B: Schematic spectrum of the complex tones. Tones were filtered between 2 and 4 kHz, and the harmonic rank was varied by varying the F0.

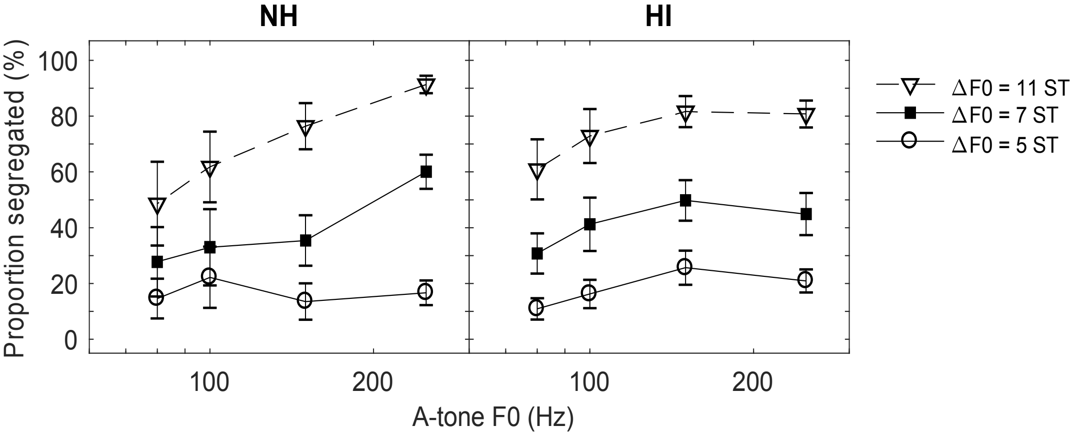


Fig. 3: Mean proportion segregated for the NH (left) and HI (right) participants as a function of the A-tone F0. The parameter is ΔF0, as indicated in the key. Error bars show ±1 standard error of the mean.

NOTCHED-NOISE EXPERIMENT

Regardless of whether stream segregation for complex tones depends on the resolvability of the harmonics or on harmonic rank *per se*, it seems reasonable to assume that performance in the stream-segregation task was dominated by the lowest-ranking harmonics in the complex tones. Therefore, auditory filter shapes were estimated at 2 kHz, the lower edge of the bandpass filter used in the stream-segregation experiment.

Method

The notched-noise method (Patterson, 1976; Rosen *et al.*, 1998) was used. The participants were asked to identify which of three successive noise bursts contained a 2-kHz pure tone. The noise had a duration of 500 ms, and the 400-ms pure tone was temporally centred in one randomly chosen noise burst. Raised-cosine ramps of 10 and 20 ms, respectively, were used to gate the noise and pure tone on and off. The signal level was fixed at 10 dB sensation level while the level of the noise was varied adaptively using a 2-up 1-down procedure. The outside edges of the noise were fixed at 400 and 3600 Hz. The notch width is specified as the deviation of each edge of the notch from the signal frequency, divided by the signal frequency. There were five symmetric conditions with notch widths of 0, 0.1, 0.2, 0.3, and 0.4, and two asymmetric conditions with notch widths of 0.2|0.4 and 0.4|0.2.

Results

A two-parameter roex(p_u, p_l) filter model (Stone *et al.*, 1992) was used to estimate the parameters p_u and p_l, which characterise the slopes of the upper and lower filter skirts, respectively. The equivalent rectangular bandwidth (ERB) of the auditory filter in Hz was estimated as $8*2000/(p_u + p_l)$ (Patterson *et al.*, 1982). As expected, the estimated ERB values were generally smaller for the NH than for the HI participants (Fig. 4). This difference was confirmed by a Welch's *t*-test [$t(14.31) = -2.38$, $p = 0.032$] and was even more pronounced when the ERB value for one NH participant with a very high ERB value was omitted [$t(16.52) = -4.74$, $p < 0.001$]. Also, results for the streaming experiment were similar, if omitting this participant [harmonic rank: $F(3,57) = 5.64$, $p = 0.002$; participant group: $F(1,18) = 0.0023$, $p = 0.96$].

DISCUSSION

While the HI participants had significantly broader auditory filters (greater ERB values) than the NH participants, stream segregation was similar for the two groups. This suggests that the effect of harmonic rank on stream segregation cannot be explained by differences in resolvability of the harmonics. If resolvability was critical, the proportion segregated should have been higher for the NH than the HI participants, at least for conditions where the lowest harmonics were only just resolved for the NH group, since the harmonics for these conditions would have been largely unresolved for the HI group. However, the ERB values varied markedly across participants, especially for the HI group. To further explore whether there was a relationship

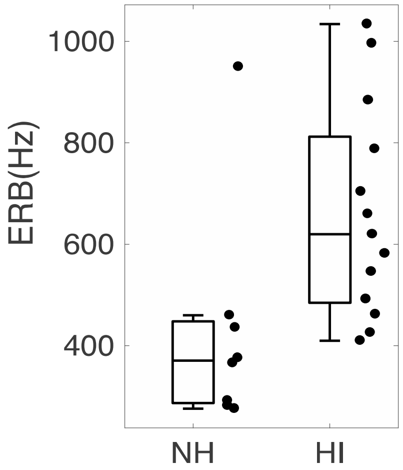


Fig. 4: Box plots of ERB values for each participant group.

between stream segregation and the resolvability of the harmonics, the average of the proportion segregated values across the conditions with the two highest F0s (250 and 150 Hz) was compared with the ERB value for each participant. If better resolvability leads to greater stream segregation, a negative correlation between these two measures would be expected. In fact, the two measures were not correlated (R^2 = 0.043, p = 0.37). A scatter plot of the proportion segregated values against the ERB values is shown in Fig. 5. This supports our conclusion that the effects of harmonic rank on the stream segregation of complex tones cannot be explained in terms of resolvability.

The finding that the stream segregation of complex tones is affected by harmonic rank is not consistent with the results of one of the three previous studies on this topic (Vliegen and Oxenham, 1999). However, Vliegen and Oxenham (1999) also presented some data from a preliminary experiment that, for some conditions and participants, showed results similar to those of the present study.

The conclusion that the effect of harmonic rank cannot be explained by resolvability contrasts with the conclusions of two earlier studies (Grimault *et al.*, 2000; Grimault *et al.*, 2001). The latter of these tested three groups: young NH, older with normal hearing for their age, and older HI. There was little difference in stream segregation between the two groups of older participants, consistent with the findings of the present study. However, they generally found more stream segregation for the young NH group than for the two other groups. This might have been due to differences in auditory filter bandwidth but is more likely to be an effect of age unrelated to frequency selectivity. Therefore, the results of Grimault *et al.* (2001) cannot be used to draw any firm conclusions about the relationship between the stream segregation of complex tones and the resolvability of the harmonics in the tones.

If it is accepted that the resolvability of the harmonics is not the critical factor governing the stream segregation of complex tones, then it appears that harmonic rank *per se* has an influence. This is consistent with the results of a study showing that F0

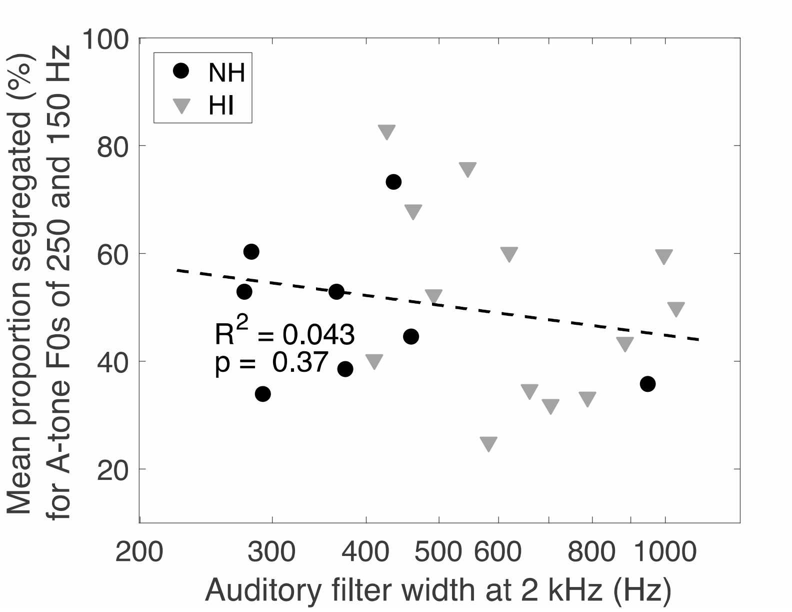


Fig. 5: Scatter plot of mean proportion segregated values across conditions with A-tone F0s of 250 and 150 Hz against ERB values at 2 kHz. The regression line and statistics were calculated from the data for all participants.

discrimination limens were similar for conditions where all harmonics were presented to both ears (diotic) and where odd harmonics were presented to one ear and even harmonics to the other ear (dichotic) (Bernstein and Oxenham, 2003). There were some conditions of this experiment where the harmonics would have been unresolved for diotic presentation but would have been resolved for dichotic presentation, because of the greater spacing of the harmonics within each ear. The lack of effect of presentation mode suggests that F0 discrimination is governed by harmonic rank and not by the resolvability of the harmonics. The effect of harmonic rank has been explained by 'place dependence', i.e., for each auditory filter there is a limited range of periodicities that can be analysed, and this range is closely tied to the centre frequency of that filter (Moore, 2003; Bernstein and Oxenham, 2005).

CONCLUSIONS

For both older near-NH and older HI participants, the proportion segregated in a sequence of complex tones increased with decreasing harmonic rank (increasing F0). Furthermore, the proportion segregated varied with harmonic rank in a similar way for the two groups. However, auditory filter bandwidths at 2 kHz estimated using the notched-noise method were, on average, greater for the HI than for the NH group. Also, the proportion segregated scores were not correlated with the auditory filter bandwidth estimates. These findings suggest that the effect of harmonic rank on stream segregation cannot be explained by the better resolution of lower than of higher harmonics, but rather reflects an effect of harmonic rank *per se*.

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. doi: 10.1121/1.1572146

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. doi: 10.1121/1.1904268

Bernstein, J.G., and Oxenham, A.J. (**2006a**). "The relationship between frequency selectivity and pitch discrimination: effects of stimulus level," J. Acoust. Soc. Am., **120**, 3916-3928. doi: 10.1121/1.2372451

Bernstein, J.G., and Oxenham, A.J. (**2006b**). "The relationship between frequency selectivity and pitch discrimination: sensorineural hearing loss," J. Acoust. Soc. Am., **120**, 3929-3945. doi: 10.1121/1.2372452

Glasberg, B.R., and Moore, B.C.J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res., **47**, 103-138.

Grimault, N., Micheyl, C., Carlyon, R.P., Arthaud, P., and Collet, L. (**2000**). "Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency," J. Acoust. Soc. Am., **108**, 263-271. doi: 10.1121/1.429462

Grimault, N., Micheyl, C., Carlyon, R.P., Arthaud, P., and Collet, L. (**2001**). "Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing," Br. J. Audiol., **35**, 173-182.

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. doi: 10.1121/1.399297

Madsen, S.M.K., Dau, T., and Moore, B.C.J. (**2015**). "Effect of harmonic rank on the streaming of complex tones," Proc. ISAAR, **5**, 477-483.

Moore, B.C.J., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta Acust., **88**, 320-333.

Moore, B.C.J. (**2003**). *An Introduction to the Psychology of Hearing, 5th Ed.* (Emerald, Bingley, UK), pp. 413.

Patterson, R.D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am., **59**, 640-654. doi: 10.1121/1.380914

Patterson, R.D., Nimmo-Smith, I., Weber, D.L., and Milroy, R. (**1982**). "The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am., **72**, 1788-1803.

Rosen, S., Baker, R.J., and Darling, A. (**1998**). "Auditory filter nonlinearity at 2 kHz in normal hearing listeners," J. Acoust. Soc. Am., **103**, 2539-2550. doi: 10.1121/1.422775

Stone, M.A., Glasberg, B.R., and Moore, B.C.J. (**1992**). "Simplified measurement of impaired auditory filter shapes using the notched-noise method," Br. J. Audiol. **26**, 329-334. doi: 10.3109/03005369209076655

Vliegen, J., and Oxenham, A. J. (**1999**). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am., **105**, 339-346. doi: 10.1121/1.424503.

Vliegen, J., Moore, B.C.J., and Oxenham, A. J. (**1999**). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," J. Acoust. Soc. Am., **106**, 938-945. doi: 10.1121/1.427140

Acoustic match to electric pulse trains in single-sided deafness cochlear implant recipients

JEREMY MAROZEAU[1,*], MARINE ARDOINT[2], DAN GNANSIA[2], AND DIANE S. LAZARD[3]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Oticon Medical CI Scientific Research, Vallauris, France*

[3] *Institut Arthur Vernes, ENT surgery, Paris, France*

Ten cochlear implant users with single-sided deafness were asked to vary the parameters of an acoustic sound played to their normal-hearing ear, in order to match its perception with that of the electric sensation of two electrodes (e14 and e20). The experiment was divided into 3 consecutive conditions in which the parameters of the acoustic sound varied. The participants had to vary i) the frequency of a pure tone (Exp. 1), ii) the center frequency and the bandwidth of a filter applied to a harmonic complex sound (Exp. 2), and iii) the based frequency (Fb) and the inharmonicity factor of a complex sound (Exp. 3). The results were averaged across participants, and compared within conditions. The pitch sensation for e14 and e20 was significantly different (Exp. 1). In Exp. 2, only the center frequencies of the band-pass filters were significantly different, not the bandwidth. In Exp. 3, the average F0s were not significantly different; The inharmonicity factor was 1.7 for both electrodes. The results of this study suggest that the sound sensation of different electrodes is more linked to a difference in timbre (brightness) than to a difference in pitch, and that the sound is more similar to an inharmonic complex sound than to a pure tone or a white noise.

INTRODUCTION

Cochlear implants (CI) can restore auditory perception in severely and profoundly deaf patients by bypassing the deficient auditory cells and electrically stimulating the auditory nerve. Over the years, technological upgrades and new coding strategies have improved speech perception and overall sound quality. Although CIs are nowadays widely used and can successfully restore speech perception, it is still unclear how the electric stimulation actually sounds like.

Vocoders have been developed to mimic the information provided to the CI user. Simulations with less than 8 channels presented to normal hearing listeners provide speech intelligibility scores in the same range as CI patients (Blamey *et al.*, 1984; Shannon *et al.*, 1995). Despite this good correspondence, some researchers argue that

*Corresponding author:jemaroz@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.

ISBN: 978-87-990013-6-1.

the vocoded information does not offer the same sound quality as that of CIs and suggest the existence of perceptual and informational discrepancies between CI stimulation and performance-matched acoustical simulations (Aguiar *et al.*, 2016; Mesnildrey *et al.*, 2016). Thus, similar level of performance obtained for both real and simulated CI may hide different patterns of errors, limiting the validity of acoustic simulations through vocoders to evaluate new coding strategies.

Some studies tried to match the perception evoked by the CI with an acoustic sound played to the non-implanted ear. Most studies focused on pitch perception (pure tone) in CI patients with residual hearing (bimodal rehabilitation) or normal hearing in the non-implanted ear (single-sided deafness) (for example Carlyon *et al.*, 2010; McDermott *et al.*, 2009), offering a valuable insight on the effect of mismatch between the frequency allocation table of the CI processor and the actual placement of the electrode-array along the cochlea. However, in the late 70s, Eddington *et al.* (1978) evoked that the sound sensation of an electric stimulation was rather a complex sound than a pure tone. Recently, this hypothesis was tested in CI users with residual hearing in the non-implanted ear (Lazard *et al.*, 2012). By modifying the fundamental frequency (F0), bandwidth, centre frequency, and the inharmonicity of the acoustic stimulus, it resulted that the percept given by the stimulation of a single apical electrode did not correspond either to a white noise or a pure tone, but more to an inharmonic complex signal. However, the "reference" ear being impaired (average hearing thresholds between the participants: 65 dB at 500 Hz), the matched acoustic sounds may have been distorted. With the emergence of patients implanted with a normal ear on the contralateral side, we had the opportunity to reproduce and extend to a more basal electrode this latter result. Our aim was to find a more precise and realistic acoustic match of a single pulse train played to an apical and a medial electrode, in patients with single-sided deafness, i.e., with a normal ear used as reference.

METHODS

Participants

Twenty-six adults with a dead ear were enrolled in a larger study about unilateral cochlear implantation. A sub-sample (n=10) was randomly selected from two French centres that participated in the whole study (Hôpital Rothschild, Paris, and Hôpital Ponchaillou, Rennes). The two projects conformed to The Code of Ethics of the World Medical Association (Declaration of Helsinki), and were approved by the Ethic committee of CPP Ile de France V. Each participant, enrolled in the present study, signed an informed consent form about the main project, and about this supplementary protocol. The experimental design of this study was largely inspired from Lazard et al. (2012).

The participants were all tested in a sound attenuated booth, they had normal or near-normal hearing in the non-implanted ear; The implanted ear was a dead ear responsible for severe tinnitus. The average hearing threshold of the non-implanted ear, calculated from the pure tone audiometric thresholds at 500, 1000, 2000, and 4000

Hz, was 24 dB ± 7 standard deviation. For six participants, testing was done after 3 months and 12 months of CI use. One participant performed the three-month session only, and the three remaining participants performed the twelve-month session only. All participants were users of Oticon Medical devices (internal part: Digisonic SP EVO, with the Saphyr Neo SP speech processor and Crystallis XDP sound-processing strategy).

Stimuli

All auditory stimuli were created using the software MAX (Cycling '74 ®), which also provided the experimental interface and enabled data collection. The CI sound processor was linked to a PC laptop via a direct connection. The electric stimulus was a pulse train with an overall duration of 900 ms, including a 100-ms ramp up and a 300-ms ramp down in level, delivered through electrodes 20 and 14 (e20 and e14), representing the most apical electrode and a medial electrode of the Oticon medical device. The stimulation rate was set to the user's regular rate of 500 pps. Each pulse was composed of an active monophasic and a balanced passive discharge using a multi-mode grounding stimulation mode (combination of 20% monopolar and 80% common ground). Acoustic stimuli were presented via insert earphones (EtymoticH, ER-4P) to the non-implanted ear. The acoustic and electric stimuli shared the same temporal envelope.

Procedure

The electric and acoustic stimulus were alternatively presented every second. The electric stimulus was fixed, and the acoustic could be varied by the participants. Their task was to find the acoustic sound that matched as similar as possible the perception of the electric stimulus. A graphical interface (Bamboo Fun pen, Wacom®) was used to adjust the acoustic signal parameters within a multi-dimensional space. The position of the pen (on virtual x and y axes) varied the incoming acoustic signal by simultaneously adjusting the values of one to two selected parameters (see below). The parameters selected at the end of one experiment were used to create the stimuli applied to the following experiment, within one trial.

The study was divided into three experiments during which different acoustic parameters were varied:

Experiment 1: Frequency of a pure tone

The participant were asked to match the frequency of a pure tone (range: 40-2200 Hz) with that of the electric stimulus. The axis (x or y) driving the F0 change varied across trials, the displacement along the other axis did not affect the tone or any other parameter.

Experiment 2: Harmonic complex tone bandpass filtered

An 11-harmonic complex sound was generated. Its fundamental frequency (F0) was the one selected during Exp. 1. This sound was filtered through a symmetrical bandpass filter. One axis controlled the centre frequency, CF, ranging from F0 to 10 times the F0 on a logarithmic scale. The other axis controlled the Q factor of this filter band, ranged from 1.4 to 100 on a logarithmic scale. The Q factor characterizes the bandwidth (Δf) of the filter relative to its centre frequency: Q = CF/Δf. Therefore, a high Q value results in a sound with a relatively small number of harmonics, whereas a low Q value results in a more complex sound.

Experiment 3: Inharmonic complex sound bandpass filtered

An 11-component complex sound was generated and filtered through the output filter selected at the end of Exp. 2. One axis controlled the based frequency (Fb) of the sound (range: 40 to 2200 Hz on a logarithmic scale), while the other axis controlled a parameter referred to as inharmonicity, i. The composite acoustic signal comprised components with frequencies defined by: $Fn=Fb*n^i$, where Fn was the frequency of each component (i.e., n was numbered 1-11), and i was the inharmonicity exponent, ranging from 0.5 to 2.8 on a linear scale. When i = 1 or 2, the sound was harmonic. Relative to this value, lower values of i resulted in a compression of the inter-component frequency spacing whereas higher values resulted in an expansion of the inter-component spacing.

Protocol

First, the presentation level of the electric stimulus was set to be comfortable by the experimenter. Then the level of the acoustic signal was adjusted to match the loudness of the electric stimulus and could be modified along the trials. Participants were first familiarized with the interface, and trained during one trial. Subsequently, Exp. 1, Exp. 2, and Exp. 3 were presented in that order, and repeated 3 times in total. In order to reduce any tendency of participants to return to the same spatial position on the interface and thereby bias the results, the settings of the interface were randomly modified before each trial of each condition by interchanging the axes (x becoming y and vice versa), and by adding offsets to the origin of the axes (up to 20% shift on each axis). The instructions were to modify the acoustic sound to create a perception similar to the perceived electric sensation. This procedure was repeated at 3 and 12 months after the first fitting, for the two electrodes.

RESULTS

Experiment 1: Frequency of a pure tone

Figure 1 shows all the individual matches for the first experiment. A mixed linear model was performed with all the individual matches as independent variables, the electrodes, the sessions and its interaction as fixed effect and the participants as random effects. Only the main effect of electrode was found significant [F(1,7.369)=8.5391, p=0.021]. On average the sensation induced by e20 matched a tone with a frequency of 506 Hz and that of e14 matched a tone with a frequency of 901 Hz. No significant effect was observed for the main factor session [F(1,6.164)=1.8367, p=0.2229] nor its interaction with the electrode [F(1,4.951)=0.2509, p=0.6379].

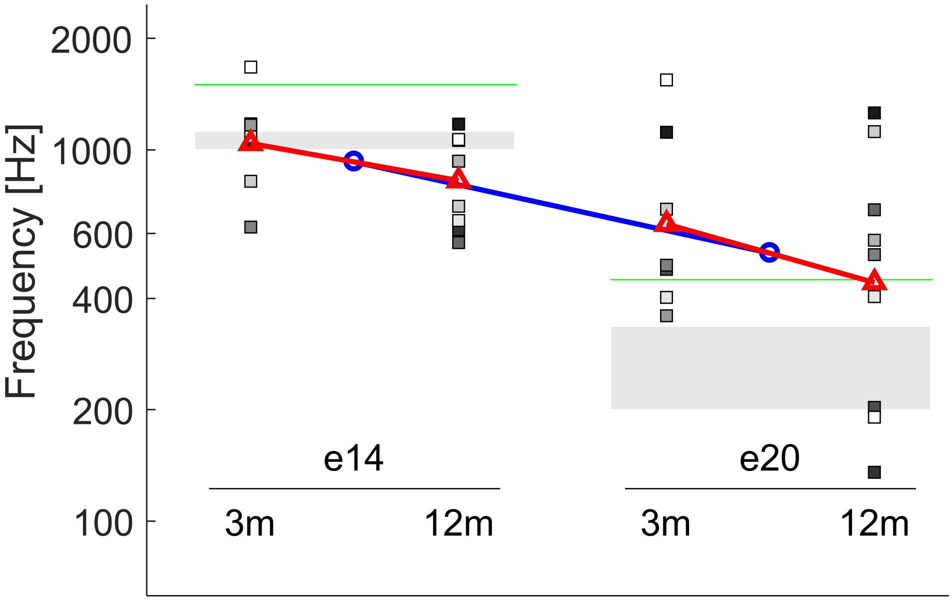


Fig. 1: Results of Exp. 1. Individual (squares) and average (red triangles and blue circles) results for the frequency matching, for e14 (left) and e20 (right) at 3 and 12 months after activation. The red triangles represent the average frequency per electrode and per session. The blue circles represent the average frequency of each electrode across both sessions. The horizontal green lines represent the supposed frequency based on the approximated place of the electrode (derived from the Greenwood function). The gray boxes outline the acoustic frequency band allocated to each electrode by the manufacturer.

Experiment 2: Harmonic complex sound bandpass filtered

The results of the characteristics of the filter applied in Exp. 2 are shown in Fig. 2. The average center frequency for the filter was 2850 Hz for e14 and 864 Hz for e20 (Fig. 2, left panel, blue circle). This difference was significant [$F(1,8.542)=18.5543$, $p=0.002$]. Similarly to Exp. 1, there was no session effect [$F(1,8.694)=0.92$, $p=0.36$], nor interaction with the electrode [$F(1,7.253)=0.001$, $p=0.95$]. The average Q factor was similar for the two electrodes, and between sessions ($p>0.05$): 5.97 and 6.21 (Fig. 2, right panel, blue circle).

Experiment 3: Inharmonic complex sound bandpass filtered

When the filter selected during Exp. 2 was applied to a complex sound**,** the task consisting of varying Fb gave similar results on average between e14 and e20 (Fig. 3 left panel, blue circles, Fb=433 Hz and 307 Hz, respectively, no statistical difference). However, an effect of session was observed, with a lower Fb for both electrodes between 3 and 12 months of CI use [$F(1,7,042)=6.7421$, $p=0.0354$]. The average results for the inharmonicity factor were also similar (n=1.77 and 1.74, respectively, no statistical difference).

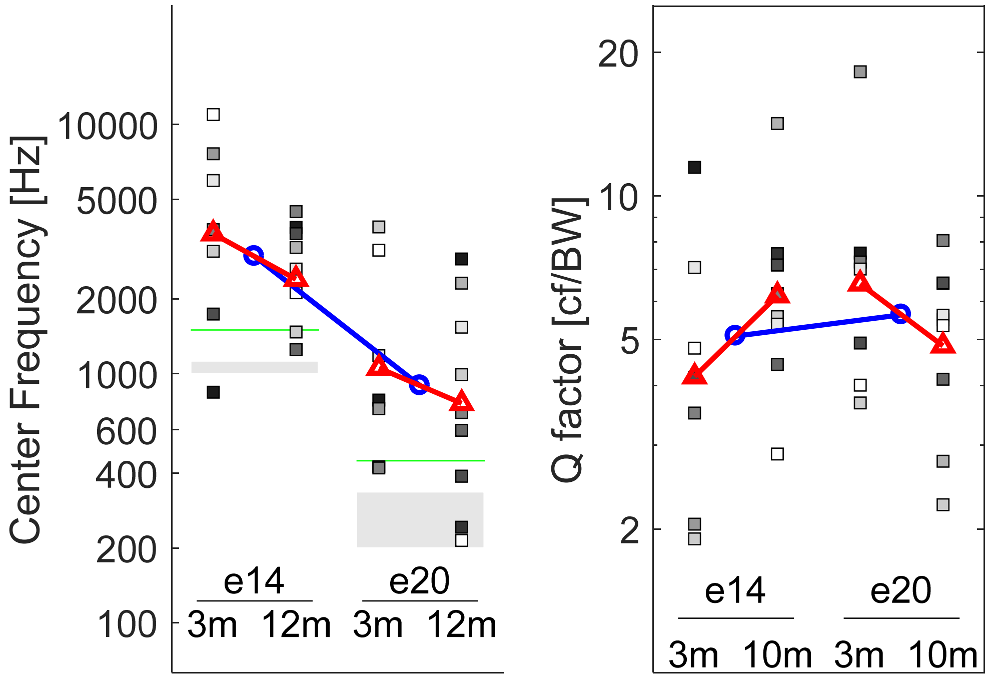


Fig. 2: Results of Exp. 2. Individual results for the center frequency and Q factor (left and right panel respectively) of the applied filter. See caption of Fig. 1 for more information.

DISCUSSIONS AND CONCLUSIONS

These experiments were designed to find an acoustic match for a single pulse train of an apical and a medial electrode. The results indicate that the best match was obtained with an inharmonic complex sound with a bright timbre for both electrodes. Exp. 1 indicates that electrode position influenced the match with a pure tone (the lower the pith, the more apical the position) and was stable after 3 months of CI use in our sample. Neither the Greenwood function nor the frequency allocation band correctly predicted what participants described. In Exp. 2, the average centre frequency of the filters matching users' perception induced by e20 and e14 was 864 Hz and 2850 Hz, respectively. As the centre frequency of a symmetrical spectrum can be considered a physical descriptor of the perceptual dimension of brightness (McAdams *et al.*, 1995), this result shows that a pulse train delivered at e14 was perceived with a brighter timbre than the same pulse train delivered at e20. In Exp. 3, the frequency of each component was set by the based frequency and the inharmonicity factor. As the resulting sound was inharmonic, the based frequency did not predict the pitch. Taken together however, the based frequency and inharmonicity influenced the tonality of the sound. Because no significant effect of electrode place was found for these parameters, it can be concluded that the electrode place influences the timbre rather than the tonality of a pulse train. This result challenges the concept of *place pitch* in cochlear implant.

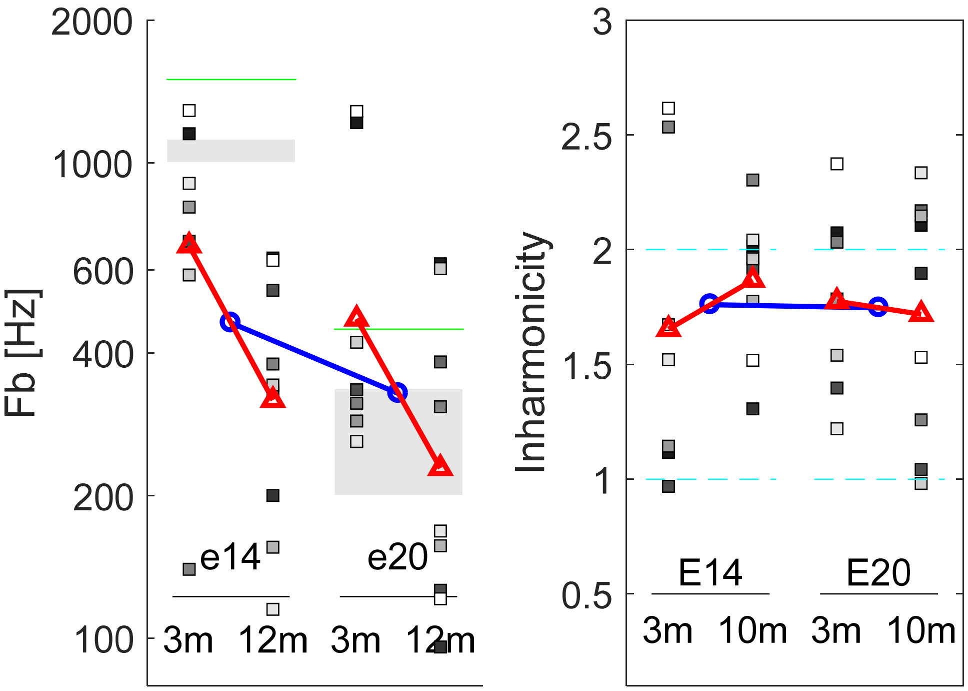


Fig. 3: Results of Exp. 3. Individual results for Fb and the inharmonicity factor (left and right panel respectively). See caption of Fig. 1 for more information. The doted lines indicate an inharmonicity factor of 1 and 2 (i.e., a harmonic sound).

ACKNOWLEDGMENTS

We would like to thank the participants for their time and enthusiasm, Dr. Christine Poncet, Prof. Bruno Frachet and Prof. Benoît Godey for giving access to their patients, and Bertrand Philippon for helping in the data collection.

REFERENCES

Aguiar, D.E., Taylor, N.E., Li, J., Gazanfari, D.K., Talavage, T.M., Laflen, J.B., Neuberger, H., *et al.* (**2016**). "Information theoretic evaluation of a noiseband-based cochlear implant simulator," Hear. Res., **333**, 185-193. doi: 10.1016/j.heares.2015.09.008.

Blamey, P.J., Dowell, R.C., Tong, Y.C., Brown, A.M., Luscombe, S.M., and Clark, G.M. (**1984**). "Speech processing studies using an acoustic model of a multiple-channel cochlear implant," J. Acoust. Soc. Am., **76**, 104-110. doi: 10.1121/1.391104

Carlyon, R.P., Macherey, O., Frijns, J.H., Axon, P.R., Kalkman, R.K., Boyle, P., Baguley, D.M., *et al.* (**2010**). "Pitch comparisons between electrical stimulation of a cochlear implant and acoustic stimuli presented to a normal-hearing contralateral ear," J. Assoc. Res. Otolaryngol., **11**, 625-640. doi: 10.1007/s10162-010-0222-7

Eddington, D.K., Dobelle, W.H., Brackmann, D.E., Mladejovsky, M.G., and Parkin, J.L. (**1978**). "Auditory prostheses research with multiple channel intracochlear stimulation in man," Ann. Otol. Rhinol. Laryngol., **87**, 1-39.

Lazard, D.S., Marozeau, J., and McDermott, H.J. (**2012**). "The sound sensation of apical electric stimulation in cochlear implant recipients with contralateral residual hearing," PLoS One, **7**, e38687. doi: 10.1371/journal.pone.0038687

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (**1995**). "Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes," Psychol. Res., **58**, 177-192.

McDermott, H., Sucher, C., and Simpson, A.M. (**2009**). "Electro-acoustic stimulation: Acoustic and electric pitch comparisons," Audiol. Neurootol., **14**, 2-7. doi: 10.1159/000206489

Mesnildrey, Q., Hilkhuysen, G., and Macherey, O. (**2016**). "Pulse-spreading harmonic complex as an alternative carrier for vocoder simulations of cochlear implants," J. Acoust. Soc. Am., **139**, 986-991. doi: 10.1121/1.4941451

Shannon, R.V, Zeng, F.G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science, **270**, 303-304.

Data-driven approach for auditory profiling

RAUL SANCHEZ[1,*], FEDERICA BIANCHI[1], MICHAL FERECZKOWSKI[1], SÉBASTIEN SANTURETTE[1,2], AND TORSTEN DAU[1]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Department of Otorhinolaryngology, Head and Neck Surgery & Audiology, Rigshospitalet, Copenhagen, Denmark*

Nowadays, the pure-tone audiogram is the main tool used to characterize hearing loss and to fit hearing aids. However, the perceptual consequences of hearing loss are typically not only associated with a loss of sensitivity, but also with a clarity loss that is not captured by the audiogram. A detailed characterization of hearing loss has to be simplified to efficiently explore the specific compensation needs of the individual listener. We hypothesized that any listener's hearing can be characterized along two dimensions of distortion: type I and type II. While type I can be linked to factors affecting audibility, type II reflects non-audibility-related distortions. To test our hypothesis, the individual performance data from two previous studies were re-analyzed using an archetypal analysis. Unsupervised learning was used to identify extreme patterns in the data which form the basis for different auditory profiles. Next, a decision tree was determined to classify the listeners into one of the profiles. The new analysis provides evidence for the existence of four profiles in the data. The most significant predictors for profile identification were related to binaural processing, auditory non-linearity and speech-in-noise perception. The current approach is promising for analyzing other existing data sets in order to select the most relevant tests for auditory profiling.

INTRODUCTION

Supra-threshold distortions can be defined as the abnormal perception of stimuli when these are audible, i.e., above the audiometric threshold. While amplification can effectively compensate for the loss of sensitivity, supra-threshold distortions may require more advanced signal processing to improve hearing, particularly, when the hearing-aid user holds a conversation in noisy environments (Kollmeier and Kiessling, 2016; Plomp, 1978). Several studies have attempted to shed light on the underlying mechanisms responsible for these distortions by means of correlations of psychoacoustic tests with speech in noise performance (Glasberg and Moore, 1989; Strelcyk and Dau, 2009; Johannesen *et al.*, 2016). While these tests could provide a better characterization of the hearing deficits, they are infeasible in a clinical set up.

[*] Corresponding author: rsalo@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

A recent study (Thorup *et al.*, 2016) proposed a test battery of new outcome measures used in a clinical set up for hearing aid candidates. Furthermore, Johannesen *et al.* (2016) investigated the influence of cochlear mechanical dysfunction, as well as the effects of temporal processing deficits and age on speech intelligibility in hearing-impaired (HI) listeners while wearing hearing aids (HA). The goal of the present study was to investigate how the characterization of hearing loss can be simplified by means of auditory profiling. For this purpose, a new analysis of these two datasets using archetypal analysis was performed. This analysis represents a useful tool for identifying patterns in the data and it has been proposed for prototyping and benchmarking (Ragozini *et al.*, 2017). The advantage of the proposed method is that the analysis involves the performance of the patient in different tests rather than correlations or regression analysis (Glasberg and Moore, 1989; Johannesen *et al.*, 2016).

Hearing deficits and auditory profiles

The characterization of hearing deficits can be complex and it needs to be simplified to efficiently explore the specific compensation strategies for the individual. Several authors have suggested classifications of the listeners to reduce this complexity. Three of such approaches served as inspiration for the hypothesis of the present study:

1. I: Plomp (1978) suggested that the hearing deficits in relation to speech intelligibility can be divided into two components: an attenuation (A) and distortion (D) component.
2. II: Lopez-Poveda (2014) reviewed the mechanisms that produce hearing loss and their perceptual consequences for speech. In a bi-dimensional space, the hearing loss can be understood as the sum of an outer-hair-cell (OHC) loss and an inner-hair-cell (IHC) loss.
3. III: Dubno *et al.* (2013) suggested four auditory phenotypes for explaining age-related hearing loss based on animal studies. Although this can be informative revealing the origin of the hearing loss, pure-tone audiometry has remained the main contributor in this classification and no information about speech or other tasks were considered.

Here, we hypothesize that any listener's hearing can be characterized along two dimensions: distortion type I and distortion type II (Fig. 1). While distortion type I can cause a loss of audibility, distortion type II is considered to reflect a non-audibility-related distortion, referred to as a *clarity loss*. In this space, four profiles may be identified: a sensitivity loss (Profile A), a sensitivity loss with associated distortions (Profile B), a sensitivity loss with a severe clarity loss (Profile C) and a mild-moderate clarity loss (Profile D). Figure 1 shows the four profiles in the two-dimensional space, where normal-hearing (NH) listeners are placed at the bottom-left corner since they would not exhibit any type of distortion. It is hypothesized that while the distortions of type I (in the vertical dimension) are due to a loss of frequency selectivity and a

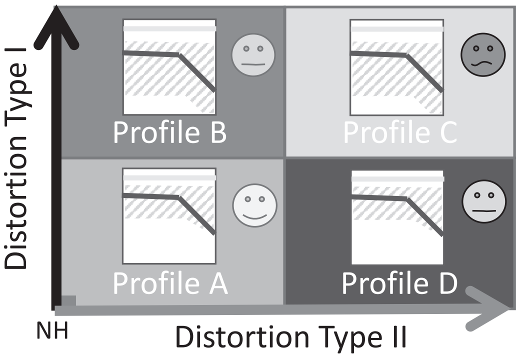


Fig. 1: Sketch of the hypothesis. Hearing deficits can happen due to two types of distortions that are independent. Distortion type I: distortions with consequent loss of sensitivity. Distortion type II: non-audibility related distortions. Profile A: sensitivity loss. Profile B: sensitivity loss and distortion. Profile D: moderate clarity loss and Profile C: severe clarity loss.

loss of cochlear compression, the distortions of type II (in the horizontal dimension) are related to inaccuracies in terms of temporal coding, in line with the conclusions from other studies (Glasberg and Moore, 1989; Strelcyk and Dau, 2009).

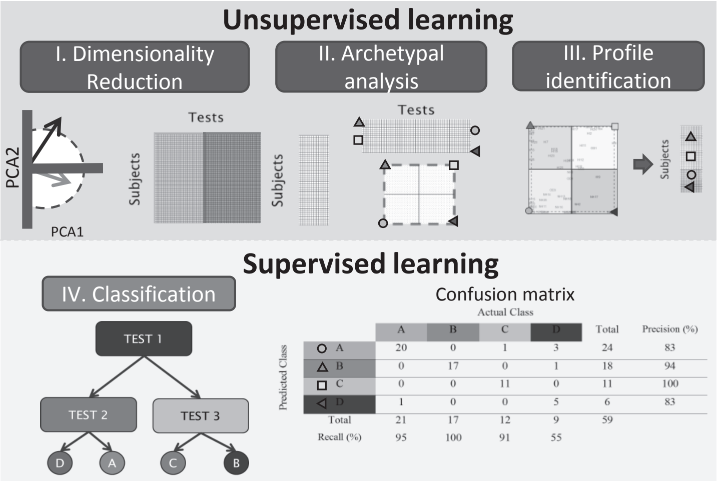
The aim of the present study was to investigate whether listeners can be grouped in the four different profiles by identifying trends in the results from the behavioral tests using a data-driven approach. The analysis is expected to help identify the underlying mechanisms and perceptual consequences of the hearing deficits that characterize an individual auditory profile. Moreover, it is of interest to reduce the test battery using only the most relevant tests for classifying the subjects in the proposed auditory profiles.

METHOD

The method used in the analysis is depicted in Fig. 2. The data-driven approach is based on two stages. First, *unsupervised learning* was used to identify the trends in the data that can be used to categorize the subjects in different profiles. The second stage consisted of *supervised learning*. Once the subjects were segmented by profiles, the data were analyzed again to find the best classification structure that can predict the identified profile.

Unsupervised learning

Unsupervised learning aims to identify patterns occurring in the data, where the output is unknown and the statistical properties of the whole dataset are explored. In contrast to linear regression, unsupervised learning does not aim to predict a specific output, for example, speech intelligibility. In the present approach, identified auditory profiles were eventually inferred using different unsupervised learning techniques.



	A	B	C	D	Total	Precision (%)
○ A	20	0	1	3	24	83
△ B	0	17	0	1	18	94
□ C	0	0	11	0	11	100
◁ D	1	0	0	5	6	83
Total	21	17	12	9	59	
Recall (%)	95	100	91	55		

Fig. 2: Sketch of the method. Upper panel shows the supervised learning techniques applied to the whole dataset.The bottom panel shows the supervised learning, which uses the original data as the input and the identified profiles from the archetypal analysis as the output.

I. *Dimensionality reduction:* According to our hypothesis, two dimensions, corresponding to the two types of distortions, should be sufficient for auditory profiling. Therefore, the subset of variables that were strongly correlated to the first principal component (PCA1) and the second principal component (PCA2) and can explain most of the variance were chosen for the next step. The optimal number of variables in each of the two principal components was chosen using a leave-one-out cross-validation in an iterative principal component analysis (PCA).

II. *Archetypal analysis:* This technique combines characteristics of matrix factorization and cluster analysis. The aim of the analysis was to identify extreme patterns in the data (archetypes). This has the advantage that the subjects are no longer defined by the quantified performance in each of the tests, but by their similarity to the extreme exemplars contained in the data, i.e., the *archetypes*.

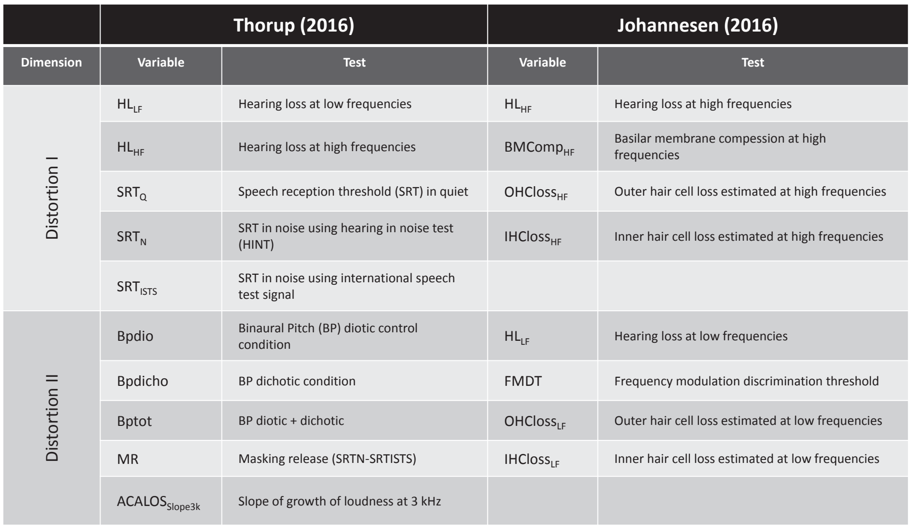
III. *Profile identification:* Based on the archetypal analysis, the subjects were placed in a simplex plot (square visualization). Here, the archetypes are located at each corner and the subjects are placed in the two-dimensional space according to the distance to each archetype. In the present analysis, it was assumed that the subjects placed close to an archetype would belong to that cluster. Consequently, each subject was labelled with the letter of an auditory profile according to the nearest archetype.

Supervised learning

Once the profiles have been identified, supervised learning can be performed. Now, the joint probability density of the dataset and the output (the identified profiles) can be used to select the tests that are most relevant for the classification of the subjects in the auditory profiles.

IV. *Classification:* Decision trees are able to predict the class that corresponds to a given observation. Here, each relevant test is used in the nodes forming a logical expression and dividing the observations accordingly (Fig. 2 IV). A classification tree needs to be trained with a subset of the data and a known output. In the present study, the identified auditory profiles (III) were used as the response variable and a 5-fold cross-validation was used for training the classifier.

RESULTS



	Thorup (2016)		Johannesen (2016)	
Dimension	Variable	Test	Variable	Test
Distortion I	HL_{LF}	Hearing loss at low frequencies	HL_{HF}	Hearing loss at high frequencies
	HL_{HF}	Hearing loss at high frequencies	$BMComp_{HF}$	Basilar membrane compession at high frequencies
	SRT_{Q}	Speech reception threshold (SRT) in quiet	$OHCloss_{HF}$	Outer hair cell loss estimated at high frequencies
	SRT_{N}	SRT in noise using hearing in noise test (HINT)	$IHCloss_{HF}$	Inner hair cell loss estimated at high frequencies
	SRT_{ISTS}	SRT in noise using international speech test signal		
Distortion II	Bpdio	Binaural Pitch (BP) diotic control condition	HL_{LF}	Hearing loss at low frequencies
	Bpdicho	BP dichotic condition	FMDT	Frequency modulation discrimination threshold
	Bptot	BP diotic + dichotic	$OHCloss_{LF}$	Outer hair cell loss estimated at low frequencies
	MR	Masking release (SRTN-SRTISTS)	$IHCloss_{LF}$	Inner hair cell loss estimated at low frequencies
	$ACALOS_{Slope3k}$	Slope of growth of loudness at 3 kHz		

Table 1: Result from the dimensionality reduction of the two datasets. Variables strongly correlated to PCA1 (distortion type I) and PCA2 (distortion type II).

The whole dataset was reduced to the variables that were strongly correlated to Dimension I (PCA1) or Dimension II (PCA2), as summarized in Table 1. In Thorup *et al.*'s study, the dimensionality reduction revealed that the binaural tests were orthogonal to most of the tests related to audibility. Principal component analysis could explain 83.82% of the variance with two components. In Johannesen *et al.*'s study, Dimension II seemed to be dominated by low-frequency processing and Dimension I by high-frequency processing. PCA could explain 67% of the variance with the chosen variables.

The archetypal analysis was used to identify four archetypes using the variables from Table 1. As shown in Fig. 3, in both studies, Profile A (archetype A) exhibits the best performance in both dimensions and Profile C the worst. Profile B shows poor performance only in Dimension I while Profile D shows poor performance only in Dimension II. Based on these archetypes, each listener was considered to belong to the auditory profile of the closest archetype. Figure 3 B1 (Thorup *et al.*'s dataset) illustrates how the listeners are divided in clusters in the two-dimensional study. However, Figure 3 B2 (Johannesen *et al.*'s dataset) shows how the listeners are spread out among the profiles.

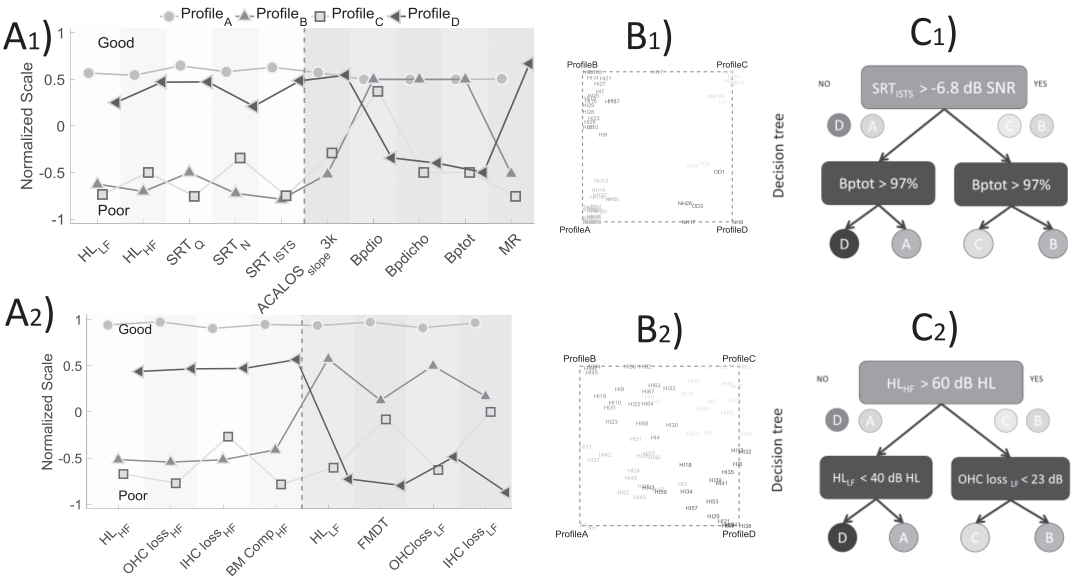


Fig. 3: Archetypes: Extreme exemplars of the different patterns found in the data. A1) Normalized performance of each of the 4 archetypes from Thorup *et al.*'s study. B1) Simplex representation of the listeners of Thorup *et al.*'s study. C1) Decision tree result of the supervised learning. A2, B2, and C2) are the same as A1, B1, and C1) but for Johannesen *et al.*'s study.

Decision trees were obtained by using the raw data as an input and the auditory profiles as the output. In Thorup *et al.*'s study, the classification tree based on $\mathrm{SRT_{ISTS}}$ and binaural pitch showed a very high precision (58 out of 59 true positives). In Johannesen *et al.*'s study, the classification was based not only on the audibility loss at high and low frequencies but also the estimate of OHC loss at low frequencies. The precision of this classifier was lower (66%).

DISCUSSION

The hypothesis in terms of the proposed four auditory profiles was evaluated in a data-driven approach. It was assumed that one dimension (distortion type I) may be related to a reduced frequency selectivity, while the second dimension would be related to temporal processing reflecting to a non-audibility related distortions (distortion type II). First, the analysis of a dataset with a population of near-normal-hearing and hearing-impaired listeners revealed that binaural processing tests were highly sensitive for the classification of the listeners and a main contributor in the distortion type II dimension. Second, the analysis of a dataset with only hearing-impaired listeners showed that the distortion type I was related to high-frequency processing and the distortion type II was related to low-frequency processing. The two analyses cannot directly be compared because the tests and the listeners differed across studies. It should be noted that Johannesen *et al.*'s study did not consider any test concerning binaural processing. Therefore, the difference in the explained variance across the two studies is mostly caused by the lack of such information in their study.

Pure-tone audiometric thresholds are used to quantify the hearing loss but they can, in fact, be the consequence of different factors. As shown in Fig. 3 A2, dimension I does not only contain the high-frequency hearing loss but also estimated cochlear compression. Dimension II contains the low-frequency hearing loss and the outcome of the frequency modulation detection task which has been suggested to reflect temporal processing abilities. Therefore, it is important to bear in mind that there are interactions between the audibility and the two types of distortions proposed here. One approach to disentangle this interaction may be made based on the effects related to the OHC vs IHC processing.

If a substantial population of IHC or neural fibers is affected, the thresholds can be elevated (Lobarinas *et al.*, 2013), leading to temporal distortions as well as degraded binaural processing (Profiles D and C). However, the temporal acuity can also be compromised while audiometric thresholds are normal or close-to-normal (Zeng *et al.*, 1999) (Profile D). OHC loss is typically associated with basilar membrane (BM) compression loss (reduced frequency selectivity) as well as elevated audiometric thresholds (Ahroon *et al.*, 1993). Although reduced compression leads to a threshold elevation (Profile B), listeners with elevated thresholds can still have a nearly-normal BM compression (Profile A).

It is likely that the two types of deficits (degraded frequency selectivity vs degraded temporal processing) affect speech perception in different ways. The signal processing strategies that can be applied to compensate for each type of impairment can be assumed to be different. Therefore, both loss of audibility and outcome measures reflecting spectral and temporal distortions should be part of a clinical test battery for characterizing hearing deficits.

Conclusion

The new analysis provides consistent evidence of the existence of two sources of distortion and different 'auditory profiles' in the data. While distortion type I was more related to audibility loss at high frequencies, the origin of distortion type II was connected to reduced binaural processing abilities and low-frequency hearing loss. The most informative predictors for the profile identification beyond the audiogram were related to temporal processing, binaural processing, compressive peripheral nonlinearity, and speech-in-noise perception. The current approach can be used to analyze other existing data and might help define a test battery to achieve an efficient auditory profiling.

REFERENCES

Ahroon, W.A., Davis, R.I., and Hamemik, R.P. (**1993**). "The role of tuning curve variables and threshold measures in the estimation of sensory cell loss," Audiology, **32**, 244-259. doi: 10.3109/00206099309072940

Dubno, J.R., Eckert, M.A., Lee, F.S., Matthews, L.J., and Schmiedt, R.A. (**2013**). "Classifying human audiometric phenotypes of age-related hearing loss from animal models," J. Assoc. Res. Otolaryngol., **14**, 687-701. doi: 10.1007/s10162-013-0396-x

Glasberg, B.R., and Moore, B.C. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl., **32**, 1-25.

Johannesen, P.T., Pérez-González, P., Kalluri, S., Blanco, J.L., and Lopez-Poveda, E.A. (**2016**). "The influence of cochlear mechanical dysfunction, temporal processing deficits, and age on the intelligibility of audible speech in noise for hearing-impaired listeners," Trends Hear., **20**. doi: 10.1177/2331216516641055

Kollmeier, B., and Kiessling, J. (**2016**). "Functionality of hearing aids: State-of-the-art and future model-based solutions," Int. J. Audiol., 1-26. doi: 10.1080/14992027.2016.1256504

Lobarinas, E., Salvi, R., and Ding, D. (**2013**). "Insensitivity of the audiogram to carboplatin induced inner hair cell loss in chinchillas," Hear. Res., **302**, 113-120. doi: 10.1016/j.heares.2013.03.012

Lopez-Poveda, E.A. (**2014**). "Why do I hear but not understand? Stochastic undersampling as a model of degraded neural encoding of speech," Front. Neurosci., **8**. doi: 10.3389/fnins.2014.00348

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am., **63**, 533-549. doi: 10.1121/1.381753

Ragozini, G., Palumbo, F., and D'Esposito, M.R. (**2017**). "Archetypal analysis for data-driven prototype identification. Stat. Anal. Data Min., **10**, 6-20. doi: 10.1002/sam.11325

Strelcyk, O., and Dau, T. (**2009**). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," J. Acoust. Soc. Am., **125**, 3328-3345. doi: 10.1121/1.3097469

Thorup, N., Santurette, S., Jørgensen, S., Kjærbøl, E., Dau, T., and Friis, M. (**2016**). "Auditory profiling and hearing-aid satisfaction in hearing-aid candidates," Dan. Med. J., **63**, A5275.

Zeng, F.G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (**1999**). "Temporal and speech processing deficits in auditory neuropathy," Neuroreport, **10**, 3429-3435. doi: 10.1097/00001756-199911080-00031

Preferred listening levels – a silent disco study

RIKKE SØRENSEN[1,*], ELIZABETH BEACH[2,3], MEGAN GILLIVER[2,3] ,
AND CARSTEN DAUGAARD[4]

[1] *Audiology Studies, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark*

[2] *National Acoustic Laboratories, Macquarie University, Australia*

[3] *HEARing Cooperative Research Centre, Melbourne, Australia*

[4] *Technical Audiological-Laboratory, Force Technology, Odense, Denmark*

Aim: To investigate preferred listening levels (PLLs) in a dance situation and compare them to typical sound levels at dance venues (90-105 dB L_{Aeq}). *Method*: Fifty-nine young people had their individually chosen sound levels measured at a silent disco event. In a separate experiment 25 participants set their PLLs for music delivered through headphones and loudspeakers respectively, and repeated measures were conducted to test intra-rater reliability. *Results:* The sound level at the silent disco event was limited to a maximum of 89-93 dB L_{Aeq}. One-third of the 59 participants expressed a preference for louder sound levels while two-thirds were satisfied with this or even softer volumes. PLLs over headphones were on average 2 dB louder than in loudspeaker mode. PLLs varied 0.8-19.1 dB within each participant for the same input, but most participants (84%) showed a personal range of less than 5 dB in 75% of their measures. *Conclusion:* Many patrons' PLLs are noticeably lower than what is typically offered at dance venues.

INTRODUCTION

The contribution of leisure noise exposure to the risk of noise-induced hearing loss (NIHL) is a growing concern in modern society (Johnson *et al.*, 2014; WHO, 2015). One source of high levels of sound exposure is discotheques, which typically offer sound levels of 90-105 dB L_{Aeq} (Tin and Lim, 2000; Sadhra *et al.*, 2002; Cassano *et al.*, 2005). In Australian nightclubs, an average of 98 dB L_{Aeq} was found, with regular attendees spending approximately 5 hours per visit (Beach, 2013).

There is no legislation in place to regulate the sound exposure of patrons on their own time, but acceptable noise limits have been defined for workplace sound environments. Since these regulations are intended to minimize the risk of NIHL, they may be used as reference limits in leisure noise situations. The limits are based on international standards which state that working more than 8 hours at a mean exposure level of 85 dB L_{Aeq} poses a risk of developing NIHL (ISO, 2013; 2014). As such, the

*Corresponding author: tegnsprog@gmail.com

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR),
Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau,
J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
© The Authors. ISBN: 978-87-990013-6-1.

90 dB L_{Aeq} seen in the lower range of dance venue exposure would be acceptable for only 2.5 hours a day, whereas 105 dB L_{Aeq} would be acceptable for less than 5 minutes. It is therefore not surprising that many patrons of high-volume venues report sound-related difficulties. Tinnitus as well as temporary hearing loss were seen in 66-88% of more than 1,000 young respondents following concert or dance venue attendance (Mercier and Hohmann 2002; Johnson *et al.*, 2014).

The general assumption seems to be that patrons prefer the music to be played at high levels. However, research has found that many young partygoers may prefer lower levels. In one study, 90% of 500 regular clubbers self-reported that they would prefer sound levels to be softer than the levels they generally experienced (Beach, 2013), while another found 42% of 700 participants to be similarly inclined (Mercier and Hohmann, 2002). Additionally, 70% of 325 participants in a third study "felt that noise levels in nightclubs should be limited to safe volumes" (Johnson *et al.*, 2014).

So far, these indications of preferred listening levels have been tested using surveys asking for people's preferences in relation to existing volumes. Less is known as to whether such responses truly represent patrons' actual preferred levels, or to what extent they may wish for sound levels to be lowered. This study aimed to investigate these questions via a party concept known as "silent disco", which gave the opportunity to test patrons' PLLs in practice. In a silent disco there is no music in the room at large, but each patron controls their own music volume emitted from a set of headphones wirelessly connected to a transmitter playing the music. As such, patrons' actual PLLs may be measured directly during or following a silent disco event.

However, it has been found that PLL may vary depending upon whether the stimulus is presented through headphones or loudspeakers, with PLLs being quoted to be 3-19 dB louder in headphone mode than in loudspeaker mode (Rudmose, 1982; Brixen, 2001). It was therefore necessary to also compare PLL for music played through the silent disco headphones against that from loudspeakers.

The study thus aimed to better understand young people's PLL in dance venues through two related experiments. The first experiment measured actual PLL under headphones in a silent disco event. The second experiment compared the PLL preferences under headphones to those in nightclub-like loudspeaker environments.

It was hypothesised that a substantial number of participants would choose softer listening levels than the 90-105 dB typically seen at dance venues, and that their PLLs would be louder for presentation through headphones than through loudspeakers.

METHOD

Both experiments had ethics approval from the Australian Hearing Human Research Ethics Committee.

Experiment 1: Silent disco event

Equipment and stimuli

Music was delivered through Samsung tablets connected to three transmitters and 59 silent disco headphones provided by a local supplier. Six popular dance songs were presented in three different sound qualities (unmanipulated, bass boosted and peak clipped) to redirect the participants' attention from sound levels. Sound levels were measured in the laboratory using a Brüel & Kjær type 2250 sound level meter connected to a Kemar mannequin. The maximum presentation level varied with headphone set and was limited to between 89 and 93 dB to comply with ethics requirements.

Participants

Fifty-nine university staff and students were recruited through convenience sampling via social media and at the university bar. Of these, 32 were male and 25 female. One participant chose the "other/unspecified" gender option in the survey, one other did not answer the question. The age span was 19-35 years with a mean age of 23.5 years. The majority (N=46) self-reported having normal hearing while 13 reported that they suspected or knew that they had 'some hearing loss'.

Procedure

Participants were given a personal set of wireless headphones that enabled them to adjust the volume of the shared music individually. To avoid bias in choosing their volume, they were initially told that the study was investigating sound quality preferences, with no mention of sound levels being made. All the participants danced together for 11.5 minutes to popular music presented through the headphones. When the music ended, they returned their headphones to the researchers with the final settings intact and filled out a written questionnaire. The survey included questions on both perceived sound quality and the satisfaction with their chosen sound levels, and the answers were linked to each participant's sound level measured in the laboratory. Duration and maximum sound levels were limited to comply with ethics requirements.

Experiment 2: Comparison study of PLL in headphones vs. loudspeakers

Equipment and stimuli

Excerpts of half a minute from five of the songs from Experiment 1 were used as stimuli. Four songs were presented in three different sound qualities (unmanipulated, bass boosted and peak clipped), giving a total of 12 individual experimental sound files. The fifth song, without sound quality manipulation, was used as a "control" to investigate intra-rater reliability.

The stimuli were presented through a Samsung tablet connected to an attenuator through a Digitor 4-Way Video Switch Box C 2505. The attenuator was connected to both a silent disco transmitter sending to the wireless headphones, and to a Marantz Integrated Stereo Amplifier PM-43 feeding two Tannoy V8 loudspeakers. A dial attached to the attenuator enabled sound levels to be manipulated in real time as the stimulus was played, in increments of 0.1 dB.

Unattenuated sound levels were measured through a Brüel & Kjær type 2250 sound level meter connected to a Kemar mannequin.

Participants

Twenty-five participants (17 female, 8 male) aged 21-36 years (mean: 26 years) and reporting normal hearing were recruited through convenience sampling.

Procedure

In a laboratory setting, participants were instructed to individually set their PLLs for a total of 16 presentations in each mode (headphone and loudspeaker) by manipulating the attenuator. For both modes, the 12 experimental sound files were each presented once, and the control sound file presented four times.

Four different presentation orders were devised to avoid order effects. The participants' chosen attenuation levels were noted and subtracted from the unattenuated maximum sound levels measured through the Kemar mannequin.

RESULTS

Experiment 1: Silent disco event

Participants' results from the silent disco were divided into four groups based on the volume setting they chose during the event. Individual sets of headphones had slightly different maxima, leading to a variation in dB L_{Aeq} within each volume setting group, as seen in Table 1. This table also shows how many people in each volume group were satisfied with the volume or felt that it should have been louder or softer.

Volume setting	Mean (dB L_{Aeq})	Range (dB L_{Aeq})	Total (n)	Were satisfied (n)	Preferred higher (n)	Preferred lower (n)
#1 (loudest)	91.1	3.7	38	17	18	2
#2	85.4	1.2	12	10	2	-
#3	79.6	1.2	5	3	2	-
#4 (softest)	73.4	2.6	4	4	-	-

Table 1: Sound level variation within volume settings and participants' satisfaction with their final sound level.

Of the 59 participants, 38 (64%) were listening to the loudest volume setting (volume #1), the remainder to softer volumes. In the softer volume groups (n=21), 81% were satisfied with the sound levels, as were 45% of the people in volume group #1. In group #1, 47% wanted it louder, 2 wanted it softer and one did not answer this question. Overall, 22 (37%) of the 59 participants wanted the music to be louder than their chosen levels, but only 18 of these were already at maximum loudness.

For the 34 participants (58%) who reported being satisfied with their final volume, the mean PLL was 86.0 dB L_{Aeq}. Overall mean PLL was 87.6 dB L_{Aeq} and mean sound level for those who would have preferred it to be louder was 89.7 dB L_{Aeq}.

Experiment 2: Comparison study of PLL in headphones vs. loudspeakers

The mean PLL for all participants and all songs was 71.3 dB L_{Aeq} in headphone mode and 69.3 dB L_{aeq} in loudspeaker mode. A paired *t*-test yielded a significant difference of 2 dB [*t*(25)=2.92, p=0.007], indicating that preferred levels under headphones were significantly higher than those heard through loudspeakers.

Intra-rater reliability

Measures from the reference song (heard four times across the experiment) showed a noticeable variation in PLLs, with intrapersonal differences in the PLL of 0.8-12.4 dB in headphone mode and 1.5-19.1 dB in loudspeaker mode. This difference was termed consistency range (CR). Seventeen participants (68%) showed CRs above 5 dB in one or both presentation modes (see Fig. 1).

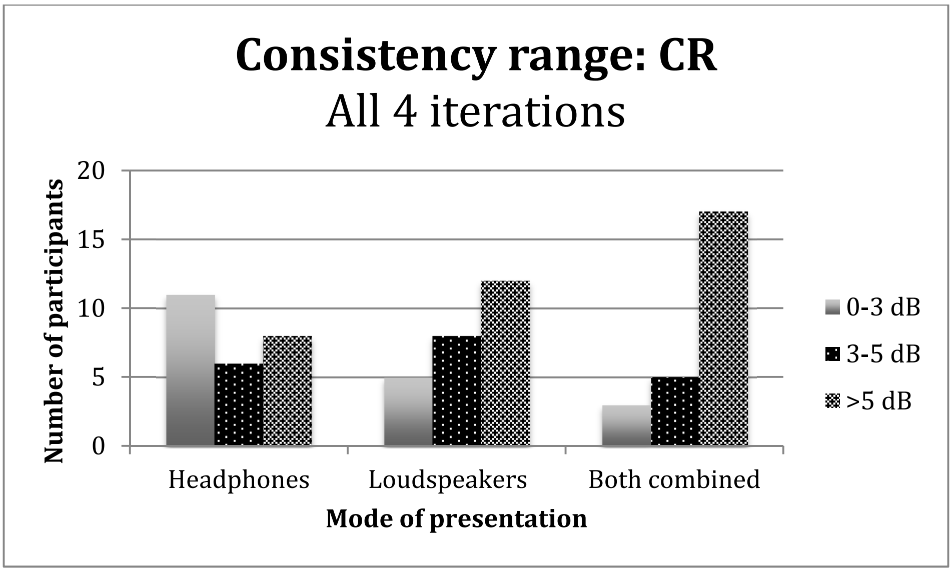


Fig. 1: The number of participants who had 0-3, 3-5 or above 5 dB L_{Aeq} difference between their loudest and softest chosen volume (CR) for the reference song, ordered by presentation mode.

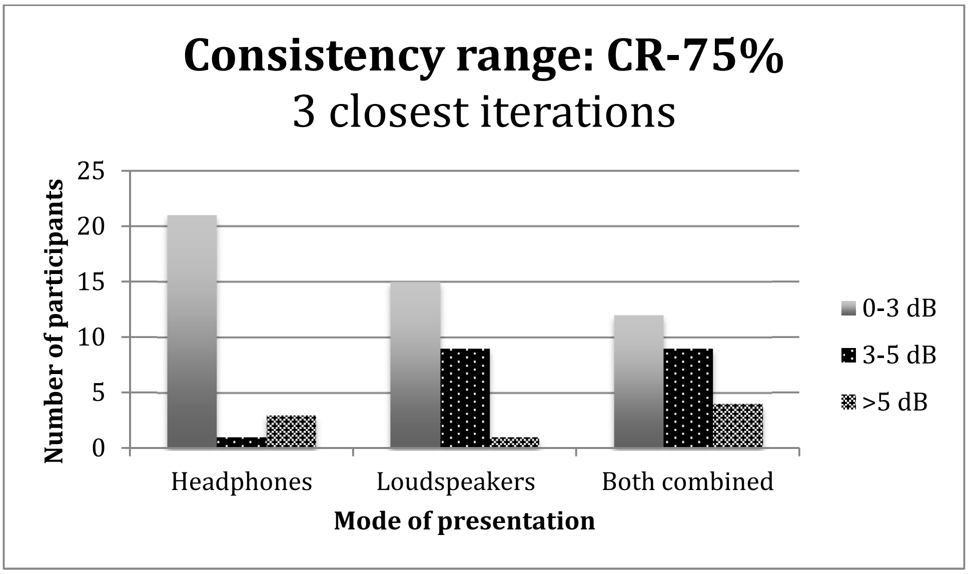


Fig. 2: The number of participants with a difference of 0-3, 3-5 or above 5 between the three PLLs for the reference song that were closest to each other, ordered by presentation mode.

Focusing on the three closest PLLs for each participant yielded CRs for 75% of the presentations, termed CR-75%. Twenty-one participants (84%) had a CR-75% below 5 dB in both presentation modes, and 12 of those (48% of all) were also below 3 dB (see Fig. 2).

DISCUSSION

PLL at dance venues

Two-thirds of the participants at the silent disco event chose the maximum volume of 89-93 dB L_{Aeq}, which is relatively low compared to the 90-105 dB L_{Aeq} usually offered at regular dance venues. The fact that half of these patrons would have preferred the music to be louder is therefore not surprising. However, the other half seemed satisfied with their sound level, as did most of those who deliberately chose even softer listening levels. This strongly indicates that two-thirds of the entire sample did in fact have lower preferred listening levels than what is typically offered at regular dance venues.

A few in the loudest volume group reported wanting softer listening levels while some in the softer volume groups said they would have preferred louder levels. It is unclear why these individuals did not simply adjust their individual volume controls to accommodate their preferences. Possible answers include that they had some difficulties using the volume controls, that they believed they were already at the loudest/softest level, that they wanted a smaller adjustment than the 6 dB offered by the equipment, or that they perhaps misread the survey when answering. Only the 18 listeners who were already at top volume and wanted it louder did not have the opportunity to self-adjust to their satisfaction.

For those satisfied with their sound levels in Experiment 1, the mean PLL under headphones was 86 dB L_{Aeq}. Experiment 2 showed that PLL for headphones were generally 2 dB higher than for loudspeakers. Taken together, the results suggest that the PLL for these participants in a standard venue may be more likely in the range of 80-90 dB L_{Aeq} than 90-105 dB L_{Aeq}.

Overall, the findings obtained from observing actual sound level choices match well with previous survey-based studies showing that 42-90% of young people prefer sound levels to be softer than what is typically offered at dance venues.

Intra-rater reliability

Participants initially appeared quite inconsistent in choosing their PLLs in Experiment 2, yielding consistency ranges of up to 19 dB with two-thirds of the participants being above 5 dB in at least one presentation mode. However, due to the number of measurements per participant and the fact that most showed a CR-75% below 5 dB, the results were deemed sufficiently valid for further analysis.

These results show that intra-rater reliability is an important factor to consider when designing studies that ask participants to set their preferred listening level, and it is highly recommended that future research includes repeated measurement for added validity and that previous research is reviewed with this in mind.

Limitations

The results presented here were obtained in a brief, volume-limited silent disco event that was specifically designed for research purposes. It may be that people's preferred listening levels would be higher if their exposure were extended to several hours as at a typical night out rather than the 11-12 minutes of this experiment. Similarly, those who consciously reported being satisfied with the maximum volume presented through the headphones might have selected a higher setting if it had been available.

CONCLUSION

The majority of young people seem to prefer sound levels noticeably softer than what is usually offered at regular dance venues. People are reasonably consistent in setting their preferred listening levels in a controlled laboratory study across repeated measures, and any single measurement should be considered in this context. Future research in this area should therefore involve repeated measures wherever possible, and caution is advised when reviewing previous research on preferred listening levels.

REFERENCES

Beach, E. (**2013**). “Everyone likes it loud… don’t they?” ENT & Audiology News, **22**, 89-90.

Brixen, E.B., and Søndergaard, M. (**2001**). “Rapport vedrørende niveauopfattelse i hovedtelefoner” (in Danish). *Technical Report KKDK 068-1-ebb-1*. Delta Akustik and Vibration, Denmark.

Cassano, F., Bavaro, P., De Marinis, G., and Aloise, I. (**2005**). “Non-occupational exposure to noise,” G. Ital. Med. Lav. Ergon., **27**, 157-159.

ISO (**2013**). *ISO 1999:2013 Acoustics – Estimation of noise-induced hearing loss.* International Organization for Standardization, Geneva.

ISO (**2014**). *ISO 9612:2009 Acoustics – Determination of occupational noise exposure – Engineering Method.* International Organization for Standardization, Geneva.

Johnson, O., Andrew, B., Walker, D., Morgan, S., and Aldren, A. (**2014**). “British university students’ attitudes towards noise-induced hearing loss caused by nightclub attendance,” J. Laryngol. Otol., **128**, 29-34.

Mercier, V., and Hohmann, B.W. (**2002**). “Is electronically amplified music too loud? What do young people think?” Noise Health, **4**, 47-55.

Rudmose, W. (**1982**). “The case of the missing 6 dB,” J. Acoust. Soc. Am., **71**, 650-659.

Sadhra, S., Jackson, C.A., Ryder, T., and Brown, M.J. (**2002**). “Noise exposure and hearing loss among student employees working in the university entertainment venues,” Ann. Occup. Hyg., **46**, 455-463.

Tin, L.L., and Lim, O.P. (**2000**). “A study on the effects of discotheque noise on the hearing of young patrons,” Asia Pac. J. Public Health, **12**, 37-40.

WHO (**2015**). *Hearing loss due to recreational exposure to loud sounds: A review.* Geneva: World Health Organization.

Estimating auditory filter bandwidth using distortion product otoacoustic emissions

ANDREAS H. RUKJÆR[1], SIGURD VAN HAUEN[1], RODRIGO ORDOÑEZ[2,*], AND DORTE HAMMERSHØI[2]

[1] *Acoustics and Audio Technology, Aalborg University, Aalborg, Denmark*

[2] *Signal and Information Processing, Department of Electronic Systems, Aalborg University, Aalborg, Denmark*

The basic frequency selectivity in the listener's hearing is often characterized by auditory filters. These filters are determined through listening tests, which estimate the masking threshold as a function of frequency of the tone and the bandwidth of the masking sound. The auditory filters have been shown to be wider for listeners with sensorineural impairment. In a recent study (Christensen *et al.*, 2017) it was demonstrated on group basis that the distortion product stimulus ratio that provided the strongest $2f_1 - f_2$ component at low frequencies had a strong correlation to the theoretical relation between frequency and auditory filter bandwidth, described by the equivalent rectangular bandwidth (ERB, Glasberg and Moore, 1990). The purpose of the present study is to test whether a similar correlation exists on an individual basis at normal audiometric frequencies. The optimal $2f_1 - f_2$ DPOAE ratio is determined for stimulus ratios between 1.1 and 1.6, at fixed primary levels ($L_1/L_2 = 65/45$ dB SPL). The auditory filters are determined using notched-noise method in a two alternative forced choice experiment with noise levels at 40 dB SPL/Hz. Optimal ratios and auditory filters are determined at 1, 2, and 4 kHz for 10 young normal-hearing subjects.

INTRODUCTION

Since the discovery of otoacoustic emissions (OAEs) by Kemp (1978), they have become a central element in auditory research as an objective measure of peripheral auditory function. Otoacoustic emissions can be used to describe the state of the inner ear, in particular, of the outer hair cells (OHC), responsible for the active processes in the cochlea and the low level sensitivity of the hearing. Distortion product otoacoustic emissions (DPOAE) are the cochlear response to a two-tone paradigm. DPOAEs are thought to be generated as a result of interaction between the excitation patterns created by the primary stimulus frequencies in the basilar membrane (BM). Research shows that cochlear frequency tuning can be related to DPOAE phase delay (Bowman *et al.*, 1998), to DPOAE suppression tuning curves (Gruhlke *et al.*, 2012), as well as to response delay from stimulus frequency OAEs (Bentsen *et al.*, 2011). These results show that frequency specific OAEs can be used to describe frequency tuning

[*] Corresponding author: rop@es.aau.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

characteristics, yet the relation between the individual measures and the cochlear tuning characteristics are rather complex, being subject to several assumptions and undergoing complex data analysis procedures. This results in measures that may be well suited for auditory research but are not suited for clinical application.

The present work is inspired by the findings of Christensen *et al.* (2015, 2017), which at low frequencies demonstrated the relation between the optimal $2f_1 - f_2$ DPOAE stimulus ratio (f_2/f_1) and the equivalent rectangular bandwidth (ERB), as defined by Glasberg and Moore (1990):

$$f_1 - f_2 = \gamma ERB(f_2), \qquad \text{(Eq. 1)}$$

where the constant γ was found experimentally to equal approximately 1.5 in Christensen *et al.* (2015, 2017) for normal hearing populations.

The auditory filter bandwidth depend on: (1) the properties of the underlying morphology, and (2) the state of health in the underlying morphology (See Ch. 1 Sec. 6B in Moore, 2012). If changes to the the optimal DPOAE stimulus ratio are correlated to changes in the auditory filters due to cochlear damage, DPOAE measurements may offer a fast alternative to the psychoacoustic test of auditory filter bandwidth. If not, it may hold individual information of the underlying morphology, and may serve as a calibration or normalisation factor for the psychoacoustic (and other) individual measurements.

The purpose of the present investigation was to further examine the individual relation between the psychoacoustic (ERB) and objective (optimal DPOAE stimulus ratio) estimates for normal hearing subjects at typical audiometric frequencies.

METHODS

Auditory filter bandwidths and optimal DPOAE ratios were determined for 10 young (18-25 years), normal hearing (hearing level, HL $<$ 20 dB, middle ear pressure, MEP $< \pm$ 100 daPa) subjects around the standard audiometric frequencies of 1, 2, and 4 kHz.

DPOAE measurements

An Etyomotic Research ER-10C probe system with a Roland UA-25EX sound card controlled through a customised MATLAB program was used to obtain the DPOAE measurements. The fixed-f_2 paradigm was utilised and stimulus levels were fixed at 65/45 dB SPL. The probe was calibrated using a Brüel & Kjær Type 4157 ear simulator with a Brüel & Kjær Type 4138 microphone. Before measurements, individual levels were adjusted using the sound card's input gain to match a 500-Hz tone measured in the ear-canal to the corresponding reference level measured in the ear simulator. During the measurements the operator could monitor the measured signal, and an amplitude rejection criteria was used to avoid noisy recordings due to swallowing or movement of the probe.

Each response was recorded at 48 kHz and 24-bits resolution. The recorded signal was analysed using an average of 10 frames of 4800 samples and a discrete Fourier transform (DFT) of the same length, giving a frequency resolution of 10 Hz. Primary frequencies were chosen so all components of interest (f_1, f_2 and $2f_1 - f_2$) had an integer number of periods in the analysis frame, so no windowing was applied. The noise level of the measurement was estimated by averaging the amplitude of all DFT bins within 1 ERB centred at a given $2f_1 - f_2$ frequency (excluding the distortion component itself).

Taking into account the possible presence of fine structure in the DPOAE levels, five measurements were made within one dip-to-dip bandwidth of the expected fine structure, according to Reuter & Hammershøi (2006). For each audiometric frequency, DPOAE levels were obtained using five primary frequency pairs linearly spaced within a 100, 160 and 320 Hz bandwidth centred at 1, 2 and 4 kHz respectively. For each set of primaries eight different ratios were used between 1.1 and 1.5, as shown in Fig. 1. In order to ensure that all major signal components have an integer number of periods in the analysis window, the primary ratios changed for the five DPOAEs around each audiometric frequency. Thus, eight individual ratios were tested for each of the five sets of primaries, within a narrow frequency band around each of the three audiometric frequencies.

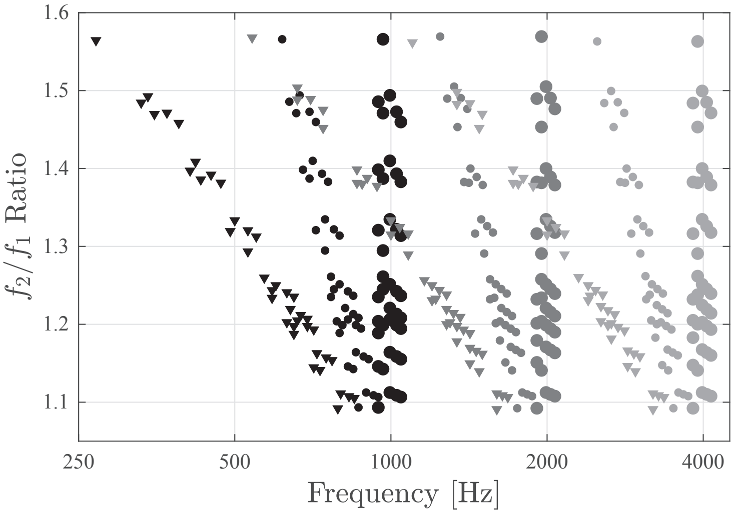


Fig. 1: Measured ratios as a function of primary and DPOAE frequencies, f_2 (large circle), f_1 (small circle), $2f_1$-f_2 (triangle)

The choice of frequencies included primary ratios that exceed 1.5 (see Fig. 1), at these high ratios the response of the $2f_1 - f_2$ component is close of $f_1/2$ and may be influenced by other distortion products. All DPOAE values obtained with primary ratios higher that 1.5 were excluded from further analysis (one case for 1 and 4 kHz and two cases for 2 kHz).

Auditory filter determination

Auditory filters were estimated using the notched noise method as described by Glasberg and Moore (1990), with relative notch widths $\Delta f/f_c$ of 0, 0.05, 0.1, 0.2, 0.3 and 0.4. The noise masker was presented simultaneously with the stimulus tone with a 100-ms Hanning ramp applied to both start and end of the combined signal. Each noise and stimulus interval was presented with 0.5-s duration and a 0.25-s pause between intervals. The masker was presented at a level of 40 dB/Hz.

Thresholds were estimated using a 2-alternative forced-choice paradigm with a 1-up 2 down tracking rule which estimates the 70.7% point on the psychometric function (Levitt, 1971). The 8-dB initial step-size was reduced to 4 and 2 dB after each reversal and a single threshold estimate was taken as an average of 6 reversals obtained with the smallest step-size. Subjects were given approximately 10 minutes under supervision for familiarisation with the procedure.

The filter shapes were derived from the notched-noise experiment data using the polynomial fitting method described by Patterson (1976). A 3$^{\text{rd}}$ order polynomial was fitted to the data and the ERB estimate was obtained from the integral of the fitted curve multiplied by $2f_c$, assuming symmetric filters.

RESULTS

DPOAE

Figure 2 shows the individual DPOAE levels for three subjects as a function of the f_2/f_1 ratio, as well as average values across subjects. The figure shows that both the individual and group results display a bell-shaped dependency to the stimulus ratio. To determine the optimal ratio, a 2$^{\text{nd}}$ order polynomial was fitted to the individual and group data obtained with primary frequency ratios below 1.35. For higher ratios, DPOAE values either decrease close to the noise floor, or show a steady increase in level. The latter seems to be related to an increasing noise floor (especially at 1 kHz), or to other artefacts as the ratio approaches 1.5. The maximum value of the fitted curves is defined as the optimal ratio. Inspection of the individual results shows that for 2 and 4 kHz all subjects, with the exception of subject 4 at 2 kHz, show the expected bell-shaped curve, for these cases the maximum DPOAE value was always found with ratios in the range of the fitted curve. In the case of 1 kHz, the results are more dependant on the levels of the emissions. Subjects with high emission levels (Subjects 1, 6, 7 and 8) have a clear bell-shaped curve and maximum DPOAEs are found with ratios in the range of the fitted curve. For subjects with low emission levels (Subjects 2, 3, 4, 5, 9 and 10), maximum DPOAE values were sometimes found at ratios above 1.35. For these subjects emission levels are in some cases within a few dB of the noise floor.

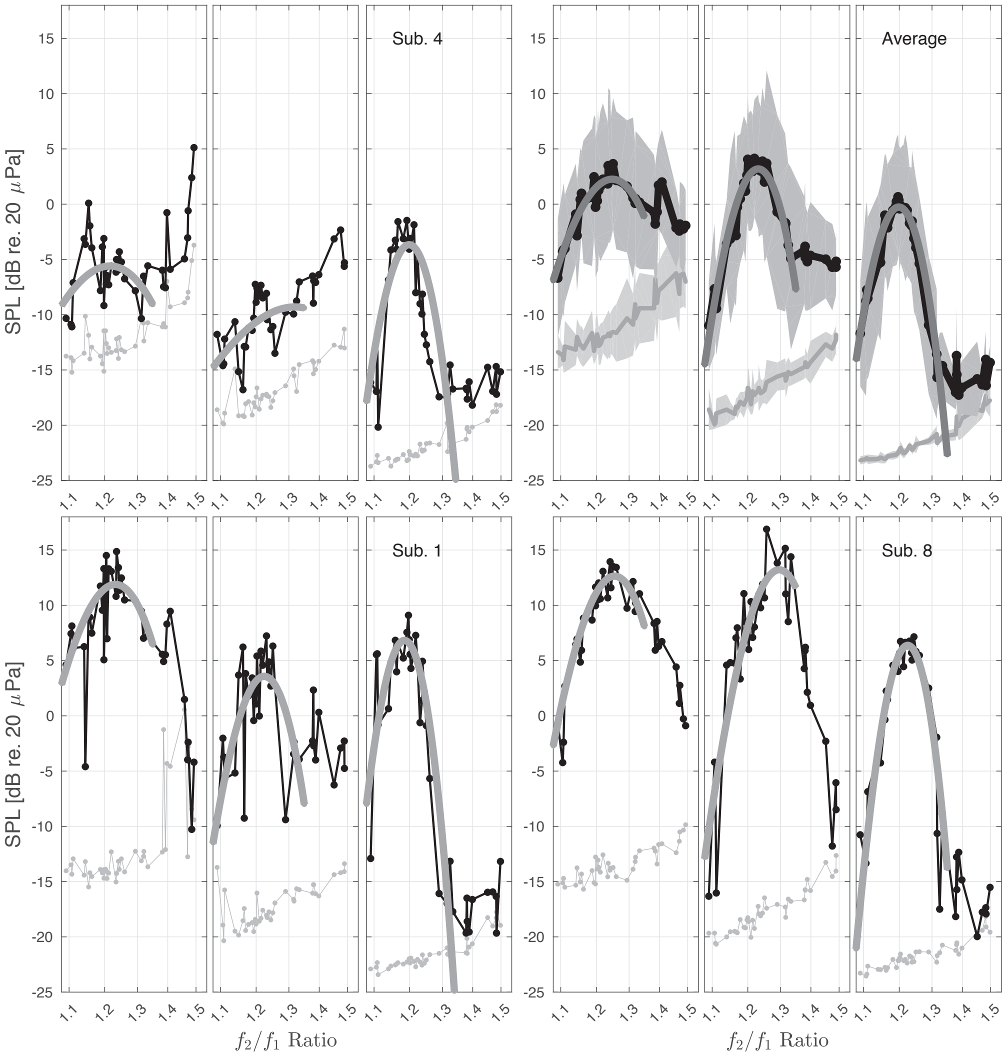


Fig. 2: Individual DPOAE levels for three subjects as a function of primary frequency ratios. Thick lines represents a 2nd order polynomial fit for data points with primary ratios between 1.1 and 1.35. Thin lines represent the measured values and the noise floor of the measurement. Top right panel, the group average in black with ±standard deviation (shaded area) and the grey line shows the 2nd order polynomial fitted to the averaged data.

Auditory filter results

The estimated notched-noise thresholds are shown in Fig. 3 for three subjects and for the group average. The figure shows that the wider the masking notch, the lower the masking effect on the stimulus tone, and that the slopes are steep, as is expected for normal-hearing individuals. There are however a few exceptions like subjects 3 and 7 at 4 kHz or subject 4 at 2 kHz.

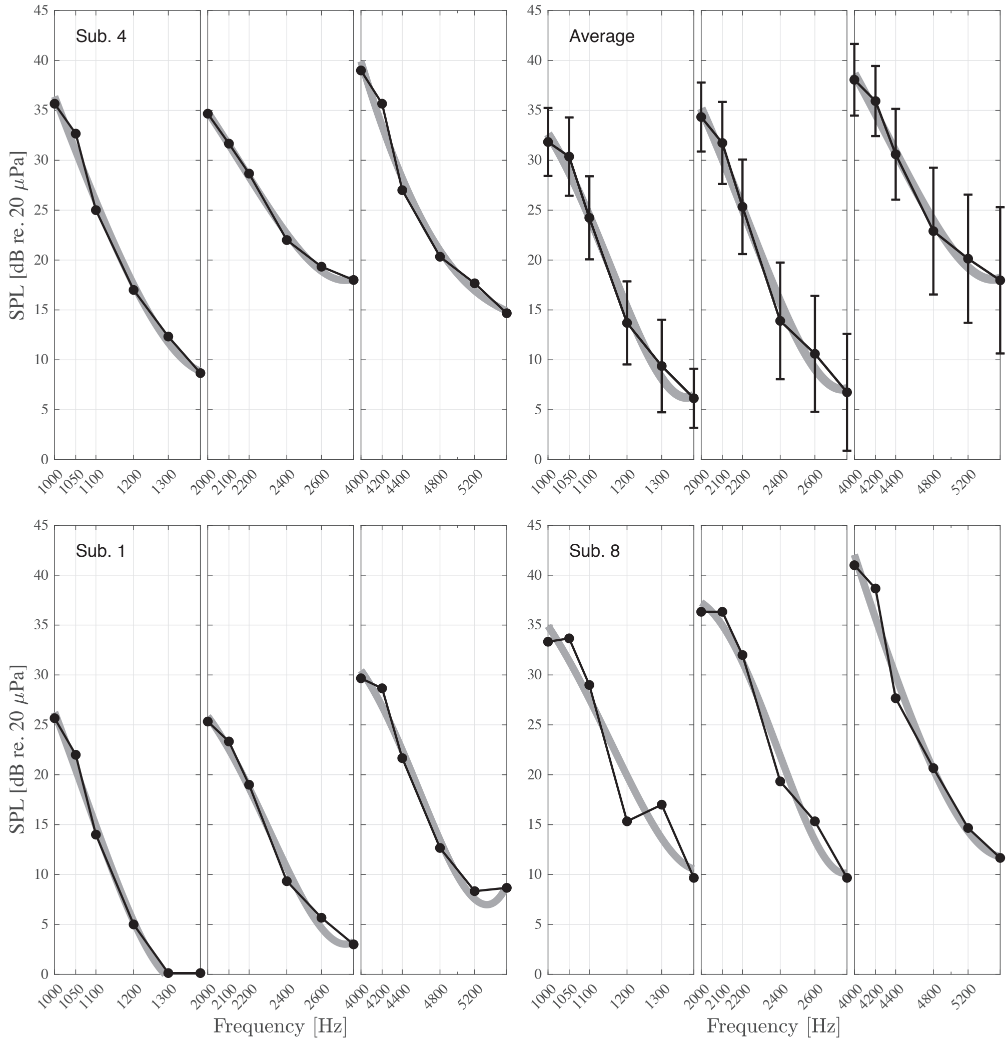


Fig. 3: Level of tone at threshold as a function of masker bandwidth (masker level 40 dB/Hz) in black (same three subjects as in Fig. 2), and the estimated auditory filter shapes using a 3$^{\text{rd}}$ order polynomial in grey. Top right panel, group average with ±standard deviation as the error bars, in black and the grey lines shows the estimated average auditory filter using a 3$^{\text{rd}}$ order polynomial.

Optimal ratio vs. ERB

Estimates of optimal ratio and equivalent rectangular bandwidths are compared for for each subject in the scatter plots of Fig. 4, with the circles. Pearson's correlation coefficient shows that there is a non-significant positive correlation at 1 kHz ($r = 0.3805$, $p = 0.2781$) and 2 kHz ($r = 0.4815$,$p = 0.1588$), and a non-significant negative correlation at 4 kHz ($r = -0.3111$,$p = 0.3817$). This result suggests that narrower ERB estimates show lower optimal ratios, with clear exceptions, as the case of the lowest optional ratio obtained at 4 kHz that shows the widest ERB estimate for that frequency band.

In order further explore the relation between frequency selectivity and DPOAE optimal ratios, optimal ratios were expressed in terms of the width DPOAE vs. ratio curve finding the point in which the fitted 2nd polynomial function drops 6 dB from its maximum value, according to the following expression:

$$OR_{span} = (OR_{max} - OR_{-6dB}), \qquad \text{(Eq. 2)}$$

where OR_{span} is the optimal ratio span representing the span of ratios that cover in main portion of the DPOAE vs. ratio estimate; OR_{max} is the optimal ratio, and OR_{-6dB} is the ratio corresponding to the -6 dB point. Pearson's correlation coefficient between OR_{span} and ERB estimates show a weak positive correlation at 1 kHz ($r = 0.6120$, $p = 0.0601$) and 4 kHz ($r = 0.5642$, $p = 0.0893$) and a non-significant positive correlation at 2 kHz ($r = 0.1409$, $p = 0.6979$). These results are shown in Fig. 4 with the diamond symbols.

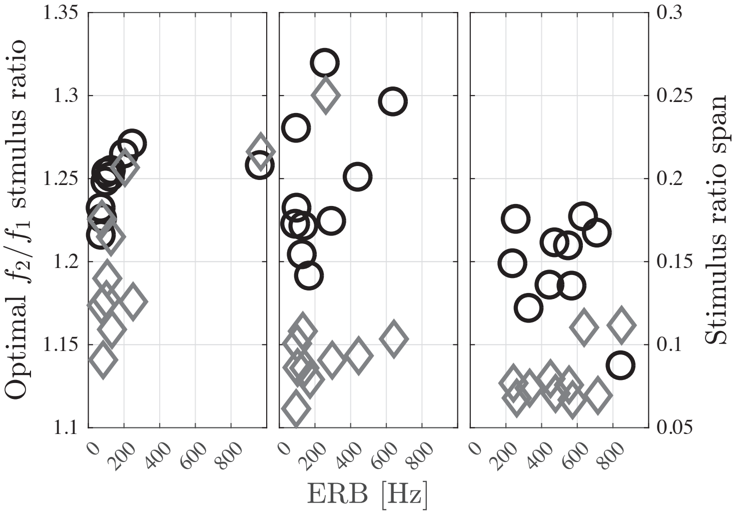


Fig. 4: Optimal DPOAE stimulus ratio estimated for each subject versus ERB (circles, left axis). Stimulus ratio span versus ERB (diamonds, right axis).

CONCLUSIONS

The present data confirms the optimal ratio relation to auditory filter bandwidth on group basis, and the relation can also be recognised to a lesser degree for individual data. The data suggests that subjects with a broad auditory filter (high ERB estimate) also have larger optimal ratios. In the same manner, subjects with low ERB estimates will have lower optimal ratios. Other estimates of frequency tuning derived from the DPOAE vs. primary ratio relationships show equal or better correlation with individual ERB estimates.

ACKNOWLEDGEMENTS

The support from the BEAR project and its partners, incl. funding from Innovation Fund Denmark is sincerely acknowledged, as is the grant from Danish Sound (the Research Talent Award 2017), and the support from the ISAAR organisation. We also appreciate the collaboration by Raúl H. Sánchez López and the Hearing Systems Group from DTU for supplying MATLAB code and material for the notched-noise experiment.

REFERENCES

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. doi: 10.1121/1.3575596

Bowman, D.M., Eggermont, J.J., Brown, D.K., and Kimberley, B.P. (**1998**). "Estimating cochlear filter response properties from distortion product otoacoustic emission (DPOAE) phase delay measurements in normal hearing human adults," Hear. Res., **119**,14-26. doi: 10.1016/S0378-5955(98)00041-0

Christensen, A.T., Ordoñez, R., and Hammershøi, D. (**2015**). "Stimulus ratio dependence of low-frequency distortion-product otoacoustic emissions in humans," J. Acoust. Soc. Am., **137**(2), 679-689. doi: 10.1121/1.4906157

Christensen, A.T., Ordoñez, R., and Hammershøi, D. (**2017**). "Distortion-product otoacoustic emission measured below 300 Hz in normal-hearing human subjects," J. Assoc. Res. Otolaryngol., **18**, 197-208. doi: 10.1007/s10162-016-0600-x

Glasberg, B.R., and Moore, B.C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res., **47**, 103-138. doi: 10.1016/0378-5955(90)90170-T

Gruhlke, A., Birkholz, C., Neely, S.T., Kopun, J., Tan, H., Jesteadt, W., Schimd, K., and Gorga, M.P. (**2012**). "Distortion-product otoacoustic emission supressiontuning curves in hearing-impaired humans," J. Acoust. Soc. Am., **135**, 3292-3304. doi: 10.1121/1.4754525

Kemp, D.T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am., **64**, 1386-1391. doi: 10.1121/1.382104

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am., **49**, 467-477. doi: 10.1121/1.1912375

Moore, B.C.J. (**2012**). *An Introduction to the Psychology of Hearing (6th Ed.)*, Emeral Group Publishing Limited, Bingley, UK, ISBN: 978-1-78052-028-4.

Patterson, R.D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am., **59**, 640-654. doi: 10.1121/1.380914

Reuter, K., and Hammershøi, D. (**2006**). "Distortion product otoacoustic emission fine structure analysis of 50 normal-hearing humans," J. Acoust. Soc. Am., **120**, 270-279. doi: 10.1121/1.2205130

Adjusting expectations: Hearing abilities in a population-based sample using an SSQ short form

PETRA VON GABLENZ[1], FABIAN SOBOTKA[2], AND INGA HOLUBE[1]

[1] *Institute of Hearing Technology and Audiology, Jade University of Applied Sciences and Cluster of Excellence "Hearing4all", Oldenburg, Germany*

[2] *Department of Health Services Research, School for Medicine and Health Sciences, Carl von Ossietzky University Oldenburg, Oldenburg, Germany*

Self-reports of hearing (dis)abilities play an important role in hearing rehabilitation. Among the large variety of questionnaires, the Speech, Spatial, and Qualities of Hearing Scale (SSQ) has become an internationally used measure to assess hearing abilities in specified everyday listening situations using a visualized scale ranging from 0 to 10. Research mainly focused on adults with impaired hearing, whereas adults with "normal" hearing were hardly considered. However, the ratings of adults out of the general population could be of particular interest when it comes to the question of score benchmarks based on different definitions of "normal" hearing. In the cross-sectional, population-based study HÖRSTAT (n=1903) the German SSQ17 short form was used along with a standardized interview and comprehensive hearing examinations. As the SSQ score distributions are extremely negatively skewed, semiparametric quantile and expectile regression analysis was performed to examine the conditional score distribution and the effects of age, gender, globally reported hearing problems, hearing loss, and social status. Though no normative cut-off values can be established from empirical findings only, the distribution of "normal" hearing abilities might align the management of expectations during the process of hearing rehabilitation.

INTRODUCTION

Since the Speech, Spatial and Qualities of Hearing Scale (SSQ) showed "promise as an instrument for evaluating interventions of various kinds" in audiological rehabilitation (Gatehouse and Noble, 2004), various short forms were developed to foster it's usability. Research focused on hearing-impaired adults, whereas 'normal'-hearing adults were included for validation in non-English versions (e.g., Banh *et al.*, 2012; Deemester *et al.*, 2012; Moulin and Richard, 2016). Recruitment of the normal-hearing participants followed audiological criteria and university students often served as the young control group. But hearing ability established by means of a questionnaire is a cognitive construct, thus shaped, e.g., by performance expectations, habitat with diverse acoustical demands, second-party opinions, and comparisons.

*Corresponding author: petra.vongablenz@jade-hs.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

Therefore, variability sources and benchmark scores derived from socially homogeneous groups are as critical as sample size. Furthermore, the score distribution is often skewed, thus the report of mean and standard deviation and the use of parametric methods is misleading.

This article sets out three objectives: First, it attempts to derive a benchmark distribution for hearing abilities in the general population using SSQ items. Assuming that ability assessment refers to a cognitive construct, it secondly aims to identify non-audiological factors such as age, gender, and education which might influence SSQ ratings. Third and finally, an innovative statistical method will be presented that copes appropriately with non-normal distributions in order to achieve the previously stated objectives.

METHODS

SSQ17 questionnaire

In general, the SSQ items describe everyday situations and a listening task. The respondents rate how well they can fulfill the task using a visualized scale ranging from 0 (not at all / a lot of effort) to 10 (perfectly / no effort). The original SSQ presented by Gatehouse and Noble (2004) comprises 50 items assigned to three subscales. Table 1 lists the items included in the German SSQ17 short form (Kießling *et al.*, 2011).

Subscale	Pragmatic subscale	Item ref. SSQ50
Speech	Speech in noise (2), speech in speech (2), multiple speech-stream processing and switching (1)	1.4, 1.5, 1.7, 1.9, 1.10
Spatial	Localization (2), distance and movement (3)	2.5, 2.6, 2.7, 2.9, 2.12
Qualities	Sound quality and naturalness (3), segregation of sounds (1), identification of sound and objects (1)	3.3, 3.4, 3.8, 3.9, 3.10

Table 1: Items (number) in the SSQ17 according to the numbering in the original SSQ (Gatehouse and Noble, 2004) and the pragmatic subscale allocation proposed by Gatehouse and Akeroyd (2006).

SSQ17 cut the subscales down to 5 items each, complemented by the items understanding speech in quiet (1.2) and listening effort (3.18). The subjects received the questionnaire together with the HÖRSTAT invitation letter and were asked to return the completed SSQ17 during the examination appointment.

Study sample

The data was derived from the cross-sectional study HÖRSTAT (2010–2012). This study was based on random samples stratified by age and gender from two medium-sized towns in Northwest Germany. The response was low in young age bands, but fairly high in the middle-aged and elderly adults from 40 to 79 years (30%), resulting in an overall response rate of 21%. At large, the study sample of 1,903 adults approximated both the national distribution by gender and age. The hearing examination included pure-tone audiometry in accordance to ISO 8253-1, the Goettingen sentence test in noise (Kollmeier and Wesselkamp, 1997), the German digit triplet test (Zokoll *et al.*, 2012), a standardized interview, and the SSQ17 questionnaire. The study design, test procedure and equipment are described in detail elsewhere (von Gablenz and Holube, 2016).

Valid data from pure-tone audiometry and the SSQ17 were inclusion criteria for this analysis leading to a sample of 1,836 adults (45% males) aged 18 to 97 years. Prevalence of hearing impairment was 16% defined as PTA4 > 25 dB HL in the better ear (PTA4: pure-tone average at 0.5, 1, 2, and 4 kHz). In total, 26% reported hearing difficulties in the standardized interview and 8% met the criterion for asymmetric hearing thresholds (interaural PTA4 difference > 10 dB). Social composition was somewhat biased towards highly educated strata if school attainment level is presumed to indicate social position. About 51% of the subjects received an advanced school education according to the traditional German educational system.

Statistical analysis

PTA4 in the better ear is used as a key parameter for the state of hearing to facilitate comparability of results, since Spearman correlation analysis showed equal to slightly better correlation coefficients between SSQ scores and better ear PTA4 ($r = -0.175$ to -0.418) than for the better ear speech reception thresholds (SRT) in the Goettingen sentence test in noise ($r = -0.170$ to -0.412). Correlation coefficients are only marginally higher if related to the worse ear PTA4 or SRT.

With regard to SSQ ratings, this analysis is based on the mean score by subject across all SSQ17 items (SSQ17) and the SSQ17 subscales. Missing values (2%) are dealt with multiple imputation through regression with error. Similarities between the subscales are used to fit generalized additive models for the imputation of one score with all remaining subscales in the predictor. The estimation of the models is then cycled and finally SSQ17 is recomputed.

Score distributions are negatively skewed for all subscales (and items). Skewness is −0.9 in the speech, −1.1 in the spatial and −1.8 in the qualities subscale. As the assumptions for parametric statistics are not met, this analysis refers to quantile regression (Koenker and Bassett, 1978), an approach on the verge of becoming a standard tool in modern regression analysis. While a simple mean regression attempts to describe the expectation of a response as a function of the covariates, the results of a quantile or expectile regression offer a much broader view. In principle, a dense set of expectiles or quantiles allows for an analysis of the complete conditional distribution of the response. This can lead to new insight into the dependency structure between the response and its covariates.

For the inclusion of nonlinear effects, an efficient semiparametric quantile regression (SPQR) is performed. Penalized splines divided into a parametric part and its random nonlinear deviations, smoothed with an additive penalty term, are used in a fast linear programming procedure. Computationally, regression quantiles with a LASSO penalty for random effects are obtained by minimizing an asymmetrically weighted absolute residuals criterion

$$\sum_{\{\mathrm{i=1}\}}^{\mathrm{n}} w_\tau\,(y_i, Z_i\beta_\tau)\,|y_i - X_i\alpha_\tau - \,B_i\gamma_\tau| + \lambda|\gamma_\tau| \qquad \text{(Eq. 1)}$$

with asymmetric weights

$$w_\tau(y, Z\beta) = \begin{cases} 1-\tau \text{ if } y < Z\beta \\ \tau \text{ if } y > Z\beta \end{cases}$$

a response y and a quantile-specific predictor $Z\beta_\tau$ consisting of the unpenalized effects $\mathrm{X}\alpha_\tau$ and the penalised random part $B\,\gamma_\tau$ for each quantile level τ. This loss function can be subject to a linear program.

RESULTS

In the general population, SSQ17 scores averaged by subject across all items are almost unchanged until the age of 50 years with 8.3 points at the median. They decline to 7.8 and 7.1 points at the age of 70 and 80 years, respectively. These results refer to a zero-model that simply regressed scores on age.

Figure 1, in contrast, shows not only SSQ17 scores as a function of age (dots), but also the score distribution from a quantile regression model that was controlled for self-reported hearing difficulties only (lines). Since these adults feel comfortable with their hearing in general, this distribution could reasonably be assumed to describe the benchmark for hearing abilities assessed using the SSQ17. The ability scores decrease with age. This decrease is somewhat more pronounced in the lower than in the upper half of the distribution. Intercept spans rounded 3 points from the 0.05 to the 0.95 quantile (5.9 to 9.0). The parametric regression coefficient, which accounts for hearing difficulties, is −1.4 at the median, that is, the median regression curve is shifted down by 1.4 points in adults reporting hearing difficulties. As expected, the effect of self-reported difficulties is greater in low scoring than in high scoring adults. The coefficients range from −2.0 at the lowest to −1.1 at the highest quantile.

Figure 2 shows the score distribution in the speech subscale as a function of PTA4 in the better ear (dots). The curves display a quantile regression model that controls the association of PTA4 and SSQ scores for age, gender, hearing difficulties, and asymmetric pure tone hearing. Thus, quantile curves refer to males with symmetric thresholds who did not report hearing difficulties as the reference distribution. The corresponding parametric regression coefficients that estimate the effect of age, female gender, asymmetric thresholds, and hearing difficulties are listed in Table 2 for selected quantiles 0.1, 0.5, and 0.9. This table also includes intercept and coefficients estimated in analogous regression models on the spatial and qualities subscale data, which are not graphically displayed.

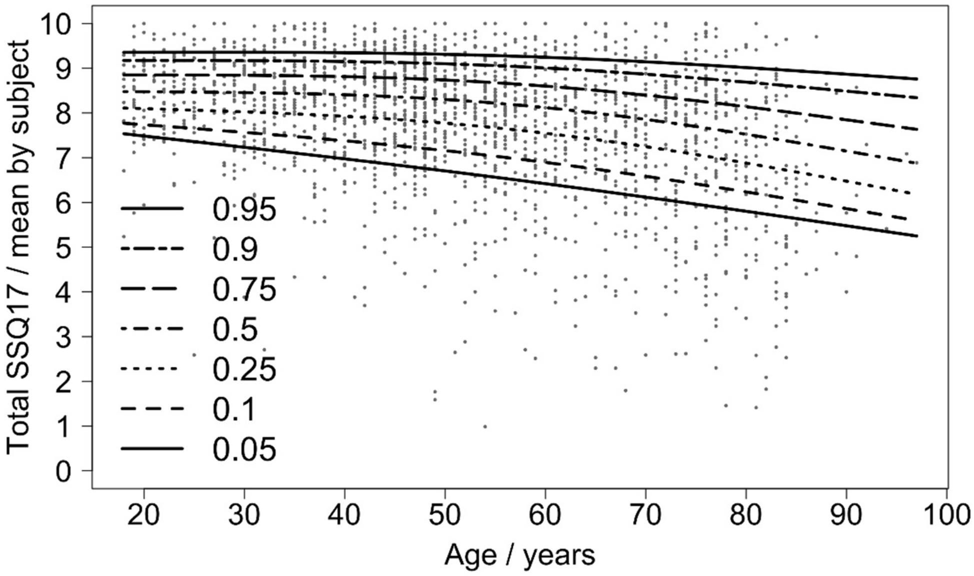


Fig. 1: SSQ17 scores averaged by subject across all items as a function of age (dots). Regression lines refer to adults without self-report of hearing difficulties.

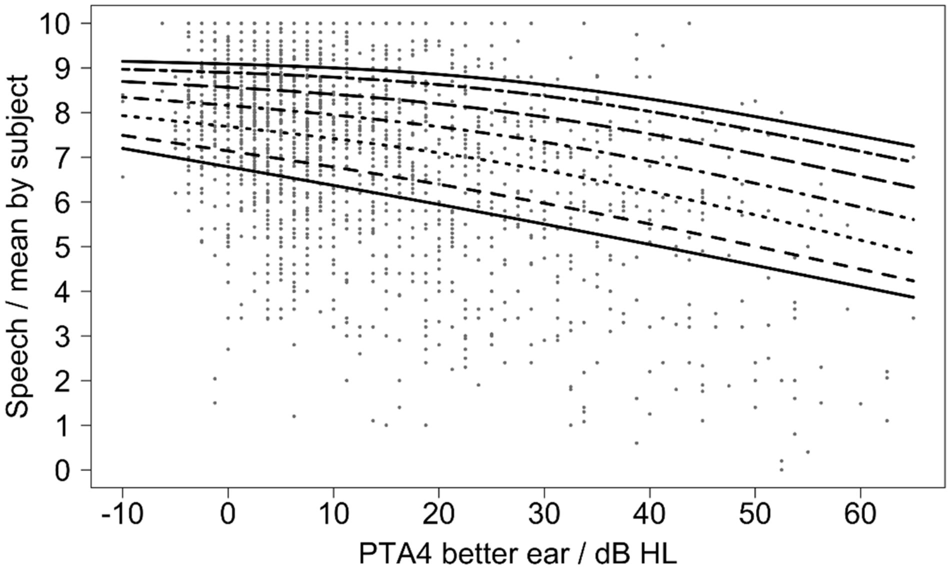


Fig. 2: Mean speech subscale SSQ17 scores by PTA4 in the better ear (dots). Regression curves refer to males with symmetric hearing thresholds without self-report of hearing difficulties.

Ability ratings are highest in the qualities subscale and lowest in the speech subscale. Age does not affect ability ratings either in high or in low scoring adults in these models. This is generally the case (models not shown) if PTA4 is controlled for. Asymmetric hearing and self-reported global hearing difficulties, in contrast, are significant factors in all subscales, with mostly higher effects in the distribution's bottom half. Self-report of hearing difficulties strongly influences the speech subscale, with only small differences between the quantiles. The corresponding coefficients are higher than in the other subscales and range from −1.5 to −1.3. Asymmetric PTA4 is the most influential factor in the spatial subscale, with coefficients between −1.1 and −0.4. Gender significantly affects the ratings on spatial items across all quantiles, though the effect is particularly strong in the bottom half. Females rate their abilities lower than males, with a maximum estimate of −0.9 points at the 0.05 quantile and −0.4 points at the median. In the speech and the qualities subscales, however, high scoring females rate their abilities slightly higher than males. Though significance is partly confirmed, gender has hardly a substantial impact because the estimated coefficients are rather small (≤ |0.2|). Extended models further revealed that high education is significantly associated with higher scores, particularly in the spatial and qualities subscales. The corresponding coefficients indicate 0.2 to 0.6 points at most quantiles.

		0.1	sd	0.5	sd	0.9	sd
Speech	Intercept	**4.73**	0.41	**6.14**	0.48	**8.04**	0.69
	Age / year	**-0.01**	0.01	**-0.00**	0.00	**0.00**	0.00
	Female gender	**-0.06**	0.11	**0.10**	0.08	**0.15**	0.07
	Asymmetric PTA4 *	**-0.83**	0.19	**-0.77**	0.18	**-0.56**	0.17
	Hearing difficulties *	**-1.46**	0.14	**-1.51**	0.11	**-1.39**	0.10
Spatial	Intercept	**5.03**	0.46	**6.55**	0.31	**7.89**	0.27
	Age / year	**0.00**	0.01	**0.00**	0.00	**0.00**	0.00
	Female gender *	**-0.80**	0.13	**-0.44**	0.08	**-0.25**	0.07
	Asymmetric PTA4 *	**-1.05**	0.26	**-0.76**	0.11	**-0.54**	0.19
	Hearing difficulties *	**-0.89**	0.18	**-0.75**	0.20	**-0.56**	0.10
Qualities	Intercept	**5.58**	0.44	**7.58**	1.23	**8.89**	0.67
	Age / year	**-0.00**	0.00	**-0.00**	0.00	**-0.00**	0.00
	Female gender	**0.00**	0.10	**0.16**	0.06	**0.14**	0.04
	Asymmetric PTA4 *	**-0.83**	0.17	**-0.39**	0.17	**-0.18**	0.14
	Hearing difficulties *	**-1.03**	0.30	**-0.73**	0.09	**-0.47**	0.07

Table 2: Parametric effects on the speech, spatial and qualities subscales of SSQ17. Regression coefficients for the 0.1, 0.5, and 0.9 quantile and standard deviation (sd). Coefficients greater than 1.96 sd are considered to be significant.

DISCUSSION

Adults who do not complain about hearing and mostly do not meet the criterion of hearing impairment rate their hearing abilities in the SSQ17 well below the scale maximum. Basically, this was already known from earlier research (e.g., Moulin and Richard, 2016; Demeester *et al.,* 2012; Banh *et al.*, 2012). The question was, rather, to estimate a reference using a population-based sample. Comparing results from different SSQ studies calls for some reservations. Whereas test administration seems not to affect scores on the SSQ systematically (Singh and Pichora-Fuller, 2010), language translation and item selection are always an issue, aggravated by different analytical approaches. This applies in particular for comparisons between the benchmark distributions derived from the general population using quantile regression to benchmark scores defined as the arithmetic mean from tailored study samples. Nevertheless, our results point towards lower benchmark score levels in young adults than reported earlier, e.g., by Demeester *et al.*, 2012; Banh *et al.*, 2012.

Chronological age does not influence ability ratings if audibility operationalized by PTA4 and interaural symmetry is controlled for. This finding is along the same lines as Agus *et al.* (2009) who did not observe a correlation between age and items addressing speech in speech and multistream listening situations, but contrasts with Banh *et al.* (2012), who established a combined effect of age and hearing impairment.

As expected, the association of SSQ ratings and pure-tone hearing is most affected by globally reported hearing difficulties. Agus *et al.* (2009) also distinguished by the self-report of hearing difficulties in their analysis. They found a group difference of 1.4 points for speech items on average which is well in line with our estimations (−1.5 to −1.4 points). Our results show, in addition, that the impact of self-reported difficulties is roughly halved for the spatial and qualities subscale.

To our knowledge, the effect of gender on spatial items was not reported for other studies though Moulin and Richard (2016) observed a possibly related trend for the differential score between the speech and spatial subscale. This effect cannot be traced back to audiological criteria from the present state of the analysis. Additionally, the impact of educational level on SSQ scores seems to be under-researched so far. Moulin and Richard (2016) reported an effect for selected items mainly in the qualities subscale, whereas our results suggest a considerably broader impact. Overall, the effect of gender, education and, in general terms, social position seem relevant enough to merit attention.

SUMMARY AND OUTLOOK

Quantile regression analysis gives an appropriate display of hearing abilities in the general population that makes a description of a benchmark distribution possible. Though de facto observations do not bear any normative power for methodological reasons, they facilitate orientation in the rehabilitation process. Furthermore, social factors influence ability ratings. This finding is only partly addressed in earlier studies. Age, however, shows no effect if audibility is controlled for. The next steps will be to extended the analysis towards item and pragmatic subscale level and to include other factors that potentially influence the SSQ ability rating. Further, two-way interaction terms reflecting potential dependencies between the covariates and the cut-off for disability will be examined and discussed for this population-based sample.

ACKNOWLEDGEMENTS

Funded by the Research Association of German Hearing Aid Acousticians, the Lower Saxony Department of Science and Culture and the European Regional Funding (EFRE) with HÖRSTAT, and the federal resources of Niedersächsisches Vorab by the Ministry of Science and Culture of Lower Saxony within the research focus “Hearing in everyday life (HALLO)”.

REFERENCES

Agus, T.R., Akeroyd, M.A., *et al.* (**2009**). “An analysis of the masking of speech by competing speech using self-report data,” J. Acoust. Soc. Am., **125**, 23-26. doi: 10.1121/1.3025915

Banh, J., Singh, G., *et al.* (**2012**). “Age affects responses on the Speech, Spatial, and Qualities of Hearing Scale (SSQ) by adults with minimal audiometric loss,” J. Am. Acad. Audiol., **23**, 81-91. doi: 10.3766/jaaa.23.2.2

Demeester, K., Topsakal, V., *et al.* (**2012**). “Hearing disability measured by the Speech, Spatial, and Qualities of Hearing Scale in clinically normal-hearing and hearing-impaired middle-aged persons, and disability screening by means of a reduced SSQ (the SSQ5),” Ear. Hearing, **33**, 615-626. doi: 10.1097/AUD.0b013e31824e0ba7

Gatehouse, S., and Noble, W. (**2004**). “The speech, spatial and qualities of hearing scale (SSQ),” Int. J. Audiol., **43**, 85-99. doi: 10.1080/14992020400050014

Gatehouse, S., and Akeroyd M. (**2006**). “Two-eared listening in dynamic situations,” Int. J. Audiol., **45,** S120-S124. doi: 10.1080/14992020600783103

Kießling, J., Grugel, L., *et al.* (**2011**). “Übertragung der Fragebögen SADL, ECHO und SSQ ins Deutsche und deren Evaluation,” Z. Audiol., **50**, 6-16.

Koenker, R.W., and Basset, G. (**1978**). “Regression Quantiles,” Econometrica, **46**, 33-50.

Kollmeier, B., and Wesselkamp, M. (**1997**). “Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment,” J. Acoust. Soc. Am., **102**, 2412-2421. doi: 10.1121/1.419624

Moulin, A, and Richard, C. (**2016**). “Sources of variability of speech, spatial, and qualities of hearing scale (SSQ) scores in normal-hearing and hearing-impaired populations,” Int. J. Audiol, **55**, 101-109. doi: 10.3109/14992027.2015.1104734

Singh, G., and Pichora-Fuller, M.K. (2010) “Older adults' performance on the speech, spatial, and qualities of hearing scale (SSQ): Test-retest reliability and a comparison of interview and self-administration methods,” Int. J. Audiol., **49**, 733-740. doi: 10.3109/14992027.2010.491097

Von Gablenz, P., and Holube, I. (**2016**). “Hearing threshold distribution and effect of screening in a population-based German sample,” Int. J. Audiol., **55**, 110-125. doi: 10.3109/14992027.2015.1084054

Zokoll, M.A., Wagener K.C., *et al.* (**2012**). “Internationally comparable screening tests for listening in noise in several European languages: The German digit triplet test as an optimization prototype,” Int. J. Audiol., **51**, 697-707. doi: 10.3109/14992027.2012.690078

An improved competing voices test for test of attention

LARS BRAMSLØW*, MARIANNA VATTI, RIKKE ROSSING,
AND NIELS HENRIK PONTOPPIDAN

Eriksholm Research Centre, Snekkersten, Denmark

People with hearing impairment find competing voices scenarios to be challenging in terms of their ability to switch attention and adapt to the situation. With the Competing Voices Test (CVT), we can explore how they can adapt and change their attention between voices. The CVT provides three male and three female speakers, played in pairs. The task of the listener is to repeat the target sentence. Three methods of cueing the listener to the target sentence were tested: a male/female cue (for male-female sentence pairs), an audio voice cue and a text cue using one word from the target sentence. The cue was presented either before or after the sentence pair playback. The CVT was evaluated on 14 moderate-severely hearing impaired listeners with four spatial conditions: summed (diotic), separate (dichotic) plus two types of ideal masks for separating the two speakers from the sum. The results show that the test is sensitive to the spatial conditions, as intended. The text cue is the most sensitive to spatial condition. The text cue has the further advantage that it can be used for, e.g., male-male speaker pairs as well. Furthermore, the applied ideal masks show test scores very close to the ideal separate spatial condition.

INTRODUCTION

Competing voices are part of the everyday challenges for a hearing aid user. This might occur for instance while attending to two voices in a restaurant or while watching TV and attending to a voice in the room at the same time. In order to test the performance of hearing aids in this user scenario, a new type of speech test has been developed.

Compared to traditional speech tests, a competing voices test would have two or more targets that are equally important to follow, and in the simplified case no masker. Tests of this type have been reported in the literature (Mackersie *et al.*, 2001; Helfer *et al.*, 2010), but no particular test has been put to common use. Furthermore, they are not available in Danish.

The purpose of the present project was to develop a competing voices test (CVT) in Danish. The proposed CVT has evolved in a number of iterations and applications using other cue timings and different speech material, this is documented in a series of posters (Bramsløw *et al.*, 2014, 2015a, 2015b, 2016b). The present paper presents the newest and improved version of the CVT.

*Corresponding author: labw@eriksholm.com

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

The aim of the present study was thus to refine and validate the competing voice test. The CVT should have the following properties:

- Be sensitive to signal processing contrasts, in this case spatial contrasts;
- Be applicable for older hearing-impaired listeners without floor and ceiling effects in the outcome measure;
- Be suitable for quick testing of multiple conditions in the laboratory.

METHOD

Speech material

In order to minimize development time, the Danish Hearing in Noise Test (HINT) was chosen as the speech corpus for the CVT, being an established and well documented natural sentences speech material (Nielsen and Dau, 2009; 2011). The Danish HINT has five words per sentence and is available with one male speaker.

Two males and three females were recruited as additional speakers. The recording was conducted as follows: The speaker was located in an audiometry booth with a microphone and a PC installed with our Danish HINT test software. The speaker would use the software to play one sentence and then repeat the sentence using the same intonation as the original speaker. This was done to ensure recordings of the same vocal quality. Each list was recorded in one take, but recorded twice to have two versions. All sentences were cut out into separate wave files, and the better of the two sentences was chosen. Each sentence was now time aligned to the original male recording by estimating the cross correlation against original recording and time shifting the new recording accordingly. Then, each sentence was level adjusted to have the same RMS value as the same original male sentence.

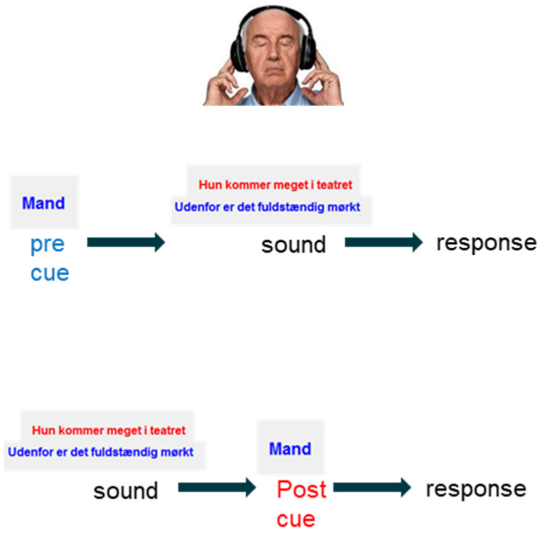

Mand
pre
cue
Hun kommer meget i teatret
Udenfor er det fuldstændig mørkt
sound
response
Hun kommer meget i teatret
Udenfor er det fuldstændig mørkt
sound
Mand
Post
cue
response

Fig. 1: Example of the Competing Voices Test with two sentences played simultaneously in a pair and the cue (male/female speaker) as showed on a screen.

Test procedure

Each CVT trial presented two sentences in pairs by selecting two different lists and randomizing the sentence order. Within each trial, all the included speaker pairs were presented in random order. In the separated (dichotic) cases, the target speaker was furthermore randomized to the left or right ear in order to make the test as unpredictable as possible for the listener. The task of the listener was to repeat the target sentence as cued by a sign on a PC monitor.

The cue could be presented before playback ('pre') or after playback ('post'). The pre cue corresponds to a classical target-masker scenario, whereas the post cue requires equal attention to both speakers, which we refer to as the 'competing voices scenario'. This is illustrated in Fig. 1.

Three cue types were tested: Audio, Text and Talker. The Audio cue is the word 'Tomato' spoken by the target speaker, thus the listener must recognise that voice in the mixture and repeat that target sentence. The Text cue is showing the first or last word from the target sentence on a screen in front of the listener: With pre cue, it will be the first word and with post cue, the last word. The words score can then be 0-4. Finally, the Talker cue uses a male-female mixture and the screen is indicating male or female to identify the target sentence.

Fourteen hearing-impaired listeners with moderate sloping hearing loss participated in the test; These are labelled Test Persons (TP) in the following.

Spatial contrasts

The sensitivity of the CVT was assessed by testing four spatial conditions: Sum (diotic), separate (dichotic), ideal binary mask (IBM) and ideal ratio mask (IRM). The ideal time-frequency masks were calculated by comparing the energy of the two clean signals in 125-Hz by 4-ms bins – as either binary masks (gain 0 or 1) or ratio masks (gain 0-1) (Naithani *et al.*, 2017). These masks were applied to the Sum signal to make an artificial separation, which was then presented dichotically. The ideal mask conditions were included to have a larger diversity of spatial conditions.

Test design

The overall test design thus consisted of the following experimental factors and levels:

- Spatial Processing: Sum, Separate, IRM, IBM
- Cuetype: Audio, Text, Talker
- Cuetime: Pre, Post
- Gender mix: Male-Female (MF), Male-Male (MM), Female-Female (FF).
- 14 test persons (TP).

The first three conditions were rotated across test persons in a balanced Latin square order, while the gender mix was varied randomly, within a given 20-pair trial. The lists order across trials was randomized such that no lists were repeated in successive trials. Finally, the sentence order within trials was randomized such that all sentences were used equally and that the initial or last words were different in the 'Text' cuetype.

RESULTS

The outcome measure from each sentence pair was a percent correct word score, based on five words (Audio cue, Talker cue) or four words (Text cue). It was then rau-transformed to provide better 'normal' distribution of the data (Studebaker, 1985). The rau scores are practically equal to %-scores in the 10-90 range and extended beyond those limits to cover the range −18 to +118.

All data were analysed using a mixed-model analysis of variance (ANOVA) with TP as a random factor and gender mix nested under cuetype (the Talker cuetype can only use the MF combinations). The ANOVA table is shown in Table 1 below.

All main effects are significant and so are the two-way interactions spatial*cuetype, spatial*gender and the three-way interaction spatial*cuetype*gender. Regarding the random factor TP effects, the TP*cuetime interaction is significant. It is also interesting to note that there are no significant interactions between cuetime and the other fixed conditions; This means that the choice between 'Pre' and 'Post' cue may be used to set the overall performance in a future application of the CVT, if both cue timing options are considered valid use cases in the given application.

	Effect (Fixed/ Random)	SS	df	MS	Syn df	Syn MS	F	p
Intercept	_Fixed_	_3383251_	_1_	_3383251_	_12.84_	_6845.43_	_494.24_	_0.00_
spatial	_Fixed_	_40281_	_3_	_13427_	_46.52_	_360.58_	_37.24_	_0.00_
cuetype	_Fixed_	_102637_	_2_	_51318_	_29.31_	_330.72_	_155.17_	_0.00_
cuetime	_Fixed_	_57586_	_1_	_57586_	_16.83_	_543.51_	_105.95_	_0.00_
gender(cuetype)	_Fixed_	_22044_	_2_	_11022_	_659.00_	_288.49_	_38.21_	_0.00_
spatial*cuetype	_Fixed_	_15756_	_6_	_2626_	_659.00_	_288.49_	_9.10_	_0.00_
spatial*cuetime	Fixed	635	3	212	659.00	288.49	0.73	0.53
cuetype*cuetime	Fixed	1574	2	787	659.00	288.49	2.73	0.07
spatial*gender	_Fixed_	_7420_	_6_	_1237_	_659.00_	_288.49_	_4.29_	_0.00_
spatial*cuetype* gender	_Fixed_	_7018_	_6_	_1170_	_659.00_	_288.49_	_4.05_	_0.00_
_TP_	_Random_	_77667_	_13_	_5974_	_18.10_	_643.71_	_9.28_	_0.00_
TP*spatial	Random	14395	39	369	659.00	288.49	1.28	0.12
TP*cuetype	Random	8677	26	334	659.00	288.49	1.16	0.27
TP*cuetime	_Random_	_8048_	_13_	_619_	_659.00_	_288.49_	_2.15_	_0.01_
Error		190113	659	288				

Table 1: Summary of Analysis of Variance (ANOVA). Significant effects ($p < 0.05$) are shown in italics.

Figure 2 shows the combined effect of spatial and cuetype. The largest sensitivity to the spatial contrast is shown for the Text cuetype, with sum score at 58 rau and the three separated conditions around 85 rau, i.e., an effect of approx. 27 rau: the Tukey HSD post-hoc test is significant at $p < 0.00002$. A smaller, but significant, contrast of 15 rau is shown for the Talker cue between Sum and Separate (Tukey HSD: $p < 0.03$). The Audio cuetype has no significant differences across the spatial conditions.

The main effect of cuetiming is also significant with mean scores at 78 rau for Pre and 60 rau for Post (not shown). Interestingly, the cue timing does not interact with any other factors than TP: Thus, cuetiming (Pre vs Post) could be used to shift the overall performance down in a given test, by altering the test paradigm from target-masker to competing voices dual attention. The only interaction with cuetiming is the TP interaction (not shown), indicating that different persons have different gains by going from Post to Pre, which can be explained by the added cognitive load for the Post timing.

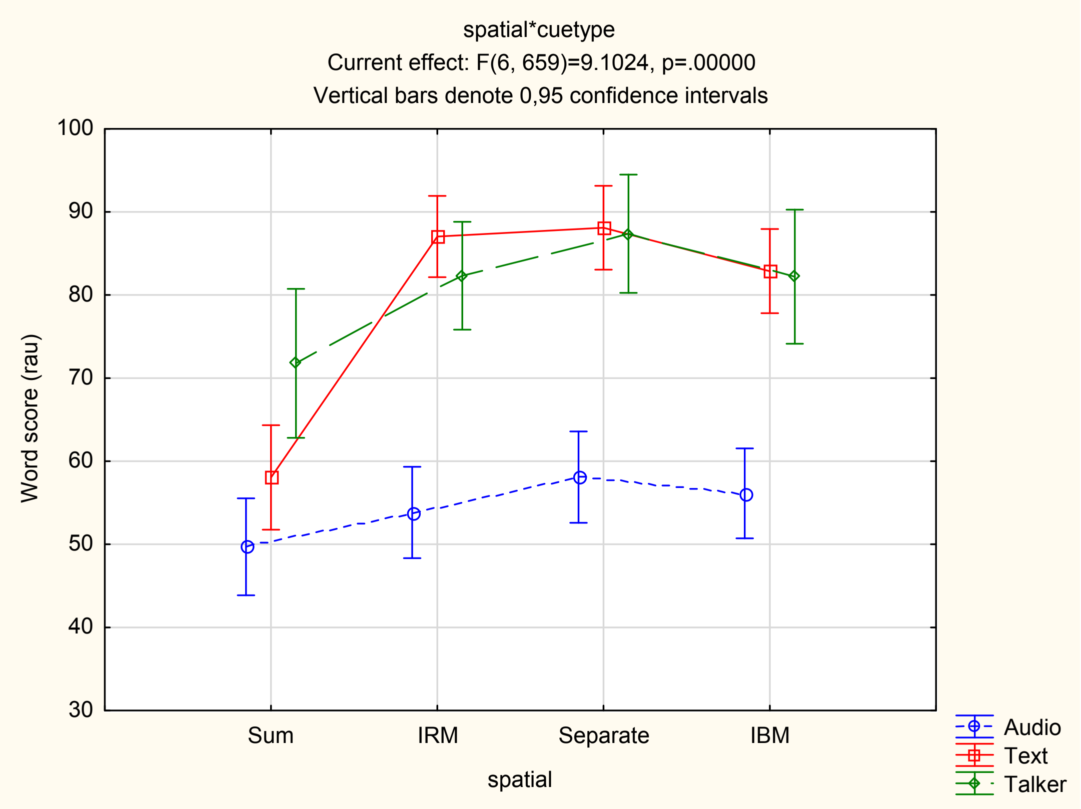


Fig. 2: The combined effect of Spatial and Cue type. The 'Text' cuetype shows the largest contrast between the spatial conditions.

Concerning the spatial modes, we find a large effect of 27 rau between Sum and the three other modes ($p < 0.00002$), and the ideal masks (IBM and IRM) are not significantly different from the perfect separation in Separate. The difference between Sum and Separate is 30 rau, which is a higher contrast than 22 rau obtained in a previous version of the CVT (Bramsløw *et al.*, 2016b).

Regarding gender mix, the test should ideally be insensitive to the gender mixes in order to have a free choice when designing new tests. These results are shown in Fig. 3. The Text cue shows no significant effect of gender mix, while the Audio cue shows a large, significant effect size going from 73 rau to 45 rau (Tukey HSD, $p < 0.00003$). The MM and FF (same gender) pairs have low scores, indicating that the two voices are easily confused when they are same gender, causing a high risk of missing what the target is. The Talker cue is robust as the Text cue, but logically only available for the male-female speaker pairs.

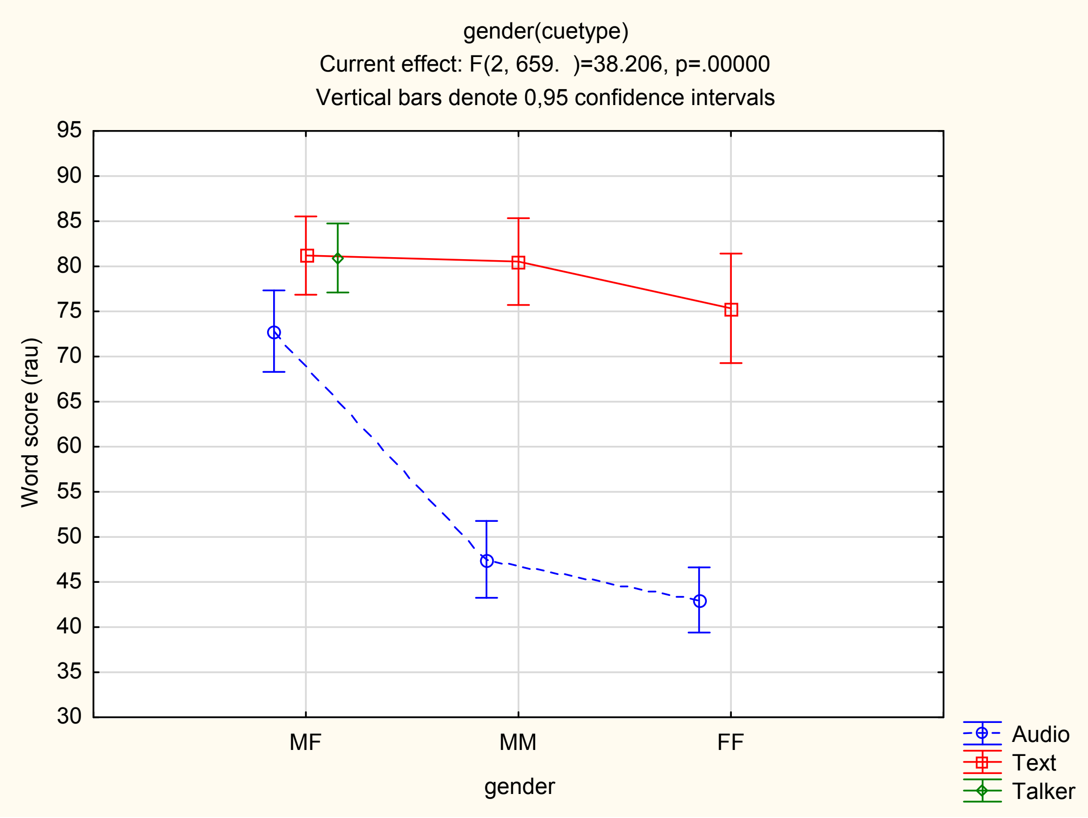


Fig. 3: The combined effect of gender mix and cuetype.

CONCLUSIONS

The CVT is now validated and may be used to evaluate signal processing algorithms such as noise reduction or speech separation algorithms. For the hearing-impaired listeners here we get scores between approx. 40 rau and 90 rau, which is in the middle between floor and ceiling. This should be compared to normal-hearing listeners, who score close to 100, i.e., close to ceiling (Bramsløw *et al.*, 2014). In general, scores do not go far below 50 rau, which may be due to listeners choosing, e.g., one ear consistently, regardless of the cue, which results in a 50% chance level if the intelligibility of a previously chosen target voice is close to 100%.

Among the three cue types Audio, Text and Talker, the Text cue was the most sensitive, providing a 30 rau contrast between Sum and Separate, compared to previously 22 rau (Bramsløw *et al.*, 2016b). The Text cue is recommended for future applications. Regarding the cue timing, the choice between Pre and Post does not affect the sensitivity to the other experimental factors, so it may in future tests be chosen to keep the scores away from floor and ceiling.

Regarding the test of ideal masks with the given time-frequency resolution, the two ideal masks, IRM and IBM are as good as the separated signals. Thus, the applied time-frequency masks are appropriate for testing of different mask-based speech separation algorithms.

When using the CVT, reuse of the ten HINT lists is unavoidable, as each trial uses two lists. Therefore, learning will take place (Bramsløw *et al.*, 2016a), and this needs to be addressed by proper balancing of the test conditions across listeners.

ACKNOWLEDGEMENTS

Thanks to Boris Søndersted, who edited all recordings into single sentence wave files.

REFERENCES

Bramsløw, L., Vatti, M., Hietkamp, R., and Pontoppidan, N.H. (**2014**). "Design of a competing voices test," International Hearing Aid Conference, Lake Tahoe, CA, USA. Retrieved from http://www.eriksholm.com/about-us/news/IHCON_2014

Bramsløw, L., Vatti, M., Hietkamp, R.K., and Pontoppidan, N.H. (**2015a**). "Binaural speech recognition for normal-hearing and hearing- impaired listeners in a competing voice test," Speech in Noise Workshop, Copenhagen, Retrieved from http://www.eriksholm.com/about-us/news/2015/SPIN_2015

Bramsløw, L., Vatti, M., Hietkamp, R., and Pontoppidan, N.H. (**2015b**). "Best application of head related transfer functions for competing voices speech recognition in hearing-impaired," International Symposium on Auditory and Audiological Research, Nyborg, Denmark.

Bramsløw, L., Simonsen, L.B., Hichou, M. El, Hashem, R., and Hietkamp, R.K. (**2016a**). "Learning effects as result of multiple exposures to Danish HINT," International Hearing Aid Conference, Lake Tahoe, CA, USA. Retrieved from http://www.eriksholm.com/about-us/news/IHCON_2016.aspx

Bramsløw, L., Vatti, M., Hietkamp, R.K., and Pontoppidan, N.H. (**2016b**). "A new competing voices test paradigm to test spatial effects and algorithms in hearing aids," International Hearing Aid Conference, Lake Tahoe, CA, USA. Retrieved from http://www.eriksholm.com/about-us/news/IHCON_2016.aspx

Helfer, K.S., Chevalier, J., and Freyman, R.L. (**2010**). "Aging, spatial cues, and single- versus dual-task performance in competing speech perception," J. Acoust. Soc. Am., **128**, 3625–3633. doi: 10.1121/1.3502462

Mackersie, C.L., Prida, T.L., and Stiles, D. (**2001**). "The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res., **44**, 19-28. doi: 10.1044/1092-4388(2001/002)

Naithani, G., Barker, T., Parascandolo, G., Bramsløw, L., Pontoppidan, N.H., and Virtanen, T. (**2017**). "Low-latency sound source separation using convolutional recurrent deep neural networks," IEEE Work. Appl. Signal Process. to Audio Acoust., New Paltz, NY.

Nielsen, J.B., and Dau, T. (**2009**). "Development of a Danish speech intelligibility test," Int. J. Audiol., **48**, 729-741. doi: 10.1080/14992020903019312

Nielsen, J.B., and Dau, T. (**2011**). "The Danish hearing in noise test," Int. J. Audiol., **50**, 202-208. doi: 10.3109/14992027.2010.524254

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. doi: 10.1121/1.408469

Studebaker, G.A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Lang. Hear. Res., **28**, 455. doi: 10.1044/jshr.2803.455

Differences in speech processing among elderly hearing-impaired listeners with or without hearing aid experience: Eye-tracking and fMRI measurements

JULIA HABICHT[1,*], OLIVER BEHLER[1], BIRGER KOLLMEIER[1], AND TOBIAS NEHER[1,2]

[1] *Medizinische Physik and Cluster of Excellence "Hearing4all", Oldenburg University, Oldenburg, Germany*

[2] *Institute of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark*

In contrast to the effects of hearing loss, the effects of hearing aid (HA) experience on speech-in-noise (SIN) processing are underexplored. Using an eye-tracking paradigm that allows determining how fast a participant can grasp the meaning of a sentence presented in noise together with two pictures that correctly or incorrectly depict the sentence meaning (the 'processing time'), Habicht *et al.* (2016, 2017) found that inexperienced HA (iHA) users were slower than experienced HA (eHA) users, despite no differences in speech recognition. To examine the influence of HA use on SIN processing further, the eye-tracking paradigm was adapted for functional magnetic resonance imaging (fMRI) measurements. Groups of eHA ($N = 13$) and iHA ($N = 14$) users matched in terms of age, hearing loss and working memory capacity participated. As before, despite no difference in speech recognition, the iHA group had longer processing times than the eHA group. Furthermore, the iHA group showed more brain activation for SIN relative to noise-only stimuli in left precentral gyrus, cerebellum anterior lobe, superior temporal gyrus and right medial frontal gyrus compared to the eHA group. Together, these results support the idea that HA experience positively influences the ability to process SIN quickly and that it reduces the recruitment of brain regions outside the core speech-comprehension network.

INTRODUCTION

To investigate the effects of cognitive-linguistic processes on speech-in-noise (SIN) processing, Wendt *et al*. (2014) developed an eye-tracking paradigm for estimating *how quickly* a participant can grasp the *meaning* of an acoustic sentence-in-noise stimulus presented concurrently with two similar pictures, only one of which depicts the sentence meaning correctly (the 'processing time'). Previously, Habicht *et al.* (2016, 2017) found that hearing-impaired (HI) listeners with HA experience had shorter processing times than HI listeners without HA experience, despite no differences in speech recognition performance or behavioral reaction times (i.e.,

*Corresponding author: julia.habicht@uni-oldenburg.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
ISBN: 978-87-990013-6-1.

button presses). Based on a literature review, Peelle and Wingfield (2016) concluded that, to compensate for their hearing deficits, HI listeners recruit regions outside the core speech-processing network (comprising middle temporal and inferior frontal gyrus) in order to achieve speech comprehension. Up until now, however, it remains unclear how interventions for hearing impairment (e.g., hearing devices) affect the neuronal processes underlying SIN processing.

The current study aimed to shed some light on how HA use may affect SIN processing abilities by investigating HA experience-related effects on brain activation. To confirm the previously observed difference in sentence processing times, we first made eye-tracking measurements with groups of experienced and inexperienced HA users. To explore differences in brain activation during speech comprehension among the two participant groups, we then performed functional magnetic resonance imaging (fMRI) measurements. For that purpose, we adapted the eye-tracking paradigm for measuring blood oxygenation level-dependent (BOLD) responses. Based on related literature findings, our hypotheses were as follows:

1. The iHA group will have longer processing times than the eHA group.
2. The iHA group will show more brain activation in areas outside the core speech-comprehension network compared to the eHA group.

METHODS

Participants

Thirteen habitual HA users with at least one year of bilateral HA experience (eHA group) and 14 inexperienced HA users with no previous HA experience (iHA group) were recruited. Inclusion criteria were (1) age from 60 to 80 yr, (2) bilateral, sloping, sensorineural hearing loss in the range from 40 to 80 dB HL between 3 and 8 kHz, (3) self-reported normal or corrected-to-normal vision, and (4) no conditions that were contraindicative for fMRI measurements (e.g., a pacemaker). The two groups were matched closely in terms of age, pure-tone average hearing loss calculated across 0.5, 1, 2 and 4 kHz and left and right ears (PTA), working memory capacity as measured using a reading span test (Carroll *et al.*, 2015) and 80%-correct speech reception threshold (SRT$_{80}$) performance (see Table 1).

	eHA	**iHA**
N	13	14
Age (yr)	68.8 (4.0)	68.8 (5.9)
PTA (dB HL)	33.9 (7.4)	31.1 (7.1)
RS (%-correct)	43.0 (11.7)	38.9 (14.2)
SRT$_{80}$ (dB SNR)	−1.6 (1.0)	−1.7 (1.0)

Table 1: Means (and standard deviations) for age, PTA, reading span (RS), and SRT$_{80}$ for the two groups of participants.

Speech-in-noise (SIN) stimuli

For the acoustic stimuli, two sentence structures of the “Oldenburg corpus of Linguistically and Audiologically Controlled Sentences” (OLACS; Uslar *et al*., 2013) were used: (1) subject-verb-object sentences with a canonical word order and therefore ‘low’ linguistic complexity, and (2) object-verb-subject sentences with a non-canonical word order and therefore ‘high’ linguistic complexity (Table 2). In each sentence, there are two characters (e.g., a dragon and a panda), one of which (the subject) performs a given action with the other (the object). In the German language, the linguistic complexity of these sentences is determined by relatively subtle grammatical or acoustic cues, e.g., “Der müde Drache fesselt den großen Panda” (meaning: “The tired dragon ties up the big panda”; low complexity) vs. “Den müden Drachen fesselt der große Panda” (meaning: “The big panda ties up the tired dragon”; high complexity). The stimuli were presented via earphones at the individual SRT$_{80}$. For the masker, stationary speech-shaped noise calibrated to a nominal sound pressure level of 65 dB was used. To ensure audibility, linear amplification in accordance with the “National Acoustic Laboratories-Revised” (NAL-R) prescription formula (Byrne *et al*., 2001) was applied using the Master Hearing Aid research platform (Grimm *et al*., 2006).

Low	De*r*$_{\text{nom}}$	müd*e*$_{\text{nom}}$	Drach*e*	fesselt	de*n*$_{\text{acc}}$	groß*en*$_{\text{acc}}$	Panda.
	Meaning: “The dragon ties up the panda.”						
High	De*n*$_{\text{acc}}$	müd*en*$_{\text{acc}}$	Drach*en*	fesselt	de*r*$_{\text{nom}}$	groß*e*$_{\text{nom}}$	Panda.
	Meaning: “The panda ties up the dragon.”						

Table 2: Examples of sentences from the “Oldenburg corpus of Linguistically and Audiologically Controlled Sentences” (Uslar *et al*., 2013) with two levels of linguistic complexity (low, high). In each case, the grammatically salient *word endings* and corresponding cases (nom = nominative; acc = accusative) are indicated, as are the English meanings.

Eye-tracking measurements

The sentence-in-noise stimuli were presented together with two similar pictures displayed on a monitor in front of the participants. The task of the participant was to identify the picture that matched the acoustic stimulus by pressing a button as fast as possible after the acoustic presentation. During the stimulus presentation, the eye movements of the participant were recorded. If a participant has understood the meaning of a sentence, (s)he will automatically start fixating the corresponding picture. In the following, the time elapsed for this to occur will be referred to as the processing time.

A total of four blocks were performed per participant. Within a block there were 30 trials based on 15 sentences with low linguistic complexity and 15 sentences with high linguistic complexity, plus seven catch trials (see Habicht et al., 2016). The different blocks were presented in randomized order across the different participants.

fMRI measurements

For the fMRI measurements, the eye-tracking paradigm was adapted. The task of the participants was to identify the target picture by pressing a button on a button pad after the presentation of the acoustic stimulus. SIN stimuli with the two levels of linguistic complexity (SIN_{low}, SIN_{high}) were presented together with the corresponding picture sets. In addition, a noise-only condition was included as baseline. In that case, only one picture of a given picture set was displayed, and the task of the participant was to identify the location of the picture (left or right) by pressing a corresponding button on the button pad.

Using this approach, BOLD responses were measured for each participant and stimulus condition (SIN_{low}, SIN_{high}, noise-only). Using the BOLD responses, different contrasts were made to investigate the main effects of stimulus type and linguistic complexity across all participants. The main effect of stimulus type was assessed by contrasting all SIN trials (SIN_{low}, SIN_{high}) with all noise trials (SIN > noise). Based on previous studies, it was expected that the SIN stimuli would lead to more activation in frontotemporal areas including bilateral temporal cortex and left inferior frontal gyrus compared to noise-only stimuli (Adank, 2012; Lee *et al.*, 2016; Rodd *et al.*, 2005). The main effect of linguistic complexity was assessed by contrasting the SIN_{high} and SIN_{low} trials ($\text{SIN}_{\text{high}} > \text{SIN}_{\text{low}}$). It was expected that high-complexity sentences would lead to more activation in frontal lobe (including left inferior frontal gyrus and middle frontal gyrus) compared to low-complexity sentences (e.g., Friederici *et al.*, 2006; Lee *et al.*, 2016; Peelle *et al.*, 2009; Rodd *et al.*, 2005). Additionally, the interaction between participant group and stimulus type was assessed by contrasting the SIN > noise contrast of the iHA group with the SIN > noise contrast of the eHA group (iHA > eHA for SIN > noise). It was expected that to achieve speech comprehension the iHA group would show more brain activation for the contrast SIN > noise in frontotemporal areas in comparison to the eHA group (Peelle and Wingfield *et al.*, 2016; Sandmann *et al.*, 2015). Furthermore, the interaction between participant group and linguistic complexity was assessed by contrasting the $\text{SIN}_{\text{high}} > \text{SIN}_{\text{low}}$ contrast of the iHA group with the $\text{SIN}_{\text{high}} > \text{SIN}_{\text{low}}$ contrast of the eHA group (iHA > eHA for $\text{SIN}_{\text{high}} > \text{SIN}_{\text{low}}$). Based on previous eye-tracking results (Habicht *et al.*, 2016; 2017), it was expected that no group differences would be apparent.

The fMRI data were recorded in one block of 150 trials. Specifically, there were 50 trials per stimulus condition (SIN_{low}, SIN_{high}, noise only). The trials from the three conditions were presented in randomized order. After the 150 trials, a structural image was acquired that served as an anatomical reference.

Test protocol

All participants attended three visits. At the first visit, the SRT_{80} measurements were performed. In addition, event-related potential measurements were carried out for another study. At the second visit, the eye-tracking measurements took place. At the third visit, the fMRI measurements were carried out. The first and second visit took 2 h each, while the third visit took 1 h.

RESULTS

Eye-tracking measurements

On average, the eHA and iHA groups achieved 91.0%-correct (standard deviation: 0.07%-correct) and 89.5%-correct (standard deviation: 0.08%-correct) picture recognition rates. An independent *t*- test revealed no significant difference in terms of picture recognition rates between the two groups ($t_{25} = -0.5, p > 0.05$).

On average, the eHA and iHA groups had longer (poorer) processing times for the sentences with high linguistic complexity (means: 1182 and 1679 ms; standard deviations: 536 and 645 ms) than for the sentences with low linguistic complexity (means: 846 and 1132 ms; standard deviations: 211 and 480 ms). Furthermore, the iHA group had longer processing times than the eHA group (means: 1406 and 1014 ms; standard deviations: 624 and 435 ms). To analyze these data further, we performed an analysis of variance with listener group as between-subject factor and linguistic complexity (low, high) as within-subject factor. Significant effects of listener group [$F(1,25) = 5.5$, $p < 0.026$, $\eta_p^2 = 0.18$] and linguistic complexity [$F(1,25) = 21.0, p < 0.0001$, $\eta_p^2 = 0.46$] were found, but no interaction ($p > 0.05$).

fMRI measurements

On average, the eHA and iHA groups achieved 88.5%-correct (standard deviation: 10.4%-correct) and 84.1%-correct (standard deviation: 4.4%-correct) picture recognition rates. An independent *t*-test revealed no significant difference in terms of picture recognition rates between the two groups ($t_{25} = -1.4, p > 0.05$).

Concerning the effect of stimulus type, the SIN stimuli led to more activation along bilateral superior temporal gyrus, frontal lobe (including left superior frontal gyrus, left inferior frontal gyrus, right middle frontal gyrus and left precentral gyrus) and bilateral middle occipital gyrus compared to the noise-only stimuli ($T = 6.27, p < 0.05$, family-wise-error (FWE) corrected). Figure 1A shows brain regions with increased activation from the SIN > noise contrast analysis.

Concerning the effect of linguistic complexity, the SIN_{high} stimuli led to more activation in bilateral frontal gyrus (including inferior and middle frontal gyrus), left precuneus, right middle occipital gyrus and left temporal lobe (including middle temporal gyrus and superior temporal gyrus) compared to the SIN_{low} stimuli ($T = 3.43$, $p < 0.001$, uncorrected). Figure 1B shows brain regions with increased activation from the $SIN_{high} > SIN_{low}$ contrast analysis.

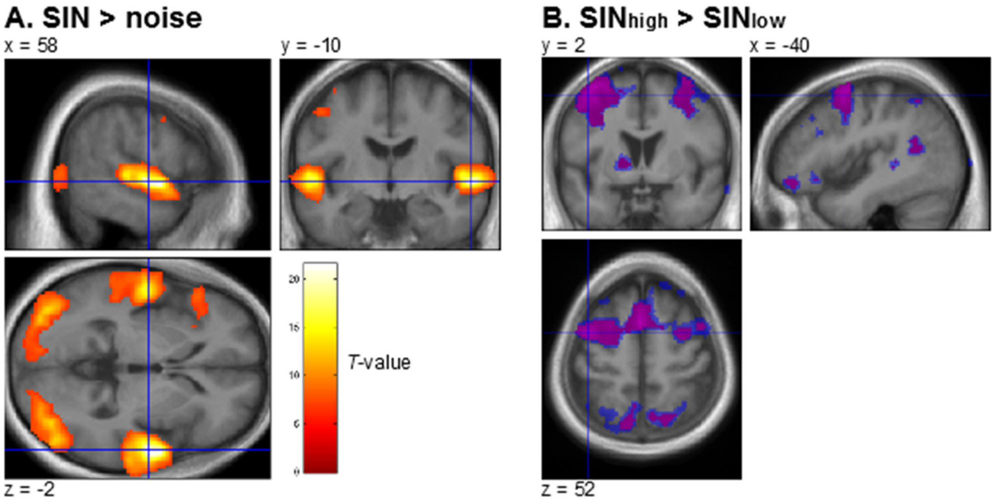


Fig. 1: Sagittal (X), coronal (Y) and axial (Z) views at the location of the global *t*-value maxima (blue crosses). A: Main effect of stimulus type for BOLD contrast SIN > noise at (FWE-corrected $p < 0.05$). B: Main effect of linguistic complexity for BOLD contrast SIN$_{high}$ > SIN$_{low}$ at (uncorrected $p < 0.001$ in purple and uncorrected $p < 0.005$ in blue).

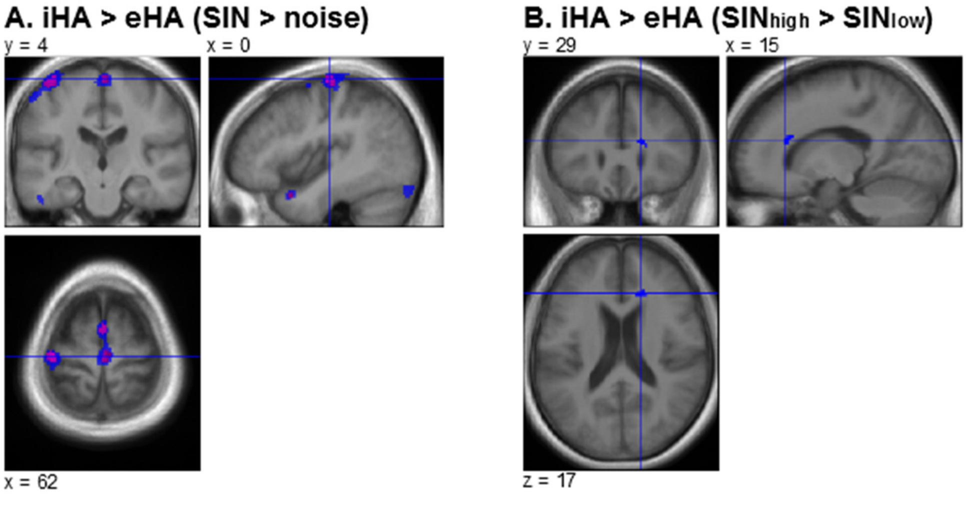


Fig. 2: Sagittal (X), coronal (Y) and axial (Z) views at the location of the global *t*-value maxima (blue crosses). A: Interaction of listener group × stimulus type for BOLD contrast iHA > eHA and SIN > noise (uncorrected $p < 0.001$ in purple and uncorrected $p < 0.005$ in blue). B: Interaction of listener group × ling. complexity for BOLD contrast iHA > eHA and SIN$_{high}$ > SIN$_{low}$ at (uncorrected $p < 0.001$ in purple and uncorrected $p < 0.005$ in blue).

Concerning the interaction between listener group and stimulus type, the iHA group showed more activation for the SIN > noise contrast in left precentral gyrus, left cerebellum anterior lobe, right medial frontal gyrus, and left superior temporal gyrus compared to the eHA group ($T = 3.5, p < 0.001$, uncorrected). Figure 2A shows brain regions with increased activation from the iHA > eHA (SIN > noise) contrast analysis.

Concerning the interaction between listener group and linguistic complexity, no significant contrasts were observed. Figure 2B shows images from the iHA > eHA (SIN_{high} > SIN_{low}) contrast analysis.

SUMMARY AND CONCLUSIONS

In the current study, a cross-sectional design was used to investigate the influence of HA experience on cognitive processes related to sentence comprehension in noise. Using the eye-tracking paradigm of Wendt *et al.* (2014), sentence-in-noise processing times were measured. Additionally, fMRI measurements were performed to measure brain activation patterns in response to SIN and noise-only stimuli. All SIN stimuli were presented at the individual SRT_{80} with individual NAL-R amplification to ensure audibility. The iHA participants had significantly longer processing times than participants matched in terms of age, PTA, working memory capacity and SRT_{80} with at least one year of bilateral HA experience. This is consistent with earlier findings and suggests poorer SIN processing due to untreated hearing loss. Regarding the fMRI measurements, sentences with high linguistic complexity activated additional brain areas in left frontal regions compared to sentences with low linguistic complexity, consistent with the literature. Furthermore, compared to the eHA group the iHA group showed more activation for SIN relative to noise-only stimuli in left precentral gyrus, left cerebellum anterior lobe, right medial frontal gyrus, and left superior temporal gyrus. This suggests that iHA users rely on additional cortical processing to compensate for their hearing deficits to achieve speech comprehension. Altogether, the current study thus confirms that HA experience leads to faster sentence-in-noise processing and also indicates that it reduces the recruitment of brain regions outside the core sentence-comprehension network.

ACKNOWLEDGMENTS

This research was funded by the DFG Cluster of Excellence EXC 1077/1 “Hearing4all”. We thank Timo Gerdes (University of Oldenburg) for performing the speech recognition and eye-tracking measurements, Stefan Uppenkamp (University of Oldenburg) for providing input to the fMRI paradigm, and Katharina Grote (University of Oldenburg) for performing parts of the fMRI measurements.

REFERENCES

Adank, P. (**2012**). “Design choices in imaging speech comprehension: an activation likelihood estimation (ALE) meta-analysis,” Neuroimage, **63**, 1601-1613.

Byrne, D., Dillon, H., Ching, T., Katsch, R., and Keidser, G. (**2001**). “NAL-NL1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures,” J. Am. Acad. Audiol., **12**, 37-51.

Carroll, R., Meis, M., Schulte, M., Vormann, M., Kießling, J., and Meister, H. (**2015**). "Development of a German reading span test with dual task design for application in cognitive hearing research," Int. J. Audiol., **54**,136-141.

Friederici, A.D., Fiebach, C.J., Schlesewsky, M., Bornkessel, I.D., and Von Cramon, D.Y. (**2006**). "Processing linguistic complexity and grammaticality in the left frontal cortex," Cerebral Cortex, **16**, 1709-1717.

Grimm, G., Herzke, T., Berg, D., and Hohmann, V. (**2006**). "The master hearing aid: a PC-based platform for algorithm development and evaluation," Acta Acust. United Ac., **92**, 618-628.

Habicht, J., Kollmeier, B., and Neher, T. (**2016**). "Are experienced hearing aid users faster at grasping the meaning of a sentence than inexperienced users? An eye-tracking study," Trends Hear., **20**. doi: 10.1177/2331216516660966

Habicht, J., Finke, M., and Neher, T. (**2017**). "Auditory acclimatization to bilateral hearing aids: Effects on sentence-in-noise processing times and speech-evoked potentials," Ear Hearing. doi: 10.1097/AUD.0000000000000476

Lee, Y.-S., Min, N.E., Wingfield, A., Grossman, M., and Peelle, J.E. (**2016**). "Acoustic richness modulates the neural networks supporting intelligible speech processing," Hear. Res., **333**, 108-117.

Peelle, J.E., Troiani, V., Wingfield, A., and Grossman, M. (**2009**). "Neural processing during older adults' comprehension of spoken sentences: age differences in resource allocation and connectivity," Cereb. Cortex, **20**, 773-782.

Peelle, J.E., and Wingfield, A. (**2016**). "The neural consequences of age-related hearing loss," Trends Neurosci., **39**, 486-497.

Rodd, J.M., Davis, M.H., and Johnsrude, I.S. (**2005**). "The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity," Cereb. Cortex, **15**, 1261-1269.

Sandmann, P., Plotz, K., Hauthal, N., de Vos, M., Schönfeld, R., and Debener, S. (**2015**). "Rapid bilateral improvement in auditory cortex activity in postlingually deafened adults following cochlear implantation," Clin. Neurophysiol., **126**, 594-607.

Uslar, V.N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T., and Kollmeier, B. (**2013**). "Development and evaluation of a linguistically and audiologically controlled sentence intelligibility test," J. Acoust. Soc. Am., **134**, 3039-3056.

Wendt, D., Brand, T., and Kollmeier, B. (**2014**). "An eye-tracking paradigm for analyzing the processing time of sentences with different linguistic complexities," PLoS ONE, **9**, e100186.

Investigating the effects of noise-estimation errors in simulated cochlear implant speech intelligibility

Abigail Anne Kressner*, Tobias May, Rasmus Malik Thaarup Høegh, Kristine Aavild Juhl, Thomas Bentsen, and Torsten Dau

Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

A recent study suggested that the most important factor for obtaining high speech intelligibility in noise with cochlear implant recipients is to preserve the low-frequency amplitude modulations of speech across time and frequency by, for example, minimizing the amount of noise in speech gaps. In contrast, other studies have argued that the transients provide the most information. Thus, the present study investigates the relative impact of these two factors in the framework of noise reduction by systematically correcting noise-estimation errors within speech segments, speech gaps, and the transitions between them. Speech intelligibility in noise was measured using a cochlear implant simulation tested on normal-hearing listeners. The results suggest that minimizing noise in the speech gaps can substantially improve intelligibility, especially in modulated noise. However, significantly larger improvements were obtained when both the noise in the gaps was minimized and the speech transients were preserved. These results imply that the correct identification of the boundaries between speech segments and speech gaps is the most important factor in maintaining high intelligibility in cochlear implants. Knowing the boundaries will make it possible for algorithms to both minimize the noise in the gaps and enhance the low-frequency amplitude modulations.

INTRODUCTION

Hochberg *et al.* (1992) reported that cochlear implant (CI) recipients typically had thresholds for speech reception in noise that were 10 to 25 dB poorer than normal-hearing listeners. Since then, there has been extensive research in the development of noise reduction algorithms and sound coding strategies in order to obtain an increased robustness to noise. Within this effort, speech intelligibility improvements have been demonstrated by applying both single-microphone noise reduction (e.g., Mauger *et al.*, 2012) and multi-microphone directional noise reduction (e.g., Hersbach *et al.*, 2013). In contrast, although many sound coding strategies have been proposed over the last few decades, none have been able to consistently produce a measured improvement in speech intelligibility in noisy environments over well-established strategies like continuous interleaved sampling (CIS) and the Advanced Combination

*Corresponding author: aakress@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.

ISBN: 978-87-990013-6-1.

Encoder (ACE^{TM}, Cochlear Ltd., New South Wales, Australia). One potential reason for this lack of success is that relatively little is known about how different kinds of errors in CI stimulation specifically influence speech intelligibility outcomes.

In an effort to improve this understanding, Qazi *et al.* (2013) investigated the effects of noise on electrical stimulation sequences and speech intelligibility in CI recipients. They suggested that noise affects stimulation sequences in three primary ways: (1) noise-related stimulation can fill the gaps between speech segments, (2) stimulation levels during speech segments can become distorted, and (3) channels which are dominated by noise can be selected for stimulation instead of channels which are dominated by speech. In order to measure the effect of each of these, Qazi *et al.* (2013) generated several artificial stimulation sequences, each of which contained different combinations of these errors. They presented these artificial stimulation sequences to CI recipients, as well as normal-hearing listeners with a vocoder, and measured speech intelligibility. Their results indicated that the most important factor for maintaining good speech intelligibility was the preservation of the low-frequency (i.e., what they called "ON/OFF") amplitude modulations of the clean speech by, for example, minimizing the noise presented in speech gaps.

Koning and Wouters (2012), however, argued that it is the information encoded in the transient parts of the speech signal that contributes most to speech intelligibility. Accordingly, they demonstrated that enhancing speech onset cues alone improves speech intelligibility in CI recipients (Koning and Wouters, 2016). By comparison, Qazi *et al.* (2013) also inherently enhanced onset and offset cues in the conditions where they removed noise in the gaps between speech, because they always identified these segments via ideal onset and offset detection. Therefore, by removing noise in the speech gaps in their experiment, they simultaneously enhanced the saliency of the onsets and offsets. Qazi *et al.* (2013) did not, however, investigate the effect of reducing noise in the gaps when the boundaries between the speech segments and speech gaps were not perfectly aligned. Therefore, it is unclear how advantageous the minimization of the noise in speech gaps is when it does not co-occur with accurate onset and offset cues. Furthermore, the importance of the separation of these two factors becomes clear when considering that realistic algorithms will not be able to perfectly identify the boundaries between speech segments and speech gaps.

The main motivation of the present study was to systematically quantify the relative impact of realistic noise-estimation errors occurring within speech segments, speech gaps, and speech transients. Specifically, this study investigated these distortions using a basic CI vocoder simulation tested with normal-hearing listeners, which provides insight into the impact of the spectro-temporal degradation in isolation from an impaired auditory system.

METHODS

A CI with an *N*-of-*M* strategy such as ACE encodes sound by first separating the input signal into *M* channels and subsequently stimulating a subset of at most *N* channels at each frame l. In this study, speech was divided into 128-sample overlapping frames, and then a Hann window and the short-time discrete Fourier Transform (STFT) was applied with $K = 128$ points to obtain the time-frequency representation of speech, $X(k,l)$. The STFT magnitudes were then combined into $M = 22$ channels using non-overlapping rectangular weights with spacing that matches Cochlear Ltd.'s (New South Wales, Australia) sound processor in order to obtain the time-frequency representation $X(m,l)$, where m represents the channel index and l represents the frame index. A new frame was calculated every 1 ms.

In the Qazi *et al.* (2013) study, sentences were divided temporally into speech segments and speech gaps. Artificial sequences were then synthesized by copying segments from the clean speech sequence and noisy speech sequence. In the present study, sentences were instead divided into three temporal regions (i.e., speech segments, speech gaps, and speech transitions). This protocol allows for the separation of the reduction of noise in the speech gaps from the encoding of the transitions. In order to do this segmentation, broadband channel activity, $A(l)$ was defined for each frame as the number of channels containing speech above a threshold:

$$A(l) = \sum_{m=1}^{M} T_{\lambda}\left(X(m,l)\right), \qquad \text{(Eq. 1)}$$

where the function $T_{\lambda}(\cdot)$ performs element-wise thresholding and returns a value of one for elements that are above 25 dB sound pressure level (i.e., the default threshold level in ACE). As in the Qazi *et al.* (2013) study, speech segment onsets were then identified as frames in which $A(l) = 0$ and $A(l+1) > 0$, and speech segment offsets were defined as frames in which $A(l) > 0$ and $A(l+1) = 0$. Speech segments with $A(l) \leq 1$ for the duration of the segment were dropped, and speech segments shorter in duration than 20 ms that were close in time to another speech segment were merged together. The merging prevented rapid switches between speech and non-speech labels. Subsequently, a transition region was defined at each onset and offset as the 10 ms before and the 10 ms after, such that a region of 20 ms in duration was created at the start and end of each speech segment. Finally, the remaining frames were labeled as speech gaps. An example stimulation sequence for a clean sentence is shown in Fig. 1(a), with the temporal regions indicated by the underlying shading. The 20-ms duration for the transition region was heuristically chosen in order to ensure the transition regions were long enough to be perceptible, but short enough to maintain a segmentation that was still comparable to the segmentation in Qazi *et al.* (2013).

Whereas Qazi *et al.* (2013) primarily manipulated channel selection and current levels within each temporal region in order to investigate the impact of noise-induced errors in sound coding strategies, the present study manipulated the gains that are applied in a preceding noise reduction stage in order to investigate the impact of noise-induced errors in noise reduction algorithms. Therefore, instead of synthesizing stimulation patterns from the clean and noisy speech, artificial gain matrices were synthesized

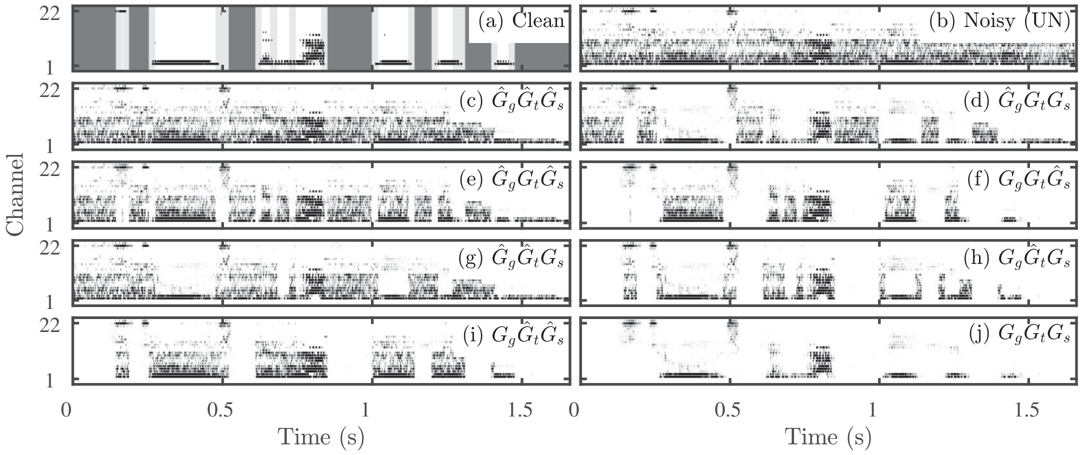


Fig. 1: Electrodograms showing (a) stimulation levels above threshold for the Danish sentence, *Stuen skal nok blive hyggelig* and (b-j) unthresholded levels for the same sentence mixed with speech-weighted noise and then de-noised using the indicated gain matrix. Speech segments, transitions, and gaps are identified in (a) by the white, light gray, and dark gray shading, respectively.

from either the *a priori* local signal-to-noise ratios (SNRs) or from estimated SNRs using a CI-optimized noise reduction algorithm (Mauger *et al.*, 2012). An underlying assumption in this study then is that a maxima selection strategy, such as ACE, will stimulate the correct set of channels if it chooses channels from a representation that has been sufficiently de-noised.

The following general signal model was thereby considered: $Y(k,l) = X(k,l) + D(k,l)$, with $X(k,l)$ representing the clean speech, $D(k,l)$ representing the noise signal, and $Y(k,l)$ representing the noisy speech signal. An estimate of the noise spectrum $\hat{D}(k,l)$ was computed from the noisy signal $Y(k,l)$ using the improved minimum controlled recursive algorithm (Cohen, 2003). $\hat{D}(m,l)$ was then computed using the same rectangular weights as were used for computing $X(m,l)$ from $X(k,l)$, and a smoothed SNR estimate $\hat{\xi}(m,l)$ was obtained using a CI-optimized smoothing technique (Mauger *et al.*, 2012). From $\hat{\xi}(m,l)$, gains $\hat{G}(m,l)$ were obtained using the CI-optimized gain function (Mauger *et al.*, 2012),

$$\hat{G}(m,l) = \left(\frac{\hat{\xi}(m,l)}{\hat{\xi}(m,l) + 2.92} \right)^{1.2}. \tag{Eq. 2}$$

Additionally, the ideal gains $G(m,l)$ were computed using the *a priori* instantaneous signal-to-noise ratio $\xi(m,l)$.

Artificial gain matrices were synthesized by concatenating segments from either $\hat{G}(m,l)$ or $G(m,l)$ for each of the three temporal regions. For example, to understand the impact of errors specifically in the speech gaps, gains from $G(m,l)$ were applied to the noisy signal $Y(m,l)$ in all of the speech gaps, and gains from $\hat{G}(m,l)$ were applied in all of the speech transitions and speech segments. This condition was named $G_g\hat{G}_t\hat{G}_s$ to signify that the estimated gains were corrected in the speech gaps, but not in the transitions and the speech segments. Accordingly, the condition $\hat{G}_gG_t\hat{G}_s$ signifies that the estimated gains were corrected in the speech transitions, and it follows that the condition $G_gG_tG_s$ signifies that the estimated gains were corrected in all of the temporal regions, which is equivalent to ideal Wiener processing with a CI-optimized gain function.

The final stimulation sequence was computed by selecting the $N = 8$ channels with the largest remaining energy. An acoustic signal was then constructed from the stimulation sequence using a 22-channel noise vocoder. Figure 1(b) shows the sequences for a noisy version of the sentence in Fig. 1(a), and Figs. 1(b-j) show the sequences after de-noising with each type of gain matrix. A visual comparison between Figs. 1(c) and 1(j) highlights the extent of the estimation errors in $\hat{G}_g\hat{G}_t\hat{G}_s$. Subsequently, the remaining figures in the left column contain the stimulation patterns for the conditions where just one of the temporal regions of the gain matrix have been corrected. Lastly, the remaining plots in the right column each show the stimulation patterns for the conditions where two of the temporal regions have been corrected.

Speech intelligibility was evaluated in six participants by obtaining speech reception thresholds (SRTs) of sentences in noise via the Danish hearing in noise test (HINT) (Nielsen and Dau, 2011). Through an adaptive procedure, HINT determines the SNR at which the participants were able to understand 50% of the sentence material. Testing was carried out in a double-walled booth, using equalized Sennheiser HD-650 circumaural headphones. Participants were at least 18 years of age, had audiometric thresholds of less than or equal to 20 dB HL in both ears (125 Hz to 8 kHz), and were native Danish speakers. All participants provided informed consent, and the experiment was approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391). The participants were paid for their participation.

At the start of the session, participants first heard vocoded sentences in quiet and then in noise to become familiar with the task. Testing subsequently commenced with either stationary speech-weighted noise (Nielsen and Dau, 2011) or the International Speech Test Signal (Holube *et al.*, 2010) (i.e., a modulated noise that is speech-like but unintelligible), and then testing proceeded with the other. The presentation order of the noise types was counterbalanced across participants. There were eight noise reduction conditions, and together with the reference, noisy condition (i.e., unity gains), there were nine test conditions for each noise type. Two SRTs were collected per condition, and the mean of the two was used for analysis. For two of the participants, only one SRT was collected for a small subset of the test conditions, and therefore, these three data points did not include test-retest averaging. None of these points were outliers. Since the Danish HINT contains only ten lists, participants heard the first nine lists multiple times, in a random order each time.

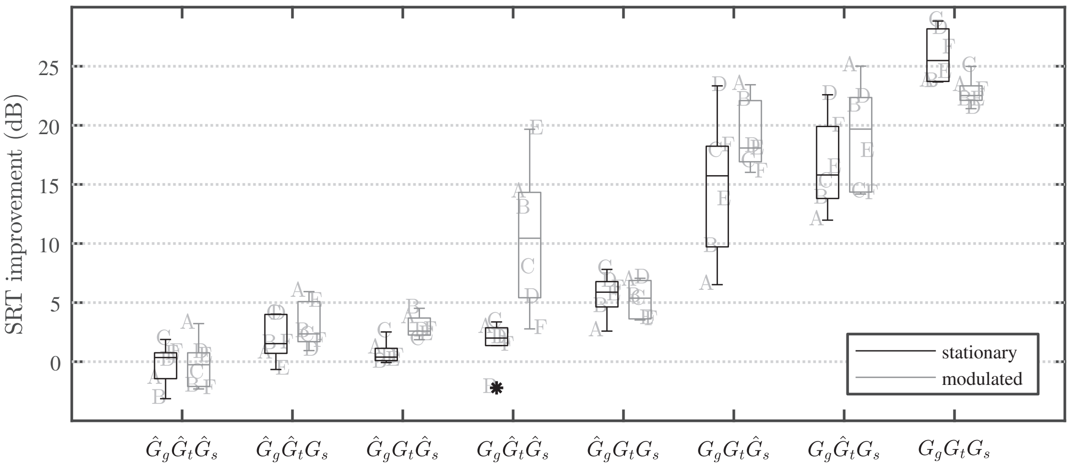


Fig. 2: Speech reception threshold (SRT) improvements relative to the reference noisy condition. Box plots show the 25^{th}, 50^{th}, and 75^{th} percentiles, together with whiskers that extend to extreme data points not considered outliers. Outliers are marked with an asterisk. Letters correspond to individual participants.

RESULTS

Figure 2 shows the improvement in SRT for each individual relative to their average SRT in the reference noisy condition. Because normal-hearing listeners generally do not benefit from single-microphone noise reduction algorithms (Hu and Loizou, 2007), it is not surprising that the CI-optimized noise reduction algorithm (i.e., $\hat{G}_g\hat{G}_t\hat{G}_s$) did not provide an SRT improvement, on average. Similarly, it is not surprising that the average SRT improvement was around 25 dB when *a priori* information about the local SNRs was used (i.e., $G_gG_tG_s$), as this was the maximum possible benefit given the constraints of the testing software.

Focusing first on the impact of errors in the speech gaps (i.e., $G_g\hat{G}_t\hat{G}_s$ versus $\hat{G}_g\hat{G}_t\hat{G}_s$), SRTs tended to improve in the stationary noise, and substantially improved in the modulated noise —though to varying degrees across participants— when the errors in the gaps were removed. This result suggests that minimizing noise-dominated stimulation in the speech gaps is an important factor for improving intelligibility, which is in line with the conclusions in Qazi *et al.* (2013).

However, in comparison to correcting the errors in the speech gaps, correcting errors in the speech segments (i.e., $\hat{G}_g\hat{G}_tG_s$) yielded, on average, a smaller SRT benefit, especially with regard to the modulated noise type. In a similar manner, correcting gain errors in the transition regions (i.e., $\hat{G}_gG_t\hat{G}_s$) yielded a relatively small SRT benefit, particularly in the stationary noise. This result was unexpected, however, given that the previous body of literature suggests that increased gain in transition regions (e.g., Vandali, 2001), or specifically at the onsets (e.g., Koning and Wouters, 2016), significantly improves speech intelligibility for CI recipients in both stationary noise and in the presence of a competing talker. Thus, it is possible that CI listeners rely more on these cues than normal-hearing listeners with a vocoder simulation. Alternatively, it may be that the detrimental effect of the sudden changes in gains in these stimuli were larger than the benefit of encoding the transitions correctly.

Despite the relatively small impact when only correcting gain errors in the transitions alone, the combination of correcting errors in the transitions and correcting errors in the gaps resulted in substantial improvements in SRTs. Furthermore, the benefit from correcting gain errors in both of these regions is much larger than the sum of the benefit from each in isolation. This result suggests that there is a strong interaction between gain errors in speech gaps and gain errors in the transitions, which implies that the potential benefit of minimizing stimulation from noise-dominated channels in speech gaps largely depends on how accurately the boundaries between the gaps and segments of speech are encoded.

CONCLUSION

Qazi *et al.* (2013) suggested that the most important factor for attaining high speech intelligibility in noise with CI listeners is to preserve the low-frequency amplitude modulations of speech across time and frequency in the stimulation patterns. In their study, both CI recipients and normal-hearing listeners tested with a vocoder simulation achieved the largest improvement in intelligibility when there was no stimulation in the gaps between speech segments. In a realistic algorithm, however, the identification of these regions will be imperfect, and the results from the current study suggest that the benefit of attenuating noise-dominated stimulation presented in speech gaps is largely diminished when the transitions between the speech and speech gaps are distorted. Although some listeners in the current study obtained a very large benefit in modulated noise with the minimization of gain errors in the gaps, even when errors in the transitions remained present, their intelligibility improvement is likely attributed to the fact that they could listen in the dips for salient onset cues. Since CI recipients are typically less able to listen in the dips (Nelson *et al.*, 2003), this benefit is likely to be less pronounced in CI listeners. Therefore, removing stimulation in the speech gaps may not itself be such a key component to improving speech intelligibility in noise in CI listeners. Instead, a more effective goal may be to identify the boundaries between the speech and gaps, so that, while minimizing the stimulation of noise-dominated channels in the gaps, it will also be possible to deliver salient cues related to the transients. These two components together seem to contribute the most to understanding speech in noise, at least with normal-hearing listeners tested with speech degraded by a vocoder simulation.

ACKNOWLEDGMENT

This work was supported by the Danish Council for Independent Research (DFF) with grant number DFF-5054-00072.

REFERENCES

Cohen, I. (**2003**). "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," IEEE Speech Audio Process., **11**, 466-475.

Hersbach, A.A., Grayden, D.B., Fallon, J.B., and McDermott, H.J. (**2013**). "A beamformer post-filter for cochlear implant noise reduction," J. Acoust. Soc. Am., **133**, 2412-2420.

Hochberg, I., Boothroyd, A., Weiss, M., and Hellman, S. (**1992**). "Effects of noise and noise suppression on speech perception by cochlear implant users," Ear Hearing, **13**, 263-271.

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.

Hu, Y., and Loizou, P.C. (**2007**). "A comparative intelligibility study of single-microphone noise reduction algorithms," J. Acoust. Soc. Am., **122**, 1777-1786.

Koning, R., and Wouters, J. (**2012**). "The potential of onset enhancement for increased speech intelligibility in auditory prostheses," J. Acoust. Soc. Am., **132**, 2569-2581.

Koning, R., and Wouters, J. (**2016**). "Speech onset enhancement improves intelligibility in adverse listening conditions for cochlear implant users," Hear. Res., **342**, 13-22.

Mauger, S.J., Arora, K., and Dawson, P.W. (**2012**). "Cochlear implant optimized noise reduction," J. Neural Eng., **9**, 1-9.

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.

Qazi, O. ur R., van Dijk, B., Moonen, M., and Wouters, J. (**2013**). "Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility," Hear. Res., **299**, 79-87.

Vandali, A.E. (**2001**). "Emphasis of short-duration acoustic speech cues for cochlear implant users," J. Acoust. Soc. Am., **109**, 2049-2061.

Contribution of low- and high-frequency bands to binaural unmasking in hearing-impaired listeners

GUSZTÁV LŐCSEI[1], SÉBASTIEN SANTURETTE[1,2], TORSTEN DAU[1], AND EWEN N. MACDONALD[1,*]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Department of Otorhinolaryngology, Head and Neck Surgery & Audiology, Rigshospitalet, Copenhagen, Denmark*

This study investigated the contribution of interaural timing differences (ITDs) in different frequency regions to binaural unmasking (BU) of speech. Speech reception thresholds (SRTs) and binaural intelligibility level differences (BILDs) were measured in two-talker babble in 6 young normal-hearing (NH) and 9 elderly hearing-impaired (HI) listeners with normal or close-to-normal hearing at and below 1.5 kHz. Target sentences were presented diotically, embedded in a stream of diotic or dichotic maskers. Both target and masker sentences were split into frequency regions above and below 1.25 kHz. In the dichotic listening conditions, the maskers were lateralized to the left side by introducing 0.68-ms ITDs in either the low-frequency band, the high-frequency band, or both bands simultaneously. BILDs were found to be similar in both listener groups when the ITDs were imposed on the low-frequency band only. ITDs in the high-frequency band alone did not produce any BILD in any of the groups. However, when ITDs were imposed in both frequency bands, the NH listeners yielded significantly greater BILDs than the HI listeners. The results suggest that, on a group level, HI listeners relied solely on ITDs in the low-frequency band while NH listeners were able to utilize envelope ITDs above 1.25 kHz to facilitate the BU of speech.

INTRODUCTION

Studies investigating the binaural unmasking (BU) of speech in noise have typically found that binaural intelligibility level differences (BILDs) are determined by interaural timing differences (ITDs) in the low-frequency domain (e.g., Levitt and Rabiner, 1967; Bronkhorst and Plomp, 1988; Edmonds and Culling, 2005) suggesting that ITD cues related to temporal fine-structure (TFS) rather than the envelope (ENV) carry critical information for the BU of speech. While hearing impaired (HI) listeners often exhibit reduced TFS sensitivity, several studies (e.g., Neher *et al.*, 2012; Santurette and Dau, 2012; Lőcsei *et al.*, 2016) have shown no or only a moderate correlation between speech intelligibility scores in spatial settings and behavioural measures of TFS ITD sensitivity. A possible explanation for the weak relationship between TFS

[*] Corresponding author: emcd@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
ISBN: 978-87-990013-6-1.

ITD sensitivity and BILDs in various conditions can be that HI listeners utilize ITD cues in the ENV of high-frequency channels.

Edmonds and Culling (2005) investigated how ITDs in isolated frequency bands contributed to BILDs in young normal-hearing (NH) listeners. Their results indicated that ITDs in the frequency regions both below and above 1.5 kHz provided some masking release, but also that the full advantage was only achieved when ITDs were present over the full spectrum. Therefore, it appears that NH listeners exploit ENV ITDs at higher frequencies to aid speech perception.

In the current study, the contribution of TFS and ENV ITDs in different frequency regions to BILD was evaluated in young NH and older HI listeners. BILDs were measured in a speech-on-speech task. The target and the interferers were divided into two independent low- and high-frequency regions, and ITDs were imposed on the interferers in the low, high, or both frequency domains.

METHODS

Listeners

Six young NH (mean age: 24.2 years, standard deviation (SD): 2.2) and 9 older HI (mean age: 69.6 years, SD: 5.5) participated in the study. For each listener, air-conduction audiometric thresholds were measured at octave frequencies between 125 Hz and 8 kHz and between 750 Hz and 6 kHz. All NH listeners had thresholds below 25 dB HL at all measured frequencies. Most of the HI listeners had normal hearing below 1.5 kHz, but a mild-to-moderate hearing loss at frequencies above 1.5 kHz. In all listeners, the hearing thresholds between the ears differed by at most 15 dB at each tested audiometric frequency. The average hearing thresholds for the HI listeners are displayed in Table 1. All listeners provided written consent and received compensation for their efforts. All but one listener were tested over a single visit lasting between two and three hours. One NH listener was tested over two visits.

			Audiometric thresholds averaged between the ears [dB HL]											Pure-tone averages		
ID	Sex	Age	125	250	500	750	1k	1.5k	2k	3k	4k	6k	8k	PTA$_{low}$	PTA$_{high}$	PTA$_{oct}$
a	f	58	5	0	2.5	2.5	2.5	-5	7.5	17.5	30	35	17.5	0.5	21.5	8.5
b	f	66	12.5	7.5	7.5	10	10	5	10	10	25	25	40	8	22	12
c	f	75	22.5	15	10	7.5	5	5	15	15	22.5	17.5	50	8.5	24	13.5
d	m	68	20	12.5	10	7.5	7.5	5	5	22.5	32.5	37.5	67.5*	8.5	33	13.5
e	f	72	10	5	10	7.5	5	15	22.5*	40	30	45	72.5	8.5	42	14.5
f	f	67	17.5*	10	12.5	7.5	7.5	12.5*	35	50	55	52.5	50	10	48.5	24
g	f	72	22.5	12.5	20	15	20	22.5*	22.5*	27.5	27.5	30	52.5	18	32	20.5
h	m	76	15	12.5	20	20	15	25	25	45	57.5	52.5	67.5	18.5	49.5	26
i	f	72	17.5	25	22.5	20	20	22.5	27.5*	37.5*	40	52.5*	65	22	44.5	27

Table 1: Gender, age, and audiometric thresholds (air-conduction, averaged across both ears), and pure-tone averages of the HI listeners. Except for cases marked by asterisks, the differences in audiometric thresholds between left and right ears were less than or equal to 10 dB.

Binaural temporal fine structure coding

The listeners' sensitivity to binaural TFS information was assessed by measuring the upper frequency limit at which listeners were able to detect an interaural phase shift of 180° (IPD_{fr}) using the same 3-interval 3-alternative forced-choice paradigm as Lőcsei *et al.* (2016). Reference and target intervals were presented at 30 dB sensation level (SL) and contained 4 tone bursts presented diotically or alternated between the diotic and dichotic presentation modes. The tone bursts were 300-ms long, gated with 50-ms raised-cosine ramps and separated by 100-ms silent gaps. The intervals were separated by 400-ms silent gaps. Six thresholds were evaluated for each listener, and the final threshold was calculated as the geometric mean of the last 3 thresholds.

Speech intelligibility tests

Speech intelligibility was assessed using the open-set DAT corpus (Nielsen *et al.*, 2014). The "Dagmar" sentences were presented via headphones as target material against a two-talker masker (TT), which consisted of sentence pairs spoken by the two other talkers of the same corpus.

The target sentences were always presented diotically. In the reference condition, the maskers were colocated with the target (TT_{co}). In the remaining conditions, the maskers were lateralized towards the right side by imposing a 0.68-ms timing delay in the left channel. This delay was either imposed on the full spectrum (TT_{bb} for "broad band"), the low spectral region (TT_{lp}, "low pass"), or the high spectral region (TT_{hp}, "high pass") of the maskers. In order to manipulate the ITD relations in low- and high-frequency bands independently, low-pass and high-pass filtered versions of the original maskers were created and the time delays were applied to the left-ear channel in the corresponding frequency regions. The resulting low- and high-frequency time signals were then added in each channel and presented to the listener. The cutoff frequencies of the low-pass and high-pass filters were set to 1173 Hz and 1332 Hz, respectively, corresponding to a 1 equivalent rectangular band (ERB) notch centered at 1.25 kHz between the low-pass filtered and high-pass filtered parts. Filter slopes were greater than 500 dB/oct in both cases in order to prevent any interactions between the two spectral regions.

In each condition, sentence correct SRTs were measured by varying the level of the maskers in 2-dB steps. The initial signal-to-noise (SNR) ratio was set to 3 dB in the TT_{co} and to 0 dB in all the other conditions. When calculating the SRTs for each list, the presentation levels of the maskers from the 5th to the hypothetical 21st sentence were averaged and subtracted from the presentation level of the target. SRTs were measured over 3 lists in each condition and the final SRT value for each condition was calculated as the average of these, expressed in SNR. Overall, 12 lists were used in the testing phase, and 3 additional lists in the training phase, two of which were presented in the TT_{co} condition and one in the TT_{bb} condition.

The frequency range of the stimuli was restricted to between 100 Hz and 10 kHz.

A 512 order finite impulse response filter was used to compensate for the frequency response of the headphones, and to simulate the frequency response of the outer ear in a diffuse-field listening scenario. This filter also compensated for the loss of stimulus audibility based on the hearing thresholds of the individuals and the long-term average spectrum of the target speech. The audibility criterion used was 15 dB at and below 3 kHz, which was reduced to 12 dB, 8 dB and 0 dB at 4, 6 and at 8 kHz and above. The target sentences were presented at a nominal level of 65 dB sound pressure level (SPL) "free field". The presentation levels were limited to at most 94 dBA. If the estimated level of a trial exceeded this level, it was scaled down in 2 dB steps to be below this upper limit.

RESULTS

The results of the IPD_{fr} experiments are shown in Fig. 1. Horizontal black lines denote the group means and the white and gray boxes indicate ±1 SD of the NH and HI listener groups, respectively. For analysis purposes, the thresholds were log-transformed. The difference in average IPD_{fr} thresholds between groups was statistically significant [$t(13) = -4.65$, $p < 0.001$].

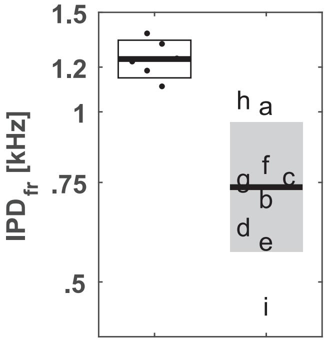


Fig. 1: The upper frequency limit for detecting a 180° interaural phase difference in the TFS of a pure tone. The dark horizontal lines with the white and gray boxes stand for the mean and ±1 SD of the NH and HI listener groups, respectively. Dots and letters denote individual thresholds within the corresponding groups.

The lowest and highest thresholds in the HI group were 456 Hz and 1056 Hz, showing a large spread of how the individual listeners performed. Within the HI group, neither age nor PTA_{low} was correlated with the IPD_{fr} thresholds. The HI listeners showed similar binaural TFS processing abilities as the HI listeners tested in Lőcsei *et al.* (2016).

Figure 2 shows the group means and standard deviations of the SRTs for the two listener groups (NH: white boxes, HI: gray boxes) in the SI experiments. SRTs in the TT_{bb} and TT_{lp} conditions were lower than in the TT_{co} condition, indicating a masking release when ITDs were imposed at least on the low-frequency part of the maskers. In contrast, SRTs were similar in the TT_{co} and TT_{hp} conditions for both listener groups. For the NH listeners. SRTs were slightly higher in the TT_{lp} condition than in the TT_{bb} condition. In contrast, HI listeners performed similarly in the two conditions.

The BILDs were calculated as the difference between SRTs in the TT_{co} and all the other conditions. The group means and standard deviations are shown in Fig. 3. Group

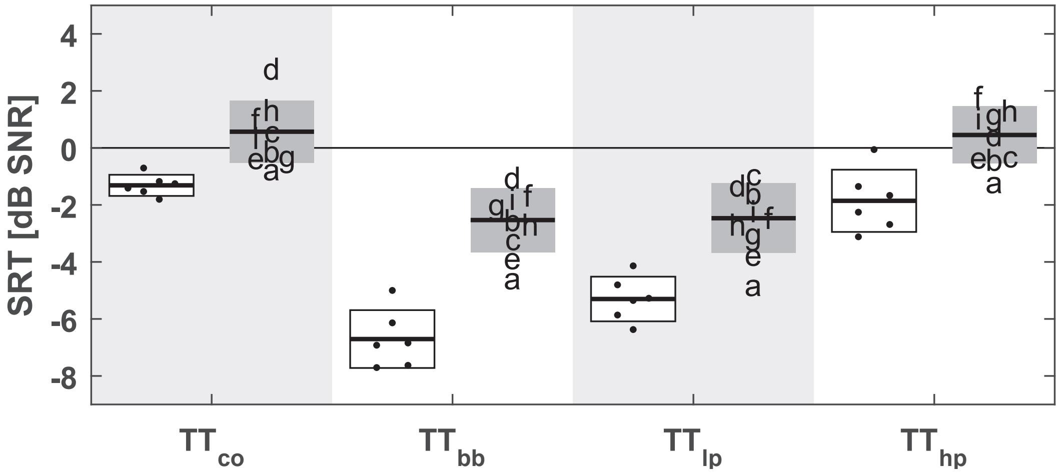


Fig. 2: Speech reception thresholds of the NH and HI listeners in the speech intelligibility tests. The target was always presented diotically. The maskers were presented either diotically (TT_{co}), or lateralized to the right side in the low or high frequency domains (TT_{lp} or TT_{hp}), or over the full frequency domain (TT_{bb}). The dark horizontal lines and the white and gray boxes indicate for the mean and ±1 SD of the NH and HI listener groups, respectively. Dots and letters denote individual thresholds within the corresponding groups.

differences in BILDs were highest in the TT_{bb} condition. Group differences were less pronounced in the TT_{lp} condition. HI listeners exhibited similar BILDs in the TT_{bb} and TT_{lp} conditions.

A mixed-design ANOVA was conducted on the BILDs obtained in the TT_{bb} and TT_{lp} conditions, with filtering as within-subject and listener group as between-subject factors. The analysis revealed a significant main effect of filtering [$F(1,13) = 5.647$, $p = 0.034$], listener group [$F(1,13) = 14.95$, $p = 0.002$], and interaction between filtering and listener group [$F(1,13) = 4.69$, $p = 0.0496$]. For the NH listeners, there was a trend towards larger BILDs in the TT_{bb} than in the TT_{lp} condition [paired t-test, $t(5) = 2.44$, $p = 0.058$]. For both groups, BILDs in the TT_{hp} were not not significantly different from 0 (one-sample t-tests, $p > 0.05$).

DISCUSSION

The results of the present study indicate that, for young NH listeners, frequencies above 1.25 kHz can contribute to the BU of speech, which is consistent with the findings of Edmonds and Culling (2005). Furthermore, it was found that young NH listeners exhibited larger BILDs than older HI listeners when ITDs were imposed on the whole frequency range. Both listener groups exhibited BU when the target and the maskers were separated by ITDs only below 1.25 kHz, and the magnitude of the

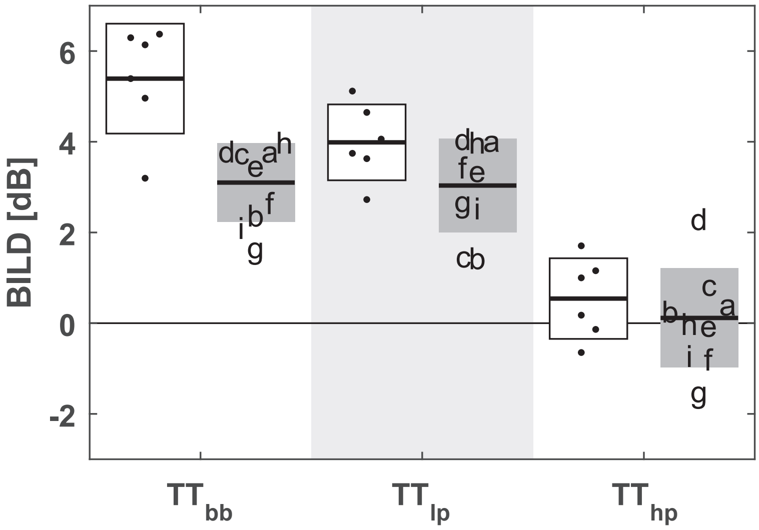


Fig. 3: Binaural intelligibility level differences of the NH and HI listeners obtained in the speech intelligibility tests. Condition notations are the same as in Fig. 2. The dark horizontal lines with the white/gray boxes show mean BILDs and ±1 SD for the NH and HI listener groups, respectively.

BILDs was comparable in the two groups. When ITDs were imposed above, but not below, 1.25 kHz, no BILD was observed. The results suggest that, both in young NH and older HI listeners, BILDs are mainly facilitated in the low-frequency region of the stimuli. This finding is consistent with the conclusions of earlier reports investigating BU (e.g., Levitt and Rabiner, 1967; Bronkhorst and Plomp, 1988; Edmonds and Culling, 2005). The contributions of ITDs at high frequencies to BILDs seem to be negligible when presented in isolation. However, in contrast to the HI listeners, NH listeners could utilize high-frequency ITD information to some degree to aid speech understanding in the $\mathrm{TT_{bb}}$ condition. For the NH listeners, the differences in BILDs between the $\mathrm{TT_{bb}}$ and $\mathrm{TT_{lp}}$ were not statistically significant, likely due to the relatively low number of listeners tested.

As the splitting frequency between the low- and high-frequency speech bands was set to 1.25 kHz, it is possible that the NH listeners with the highest $\mathrm{IPD_{fr}}$ thresholds had some limited access to TFS information in the high-frequency band. This could explain why the group differences in BILDs were greater in the $\mathrm{TT_{bb}}$ than in the $\mathrm{TT_{lp}}$ condition. Edmonds and Culling (2005) utilized a similar paradigm in the presence of brown noise or a single interfering talker, separating the low- and high-frequency bands at either 750 Hz or 1.5 kHz. Their results showed that, for young NH listeners, changing the splitting frequency did not affect BILDs elicited by the low-frequency band or by both bands. Since they tested young NH listeners, it is likely that the listeners' access to TFS cues was drastically reduced when the cut-off frequency was lowered from 1.5 kHz to 750 Hz; Yet, the BILDs in these two lateralized conditions remained similar. Therefore, it is unlikely that the differences in BILDs between the NH and the HI groups were driven by the NH listeners' ability to utilize TFS ITDs above 1.25 kHz.

There are several possibilities why the HI listeners, compared to the NH group, showed greater deficits in BILDs in the TT_{bb} than in the TT_{lp} condition. First, aging has been associated with a general reduction in temporal coding abilities, degrading TFS and ENV processing simultaneously (He *et al.*, 2008; King *et al.*, 2014). In terms of ENV processing, aging has also been shown to affect performance both in monaural tasks, like gap detection or amplitude modulation detection (Strouse *et al.*, 1998; He *et al.*, 2008), and in binaural tasks like interaural phase discrimination (King *et al.*, 2014). Therefore, it is possible that, besides their impoverished binaural TFS coding ability, the older HI listeners were less sensitive to ENV ITDs than the young NH listeners, rendering the relatively small contribution of ITDs at high-frequencies ineffective. In contrast, the reduced binaural TFS coding abilities might still have allowed for a reasonable amount of binaural information to facilitate BILDs both in the TT_{bb} and TT_{lp} conditions. As sensitivity to binaural temporal cues at higher frequencies was not measured in the current study, it is unclear whether the older HI listeners indeed had a reduced sensitivity to ENV ITDs. Second, the reduced sensation level at which the HI listeners received the stimuli could also have affected the contribution of ENV ITDs in the BILDs. Even though elevated hearing thresholds are not necessarily related to greater-than-normal ENV ITD detection thresholds when stimuli are presented at a fixed sensation level (King *et al.*, 2014), thresholds tend to increase with decreasing SL even for NH listeners (Lacher-Fougère and Demany, 2005). In the current study, stimulus audibility was controlled by compensating for elevated hearing thresholds. Nevertheless, the HI listeners generally received the speech stimuli at lower sensation levels than the NH listeners, especially at higher frequencies where the audibility criterion was gradually reduced. Thus, it is possible that, for the HI listeners, stimulus audibility was not sufficient to contribute to BILDs. Finally, a combination of both reduced temporal processing abilities and reduced stimulus audibility is also possible. In any case, the data demonstrate that, in contrast to their NH peers, the HI listeners could not utilize ITD cues above 1.25 kHz to facilitate BILDs.

CONCLUSIONS

BILDs were found to be similar for a group of young NH and older HI listeners when elicited by ITDs below 1.25 kHz, despite the fact that the the latter group showed a clear reduction in binaural TFS coding abilities. BILDs were slightly lower for the HI group when triggered by ITDs over the full frequency range of the stimuli. When ITDs were imposed above 1.25 kHz only, no BILDs were found in any of the groups. Overall, the results suggest that, while the young NH listeners might have utilized both TFS ITDs at low frequencies and ENV ITDs at high frequencies to facilitate BU, older HI listeners relied exclusively on ITDs at the low frequencies. It still remains possible that BILDs were affected by the sensitivity to ENV ITDs in the low frequency region.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7-PEOPLE-2011-290000.

REFERENCES

Bronkhorst, A.W., 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. doi: 10.1121/1.395906

Edmonds, B.A., and Culling, J.F. (**2005**). "The spatial unmasking of speech: evidence for within-channel processing of interaural time delay," J. Acoust. Soc. Am., **117**, 3069-3078. doi: 10.1121/1.1880752

He, N.-j., Mills, J.H., Ahlstrom, J.B., and Dubno, J.R. (**2008**). "Age-related differences in the temporal modulation transfer function with pure-tone carriers," J. Acoust. Soc. Am., **124**, 3841-3849. doi: 10.1121/1.2998779

King, A., Hopkins, K., and Plack, C.J. (**2014**). "The effects of age and hearing loss on interaural phase difference discrimination," J. Acoust. Soc. Am., **135**, 342-351. doi: 10.1121/1.4838995

Lacher-Fougère, S., and Demany, L. (**2005**). "Consequences of cochlear damage for the detection of interaural phase differences," J. Acoust. Soc. Am., **118**, 2519-2526. doi: 10.1121/1.2032747

Levitt, H., and Rabiner, L.R. (**1967**). "Predicting binaural gain in intelligibility and release from masking for speech," J. Acoust. Soc. Am., **42**, 820-829. doi: 10.1121/1.1910654

Lőcsei, G., Pedersen, J.H., Laugesen, S., Santurette, S., Dau, T., and MacDonald, E.N. (**2016**). "Temporal fine-structure coding and lateralized speech perception in normal-hearing and hearing-impaired listeners," Trends Hear., **20**, 1-15. doi: 10.1177/2331216516660962

Neher, T., Lunner, T., Hopkins, K., and Moore, B.C.J. (**2012**). "Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners (L)," J. Acoust. Soc. Am., **131**, 2561-2564. doi: 10.1121/1.3689850

Nielsen, J., Dau, T., and Neher, T. (**2014**). "A Danish open-set speech corpus for competing-speech studies," J. Acoust. Soc. Am., **135**, 407-420. doi: 10.1121/1.4835935

Santurette, S., and Dau, T. (**2012**). "Relating binaural pitch perception to the individual listener's auditory profile," J. Acoust. Soc. Am., **131**, 2968-2986. doi: 10.1121/1.3689554

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. doi: 10.1121/1.423748

Lateralized speech perception with small interaural time differences in normal-hearing and hearing-impaired listeners

Gusztáv Lőcsei[1], Sébastien Santurette[1,2], Torsten Dau[1],
and Ewen N. MacDonald[1,*]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Department of Otorhinolaryngology, Head and Neck Surgery & Audiology, Rigshospitalet, Copenhagen, Denmark*

Spatial release from masking (SRM) elicited by interaural timing differences (ITDs) only can be almost normal for listeners with symmetrical hearing loss. This study investigated whether elderly hearing-impaired (HI) listeners still achieve similar SRMs as young normal-hearing (NH) listeners, when SRMs are elicited by small ITDs. Speech reception thresholds (SRTs) and SRM due to ITDs were measured over headphones for 10 young NH and 10 older HI listeners, who had normal or close-to-normal hearing below 1.5 kHz. Diotic target sentences were presented in diotic or dichotic speech-shaped noise or two-talker babble maskers. In the dichotic conditions, maskers were lateralized by delaying the masker waveforms in the left headphone channel. Multiple magnitudes of masker ITDs were tested in both noise conditions. Although deficits were observed in speech perception abilities in speech-shaped noise and two-talker babble in terms of SRTs, HI listeners could utilize ITDs to a similar degree as NH listeners to facilitate the binaural unmasking of speech. A slight difference was observed between the group means when target and maskers were separated from each other by large ITDs, but not when separated by small ITDs. Thus, HI listeners do not appear to require larger ITDs than NH listeners do in order to receive a benefit from binaural unmasking.

INTRODUCTION

If a target and maskers are separated in space, the intelligibility of the target typically improves, a phenomenon termed spatial release from masking (SRM). While SRM is mainly facilitated by better-ear listening, binaural unmasking (BU) can also play a role. Several studies have found normal or close-to-normal binaural intelligibility level difference (BILDs) in hearing-impaired (HI) listeners with symmetrical hearing loss (Bronkhorst and Plomp, 1989; Strelcyk and Dau, 2009; Lőcsei *et al.*, 2016). These results are surprising given that HI listeners usually exhibit degraded temporal fine

[*] Corresponding author: emcd@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

structure (TFS) processing. However, most studies that investigate BILDs in normal-hearing (NH) and HI listeners use relatively large interaural time differences (ITDs).

In the present study, binaural intelligibility level differences (BILD) were measured for speech stimuli embedded in noise and separated by either large or small ITDs for a group of young NH listeners and older HI listeners in a series of headphone experiments. The hypothesis was that deficits in BU abilities in HI listeners, as measured by BILDs, should be more prominent when triggered by small ITDs than by large ITDs. In addition to BILDs, TFS interaural phase difference (IPD) thresholds were measured in pure-tone carriers over a range of frequencies. BILDs in the large and small ITD conditions were compared between the listener groups and the resulting IPD threshold profiles were contrasted with the size of BILDs in both cases.

METHODS

Participants

Ten young NH (20-27 years, mean: 23, standard deviation (SD): 2.31) and ten older HI (50-76 years, mean: 66.9, SD: 7.48) listeners participated in the study (see Table 1).

			Audiometric thresholds averaged between the ears [dB HL]										
ID	Sex	Age	125	250	500	750	1k	1.5k	2k	3k	4k	6k	8k
a	m	60	7.5	0	5	5	0	0	7.5	25	32.5	27.5	35^{+}
b	m	50	2.5	0	5	5	0	0	10	25	37.5	35	45
c	m	67	22.5	15	7.5	2.5	2.5	0	2.5	22.5	30	32.5^{*}	65
d	f	65	7.5	2.5	5	10	10	7.5	7.5	12.5	22.5	32.5^{*}	45
e	f	72	10	5	7.5	5	7.5	15	22.5	40	32.5	52.5	67.5
f	f	66	15	15	12.5	15	12.5	20	37.5*	50	45	50	65
g	f	72	22.5	12.5	15	15	17.5	17.5^{*}	20^{+}	25	27.5	35	52.5
h	f	69	5	5	12.5	17.5	22.5	32.5	40	47.5	52.5	55	60
i	m	76	15	12.5	20	20	15	25	25	45	57.5	52.5	67.5
j	f	72	15	25	22.5	20	20	27.5	32.5	32.5	40	45	65

Table 1: Gender, age, and audiometric thresholds (air-conduction, averaged across both ears) of the HI listeners. In some cases, differences in audiometric thresholds between left and right ears were as large as 15 dB (*) or 20 dB ($^{+}$). In all other cases, these differences were less than or equal to 10 dB.

Binaural fine structure processing

In the measurements assessing sensitivity to binaural TFS information, the task of the listeners was to detect IPDs of pulsating pure-tones at different frequencies. Thresholds were estimated using a 3-interval 3-alternative forced-choice paradigm. Each interval contained a sequence of four 200-ms pure tones presented at the same frequency, separated by 100-ms silent gaps. The gaps between presentation intervals were 400 ms long. In the reference intervals, all of the tones were presented diotically. In the target interval, the first and third tones were presented with zero IPD, and the second and fourth tones with a starting phase of $-\frac{\Delta\varphi}{2}$ and $\frac{\Delta\varphi}{2}$ in the left and right channels of the headphones, respectively, yielding a total IPD of $\Delta\varphi$.

For each listener, the frequency range at which an IPD of 180° could be detected ($\mathrm{IPD_{fr}}$) was measured first. Thereafter, IPD thresholds at fixed frequencies ranging from 250 Hz up to $\mathrm{IPD_{fr}}$ were measured in 250-Hz steps ($\mathrm{IPD_{lf}}$ experiments). Presentation levels were set to 30 dB sensation level (SL).

Speech perception in noise

Speech intelligibility was evaluated using the DAT corpus (Nielsen *et al.*, 2014), both in speech-shaped noise (SSN) and in an interfering two-talker background (TT). In the SSN condition, the "Dagmar" sentences were used as target material, and the long-term average spectrum of the noise was matched to that of the "Dagmar" sentences. To avoid repeating any lists within the experiment, the "Asta" sentences were used in the TT condition. In these cases, sentences spoken by the two other talkers were applied as maskers. No spectral matching was applied between target and maskers in the TT conditions. The SSN tokens were semi-randomly chosen from a pool of fifty 5-second noise samples, which were then truncated to match with the length of the target sentence. The TT maskers started at the same time as the target but could end earlier or later than the target. The target sentences were always presented diotically while the maskers were delivered in one of the following lateralization settings: (1) diotic presentation, colocated with the target ($\mathrm{SSN_{co}}$ and $\mathrm{TT_{co}}$), (2) lateralized to the side through large ITDs of 0.68 ms ($\mathrm{SSN_{lrg}}$ and $\mathrm{TT_{lrg}}$), or (3) lateralized to the side through small ITDs of 0.27 ms ($\mathrm{SSN_{sm}}$). SRTs were measured adaptively using 20 sentence lists. In the $\mathrm{TT_{sm}}$ condition, instead of measuring SRTs at a fixed ITD, the 50% sentence-correct point was tracked as a function of ITD at a fixed SNR. The SNR for this condition was set to 3 dB lower than each individuals' SRT in the $\mathrm{TT_{co}}$ condition. Thus, $\mathrm{TT_{sm}}$ tracks the ITD needed to achieve a BILD of 3 dB.

All stimuli were delivered via headphones. Audibility of the stimuli was restored by applying individualized linear gains based on the individual listeners' audiogram and on the long-term average spectrum of the "Dagmar" sentences. The audibility criterion was set to be 15 dB above the individual hearing thresholds for one-third octave bands between 110 Hz and 3 kHz, which was reduced to 12, 8, and 0 dB at 4, 6, and 8 kHz. Then, the target stimulus was scaled to a nominal level of 65 dB SPL when measured at the eardrums of a HATS and mixed with the scaled maskers. The individualized gains were applied to this mixture amplifying both target and maskers. These filters also compensated for the headphone frequency response. Presentation levels were limited to 94 dBA and if the estimated overall presentation level of a stimulus exceeded this, it was downscaled in 2-dB steps.

RESULTS

The IPD thresholds measured at fixed frequencies are shown in Fig. 1 for both the NH (dots) and HI (letters) listeners. The solid horizontal black lines denote the group means and the corresponding boxes represent ±1 SD. Significant differences were confirmed between the log-transformed group means for IPD_{250} [$t(18) = -2.79$, $p = 0.012$]. Note, however, that in the $\mathrm{IPD}_{\mathrm{lf}}$ tests at frequencies at or above 750 Hz, thresholds could not be measured for all of the HI listeners, biasing the group means towards lower values than the true group average. This is also clearly reflected in Fig. 2, which shows the results of the $\mathrm{IPD}_{\mathrm{fr}}$ experiment. Differences in group means were significant for the $\mathrm{IPD}_{\mathrm{fr}}$ [$t(18) = 5.67$, $p < 0.001$] thresholds, and also for the $\mathrm{ITD}_{\mathrm{min}}$ thresholds [$t(10.16) = -3.234$, $p < 0.009$].

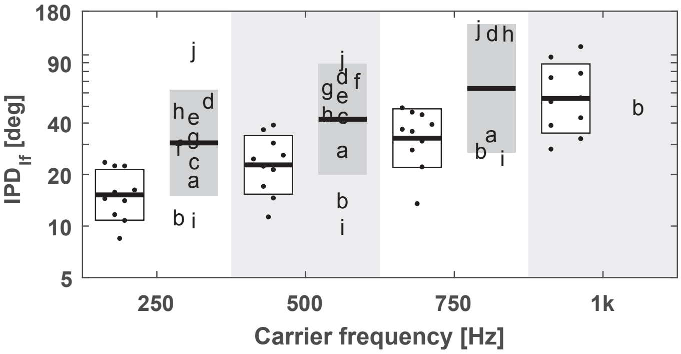


Fig. 1: $\mathrm{IPD}_{\mathrm{lf}}$ thresholds for the NH (dots) and HI (letters) listener groups. Black horizontal lines mark group means and the boxes denote ±1 SD of the corresponding groups. The shading of the background is according to the conditions with different carrier frequencies.

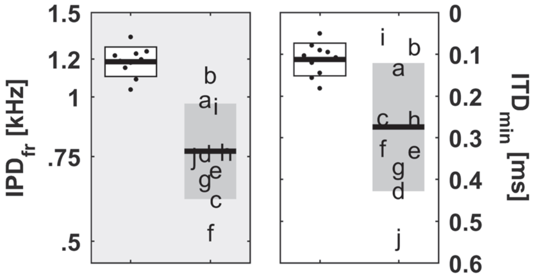


Fig. 2: Maximum frequency for detection of 180° IPD and minimum ITD thresholds ($\mathrm{ITD}_{\mathrm{min}}$) of the NH (dots) and HI (letters) listeners. Black horizontal lines mark group means and the boxes denote ±1 SD of the corresponding groups. Note that the y-axis in the right panel is reversed, so that data points located further towards the top of each panel represent better performance.

Figure 3 shows the SRTs for the NH and the HI listeners obtained in the fixed-ITD conditions. A mixed-design ANOVA was conducted on the SRT data for the $\mathrm{SSN}_{\mathrm{co}}$, $\mathrm{SSN}_{\mathrm{lrg}}$, $\mathrm{TT}_{\mathrm{co}}$ and $\mathrm{TT}_{\mathrm{lrg}}$ conditions. The model contained the SRTs as the dependent variable, and used noise type (SSN or TT) and lateralization (co or lrg) as within-subject factors and listener group (NH or HI) as between-subject factors. All main effects and two-way interactions were significant.

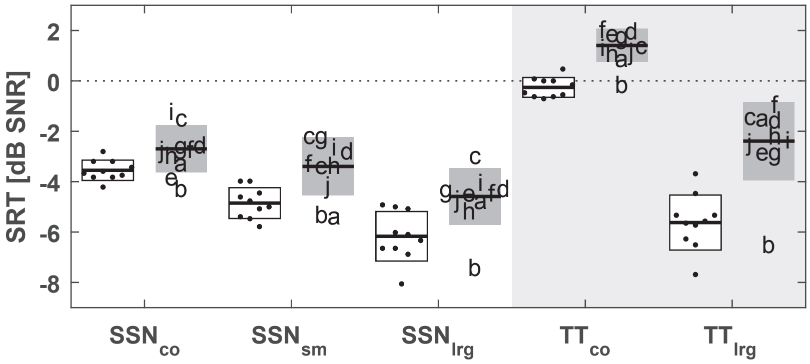


Fig. 3: SRTs in SSN and two-talker babble (TT) for NH (dots) and HI (letters) listeners. Solid black horizontal lines indicate group means and the boxes denote ±1 SD. The background shadings indicate condition groups using the same type of background noise. In each condition the target was presented diotically. The different test conditions are denoted on the x-axis. Subscripts indicate the ITD configuration of the masker: **co**: diotic presentation, colocated with the masker; **sm**: masker lateralized with a small ITD (0.27 ms); **lrg**: masker lateralized with a large ITD (0.68 ms).

The measures characterizing BU of speech are plotted in Fig. 4. In the left panel, the BILDs due to masker lateralization are plotted, which were calculated as the difference in SRTs between the co and the sm or lrg conditions. The right panel indicates the results obtained in the $\mathrm{TT_{sm}}$ condition, which indicates the ITD needed to achieve a BILD of 3 dB. In general, the NH listeners showed a slightly better performance than the HI listeners in all conditions. For BILDs at fixed ITDs, a statistically significant interaction between lateralization and listener group $[F(1,18) = 8.81, p = 0.008]$ was observed in the ANOVA model. Most listeners benefitted from masker lateralization in all of the tested conditions. While BILDs were small in the $\mathrm{SSN_{sm}}$ condition, they increased as the ITD magnitudes of the maskers increased from 0.27 to 0.68 ms. The benefit was greatest in the $\mathrm{TT_{lrg}}$ condition, where it reached 5.4 dB and 3.8 dB for the NH and HI listeners, respectively. It appears that NH listeners exhibited greater BILDs in the conditions with fixed ITDs and a 3-dB BILD at smaller ITDs than the HI listeners. However, independent t-tests on the BILD data indicated that the group differences were only statistically significant in the $\mathrm{TT_{lrg}}$ condition $[t(18) = 3.03, p = 0.007]$.

Pearson's correlations were calculated between each of the four measures of BU and the $\mathrm{ITD_{min}}$ or $\mathrm{IPD_{fr}}$ results within the group of HI listeners. None of the correlations were significant.

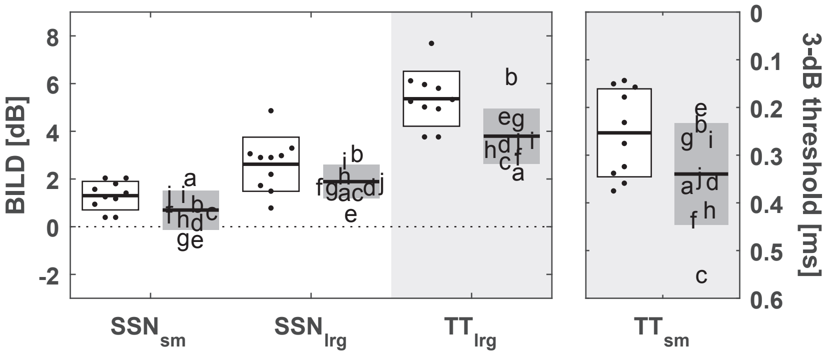


Fig. 4: BILDs at fixed ITDs in SSN and two-talker babble (TT) and the ITD threshold needed to yield a fixed 3-dB BILD in the TT noise (i.e., a 3-dB decrease in SRTs as compared to the $\mathrm{TT_{co}}$ condition). Solid horizontal black lines and the boxes around denote group means and ±1 SD for the NH (dots) and HI individuals (letters). Background shadings mark condition groups with the same noise type. Note that the first 3 conditions to the left are expressed in dB, while the last condition in ms. Condition notations are the same as in Fig. 3.

DISCUSSION

In the present study, HI listeners exhibited poorer thresholds compared to NH listeners for the binaural measures of TFS processing $\mathrm{IPD_{fr}}$, $\mathrm{IPD_{lf}}$ and $\mathrm{ITD_{min}}$. These results are consistent with previous studies (Ross *et al.*, 2007; Hopkins and Moore, 2011; Neher *et al.*, 2011; King *et al.*, 2014).

In the speech experiments, both groups exhibited lower average SRTs in SSN than in TT noise. Listeners in both groups showed a clear benefit when the maskers were lateralized to the side, indicating the presence of an active BU mechanism. The amount of BILDs differed slightly between the two groups, and this difference was only statistically significant in the $\mathrm{TT_{lrg}}$ condition. Therefore, the results obtained in the TT conditions do not support the hypothesis that the HI listeners' processing deficits in BU are more pronounced when triggered by small rather than by large ITDs. Rather, the deficits in the BU of speech manifested themselves mainly by reducing the overall benefit HI listeners could achieve when target and maskers were separated by large ITDs. Nonetheless, the SRTs obtained in the $\mathrm{SSN_{lrg}}$ and $\mathrm{TT_{lrg}}$ conditions suggest the possibility that BILDs in the $\mathrm{TT_{lrg}}$ condition were, at least partly, affected by monaural deficits in temporal processing. The SRTs in the TT conditions were different from those in the SSN conditions as the two maskers differ in the amount of modulation and informational masking. While the NH listeners yielded similar SRTs in the $\mathrm{TT_{lrg}}$ and $\mathrm{SSN_{lrg}}$ conditions, the HI listeners had about 2-dB higher SRTs in the $\mathrm{TT_{lrg}}$ condition than in the $\mathrm{SSN_{lrg}}$ condition. However, informational masking is substantially reduced when target and maskers are spatially separated (Arbogast *et* *al.*, 2002). Therefore, the performance in the $\mathrm{TT_{lrg}}$ condition can be assumed to be limited by factors other than informational masking (c.f. Best *et al.*, 2002). Several studies have shown that HI listeners are more susceptible to modulation masking than NH listeners, which manifests itself in less-than-normal fluctuating-masker benefit when modulations are imposed on a stationary masker (Festen and Plomp, 1990; Strelcyk and Dau, 2009). Therefore, it is possible that, compared to the NH listeners, the HI listeners would have elevated thresholds in the $\mathrm{TT_{lrg}}$ condition due to their susceptibility to modulation masking, even if they had intact binaural processing abilities. The extent to which such monaural factors might have contributed to the reduced BILDs in the current study is nonetheless difficult to evaluate, as it is likely that both informational and modulation masking are involved in the $\mathrm{TT_{co}}$ and $\mathrm{TT_{lrg}}$ conditions.

The limitations of the experimental paradigm utilized in the $\mathrm{TT_{sm}}$ condition deserve some further attention. This condition assessed the sharpness of spatial tuning due to BU by measuring the amount of ITDs by which target and maskers had to be separated in order to give raise to a BILD of 3 dB. First, assuming that the magnitude of the BILD monotonically increases with increasing ITD, this paradigm is only plausible if one assumes that listeners can obtain a 3 dB benefit at the largest ITDs applied. While this was clearly the case for the NH listeners, who showed a BILD of at least 3.7 dB, and about 5.4 dB on average, three listeners from the HI group (listener *a*, *c*, and *f*) had a BILD lower than 3 dB in the $\mathrm{TT_{lrg}}$ condition. Theoretically, for these listeners, the thresholds in the $\mathrm{TT_{sm}}$ conditions should be greater than 0.68 ms. Thus, even though these listeners had the greatest thresholds in the $\mathrm{TT_{sm}}$ condition, their results should be treated with caution. Furthermore, the average BILDs of the HI listeners in the $\mathrm{TT_{lrg}}$ condition was about 4 dB, while the thresholds in the $\mathrm{TT_{sm}}$ condition were assessed for a fixed BILD of 3 dB. This means that the differences in performance criteria between these two conditions were relatively small. A possible modification of the existing paradigm to alleviate these issues would be to use identical talkers for the target and the maskers, which would likely increase the BILDs for all listeners.

CONCLUSIONS

HI listeners showed a reduction in binaural TFS coding abilities compared to NH listeners, as reflected in a reduction of the $\mathrm{IPD_{fr}}$ and an increase of the $\mathrm{ITD_{min}}$ thresholds. Although deficits were observed in speech perception abilities in SSN and two-talker babble in terms of SRTs, HI listeners could utilize ITDs to a similar degree as NH listeners to facilitate the binaural unmasking of speech. A slight difference was observed between the group means when target and maskers were separated from each other by large ITDs, but not when separated by small ITDs. Therefore, HI listeners did not experience greater difficulties in terms of reduced BILDs when spatial differences between target and maskers were induced by small ITDs.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7-PEOPLE-2011-290000.

REFERENCES

Arbogast, T.L., Mason, C.R., and Kidd, G. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am., **112**, 2086-2098. doi: 10.1121/1.1510141

Best, V., Thompson, E.R., Mason, C.R., and Kidd, G. (**2013**). "An energetic limit on spatial release from masking," J. Assoc. Res. Otolaryngol., **14**, 603-610. doi: 10.1007/s10162-013-0392-1

Bronkhorst, A.W., and Plomp, R. (**1989**). "Binaural speech intelligibility in noise for hearing-impaired listeners," J. Acoust. Soc. Am., **86**, 1374-1383. doi: 10.1121/1.398697

Festen, J.M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am., **88**, 1725-1736. doi: 10.1121/1.400247

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.

King, A., Hopkins, K., and Plack, C.J. (**2014**). "The effects of age and hearing loss on interaural phase difference discrimination," J. Acoust. Soc. Am., **135**, 342-351. doi: 10.1121/1.3585848

Lőcsei, G., Pedersen, J.H., Laugesen, S., Santurette, S., Dau, T., and MacDonald, E.N. (**2016**). "Temporal fine-structure coding and lateralized speech perception in normal-hearing and hearing-impaired listeners," Trends Hear., **20**, 1-15. doi: 10.1177/2331216516660962

Neher, T., Jensen, N.S., and Kragelund, L. (**2011**). "Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities?" J. Acoust. Soc. Am., **130**, 1542-1558.

Nielsen, J., Dau, T., and Neher, T. (**2014**). "A Danish open-set speech corpus for competing-speech studies," J. Acoust. Soc. Am., **135**, 407-420. doi: 10.1121/1.4835935

Ross, B., Tremblay, K.L., and Picton, T.W. (**2007**). "Physiological detection of interaural phase differences," J. Acoust. Soc. Am., **121**, 1017-1027. doi: 10.1121/1.2404915

Strelcyk, O., and Dau, T. (**2009**). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," J. Acoust. Soc. Am., **125**, 3328-3345. doi: 10.1121/1.3097469

Extending a computational model of auditory processing towards speech intelligibility prediction

HELIA RELAÑO-IBORRA*, JOHANNES ZAAR, AND TORSTEN DAU

Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

A speech intelligibility model is presented based on the computational auditory signal processing and perception model (CASP; Jepsen *et al.*, 2008). CASP has previously been shown to successfully predict psychoacoustic data obtained in normal-hearing (NH) listeners in a wide range of listening conditions. Moreover, CASP can be parametrized to account for data from individual hearing-impaired listeners (Jepsen and Dau, 2011). In this study, the CASP model was investigated as a predictor of speech intelligibility measured in NH listeners in conditions of additive noise, phase jitter, spectral subtraction and ideal binary mask processing.

INTRODUCTION

Computational models of the auditory system are a powerful tool to investigate the ability of humans to hear, process and encode acoustic stimuli. These models provide information about the mechanisms involved in the perception of acoustic signals. Moreover, they can provide insights about the effects of hearing loss in the impaired system. Recently, a model termed correlation-based speech-based Envelope Power Spectrum Model (sEPSM$^{\mathrm{corr}}$; Relaño-Iborra *et al.*, 2016) was presented, which employs the auditory processing of the multi-resolution speech-based Envelope Power Spectrum Model (mr-sEPSM; Jørgensen *et al.*,2013) and combines it with the correlation back end of the Short-Time Objective Intelligibility measure (STOI; Taal *et al.*, 2011). The sEPSM$^{\mathrm{corr}}$ was shown to accurately predict NH data for a broad range of listening conditions, e.g., additive noise, phase jitter and ideal binary mask processing. The main idea behind the sEPSM$^{\mathrm{corr}}$ is that the correlation between the envelope representations of the clean speech and the degraded speech is a strong predictor of intelligibility. However, recent studies have shown that the mr-sEPSM preprocessing is limited with respect to predicting intelligibility data from hearing-impaired (HI) listeners (Scheidiger *et al.*, 2017). Specifically, while sensitivity loss and loss of frequency selectivity can functionally be incorporated, the crucial level-dependent effects and nonlinearities that are typically strongly affected by hearing loss cannot be successfully simulated using this framework.

The finding from the Relaño-Iborra *et al.* (2016) study that the correlation between the clean and degraded speech in the modulation power domain can be a reliable predictor of intelligibility was further investigated here using a more realistic auditory

*Corresponding author: heliaib@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.

ISBN: 978-87-990013-6-1.

preprocessing front end. In particular, the front end of the computational auditory signal processing and perception model (CASP; Jepsen *et al.*, 2008) was considered. CASP has been shown to successfully predict psychoacoustic data of normal-hearing (NH) listeners obtained in conditions of, e.g., spectral masking, amplitude-modulation detection, and forward masking. Furthermore, the model can be adapted to account for data obtained in individual HI listeners in different behavioural experiments (Jepsen and Dau, 2011).

In this study, the CASP model was extended to investigate its potential use as a predictor of speech intelligibility data. In order to adapt CASP to function as a speech intelligibility prediction model, the speech-based CASP (sCASP) introduces modifications in the model's back-end processing and decision metric. The model was validated as a predictor of intelligibility of Danish sentences measured in NH listeners in conditions of additive noise, phase jitter, spectral subtraction and ideal binary mask processing.

THE sCASP MODEL

General structure

The proposed sCASP implementation maintains most of the structure of the original CASP model, albeit with some minor changes required due to the use of speech stimuli. The model receives the unprocessed clean speech and the noisy or degraded speech mixture as inputs (i.e., it has a-priori knowledge of the speech signal). Both inputs are processed through outer- and middle-ear filtering, a nonlinear auditory filterbank, envelope extraction, expansion, adaptation loops, a modulation filterbank, and a second-order envelope extraction for modulation channels above 10 Hz. The internal representations produced at the output of these stages are analyzed using a correlation-based back end. Figure 1 shows a diagram of the main model stages.

Modelling of the auditory preprocessing

The first stage is an outer- and middle-ear filtering stage implemented as two finite impulse response filters as in Lopez-Poveda and Meddis (2001); the output of this stage can be related to the peak velocity of vibration at the stapes as a function of frequency. Afterwards, the inputs pass through the dual-resonance nonlinear filterbank (DRNL; Lopez-Poveda and Meddis, 2001). Within this auditory filterbank, the signals are processed in two independent parallel paths, where the linear path applies a linear gain, a cascade of gammatone filters and a lowpass filter, and the nonlinear path applies a cascade of gammatone filters and a broken-stick nonlinearity followed by another cascade of gammatone filters and a lowpass filter. The summed signal of the two paths includes the effects of the nonlinear basilar-membrane processing, which accounts for level-dependent compression and auditory-filter tuning. This is followed by an envelope extraction stage, realized by half-wave rectification and second order low-pass filtering (f_c = 1 kHz). The envelopes are then expanded quadratically into an intensity-like representation. Afterwards, effects of adaptation are modelled using

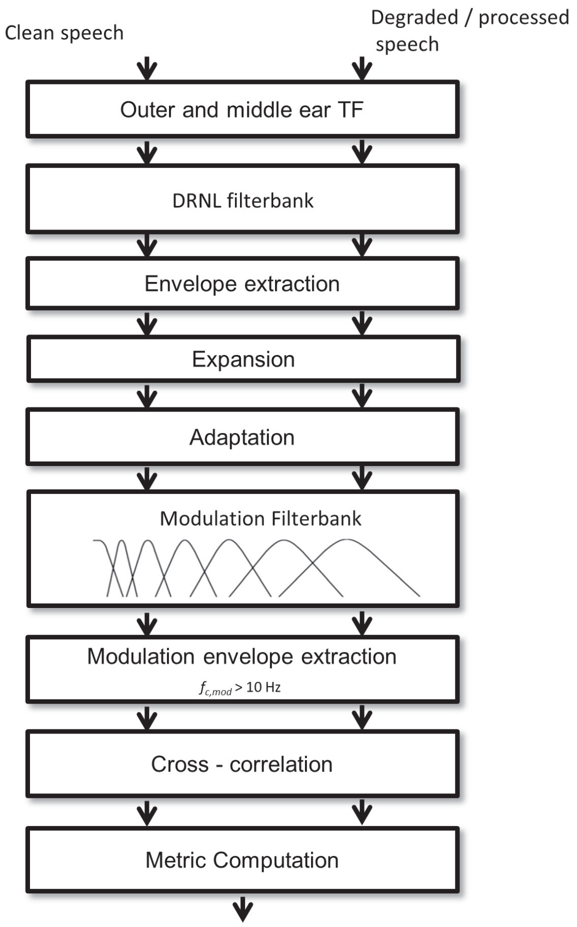


Fig. 1: Modelling stages of the speech-based CASP model.

a chain of five feedback loops (Dau *et al.*, 1996). Finally, modulation processing is included in the model using a bank of frequency-shifted first-order low-pass filters (i.e, they act as band-pass filters) in parallel with a second-order low-pass filter. For modulation filters centered below 10 Hz, the real part is considered, and for modulation filters above 10 Hz, the absolute ouput is considered, in order to account for decrease of a modulation phase sensitivity (Dau, 1996). For more details, reference is made to Jepsen *et al.* (2008) and Jepsen and Dau (2011).

Back end and decision metric

The resulting three-dimensional internal representations (as a function of time, audio frequency and modulation frequency) are analysed by cross-correlating the time signals obtained in each combination of modulation and auditory channel. The cross-correlation is performed in short time windows in a similar way as in the sEPSM$^{\mathrm{corr}}$ model (Relaño-Iborra *et al.*, 2016), with the window length defined by the inverse of the modulation frequency. In order to obtain a unique model output for each pair of input signals, the correlation values are averaged across time, audio-frequency and modulation channel. This differs from the calculation in the sEPSM$^{\text{corr}}$, since it does not require the summation of correlation values across time windows (Relaño-Iborra *et al.*, 2016, Eq. 3) but instead averages the correlation values across time. The sCASP back end also differs from the original CASP model, where (i) decisions are based on the correlation of the normalized difference between the internal representation of the masker plus a suprathreshold signal (considered as template) and that of the masker alone (Dau *et al.*, 1996) and (ii) no short-term processing is applied.

METHODS

Test conditions

The model was validated in four different listening conditions: speech in the presence of additive interferers, noisy speech under reverberation, phase jitter and ideal binary mask (IBM) processing. For the latter, the Dantale II corpus (Wagener *et al.*, 2003) was used, and for all other conditions the CLUE corpus was used (Nielsen and Dau, 2009).

Three additive noises were considered for the first experiment: (i) speech-shaped noise ('SSN'), (ii) an 8-Hz sinusoidally amplitude-modulated SSN with a modulation depth of 1 ('SAM'), and (iii) the speechlike but non-semantic, international speech test signal ('ISTS'; Holube *et al.*, 2010). Signal-to-noise ratios (SNRs) ranging from -27 to 3 dB with a step size of 3 dB were used. Model predictions were compared to human data obtained under the same conditions by Jørgensen *et al.* (2013).

Phase jitter was applied to sentences mixed with SSN at a fixed SNR of 5 dB as follows:

$$r(t) = \mathrm{Re}\{s(t)e^{j\Theta(t)}\} = s(t)\cos(\Theta(t))) \qquad \text{(Eq. 1)}$$

where $s(t)$ represents the non-processed mixture, $r(t)$ the resulting jittered stimulus and $\Theta(t)$ denotes a random process with a uniform probability distribution between $[0, 2\alpha\pi]$ with α ranging between 0 and 1 (Elhilali *et al.*, 2003). The simulations were compared to the data obtained in Chabot-Leclerc *et al.* (2014).

For the spectral subtraction experiment, the sentences were mixed with SSN at SNRs from -9 to 9 dB, in 3 dB steps. Spectral subtraction was applied to each mixture following:

$$\widehat{S(f)} = \sqrt{P_Y(f) - \kappa\widehat{P_N}(f)} \qquad \text{(Eq. 2)}$$

where $\widehat{S}$ is the enhanced magnitude spectrum of the noisy mixture after spectral subtraction. $\widehat{P_N}$ and P_Y are the averaged power spectra of the noise alone and the original speech-plus-noise mixture, respectively. Values for the over-subtraction factor, κ, of 0, 0.5, 1, 2, 4, and 8 were considered. The simulations were compared with data collected by Jørgensen and Dau (2011).

Finally, IBMs where applied to two SNR mixtures(corresponding to 20% and 50% understanding) of Dantale II sentences with four different interferers: SSN, car-cabin noise ('Car'), noise produced by bottles on a conveyor belt ('Bottle'), and two people speaking in a cafeteria ('Café'). Different IBMs were built as follows:

$$\mathrm{IBM}(t,f) = \begin{cases} 1 & \text{if } \mathrm{SNR}(t,f) > \mathrm{LC} \\ 0 & \text{otherwise} \end{cases} \tag{Eq. 3}$$

where LC corresponds to the local criterion, which defines the density of the mask. Eight different values of LC were considered, and discussed here in terms of the relative criterion defined as $RC = LC - RC$ as in the reference study of Kjems *et al.* (2009).

Model fitting

The correlation-based output of the proposed model is monotonically related to the SNR of the input mixture. In order to convert the model output to intelligibility scores, a fitting condition is required. The transformation is performed by applying a logistic function to the model outcome χ:

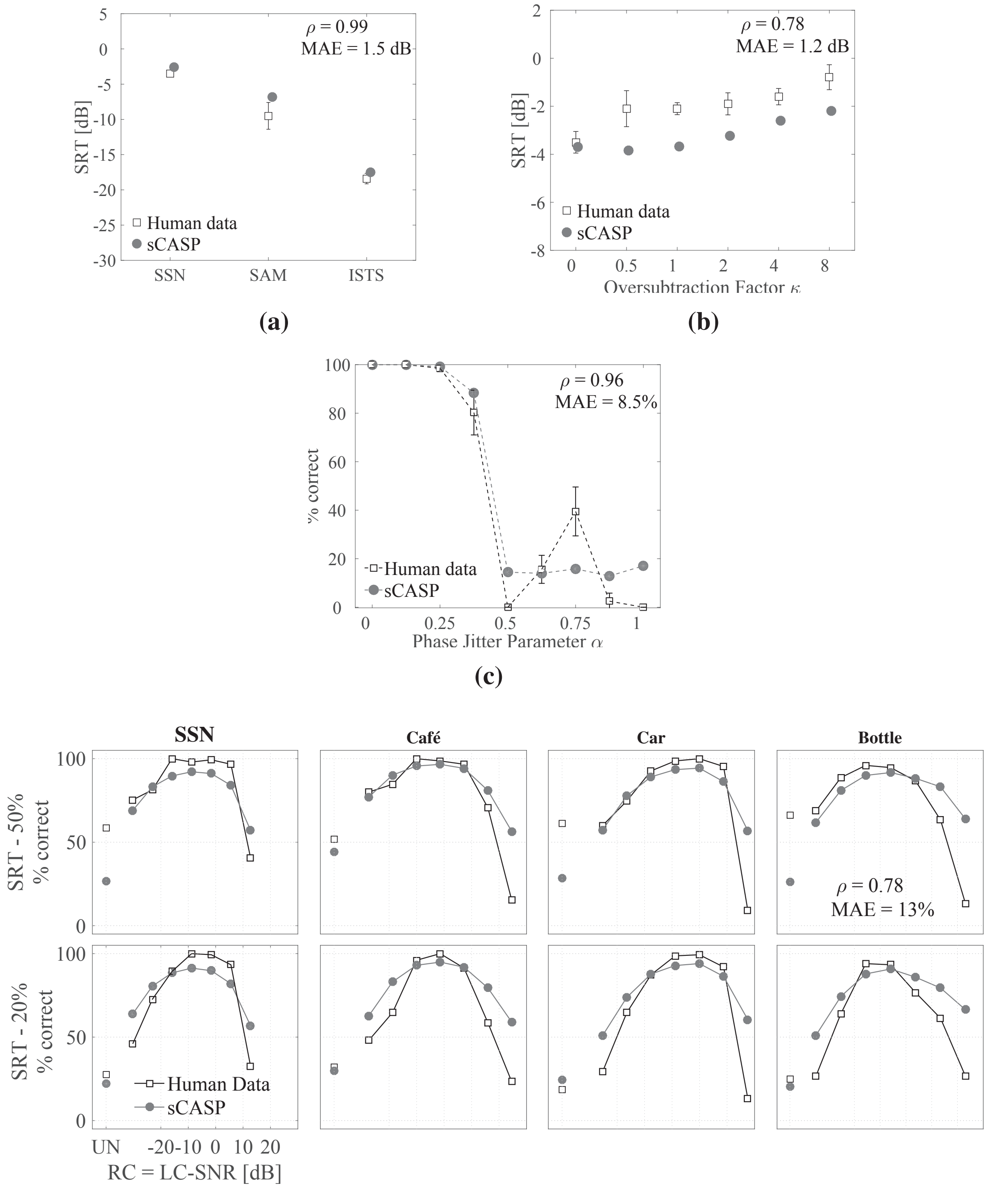
$$\Phi(\chi) = \frac{100}{1 + e^{a\chi + b}} \tag{Eq. 4}$$

where a and b are free parameters adjusted to map the model output to intelligibility scores in the fitting condition. The model was calibrated twice in the present study, once per speech material, using SSN at different SNRs. Thus, the mapping accounts for the intelligibility of the speech material but implies no a-priori knowledge about the degradations tested, other than the degradation induced by the SSN.

RESULTS AND DISCUSSION

Figure 2 shows the human data, in open squares, and the corresponding model predictions, in gray circles, for the four conditions under consideration. The accuracy of the model predictions was measured in terms of their Pearson's correlation and mean average error (MAE) with the human data.

The model can account for the changes in intelligibility reported by the listeners for speech in the presence of different additive noises, as seen in panel (a), with $\rho = 0.99$ and MAE = 1.5 dB. Regarding the non-linear conditions, i.e., spectral subtraction, phase jitter and IBM (panels b, c and d), the model can account fairly well for the data with $\rho = 0.78$ and MAE = 1.2 dB, $\rho = 0.96$ and MAE = 8.5% and $\rho = 0.78$ and MAE = 13%, respectively.



(a)

(b)

(c)

(d)

Fig. 2: Panels (a) - (d) show human data (open squares) and model predictions (gray circles) for conditions of speech in the presence of additive noise (a), noisy speech with spectral substraction processing (b), noisy speech distorted with phase jitter (c) and IBM-processed noisy speech (d). Human data from Jørgensen *et al.* (2013), Jørgensen and Dau (2011), Chabot-Leclerc *et al.* (2014) and Kjems *et al.* (2009), respectively.

However, in the phase jitter condition (panel c), it can be observed that the model exhibits some flooring effects, such that it does not predict intelligibility scores below 15% and underestimates the recovery of intelligibility reported by listeners for $\alpha = 0.75$. This might be an indication of the need for an across-frequency analysis as suggested by Chabot-Leclerc *et al.* (2014). Furthermore, although not shown here, the sCASP model does not account for effects of reverberation on speech intelligibility (as also reported for the sEPSM$^{\text{corr}}$).

Overall, these results are very similar to those reported in Relaño-Iborra *et al.* (2016) (see Table 1), despite the changes both in the front-end and the back-end processing. It thus appears that the pre-processing of CASP, which includes an adaptation stage that emphasizes the higher-frequency envelope content, does not require the accumulation process used in the sEPSM$^{\text{corr}}$ back end in order to replicate its performance (i.e., a linear average of the correlation values across time windows suffices). Still, the main finding of Relaño-Iborra *et al.* (2016) holds, namely that the correlation in the modulation domain can account for speech intelligibility.

	sCASP		**sEPSM$^{\text{corr}}$**	
	ρ	MAE	ρ	MAE
Additive Interferers	0.99	1.5 dB	0.97	1.85 dB
Spectral Subtraction	0.78	1.2 dB	0.82	0.6 dB
Phase Jitter	0.96	8.5%	0.97	19.0%
Ideal Binary Mask Processing	0.78	13%	0.79	12.1 %

Table 1: Comparison of the accuracy of the predictions for the proposed sCASP model and the referenced sEPSM$^{\text{corr}}$ (Relaño-Iborra *et al.*, 2016). ρ denotes the Pearson's correlation between human data and model predictions and MAE stands for mean average error.

CONCLUSION

The sCASP model shows promising results in terms of predicting NH intelligibility in a wide range of listening conditions. Combined with the original CASP model's ability to account for individual HI psychoacoustic data, this provides a strong basis for a framework investigating consequences of hearing loss on speech intelligibility.

ACKNOWLEDGEMENTS

This research was supported by the Oticon Centre of Excellence for Hearing and Speech Sciences (CHeSS).

REFERENCES

Chabot-Leclerc, A., Jørgensen, S., and Dau, T. (**2014**). "The role of auditory spectro-temporal modulation filtering and the decision metric for speech intelligibility prediction," J. Acoust. Soc. Am., **135**, 3502-3512.

Dau, T. (**1996**). *Modeling Auditory Processing of Amplitude Modulation.* Doctoral dissertation. Retrieved from Bibliotheks- und Informationssystem der Universität Oldenburg.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the effective signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am., **99**, 3615-3622.

Elhilali, M., Chi, T., and Shamma, S.A. (**2003**). "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," Speech Commun., **41**, 331-348.

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.

Jepsen, M.L., Ewert, S.D., and Dau, T. (**2008**). "A computational model of human auditory signal processing and perception," J. Acoust. Soc. Am., **124**, 422-438.

Jepsen, M.L., and Dau T. (**2011**) "Characterizing auditory processing and perception in individual listeners with sensorineural hearing loss," J. Acoust. Soc. Am., **129**, 262-281.

Jørgensen, S., and Dau, T. (**2011**). "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing," J. Acoust. Soc. Am., **130**, 1475-1487.

Jørgensen, S., Ewert, S.D., and Dau, T. (**2013**). "A multi-resolution envelope-power based model for speech intelligibility," J. Acoust. Soc. Am., **134**, 436-446.

Kjems, U., Boldt, J.B., Pedersen, M.S., Lunner, T., and Wang, D.L. (**2009**)."Role of mask pattern in intelligibility of ideal binary-masked noisy speech," J. Acoust. Soc. Am. , **126**, 1415-1426.

Lopez-Poveda, E.A., and Meddis, R. (**2001**). "A human nonlinear cochlear filterbank," J. Acoust. Soc. Am., **110**, 3107-3118.

Nielsen, J.B., and Dau, T. (**2009**). "Development of a Danish speech intelligibility test," Int. J. Audiol., **48**, 729-741.

Relaño-Iborra, H., May, T., Zaar, J., Scheidiger, C., and Dau, T. (**2016**). "Predicting speech intelligibility based on a correlation metric in the envelope power spectrum domain," J. Acoust. Soc. Am., **140**, 2670-2679.

Scheidiger, C., Zaar, J., Swaminathan, J., and Dau, T. (**2017**). "Modeling speech intelligibility based on neural envelopes derived from auditory nerve spike trains". Association for Research in Otolaryngology Mid-Winter Meeting, Baltimore.

Taal, C.H., Hendriks, R.C., Heusdens, R., and Jensen, J. (**2011**). "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," IEEE Trans. Audio Speech Lang. Process., **19**, 2125-2136.

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.

Zilany, M.S.A., and Bruce, I.C. (**2007**). "Predictions of speech intelligibility with a model of the normal and impaired auditory-periphery," 3rd Int. IEEE/EMBS Conf. Neural Eng., **1-2**, 481.

Effects of non-stationary noise on consonant identification

JOHANNES ZAAR*, BORYS KOWALEWSKI, AND TORSTEN DAU

Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

Consonant perception has typically been measured using consonant-vowel (CV) syllables presented in a stationary noise masker at various signal-to-noise ratios (SNRs). Recently, a microscopic speech perception model was proposed (Zaar and Dau, 2017) and shown to account well for consonant perception data obtained in stationary noise. However, unlike stationary noise, real-life interfering sounds typically exhibit strong fluctuations. The present study therefore investigated the effects of highly non-stationary noise on consonant perception and assessed the predictive power of the model in such conditions. Normal-hearing listeners were presented with 15 Danish CVs in 5-Hz interrupted noise at SNRs of −20, −10, 0, and 10 dB. Five different CV onset times with respect to the noise bursts were considered, differing in the amount of induced simultaneous and forward masking. As expected, the consonant recognition scores were inversely related to the amount of simultaneous masking. However, even with minimum simultaneous masking, a substantial loss of consonant recognition was observed at low SNRs, suggesting a forward masking effect. The model, which employs adaptive processes in the front end, accounted for these experimental data to a large extent. The experimental paradigm and the model may be useful for assessing temporal effects of hearing-aid algorithms on consonant perception.

INTRODUCTION

Speech perception has often been measured using sentences as target signals, thus typically providing listeners with context and lexical information that can be exploited to compensate for the sparse acoustic information available in acoustically adverse conditions. To exclude such effects of high-level linguistic processing and, instead, focus solely on the relationship between the available acoustic cues and the speech percept, consonant perception has been measured, typically using consonant-vowel combinations (CVs, e.g., /ta, ba/) at various signal-to-noise ratios (SNRs) in stationary noise (e.g., Miller and Nicely, 1955; Phatak and Allen, 2007; Zaar and Dau, 2015). The resulting consonant recognition and confusion data are useful for investigating the characteristics and confusability of consonant cues. Furthermore, consonant perception tests have been shown to be particularly useful for assessing hearing-aid processing due to the consonants' short-term and high-frequency characteristics (e.g.,

*Corresponding author: jzaar@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.

ISBN: 978-87-990013-6-1.

Schmitt *et al.*, 2016). Zaar and Dau (2017) proposed a microscopic speech perception model to account for consonant perception data, which combines an auditory processing front end proposed by Dau *et al.* (1997) with a correlation-based, temporally dynamic template-matching back end. The model was shown to account well for the effects of stationary noise (Zaar and Dau, 2017) as well as for spectral effects of hearing-instrument signal processing (Zaar *et al.*, 2017) on consonant recognition and confusions.

In contrast to the stationary masking noise employed in the above-mentioned consonant perception studies, real-life interfering sounds are typically highly non-stationary. While stationary noise introduces only simultaneous masking, non-stationary noise may additionally lead to forward and backward masking of consonant cues. As fine temporal differences presumably play an important role in this context, perceptual effects induced by non-stationary interferers may be particularly useful for evaluating the temporal effects of hearing-aid processing. In the present study, the effect of highly non-stationary noise on consonant identification was measured in normal-hearing (NH) listeners. Special attention was paid to the temporal positioning of the considered CV speech tokens relative to the noise envelope's minima and maxima. Furthermore, the predictive power of the microscopic speech perception model by Zaar and Dau (2017) was evaluated for non-stationary interferers based on the experimental stimuli and the collected data.

EXPERIMENTAL METHOD

Stimuli

The target speech consisted of fifteen consonant-vowel (CV) tokens: /bi, di, fi, gi, hi, ji, ki, li, mi, ni, pi, si, ʃi, ti, vi/ spoken by one male and one female talker (thirty utterances in total). The speech tokens were a subset of the ones employed in a previous study (Zaar and Dau, 2015) and were selected based on maximum intelligibility in stationary noise. The noise was composed of five 100-ms long bursts with 1-ms raised-cosine ramps, separated by 100-ms silent gaps (corresponding to a 5-Hz repetition rate). White noise was chosen as a carrier in order to maximize masking of high-frequency consonants. The presentation level was 65 dB SPL, defined as the level of the noise bursts. Thirty noise waveforms (one per CV utterance) were pre-generated and stored as .wav-files. Each utterance was always presented in combination with the same noise recording. This was done in order to limit the across-repetition variability due to the random fluctuations in the Gaussian noise carrier, whilst preventing noise-learning effects that could occur if only one noise-waveform was used for all utterances (cf. Zaar and Dau 2015). The speech tokens were mixed with the fixed-level noise at four presentation levels: 45, 55, 65, and 75 dB SPL, corresponding to broadband SNRs of -20, -10, 0, and 10 dB. The onsets of the CV tokens were positioned at five different onset times relative to the noise: 400, 450, 500, 525, and 550 ms after the initial noise onset, as shown in Fig. 1. To investigate whether the speech tokens *per se* were sufficiently intelligible at the considered speech levels, two additional conditions with speech in quiet at presentation levels of 45 and 65 dB SPL (termed Q65 and Q45) were considered.

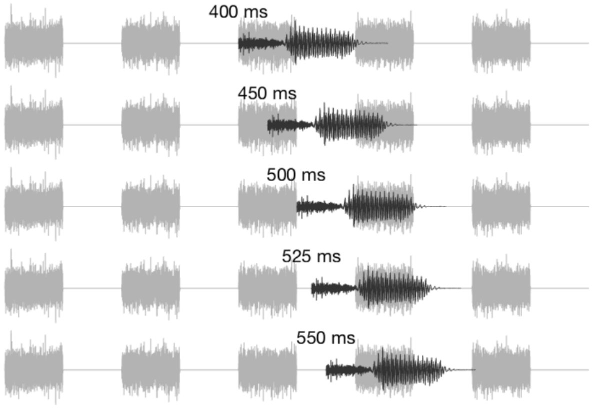


Fig. 1: Stimulus generation. Example of the CV /ki/ (black waveforms) in 5-Hz interrupted noise (gray waveforms) for the five considered CV onset times, as indicated above the respective waveforms.

Listeners and procedure

Twelve young NH native Danish listeners aged 19-26 (average age: 21.7 years) were tested. The normal-hearing status was established based on pure-tone thresholds lower than 20 dB HL in the 250 – 8000 Hz range. The listeners were seated in a sound-attenuating listening booth, monaurally presented with the stimuli via headphones, and asked to indicate the consonant they heard on a graphical user interface. The stimulus presentation could not be repeated and no feedback was provided to the listeners. Six different experimental blocks were defined based on the two quiet conditions and four SNRs (order: Q65, Q45, SNR = 10, 0, −10, −20 dB). A short training run was provided at the beginning of each block. Within each block, the order of presentation was randomized. In each condition, each stimulus was presented to the listeners five times.

MODELING

Model description

The consonant perception model of Zaar and Dau (2017) was considered to predict the perceptual data obtained in the experiment. Figure 2 shows the model, which combines the auditory model front end of Dau *et al.* (1997) with a temporally dynamic correlation-based back end. The auditory model consists of (i) a bank of 15 fourth-order gammatone filters with center frequencies logarithmically spaced between 315 Hz and 8 kHz, (ii) an envelope extraction stage (realized by half-wave rectification and lowpass filtering at 1 kHz), (iii) a chain of five adaptation loops (designed to mimic adaptive properties of the auditory periphery), and (iv) a bank of four modulation filters, implemented as a 2-Hz lowpass filter in parallel with three second-order bandpass filters with a Q-factor of 1 and center frequencies of 4, 8, and

16 Hz, respectively. For a given noisy speech signal, the temporal pattern of the noise alone (after the preprocessing stages) is subtracted from the corresponding temporal pattern of the noisy speech. The resulting model representations of the test signal (R_{test}) and of a set of templates (R_{t_1}, R_{t_2}, …, R_{t_N}) are then aligned in time using a dynamic time warping (DTW) algorithm. Finally, the cross-correlation coefficients between the time-aligned test-signal representation ($\hat{R}_{test}$) and the time-aligned template representations ($\hat{R}_{t_1}, \hat{R}_{t_2}, \ldots, \hat{R}_{t_N}$) are calculated and, after adding a constant-variance internal noise to limit the model's resolution, converted to response percentages.

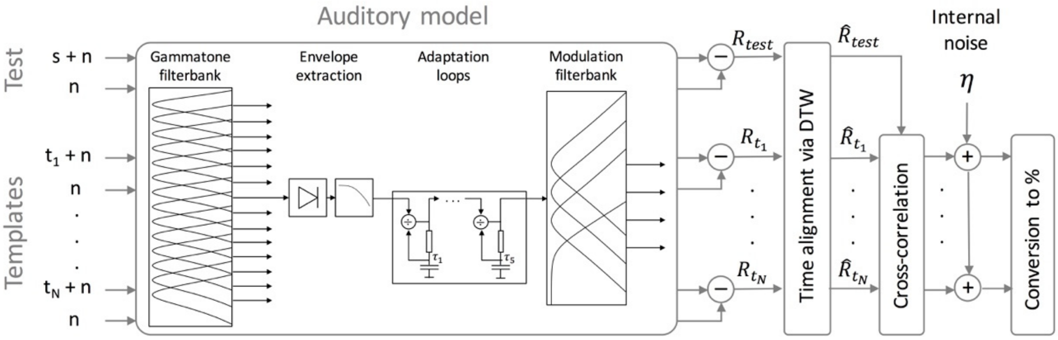


Fig. 2: Scheme of the considered microscopic speech perception model (from Zaar and Dau, 2017).

Simulation procedure

The experimental stimuli employed in the noise conditions were fed to the model as test signals. The templates were created by mixing the fifteen available CV tokens from the test-signal talker with randomly generated interrupted noise, using the same speech level and CV onset time as in the test signal. The "correct" template contained the same speech token as the test signal, whereas the noise signals differed. Randomly generated interrupted noise was considered as "noise alone". This is different from Zaar and Dau (2017), where the noise waveform in the test signals and templates was identical to the "noise alone", and was modified here to simulate potential informational contributions of the noise (i.e., noise bursts being mistaken for consonant cues). Five templates were generated for each speech token, SNR, and CV onset time, each using a different randomly generated interrupted noise waveform. The test signals and templates were passed through the model front end; only the consonant portions of the resulting internal representations, i.e., the portions of the CV tokens between consonant onset and vowel onset, were further considered in the back end. Whereas Zaar and Dau (2017) had considered the entire CV tokens, this modification was necessary here to prevent the model from being influenced by the task-irrelevant vowel portions, in particular when positioned in a noise gap. After obtaining the correlation coefficients between the internal representations of each test signal and the corresponding templates, the internal noise was added. Consistent with Zaar and Dau (2017), the variance of the internal noise was 0.05 and was held constant across the considered conditions. The model response for each iteration was defined as the template showing the largest correlation with the test signal.

RESULTS AND ANALYSIS

Experimental results

The measured consonant recognition scores were averaged across consonants and talkers. Figure 3 shows the recognition scores in terms of the mean and standard deviations across listeners as a function of speech level for the considered conditions. It can be observed that consonant recognition was at ceiling for the speech tokens presented in quiet (crosses), both for presentation levels of 45 and 65 dB SPL, albeit with a larger standard deviation at 45 dB SPL. Thus, it can be concluded that (i) the speech tokens were perfectly identifiable in quiet and (ii) that audibility was sufficient even at the lowest speech level considered. A two-tailed paired-sample t-test confirmed the latter observation, indicating no significant effect of presentation level in quiet ($p = 0.143$).

The remaining symbols in Fig. 3 depict the recognition scores obtained for the CVs mixed with the fixed-level interrupted noise according to the different CV onset times (cf. Fig. 1). As expected, consonant recognition generally decreased with decreasing speech level. Moreover, a clear effect of CV onset time can be observed. Specifically, the earliest CV onset time of 400 ms resulted in the lowest recognition scores (circles) and increasing CV onset times generally induced increasing recognition scores. However, this trend did not persist for the CV onset time of 550 ms (upward facing triangles), which induced lower recognition scores than the CV onset time of 525 ms (downward facing triangles). Furthermore, the recognition scores obtained for CV onset times of 500 ms and 525 ms were almost identical at the two lowest speech levels. Most of the reduction in recognition scores can be attributed to the degree that the consonant cues were simultaneously masked: As some consonant cues last up to around 100 ms, simultaneous masking was – depending on the CV onset time – induced by the third (CV onset times of 400 and 450 ms; circles and squares) or the fourth (CV onset times of 525 and 550 ms; upward and downward facing triangles) noise burst (cf. Fig. 1). Nonetheless, an effect of forward masking was clearly also present, as the recognition scores obtained in the condition with the least amount of simultaneous masking (CV onset time of 500 ms; diamonds) were much lower than in quiet (crosses) and somewhat lower than in the conditions with more simultaneous masking (CV onset time of 525 and 550 ms; downward and upward facing triangles, respectively). Two-tailed paired-sample t-test comparing all ten combinations of the five CV onset-time conditions were conducted after collapsing the recognition scores across speech level. In accordance with the previous observations, the results indicated highly significant ($p < 0.0001$) differences between all conditions except between CV onset times of 500 and 550 ms ($p = 0.568$). The latter two conditions did, however, exhibit highly significant ($p < 0.0001$) differences at a speech level of 65 dB SPL.

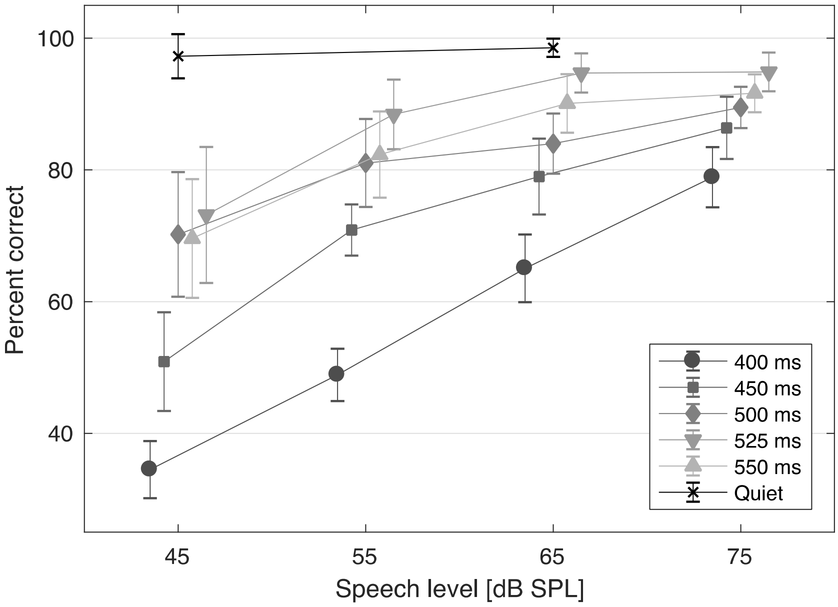


Fig. 3: Average consonant recognition scores as a function of speech level. The crosses represent the quiet conditions. The other symbols represent the different noise conditions, as indicated by the CV onset times in the legend. The noise level was 65 dB SPL. The error bars depict the standard deviation across listeners. A slight horizontal jitter was applied for visual clarity.

Model predictions

The predicted recognition scores obtained for the experimental stimuli in the noise conditions are depicted in the left panel of Fig. 4. Comparing these predictions with the data shown in Fig. 3, it can be observed that the model predictions were generally similar to the measured data as (i) the recognition scores globally decreased with decreasing speech level and (ii) the loss of consonant recognition was proportional to the amount of simultaneous masking. However, the model did not predict the extent of consonant recognition loss induced by the predominantly forward-masking based condition (CV onset time of 500 ms; diamonds), i.e., the effect of forward masking was smaller in the model than measured in listeners. Accordingly, the mean average error between predicted and measured recognition scores was relatively large for the CV onset time of 500 ms (15.8%) and much smaller for the remaining conditions (4.8% on average). Nonetheless, the recognition scores shown in Figs. 3 and 4 (left panel) were overall strongly correlated (Pearson's *r* of 0.94), as can be seen in the scatter plot presented in the right panel of Fig. 4.

So far, only average consonant recognition scores have been considered. However, different consonants are typically very differently affected by the masking noise (cf. Zaar and Dau, 2015). To investigate whether the model predicted the trends across consonants correctly, the measured and predicted consonant-specific recognition scores were averaged across speech levels and their Pearson's correlation across consonants was computed. Consistent with the model predictions reported for stationary-noise conditions (Zaar and Dau, 2017), the correlations were large ($r > 0.75$) and highly significant ($p < 0.001$) for all considered conditions except for the CV onset time of 500 ms, for which, nonetheless, significant correlation was found ($r = 0.5$, $p < 0.05$).

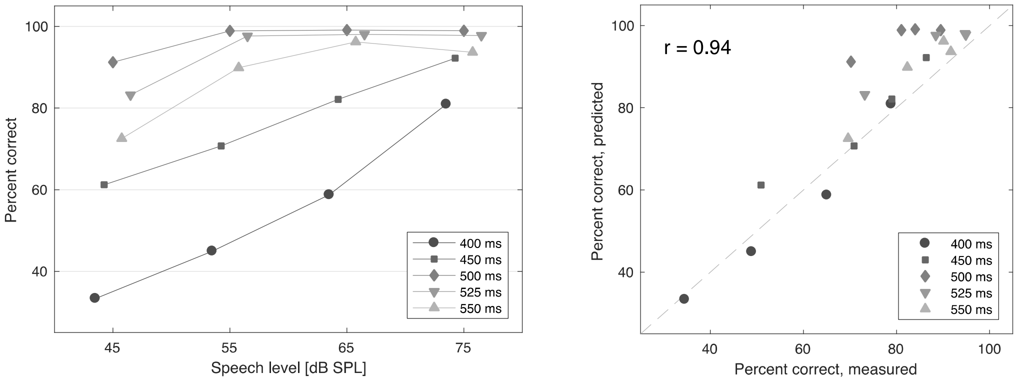


Fig. 4: Left panel: Model predictions of average consonant recognition scores as a function of speech level, corresponding to Fig. 3. Right panel: Model predictions of average consonant recognition scores (shown in the left panel) as a function of their measured counterparts (shown in Fig. 3).

DISCUSSION AND OUTLOOK

The perceptual effects measured in the present study suggest that consonant perception in non-stationary noise strongly depends on the position of the consonant cue relative to the noise envelope's minima and maxima and thus on the amount of simultaneous masking. This is consistent with the well-established observation that listeners make use of "glimpses" of the target speech in fluctuating interferers (e.g., Cooke, 2006). However, since the present study considered CVs as target signals, the experimental paradigm revealed more detailed effects, including a clear effect of forward masking. Thus, the paradigm may be useful for revealing temporal effects of hearing-aid processing, as demonstrated in a related study by Kowalewski *et al.* (2017), which applied the paradigm to investigate the effects of slow- vs. fast-acting compression in hearing-impaired listeners. While only consonant recognition has been discussed here, an additional analysis of the consonant confusions in the data may reveal the interaction between noise and speech tokens in more detail. For instance, it is possible that the noise bursts acted not only as a masker but were even mistaken for consonant cues, thus adding an informational component. It needs to be further investigated, however, whether the detailed effects measured with the considered nonsense speech tokens and artificial interrupted noise also play a role in more realistic conditions.

The model predictions obtained in the present study show that the large predictive power of the model, which had previously been demonstrated in conditions of stationary noise (Zaar and Dau, 2017) and spectral aspects of hearing-instrument processing (Zaar *et al.*, 2017), also extends to conditions of non-stationary noise. Only consonant recognition was considered here and it needs to be investigated whether the model also can account for consonant confusions in the data. Despite the overall accurate predictions, the model was found to be not sensitive enough to the effects of forward masking. While the underlying auditory model (Dau *et al.*, 1997) contains an adaptation stage and does account for "classical" forward-masking data (for narrowband signals), the present speech-based configuration does not seem to fully capture the reported forward-masking effects measured with speech signals. Thus, it may be useful to adapt the model such that it better accounts for this aspect of the data, for instance by modifying the time constants in the adaptation loops or integrating a simulation of the cochlear nonlinearities. Overall, the model may be useful as a tool for analysing temporal effects of hearing-aid processing, in particular when combined with simulations of individual hearing loss.

REFERENCES

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am., **119**, 1562-1573. doi: 10.1121/1.2166600

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation: I. Detection and masking with narrow band carrier," J. Acoust. Soc. Am., **102**, 2892-2905. doi: 10.1121/1.420344

Kowalewski, B., Zaar J., Fereczkowski, M., MacDonald, E., Strelcyk, O., May, T., and Dau T. (**2017**) "Effects of slow- and fact-acting compression on hearing-impaired listeners' consonant-vowel identification in interrupted noise," Proc. ISAAR, **6**, 375-382.

Miller, G.A., and Nicely, P.E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am., **27**, 338-352. doi: 10.1121/1.1907526

Phatak, S.A., and Allen, J.B. (**2007**). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am., **121**, 2312-2326. doi: 10.1121/1.2642397

Schmitt, N., Winkler, A., Boretzki, M., and Holube, I. (**2016**). "A phoneme perception test method for high-frequency hearing aid fitting," J. Am. Acad. Audiol., **27**, 367-379. doi: 10.3766/jaaa.15037

Zaar, J., and Dau, T. (**2015**). "Sources of variability in consonant perception of normal-hearing listeners," J. Acoust. Soc. Am., **138**, 1253-1267. doi: 10.1121/1.4928142

Zaar, J., and Dau, T. (**2017**). "Predicting consonant recognition and confusions in normal-hearing listeners," J. Acoust. Soc. Am., **141**, 1051-1064. doi: 10.1121/1.4976054

Zaar, J., Schmitt, N., Derleth, R.-P., DiNino, M., Arenberg, J.G., and Dau, T. (**2017**). "Predicting effects of hearing-instrument signal processing on consonant perception," J. Acoust. Soc. Am., *under review.*

Predicting the benefit of binaural cue preservation in bilateral directional processing schemes for listeners with impaired hearing

THOMAS BRAND[1,*], CHRISTOPHER F. HAUTH[1], KIRSTEN C. WAGENER[2], AND TOBIAS NEHER[1,3]

[1] *Medizinische Physik and Cluster of Excellence "Hearing4All", Universität Oldenburg, Oldenburg, Germany*

[2] *Hörzentrum Oldenburg GmbH and Cluster of Excellence "Hearing4All", Oldenburg, Germany*

[3] *Institute of Clinical Research, University of Southern Denmark, Odense, Denmark*

Linked pairs of hearing aids offer various possibilities for directional processing providing adjustable trade-off between improving signal-to-noise ratio and preserving binaural listening. The benefit depends on the processing scheme, the acoustic scenario, and the listener's ability to exploit binaural cues. Neher *et al.* (2017) investigated candidacy for different bilateral processing schemes for 20 elderly listeners with symmetric and 19 age matched listeners with asymmetric hearing thresholds below 2 kHz. The acoustic scenarios consisted of a frontal target talker presented against two intelligible or unintelligible speech maskers from ±60° azimuth. In this study, the speech reception threshold (SRT) data were compared to predictions of the binaural speech intelligibility model (BSIM; Beutelmann *et al.*, 2010), which was used to model pure better-ear-glimpsing as well as additional binaural unmasking. The speech intelligibility index (SII), which served as backend of BSIM, was calibrated to an individual reference value at the SRT for each listener. This reference value mirrors the amount of acoustical information needed by the listener to achieve the SRT and correlated with the listeners' ability to process temporal fine structure. BSIM revealed a benefit due to binaural processing in well-performing listeners when processing provided low-frequency interaural timing cues.

INTRODUCTION

Due to wireless across-device links, bilateral processing schemes have become applicable in commercial hearing aids (HA). This allows improving the signal-to-noise-ratio (SNR) by exploiting interaural differences between target speech and interferers. This mimics the human binaural auditory system that is known to exploit interaural differences for binaural release from masking. Some bilateral processing schemes sacrifice interaural differences for the sake of SNR improvement. However,

[*] Corresponding author: thomas.brand@uol.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

for the individual listener with impaired hearing it is unclear in advance if binaural release from masking provided externally by bilateral HA processing is beneficial or if the listener is able to achieve the same benefit using his or her own binaural processing. The latter has the advantage that binaural cues are preserved, enabling more natural listening including localisation and source separation. Neher *et al.* (2017) investigated the suitability of different bilateral directional processing schemes for listeners with hearing impairment and for both intelligible and unintelligible speech maskers. As the hearing loss of all listeners was compensated for by providing amplification in accordance with the NAL-R prescription rule, supra-threshold (e.g., binaural) processing played a major role.

Neher *et al.* used the binaural intelligibility level difference (BILD) for assessing the listeners' binaural processing abilities for speech in noise. The BILD is defined as the difference between two speech reception threshold (SRT) measurements obtained for a speech source located at 0° azimuth in presence of a noise source located at 90° (or −90°) azimuth. For the first SRT only the ear which benefits from the head shadow is used, for the second SRT both ears are used which enables binaural processing. Neher *et al.* found that listeners with BILDs larger than about 2 dB showed a larger benefit from preserved binaural cues at low frequencies compared to greater SNR improvement achieved by the beamforming algorithms used in their study. The opposite was true for listeners with smaller BILDs. Audiometric asymmetry reduced the influence of binaural hearing only slightly. Furthermore detection performance of an interaurally phase inverted 500-Hz sinusoid in interaurally coherent noise (N0Sπ) was an effective predictor of the benefit from preserved low-frequency binaural cues.

In this study, the data of Neher *et al.* (2017) were reanalysed using the binaural speech intelligibility model (BSIM) of Beutelmann *et al.* (2010), which combines the equalization-cancellation (EC) process (Durlach, 1963) as a model of the effective binaural processing with the speech intelligibility index (SII; ANSI, 1997) to predict binaural speech intelligibility in different acoustic scenarios. This model also considers the individual hearing status by taking the audiogram into consideration. BSIM can be used to analyse the relative contribution of binaural processing and better-ear-glimpsing (i.e., using in each frequency channel and each time frame the ear with the better SNR; Brungart and Iyer, 2013) on the predicted SRTs. Therefore, it can also be used to investigate whether or not individual listeners rely mainly on their better-ear to understand speech and to which extent they benefit from their own binaural processing. Furthermore it is evaluated if BSIM is able to predict the correlations to BILD and binaural masking level difference (BMLD) found by Neher *et al.* (2017).

The main research questions of this study were: (1) Is BSIM applicable to intelligible and unintelligible speech maskers? (2) Does binaural processing play a role in aided patients or does better-ear-glimpsing explain most of the benefit in spatial listening conditions? (3) Is BSIM able to separate the benefit due to processing of interaural differences of temporal fine structure from the benefit due to better-ear-glimpsing?

METHOD

Data

The data described in Neher *et al.* (2017) were used. Twenty elderly listeners (age: 63-80 years) with symmetric hearing thresholds (PTA4: 52 dB HL) and 19 elderly listeners (age: 62-80 years) with asymmetric hearing thresholds (PTA4: 53 dB HL) below 2 kHz took part in the experiment. Listeners were matched for age, hearing loss, and selective attention. Furthermore the listeners were split into a group with high BILD and a group with low BILD in this study.

The aided SRTs were measured using the Oldenburg sentence test (OLSA; Wagener *et al.*, 1999) with a frontal target talker and two speech maskers located at ±60° azimuth. To create the spatial arrangement of target and interfering signal, the signals were convolved with head related impulse responses (HRIRs; Kayser *et al.*, 2009), recorded with the microphones of two behind-the-ear (BTE) hearing aid dummies, which were equipped to a head-and-torso simulator.

The first masker with high informational masking (IM) consisted of Oldenburg sentences uttered by another male talker. The second masker with low IM was generated by transforming the unintelligible international speech test signal (ISTS; Holube *et al.*, 2010) to male pitch and vocal track length.

Bilateral directional processing simulating a linked pair of completely occluding BTE HAs was applied. The processing schemes differed in the trade-off between SNR improvement and binaural cue preservation. Scheme “pinna” simulated the directivity of the human pinna without any bilateral processing. Scheme “beamfull” simulated a bilateral beamformer steered towards the frontal direction which sacrificed all interaural cues to improve the SNR.

Binaural speech intelligibility model (BSIM)

The BSIM of Beutelmann *et al.* (2010) is shown in Fig. 1. It uses a gammatone filterbank with 30 frequency channels ranging from 143 to 8346 Hz to separate the input signals into different frequency bands. The individual hearing loss is considered after peripheral filtering by adding uncorrelated noises to the left and right ear to the interfering noise. In each band, an independent EC process according to Durlach (1963) is performed. Durlach’s model assumes that the left and right ear signals are subtracted after an equalization of interaural level and time differences. As such, this approach implicitly requires an analysis of the temporal fine structure in the auditory system. Normally distributed internal processing errors are assumed, which limit the EC processing to be performed effectively at frequencies below 1500 Hz; At higher frequencies BSIM effectively performs better-ear-glimpsing.

The final step in the BSIM framework is the transformation of the SII value to an SRT value. This transformation belongs to the SII concept (ANSI, 1997) and defines the SII value representing the amount of acoustical information (expressed in the form of a frequency-weighted SNR) which is required to understand 50% of the speech, i.e., to reach the SRT. This SII reference is dependent on the speech material. The model’s processing is performed in 23-ms time frames, and the final SRT is calculated by averaging the short-time SRTs across time, which takes into account that the used speech maskers fluctuate over time. To test the hypothesis that better-ear-glimpsing alone can explain the binaural benefit, BSIM was used both in "better-ear-glimpsing only" mode and in equalization-cancellation (EC) mode, respectively. In the first mode, in each 23-ms time frame and each auditory frequency band the SNR from the better ear was used for intelligibility prediction. In the EC mode, BSIM additionally incorporates interaural processing of the left and right ear signals according to the EC model.

An extension of BSIM incorporating individual internal processing errors derived from individual BMLD measurements was introduced by Hauth *et al.* (2017). Neher *et al.* (2017) measured BMLDs at 500 and 1000 Hz. The possible improvement of the individualized internal processing errors was evaluated by comparing the original BSIM (see Figs. 2 to 4) with the extended BSIM (see Fig. 5).

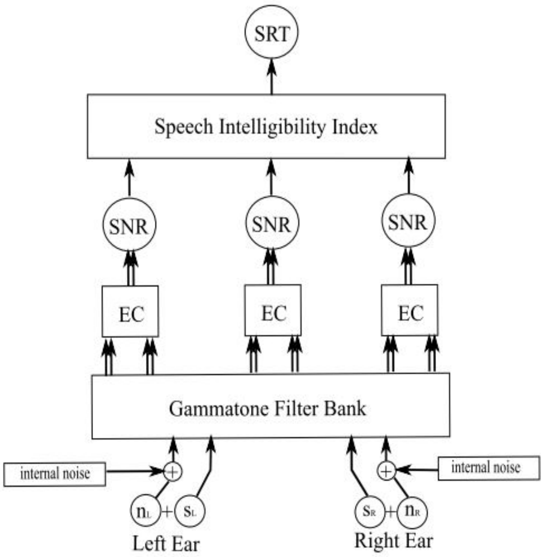


Fig. 1: Binaural Speech Intelligibility Model (BSIM) according to Beutelmann *et al.* (2010).

RESULTS AND DISCUSSION

A first analysis showed that the SII reference varied strongly across listeners. First, a fixed SII criterion was chosen to predict SRTs. This approach successfully predicted the effect of the audiogram on SRTs in hearing-impaired listeners (Beutelmann *et al.*, 2010). However, the stimuli analyzed in this study were amplified according to the audiogram (NAL-R) and audibility played a minor role. As a consequence, using a fixed SII criterion did not successfully predict the individual SRT in this study. Instead, it can be assumed that the observed variability of SRTs is mainly caused by supra-threshold and cognitive processing differences of the individual listeners. In order to test this hypothesis, one individual SII reference was fitted for each listener

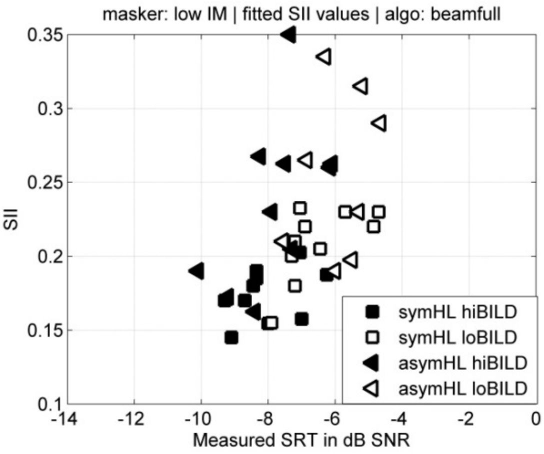


Fig. 2: Individual SII values (y-axis) for unintelligible speech maskers at the individual SRT (x-axis) for the diotic (“beamfull”) algorithm. These SII values were used as individual SII references in the following predictions.

for the diotic situation (“beamfull” algorithm) and the unintelligible speech masker (low IM) and it was evaluated how BSIM using this individual SII reference is able to predict the SRTs for the other HA algorithms and for the unintelligible speech masker. Note that the algorithm used for fitting the SII reference does not provide any binaural cues and can thus be regarded as diotic (despite the compensation of hearing loss according to NAL-R which differed across ears). Consequently, the individual SII reference is independent of the listener’s binaural processing. In Fig. 2, the resulting SII references for the different listeners are shown on the y-axis with the corresponding measured SRT on the x-axis.

Note that the individual SII reference value correlates with both the symmetry of the hearing loss and the the BILD. The lowest SII values were obtained for listeners with symmetric hearing loss and large BILDs. Slightly higher SII values were obtained for listeners with symmetric hearing losses and small BILDs. Listeners with asymmetric hearing losses showed higher SII values and a larger spread of these values across listeners compared to listeners with symmetric hearing losses. Note that the BILD is a binaural measure whereas the SRTs analyzed here were obtained with diotic stimuli. However, the BILD requires intact processing of temporal fine structure and was found to be correlated to monaural temporal fine structure sensitivity (Neher *et al.*, 2017).

Figure 3 shows BSIM predictions for the unintelligible speech masker (low IM) assuming only better-ear-glimpsing (left panel) or additional EC processing (right panel). The “pinna” algorithm was used here, meaning that binaural cues were available to the listeners. The diagonal line corresponds to perfect match between predicted and measured data; Points below the diagonal line represent underestimated SRTs and points above the diagonal overestimated SRTs. The left panel shows that using only better-ear-glimpsing results in an overestimation of the obtained SRTs. If the EC mechanism is applied, the predicted SRTs decrease by 1-2 dB, leading to a somewhat better agreement with the perceptual data.

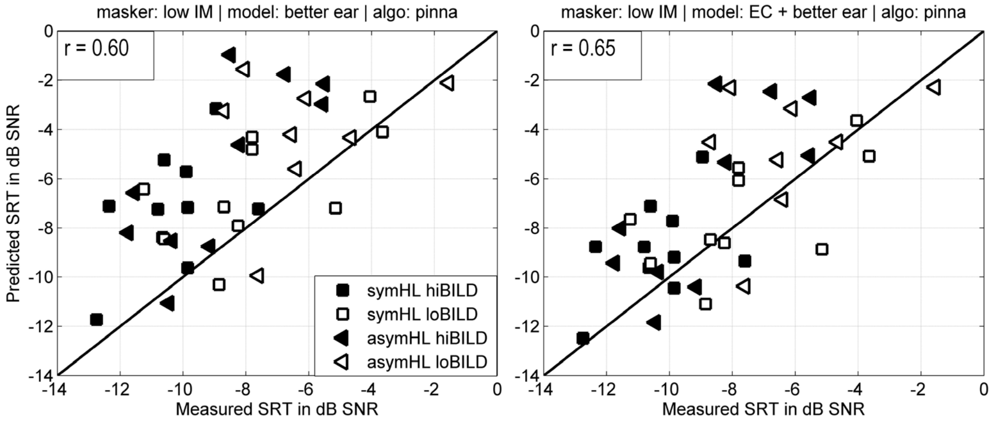


Fig. 3: The left panel shows predicted SRTs (y-axis) vs. measured SRTs (x-axis) for the *unintelligible* speech maskers (low IM) and the “pinna” condition (full binaural information available) for the case of better-ear-glimpsing only (no binaural interaction). The right panel shows corresponding predictions with EC processing included in the BSIM.

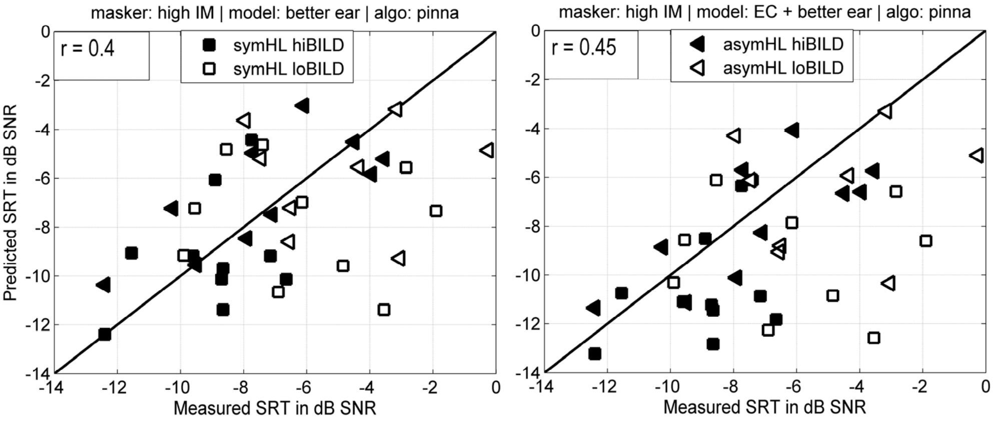


Fig. 4: Identical display as shown in Figu. 3 for the *intelligible* speech maskers (high IM).

Figure 4 shows BSIM predictions for the intelligible speech maskers (high IM). The left panel shows SRTs predicted using better-ear-glimpsing, the right panel shows SRTs predicted using both EC processing and better-ear-glimpsing. The predicted SRTs were mostly underestimated, which can be explained by the stronger IM of the intelligible speech maskers (which consisted of the same types of sentences as the target speech). This is not taken into account by BSIM as the SII was calibrated to the unintelligible speech maskers. Especially SRTs obtained at higher SNRs were underestimated for both better-ear-glimpsing only and better-ear-glimpsing plus EC.

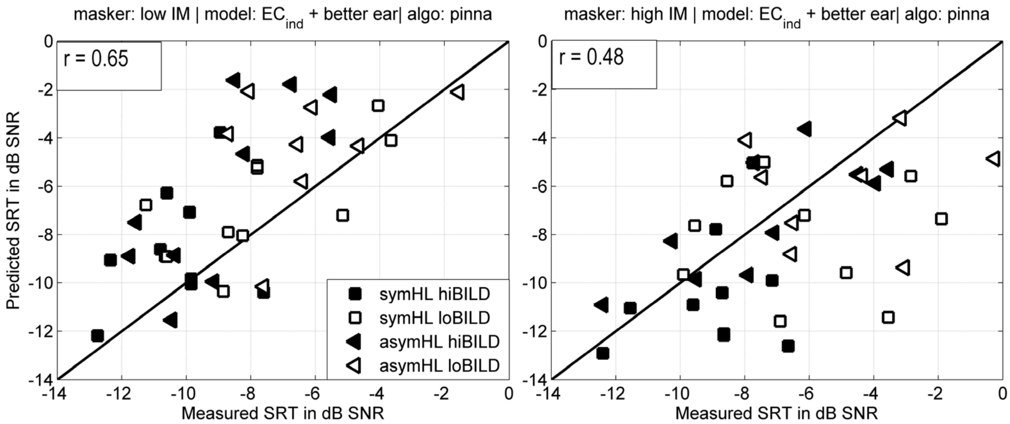


Fig. 5: Predictions obtained with extended BSIM incorporating individual binaural processing errors estimated from BMLDs measured at 500 Hz and 1000 Hz. The left panel shows results for the speech maskers with low IM and the right panel shows results for the speech maskers with high IM.

Figure 4 (high IM) shows that for several listeners with low BILD (open symbols) the observed SRT for the intelligible masker is much higher (worse) than predicted which is not the case for the unintelligible masker as shown in Fig. 3 (low IM). This suggests that the additional IM of the intelligible masker is more detrimental to these listeners than it is for listeners with large BILDs. Note that Neher et al. (2017) showed a significant correlation of BILDs with temporal fine structure sensitivity which might be relevant in scenarios with high IM, where target and interferer can better be separated, for instance, using pitch information.

Figure 5 shows predictions of the extended BSIM (Hauth *et al.*, 2017) incorporating individualized binaural processing errors estimated from BMLDs measured at 500 and 1000 Hz for characterizing supra-threshold binaural processing deficits. In general, larger processing errors were found compared to normal-hearing data. As a consequence, the predicted benefit from binaural processing is reduced, with better-ear-glimpsing defining the upper bound.

The additional individualization of binaural processing errors led to only slight improvement of the predictions. But based on the findings from Neher *et al.* (2017) it can be assumed that both SII and EC individualization mirror supra-thresholds processing deficits in temporal fine structure and are, therefore, highly correlated.

CONCLUSIONS

BSIM is applicable to speech maskers. This was achieved by calibrating the SII back-end of BSIM to a diotic condition comprising an unintelligible speech masker individually for each listener. This took into account that different listeners required different amounts of acoustical information (as quantified by the SII) to reach the SRT in a diotic condition. Using this "monaural" individualization, significant correlations between predictions and observations were achieved for an intelligible masker as well as for bilateral processing schemes that also included binaural processing by the listeners.

BSIM predicted a benefit due to 'true' binaural processing for aided listeners with impaired hearing.

Virtually no improvement of prediction accuracy was achieved, when additionally to monaural individualization the binaural processing errors in BSIM were individualized based on BMLDs measured at 500 and 1000 Hz. This might be due to the fact that the BMLDs were also correlated to the diotic SRTs, so that the binaural individualization did not add information.

ACKNOWLEDGMENTS

This study was supported by the DFG Cluster of Excellence EXC 1077/1 "Hearing4All". The research reported in (Neher *et al.*, 2017) was additionally supported by Sonova AG, Switzerland.

REFERENCES

ANSI (**1997**). *Methods for Calculation of the Speech Intelligibility Index (S3.5-1997).* American National Standard Institute, New York, NY

Beutelmann, R., Brand, T., and Kollmeier, B. (**2010**). "Revision, extension, and evaluation of a binaural speech intelligibility model," J. Acoust. Soc. Am., **127**, 2479-2497. doi: 10.1121/1.3295575

Brungart D.S., and Iyer N. (**2012**) "Better-ear glimpsing efficiency with symmetrically-placed interfering talkers," J. Acoust. Soc. Am., **132**, 2545-2556.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking level differences," J. Acoust. Soc. Am. **35**(8), 1206–1218. doi: 10.1121/1.1918675.

Hauth, C.F., Brand, T., and Kollmeier B. (**2017**). "Modelling the frequency dependency of binaural masking level difference and its role for binaural unmasking of speech in normal hearing and hearing impaired listeners," J. Acoust. Soc. Am. **141**, 3638. doi: 10.1121/1.4987846

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. doi: 10.3109/14992027.2010.506889

Kayser, H., Ewert, S.D., Anemüller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B. (**2009**). "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," EURASIP J. Adv. Signal Process., 1-10.

Neher, T., Wagener, K.C., and Latzel, M. (**2017**). "Speech reception with different bilateral directional processing schemes: Influence of binaural hearing, audiometric asymmetry, and acoustic scenario," Hear. Res. **353**, 36-48.

Wagener, K., Brand, T., and Kollmeier, B. (**1999**). "Development and evaluation of a sentence test for the German language. I-III: Design, optimization and evaluation of the Oldenburg sentence test," Z. für Audiol. Audiol. Acoust., **38**, 86e95, 4-15, 44-56.

Innovative methods and technologies for spatial listening and speech intelligibility using hearing implants

Anja Chilian[2], Maria Gadyuchko[1], András Kátai[2], Florian Klein[1], Thomas Sattel[1], Verena G. Skuk[3], and Stephan Werner[1,*]

[1] *Technische Universität Ilmenau, Ilmenau, Germany*

[2] *Fraunhofer IDMT, Ilmenau, Germany*

[3] *HNO-Klinik, Universitätsklinikum Jena, Jena, Germany*

The proportion of the population with acquired hearing loss is increasing worldwide. Specific types of hearing loss require the treatment with hearing implants. Cochlear implants and bone conduction hearing implants are two examples. The present contribution is a prospect of the underlying project in its early stadium. The project addresses new methods and technologies that improve spatial hearing with such implants. The methods are adjusted specifically for both types of hearing implants. For cochlear implants bio-inspired signal processing methods are applied. For bone conduction implants new working principles for mechanical stimulation based on piezoelectric transducers are investigated. To evaluate the developments perceptional experiments are conducted, which investigate spatial hearing and speech intelligibility with normal-hearing and hearing-impaired persons. For this purpose a virtual listening environment is applied to synthesize different room acoustics, source positions, audio signals, and acoustic scenes with different complexity. Cochlear implants and a custom-made bone conduction device are used as playback systems. The bone conduction device generates the mechanical input and transmits mechanical oscillations via the temporal bone to the cochlea. Listening tests assess speech intelligibility with spatially distributed background noise and localization abilities.

INTRODUCTION AND MOTIVATION

Due to extensive research in the field of cochlear implants (CI) and bone conduction hearing implants (BCI), significant improvements have been achieved for hearing-impaired patients in the last decades. The range of indications for these implants is constantly extending because of diverse continuance of system developments. However, as hearing cannot be fully restored by such implants, hearing implant users are faced several challenges in everyday situations. For example, a reduced or missing capability to localize sound objects leads to reduced speech intelligibility in noisy situations.

We aim to contribute to the development of novel technologies for hearing implants and new methods for improved spatial hearing. The methods are evaluated, validated

[*] Corresponding author: stephan.werner@tu-ilmenau.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. ISBN: 978-87-990013-6-1.

and in particular adjusted for the two types of hearing implants, CI and BCI. In case of CI research, we work on improvements of a biologically inspired speech processing strategy, which will be evaluated in tests with bilaterally implanted CI users. In the field of BCI, our aim is to investigate piezoelectric transducers for mechanical stimulation. For listening tests with normal-hearing persons a custom-made device for bilateral percutaneous bone conduction will be used. Furthermore, a virtual listening environment is created to synthesize different room acoustics, source positions, audio signals, and acoustic scenes with different complexity. Cochlear implants and the head-mounted bone conduction hearing device are used as playback systems.

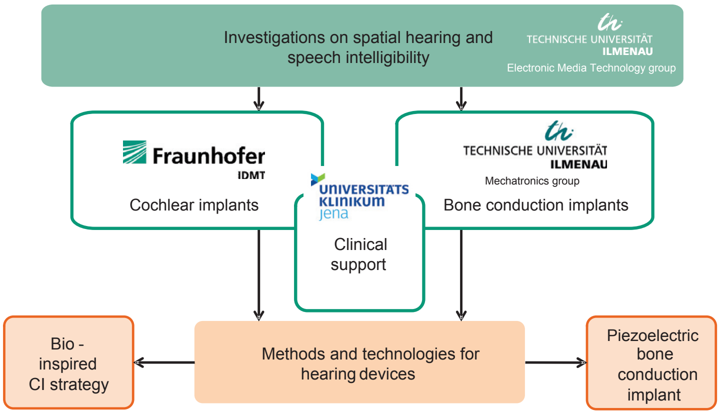

Investigations on spatial hearing and speech intelligibility
TECHNISCHE UNIVERSITÄT ILMENAU
Electronic Media Technology group
Fraunhofer IDMT
Cochlear implants
UNIVERSITÄTS KLINIKUM jena
Clinical support
TECHNISCHE UNIVERSITÄT ILMENAU
Mechatronics group
Bone conduction implants
Bio - inspired CI strategy
Methods and technologies for hearing devices
Piezoelectric bone conduction implant

Fig. 1: Composition of the research group: The four upper boxes show the expertise of each institution. The three boxes at the bottom stand for the three main goals of the project.

A schematic overview of our research group and our aims is given in Fig. 1. In a first step, methods and technologies for both CI and BCI are developed. Secondly, investigations on spatial hearing and speech intelligibility are performed for both types of hearing implants under the supervision of clinicians. Finally, the analysis of listening tests data is the base for further improvements of the two hearing device technologies.

COCHLEAR IMPLANT STRATEGIES

One factor influencing performance of CI users is the signal processing algorithm (also called CI strategy), which translates acoustic signals into electrical stimuli. Common CI strategies usually rely on linear filter banks and therefore mimic the processes of normal hearing only to a limited extent. As a result, CI users usually reach good speech intelligibility in quiet, but have deficits in noisy environments and music perception. Moreover, various studies show that spatial hearing is very limited, due to inadequate transmission of temporal fine structure in current CI strategies (Ching *et al.*, 2017; Wilson *et al.*, 2003).

To improve CI user performance, we developed a novel bio-inspired CI strategy called Stimulation based on Auditory Modeling (SAM, Harczos *et al.*, 2013a). It is based on a model of the human peripheral auditory system. By closer mimicking the normal human cochlea, a more natural hearing impression should be achieved. A pilot study demonstrated the potential of SAM for improvements in hearing perception (Harczos *et al.*, 2013b). Furthermore, simulations showed possible advantages of SAM regarding sound source localization (Harczos *et al.*, 2011).

For further enhancements of the new bio-inspired strategy, SAM will be adapted for bilateral usage. The aim is to investigate main characteristics of the SAM strategy for spatial hearing and speech intelligibility to optimize signal processing in respect to computational effort and perceived quality. For this purpose, the processing stages of SAM (shown in Fig. 2) will be modified and preprocessing algorithms for improved spatial hearing will be included into the strategy.

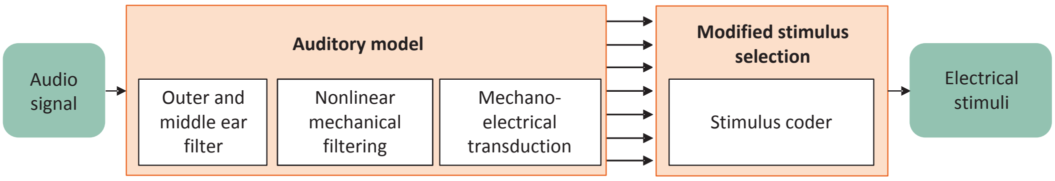

Audio signal
Auditory model
Outer and middle ear filter
Nonlinear mechanical filtering
Mechano-electrical transduction
Modified stimulus selection
Stimulus coder
Electrical stimuli

Fig. 2: Processing stages of the novel bio-inspired CI strategy SAM (Stimulation based on Auditory Modeling).

BONE CONDUCTION TECHNOLOGY

In today’s bone conduction (BC) hearing devices, like BAHA® or BONEBRIDGE™, electromagnetic transducers are commonly used for exciting bone vibrations. Due to the bandpass characteristic of their electro-magneto-mechanical impedance between output force and input voltage current devices are limited to BC-excitation frequencies less than 7 kHz. This restricts the intelligibility of fricatives and spatial hearing. Moreover, the functional principle leads to an electromechanical resonance in the auditory frequency range, which requires electronic compensation. An idea to avoid such disadvantages is the use of piezoelectric transducers. The principle of operation is based on a clamping mechanism instead of using the principle of inertia. Electromechanical resonances in the auditory frequency range can thus be avoided. The proposed concept is illustrated in Fig. 3. The long-term objective is to investigate possibilities and limitations of the use of piezoelectric transducers as implantable actuators. These should be embedded in the mastoid completely under the skin.

We will study the transmission characteristic of piezoelectric transducers and compare it with data of their electromagnetic counterparts. Since there are no subjects with implanted piezoelectric actuators, listening tests are carried out with normal hearing listeners using a custom-made device for bilateral percutaneous bone conduction.

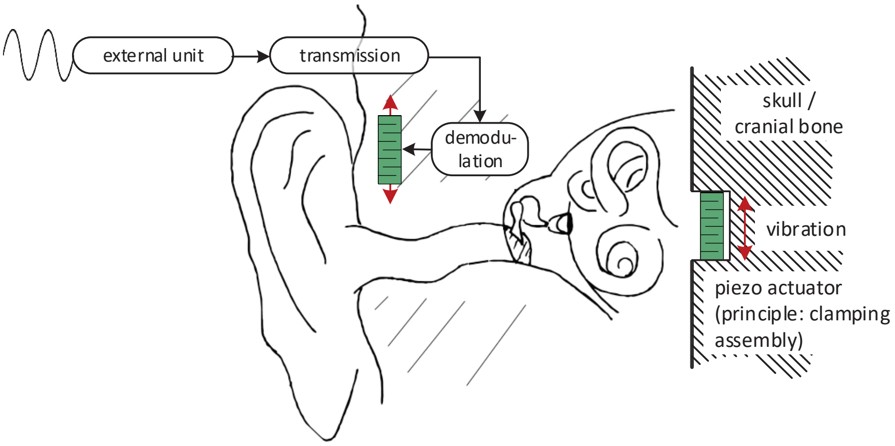


Fig. 3: Basic principle of implanted piezoelectric transducers for bone conduction.

We designed and built the device to produce comparable and reproducible listening test results. It has options to adjust the contact pressure and contact position of the piezoelectric transducer to the skin, see Fig. 4.

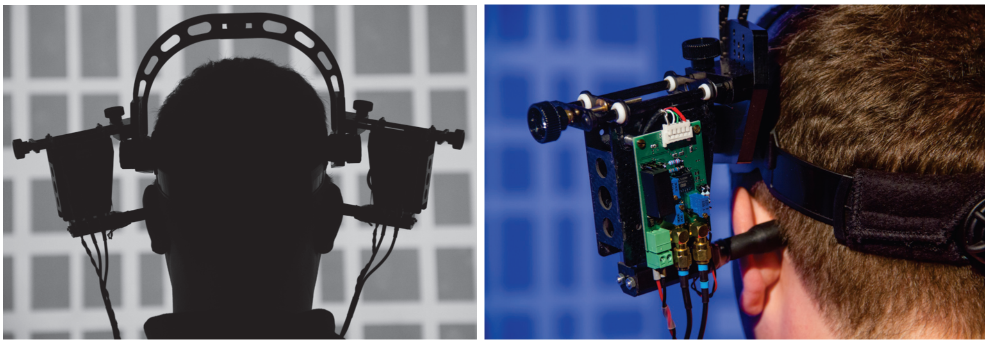


Fig. 4: Custom-made device for bilateral percutaneous bone conduction using piezoelectric transducers.

Only a few studies exist investigating spatial hearing via bone conduction. Regardless of the transducer principle, there are some restrictions on spatial listening. One reason for that is binaural cross-talk, which interferes with the interpretation of naturally occurring interaural level and time differences (ILD and ITD). Compared to air conduction, larger ILDs and ITDs are necessary for bone conduction to realize the same sound sources lateralization (Stenfelt and Zeitooni, 2013). This fact is probably the reason for the results of Barde *et al.*, 2016, who report elevated levels for the minimum discernible angular difference when using bone conduction and binaural synthesis. Another difficulty is the equalization of bone conduction devices, because the speed of sound depends on frequency and position. Beside of studies on localization, no other publications investigating other parameters of spatial hearing, such as distance or externality, could be found. Therefore, we aim to evaluate a wider range of spatial auditory parameters.

BINAURAL SYNTHESIS

Perceptional investigations and experiments for spatial hearing with CI and BC are realized using a binaural synthesis system. Binaural room impulse responses (BRIRs) have been recorded in real rooms using a head and torso simulator (KEMAR) (Klein *et al.*, 2017). BRIRs for several source-to-receiver positions for three different rooms are available. This allows to create numerous combinations of different room acoustics, source positions, audio signals, and acoustic scenes with different complexity for the planned listening tests. The technical and perceptional principles of binaural synthesis using airborne sound and playback via headphones are well understood (Lindau, 2014; Werner *et al.*, 2016). The challenges within the project lie in the measurement of system characteristics and development of adequate signal processing approaches for CI and BCI. The addressed characteristics for bone conduction are the estimation of transfer functions of the custom-made device for bilateral percutaneous bone conduction using equal loudness level contours. Furthermore, the binaural cross-talk is estimated in an indirect way using localization tests varying magnitudes of the used interaural cues of the binaural synthesis system.

LISTENING TESTS

Listening tests with bilateral cochlear implant users

In order to evaluate the novel CI strategy SAM, twenty adult patients with bilaterally implanted CIs, who have no apparent neurological or psychiatric disorders, will be invited to several listening tests. All tests described in the section below will be conducted as a comparison between SAM and a commercial strategy. Furthermore, the tests are performed iteratively in order to accompany the enhancements of SAM. Before each test, the CI users complete a habituation procedure to the coding strategy used subsequently.

Listening tests with the bone conduction device

To evaluate the novel bilateral piezoelectric bone conduction device (Fig. 4) a total number of 40 adult listeners (age between 18 and 25 years) with normal hearing and no apparent neurologic or psychiatric disorder will be tested. Participants wear the bone conduction device and in-ear hearing protectors whenever required, to minimize hearing of airborne sound. To calibrate the new device, curves of equal loudness will be measured for both airborne sound and bone conduction. The participants will perform the tests A to C, as described below, to investigate the properties of spatial hearing with the new device. To ensure equal loudness levels for all participants, the contact pressure of the piezoelectric actuators is adjusted accordingly.

Description of tests

In **Test A** the Oldenburg Sentence Test (OLSA, Wagener *et al.*, 1999) is conducted with spatially distributed sound sources. Speech recognition thresholds (SRT) are assessed in noise and in quiet. **Test B** describes a localization tests with varying spatial positions (see right side of Fig. 5) on the horizontal plane around the head of the subject. We measure the accuracy as the angular difference and as number of missed trials (no direction perceivable). Speech signals are used as test signals. **Test C** is a relative distance perception test at four directions. The two stimuli of each trial vary in distance to the listeners, who must tell in a 2-alternative-forced-choice task (2-AFC with response options for "first stimulus" or "second stimulus") which of the two stimuli is perceived as being closer. As additional experimental factors, the reverberation of the synthesized listening room is either dry or reverberant. The left side of Fig. 5 visualizes this test procedure. Speech signals are used as test stimuli.

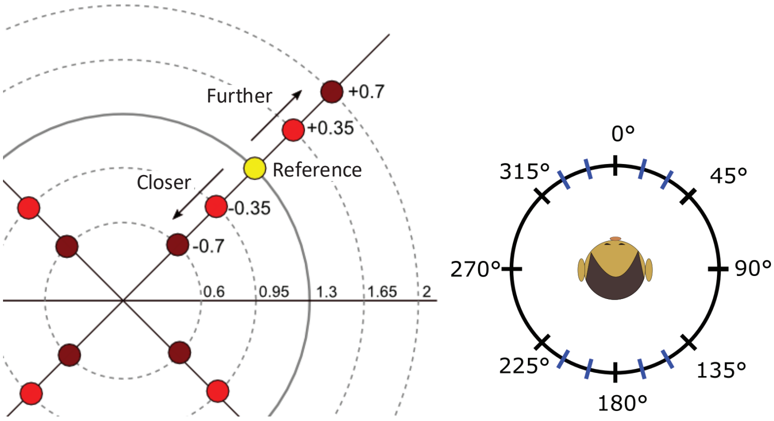


Fig. 5: Visualization of the direction and distances used for the spatial hearing listening tests.

Test D is designed to determine the just noticeable difference in the perception of pitch, direction, and distance. The 3-AFC 1-up-2-down method (Levitt, 1971) will be used. The listener has to determine which of the three stimuli is perceived as "different", or "odd". **Test E** aims to mimic a cocktail party situation. This competing talker test, measures the 50% SRT with OLSA sentences material. This time, however, the noise is replaced by two concurrent additional talkers. Possible combinations of positions are shown in Fig. 6. **Test F** is a test on prosody perception (Kuhnke *et al.*, 2015). The participants must discriminate in a 2-AFC procedure if a sentence is spoken with the intonation of a question or a declaration.

CONCLUSION

While the presented research plans promise great advantages for future users of the two types of hearing implants (CI and BCI), there are many open issues. Listening test evaluating new methods and technologies for hearing implants are difficult: For an

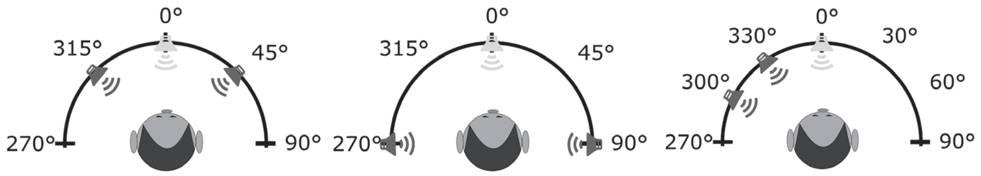


Fig. 6: Different sound source combinations for the cocktail party test. The target speech is marked light grey and the concurrent speakers dark grey.

objective comparison of different CI strategies adaptation processes have to be taken into account. The habituation of a CI patient to an new strategy is usually a long-term process and can not be covered within the short test periods. In case of BCI, the new technology can currently only be evaluated with normal hearing listeners. Thus the comparability with currently used bone conduction implants is limited. For implantation of the new technology there are still many challenges to overcome.

ACKNOWLEDGMENT

The research group and the underlying project is funded by the Free State of Thuringia and the European Social Fund. Many thanks to all members of the advisory board and to the students, which contributed to the work.

REFERENCES

Barde, A., Helton, W.S., Lee, G., and Billinghurst, M. (**2016**) "Binaural spatialisation over a bone conduction headset: Minimum discernable angular difference," Proc. 140th AES Convention, Paris, France.

Ching, T.Y.C., van Wanrooy, E., and Dillon, H. (**2007**) "Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review," Trends Amplif., **11**, 161-192.

Harczos, T., Chilian, A., and Katai, A. (**2011**) "Horizontal-plane localization with bilateral cochlear implants using the SAM strategy," Proc. ISAAR, **3**, 339-345.

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.

Harczos, T., Chilian, A., Kátai, A., Klefenz, F., Baljic, I., Voigt, P, and Husar, P. (**2013**) "Making use of auditory models for better mimicking of normal hearing processes with cochlear implants: First results with the SAM coding strategy," Proc. ISAAR, **4**, 317-324.

Klein, F., Werner, S., Chilian, A., and Gadyuchko, M. (**2017**) "Dataset of in-the-ear and behind-the-ear binaural room impulse responses used for spatial listening with hearing implants," Proc. 142nd AES Convention, Berlin, Germany.

Kuhnke, F., Jung, L., and Harczos, T. (**2015**) "Compensating for impaired prosody perception in cochlear implant recipients: A novel approach using speech preprocessing," Proc. ISAAR, **5**, 309-316.

Levitt, H. (**1971**) "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am., **49**, 467-477. doi: 10.1121/1.1912375

Lindau, A. (**2014**) *Binaural resynthesis of acoustical environments - Technology and perceptual evaluation.* Ph.D. thesis, Technische Universität Berlin, Fakultät I - Geisteswissenschaften. doi: 10.14279/depositonce-4085

Stenfelt, S., and Zeitooni, M. (**2013**) "Binaural hearing ability with mastoid applied bilateral bone conduction stimulation in normal hearing subjects," J. Acoust. Soc. Am., **33**, 481-493. doi: 10.1121/1.4807637

Wagener, K., Kühnel, V., and Kollmeier, B. (**1999**) "Entwicklung und Evaluation eines Satztests für die deutsche Sprache, Teil 1: Design des Oldenburger Satztests," Zeitschrift für Audiologie, **38**, 4-15.

Werner, S., Klein, F., Mayenfels, T., and Brandenburg, K. (**2016**) "A summary on acoustic room divergence and its effect on externalization of auditory events," Proc. 8th International Conference on Quality of Multimedia Experience (QoMEX), Portugal. doi: 10.1109/QoMEX.2016.7498973

Wilson, B.S., Lawson, D.T., Muller, J.M., Tyler, R.S., and Kiefer, J. (**2003**) "Cochlear implants: Some likely next steps," Annu. Rev. Biomed. Eng., **5**, 207-249.

Measuring hearing instrument sound modification using integrated ear-EEG

FLORIAN DENK[1,3,*], MARLEEN GRZYBOWSKI[1,3], STEPHAN M. A. ERNST[1,3,4], BIRGER KOLLMEIER[1,3], STEFAN DEBENER[2,3], AND MARTIN BLEICHNER[2,3]

[1] *Medizinische Physik, University of Oldenburg, Oldenburg, Germany*

[2] *Neuropsychology Group, University of Oldenburg, Oldenburg, Germany*

[3] *Cluster of Excellence Hearing4all, Oldenburg, Germany*

[4] *Present address: University Hospital Gießen and Marburg, Germany*

We integrated ear electrodes into a live hearing system and evaluated the feasibility of recording electroencephalography (EEG) features with this setup using an auditory discrimination experiment. The long-term goal is to construct a closed-loop brain-computer-interface that is integrated in a mobile research hearing system. Here, the EEG setup consists of 3 electrodes embedded in the earmoulds of an experimental hearing system and 10 flex-printed electrodes positioned around each ear, all connected to a wireless EEG amplifier. Four consecutive identical broadband stimuli were played in headphones while the spectral profile of sounds arriving at the eardrum was altered by switching the signal processing setting of the hearing system. Such switches were made between presentation of the third and the fourth stimulus, in half of all epochs. Seventeen normal hearing subjects participated and were instructed to indicate whether the last stimulus sounded different. The behavioural data verified clear audibility of the switches. The EEG analysis revealed differences between switch and no-switch trials in the N1 and P3 latency range. Importantly, changes in the spectral content of the noise floor of the hearing device were already sufficient to elicit these responses. These results confirm that stimulus-related brain signals acquired from ear-EEG during real-time audio processing can be successfully derived.

INTRODUCTION

Neuro-control of hearing devices has the potential to improve hearing support by taking the electroencephalography (EEG) derived listening intent of a person into account for better control of algorithms and hearing aid setting. For instance, it has been demonstrated that the direction of attention in a competing talker situation can be decoded from EEG signals within a reasonable duration (O'Sullivan *et al.*, 2015; Mirkovic *et al.*, 2016). Such information about the direction of attention may be utilized to enhance the speech signal of the desired speaker by, for example, steering a beamformer to the position of the attended speaker (Doclo *et al.*, 2015).

*Corresponding author: florian.denk@uni-oldenburg.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.

ISBN: 978-87-990013-6-1.

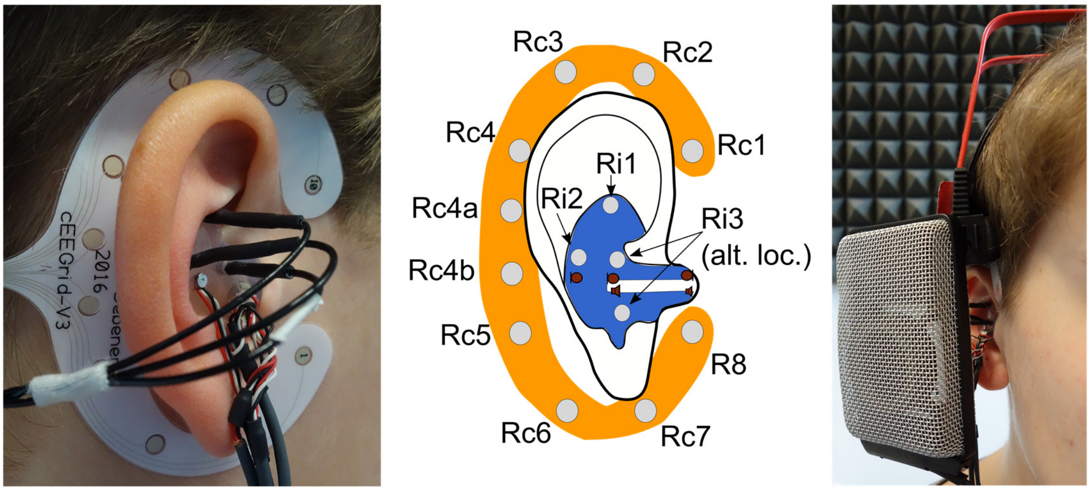


Fig. 1: Left: Photograph of the setup in the ear of a participant. In the concha, the earmould containing the hearing system and 3 electrodes (black sticks) are placed. The cEEGrid (www.ceegrid.com) is glued around the ear. Centre: Schematic view of the layout in the ear. Grey circles indicate electrodes with their according nomenclature; Positions of electro-acoustic transducers are marked by according symbols. The shaded area marks the part of the earmould which is inserted into the ear canal. Right: Participant wearing super-aural headphones (*AKG K-1000*) over the in-ear setup.

For an integration of EEG with hearing devices, however, it is necessary that the EEG signal can be acquired in a socially acceptable manner, with as little inconvenience for the hearing aid user as possible (Bleichner and Debener, 2017). In order to achieve such a transparent EEG acquisition, several ear-EEG approaches have been presented that allow to record EEG reliably in and around the ears (Looney *et al.*, 2011; Bleichner *et al.*, 2015; Bleichner and Debener, 2017). It has been shown that ear-EEG can be used to capture a wide variety of auditory perception-related processes: auditory steady state responses (Kidmose *et al.*, 2012), auditory onset responses, mismatch negativity as well as alpha attenuation (Mikkelsen *et al.*, 2015). Moreover, ear-EEG can also be used to detect the direction of auditory attention (Mirkovic *et al.*, 2016; Bleichner *et al.*, 2016).

The next step towards closing the loop between the hearing device and EEG is the integration of ear-centred EEG hardware into a live hearing system. We present a feasibility study of this combination, where electrodes are placed in and around the ear of a participant, integrated with an experimental hearing system (Denk *et al.*, 2017). To our best knowledge, this setup is the closest to a functional hearing device with integrated EEG electrodes that has been reported. Our goal here was to determine whether an audible switch in the hearing device processing is reflected in auditory evoked potentials (AEP) measured with ear-EEG.

METHODS

Setup

The participants were equipped with a prototype hearing system as presented by (Denk *et al.*, 2017), consisting of an individual silicone earmould that contains a set of electro-acoustic transducers shown in Fig. 1. External sound is captured with a microphone located in the concha, processed, and played back via an included receiver. Real-time processing is performed on a laptop running the Master Hearing Aid (MHA) platform (Grimm *et al.*, 2006), which is connected to the transducers through a *Multiface II* soundcard (RME, Haimhausen, Germany) with an input-output delay of 7.8 ms. By means of an individual in-situ calibration routine, the processing chain (here a finite impulse response filter) is adapted in a way that the superposition of electro-acoustically generated sound and sound leaking through the vented earpiece approximates the pressure at the eardrum that is observed with an open ear. Hence, acoustically transparent reproduction of the acoustic environment is provided while having the possibility to modify the presented sound in a desired manner by changing the output filter F.

EEG was acquired with ear-centred electrodes. In each ear three cylindrical electrodes (2 x 4 mm, Ag/AgCl, EasyCap, Herrsching, Germany; cf. Bleichner *et al.*, 2015) were distributed in the cavum concha by insertion into bores in the earmould. Additionally, ten printed Ag/AgCl electrodes were placed around the ear using the commercially available cEEGrid system (www.ceegrid.com), which is a flex-printed C-shaped electrode array placed around the ear (Debener *et al.*, 2015). After skin preparation with an abrasive gel and alcohol, a small amount of electrolyte gel (Abralyt HiCl, Easycap GmbH, Germany) was applied to the electrodes and the cEEGrids were placed with a double-sided adhesive tape around the ear. The cylindrical electrodes were inserted into the earmould after a drop of electrolyte gel was administered into the bores. The whole setup in the ear of a subject, as well as the schematic layout, is shown in Fig. 1. All electrodes were connected to a portable wireless 24-channel EEG amplifier attached to the subjects' heads (SMARTING, mBrainTrain, Belgrade, Serbia) and recording EEG signals with a sampling rate of 500 Hz and 24-bit resolution. A Bluetooth connection enabled wireless EEG recording on a separate computer. Although the system is a laboratory-state prototype, the suggested electrode layout is readily applicable in a real hearing system, or a fully mobile prototype.

The participants performed all tasks autonomously using graphical interfaces shown on a laptop that also controlled the hearing device while participants were seated in a sound-proof booth. Auditory stimulation and experimental control was implemented in MATLAB on the same laptop, which was also used to send EEG triggers synchronously to audio stimulation via Lab Streaming Layer (LSL; Kothe, 2015). On an additional computer located outside the booth, the Bluetooth EEG signal was recorded together with the trigger stream and a mirror of the acoustic stimuli.

Stimuli were presented via super-aural headphones (K-1000, AKG, Vienna, Austria), which are shown in Fig. 1. The special design assures that neither the electrodes nor the hearing device was touched by the headphone. Whereas EEG was recorded at both ears, the stimuli were presented monaurally to the right ear. Consequently, only the right ear was equipped with a hearing device and the left ear was fully occluded.

Paradigm and stimuli

Two different listening conditions were implemented by variable operation modes of the hearing device, while in all compared trials the identical stimulus waveform was played on the headphones. In one adjustment, the output filter of the hearing device was adjusted by individual calibration prior to the main experiment (filter F1). In the other condition, the output filter resulting from equivalent calibration of the system on a dummy head was used (F2), which results in a notable difference in the spectral profile arriving at the eardrum. Alternative cues that may arise from differences in loudness were compensated through an additional broadband gain applied to the dummy head filter, which was adjusted by means of an adaptive 1-up 1-down procedure prior to the main experiment.

Three types of stimuli were included: white noise ("*Noise*"), a logatome spoken by a female voice referred to as "*Speech*" (*Sass*, from the OLLO corpus; Meyer *et al.*, 2010), and the superposition of both with an SNR of 5 dB ("*Speech-In-Noise*"). To all stimuli, bandpass filtering between 0.1 and 12 kHz was applied.

Four identical stimuli were presented sequentially, in 50% of the trials the last stimulus was presented with a different filter setting (deviant condition, e.g., F1 F1 F1 F2) in 50% with the same filter (identical condition, e.g., F2 F2 F2 F2). The onset of the *n*-th stimulus is referred to as T*n*. Each stimulus was 500-ms long, separated by 300-ms breaks. To assure the participants' attention, they were asked to indicate whether the last stimulus was perceived as identical to the three prior sounds or not by pressing buttons on the laptop (y/n) guided by a graphical user interface. The response time window was limited to 1 second to get a spontaneous response from the participants, followed by a pause lasting randomly between 2.5 and 3.5 seconds.

The waveform of the *Speech-In-Noise* stimulus is shown together with the AEPs in Fig. 3 (Results section). Since a real-time hearing device was used, a noise floor was perceivable in silence, originating mainly from the microphone. Aiming to avoid sudden audible modification in noise timbre when the output filter was switched, the hearing device output was briefly deactivated while switching the output filter (or not), 120 ms after presentation of every trial (20-ms pause, with 10-ms ramps).

For each of the three stimuli, 16 deviant epochs in both possible orders (F1F2, F2F1), and the same number of non-deviant epochs in either filter setting (F1F1, F2F2) were presented. Hence, 192 sequences of stimuli were presented in randomized order, subdivided into four blocks of equal case distribution. The experiment included further conditions with a comparable number of trials, which are not considered here. Seventeen participants without any self-reported history of hearing disorder participated in the study. Including calibration of the hearing device and loudness matching of the presentation conditions prior to the main experiment, the experiment lasted about 90 minutes, separated by four small breaks between the experimental blocks.

EEG analysis

The offline analysis was performed with EEGlab (Delorme and Makeig, 2004) and MATLAB (Mathworks, Natick, MA). The data from each block was filtered between 0.1 and 12 Hz with consecutive high-pass and low-pass filters. Epochs were extracted for the entire trial (−1000 ms to 4000 ms relative to T1) as well as to the onset of the device before the last stimulus (−500 ms to 1000 ms). Epochs dominated by artefacts were identified using the probability criteria implemented in EEGLAB (standard deviation: 2) and rejected from further analysis. The grand average AEP over all trials and all participants was computed.

RESULTS AND DISCUSSION

Behavioural results

The behavioural discrimination results are shown in Fig. 2. Generally, the participants were able to discriminate well between identical and deviant trials. On average, the correct response was given in 90% and 93% of all epochs, respectively. Thus, the listening results verify the desired audibility of the difference between the two filter settings.

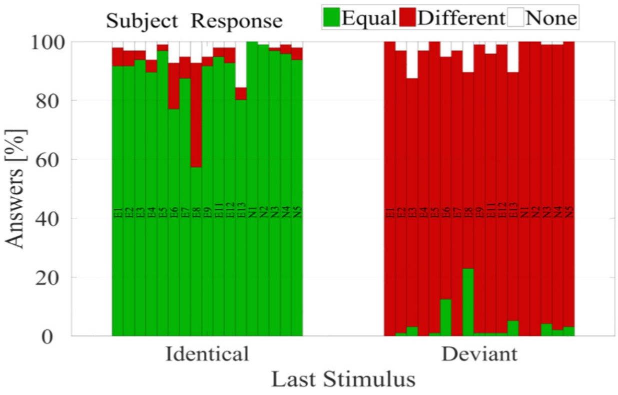


Fig. 2: Behavioural results, pooled over stimuli. Subjects E6, E8 and E13 were excluded from further analysis due to results indicating poor attention.

Some participants performed very clearly below average, which may be attributed to poor vigilance or task compliance. To avoid compromising the physiological results, data from participants were discarded if the following criteria were not fulfilled:

1. Identical stimuli sequences indicated as “identical” in more than 80 % of all epochs;
2. Deviant stimuli sequences marked as “identical” in less than 20 % of all epochs;
3. Answer given in more than 90% of all trials.

Consequently, the data from subjects E6, E8 and E13 were excluded from further analysis.

Auditory evoked potentials

Extensive pilot studies, including stimulation over distant loudspeakers with the hearing device deactivated, verified that the signals obtained in the electrodes originate from neural activity and not due to crosstalk from the audio transducers or connections.

The grand average AEP is shown in Fig. 3 together with the recording of the *Speech-in-Noise* stimulus. For the latter, the sound pressure measured at the eardrum of a dummy head is shown together with the output voltage of the hearing device’s receiver. The shown AEPs were measured for electrode Rc3 referenced to Rc6 (see Fig. 1). Clearly apparent is the negative deflection (N1) around 150 ms after stimulus onset (for T1). Note that the sound onset of the *Speech* is later (~200 ms) than in the noise conditions and that the N1 is shifted accordingly. Also apparent is the amplitude reduction of the N1 for T2 and T3 relative to T1 for all conditions. Likewise, all stimulus types evoked a negative deflection prior to stimulus onset with a latency that matches the onset of the idle noise when the device was first switched on. When comparing the identical and deviant last tones (T4) we observed a condition difference with a larger N1 amplitude followed by a larger P3 amplitude (at around 2700 ms) for the deviant stimuli. This difference was most pronounced for *Speech*, but was also observed for the other stimuli. Importantly, the peak latency of the N1 did not fit to the onsets of the stimuli, but matched the last switch (Off/On mark) of the hearing device filter. The explanation is most probably that the subjects perceived the difference in hearing device filter setting already in the idle noise.

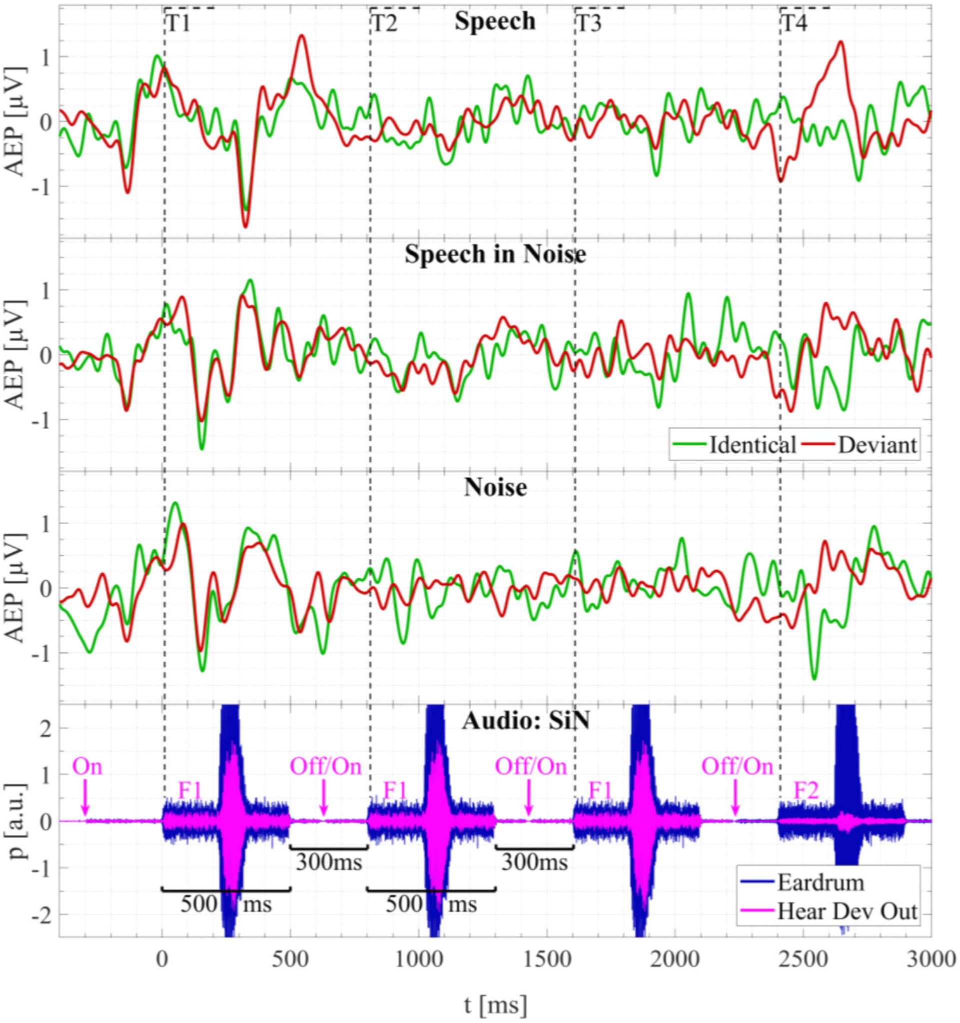


Fig. 1: AEPs averaged over subjects for all stimuli individually, and the recording of the *Speech-In-Noise* stimulus made in a dummy head.

Figure 4 shows the AEP amplitudes relative to the idle noise onset prior to T4 averaged over the identical and deviant stimulus types, respectively. A clear difference was observed in the average AEP, where a N1 and P3 was identified for the deviant, but not in the identical condition. The N1 amplitudes were averaged for the time window between 142 and 182 ms, and the P3 in the time window between 270 and 470 ms. A significant difference between conditions was evident for N1 ($p = 0.0046$) and P3 amplitudes ($p = 0.0078$).

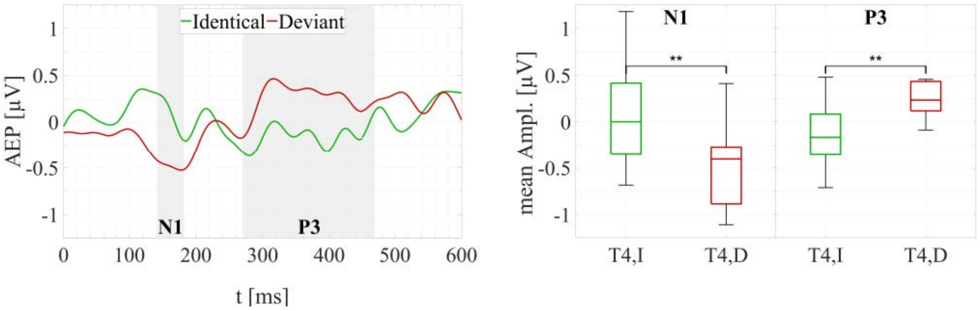


Fig. 4: Left: AEP on device onset prior to T4 (=0ms), averaged over all identical and deviant stimuli, respectively. Shaded areas indicate the time ranges where the average amplitudes for the N1 and P3 were obtained. Right: Boxplot of the N1 and P3 amplitudes for identical (I) and different (D) T4. Whiskers indicate the whole data range, boxes the 25% to 75% quantiles and the median.

SUMMARY AND CONCLUSION

We demonstrated a successful integration of ear-EEG acquisition with live hearing device processing. Using ear-centred electrodes, AEPs could be measured while the hearing device inserted into the same ear was active. This result, along with extensive pilot studies not reported here, demonstrate that potential practical obstacles, such as electro-magnetic crosstalk between audio transducers and EEG electrodes that stand in the way of integrating ear-EEG and hearing devices can be overcome.

It was possible to verify perceived differences in the hearing device processing with AEP differences. The timing of the AEPs with respect to the audio signals revealed that the participants were able to detect the change in filter settings already based on the idle noise of the hearing device. Despite this unforeseen effect we could show that the ear-centred EEG electrode placement in combination with a wireless EEG amplifier and a hearing device, provides conclusive information about auditory perception in this context. Furthermore, the EEG analysis provided additional insight in the perception process that was not apparent from the psychoacoustic results and clearly demonstrates that special considerations are necessary when studying AEPs to stimulation with a live hearing device. Future work will include further evaluation of the current dataset, particularly a quantification of the importance of electrode positioning and the evaluation of single-subject and single-trial data.

ACKNOWLEDGEMENTS

This work was supported by the Cluster of Excellence *Hearing4all* and research unit FOR1732 *Individualized Hearing Acoustics,* both funded by the DFG.

REFERENCES

Bleichner, M.G., Lundbeck, M., Selisky, M., Minow, F., *et al.* (**2015**). "Exploring miniaturized EEG electrodes for brain-computer interfaces. An EEG you do not see?" Physiol. Rep., **3**, e12362. doi: 10.14814/phy2.12362

Bleichner, M.G., Mirkovic, B., and Debener, S. (**2016**). "Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison," J. Neural Eng., **13**, 066004. doi: 10.1088/1741-2560/13/6/066004.

Bleichner, M.G., and Debener, S. (**2017**). "Concealed, unobtrusive ear-centered EEG acquisition: cEEGrids for transparent EEG," Front. Hum. Neurosci., **11**, doi: 10.3389/fnhum.2017.00163

Debener, S., Emkes, R., De Vos, M., and Bleichner, M. (**2015**). "Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear," Sci. Rep., **5**, 16743. doi: 10.1038/srep16743.

Delorme, A., and Makeig, S. (**2004**). "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis", J Neurosci. Methods **134**, 9-21, doi: 10.1016/j.jneumeth.2003.10.009

Denk, F., Hiipakka, M., Kollmeier, B., and Ernst, S.M.A. (**2017**). "An individualised acoustically transparent earpiece for hearing devices," Int. J. Audiol., 1-9. doi:10.1080/14992027.2017.1294768.

Doclo, S., Kellermann, W., Makino, S., and Nordholm, S.E., (**2015**). "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," IEEE Sig. Proc. Mag., **32**, 18-30. doi: 10.1109/MSP.2014.2366780

Grimm, G., Herzke, T., Berg, D., and Hohmann, V. (**2006**). "The master hearing aid: a PC-based platform for algorithm development and evaluation," Acta Acust. United Ac., **92**, 618-628.

Kothe, C., Schwartz Centre for Computational Neuroscience (**2015**). "Lab Streaming Layer (LSL)". Available at https://github.com/sccn/labstreaminglayer.

Kidmose, P., Looney, D., and Mandic, P. (**2012**). "Auditory evoked responses from ear-EEG recordings," Proc. IEEE EMBS, San Diego, USA, 586-589. doi: 10.1109/EMBC.2012.6345999.

Looney, D., Park, C., Kidmose, P., Rank, M.L., Ungstrup, M., Rosenkranz, K., and Mandic, D.P. (**2011**). "An in-the-ear platform for recording electroencephalogram," Proc. IEEE EMBS, Boston, USA, 6882-6885. doi: 10.1109/IEMBS.2011.6091733

Meyer, B.T., Jürgens, T., Wesker, T., Brand, T., and Kollmeier, B. (**2010**). "Human phoneme recognition depending on speech-intrinsic variability," J Acoust. Soc. Am., **128**, 3126-3141. doi: 10.1121/1.3493450

Mikkelsen, K.B., Kappel, S.L., Mandic, D.P., and Kidmose, P. (**2015**). "EEG recorded from the ear: Characterizing the Ear-EEG Method," Front. Neurosci., **9**, doi: 10.3389/fnins.2015.00438

Mirkovic, B., Bleichner, M.G., De Vos, M., and Debener, S., (**2016**). "Target speaker detection with concealed EEG around the ear," Front. Neurosci., **10**, doi: 10.3389/fnins.2016.00349

O'Sullivan, J.A., Power, A.J., Mesgarani, N., Rajaram, S., *et al.* (**2015**). "Attentional selection in a cocktail party environment can be decoded from single-trial EEG," Cereb. Cortex, **25**, 1697-1706. doi: 10.1093/cercor/bht355

Optimizing the microphone array size for a virtual artificial head

Mina Fallahi[1,*], Matthias Blau[1], Martin Hansen[1], Simon Doclo[2], Steven van de Par[2], and Dirk Püschel[3]

[1] *Institut für Hörtechnik und Audiologie, Jade Hochschule Oldenburg, Oldenburg, Germany*

[2] *Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany*

[3] *Akustik Technologie Göttingen, Göttingen, Germany*

As an alternative to traditional artificial heads, individual head-related transfer functions (HRTFs) can be synthesized with a virtual artificial head (VAH) consisting of a multi-microphone array in combination with filter-and-sum signal processing. The accuracy of the synthesis depends, amongst others, on the number of microphones in the array and on its topology (array size and microphone positions). In this study the effect of microphone array size on the synthesis accuracy was investigated. Five simulated microphone arrays of different sizes were used to synthesize individual HRTFs in the horizontal plane. Objective results in terms of spectral distortion and ILD deviation as well as subjective results with 10 participants showed that array size has a major effect and that the synthesis accuracy can be improved by carefully choosing an appropriate array size.

INTRODUCTION

An established method to include the spatial properties of the sound within a binaural reproduction is the use of so-called artificial heads. With anthropometric characteristics of an average human head and torso, an artificial head preserves the spatial information in the sound field, which is crucial for sound source localization. However, the non-individual anthropometric geometries of these artificial heads often lead to perceptible deficiencies. Alternatively, a microphone array can be used as a filter-and-sum beamformer to synthesize individual head-related transfer functions (HRTFs). The major advantage of this system, referred to as a virtual artificial head (VAH), is the possibility to modify the individual calculated filter coefficients and to process the same recording post-hoc for individual HRTFs, both statically as well as dynamically (using head tracking). Another potential benefit of the VAH is its flexibility due to the smaller size and weight. One decisive aspect for the accuracy of the VAH is the choice of microphone array topology (including its size and microphone positions). Rasumow *et al.* (2016) developed a VAH as a planar

[*] Corresponding author: mina.fallahi@jade-hs.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

microphone array consisting of 24 microphones and showed that a regularized least-squares cost function approach could be used to synthesize individual binaural HRTFs accurately in the horizontal plane (c.f., also, Rasumow *et al.*, 2011, 2017). In accordance with this approach and with some modifications for increasing the spatial resolution of the VAH (c.f. Fallahi *et al.*, 2017)), the present study investigated the effect of the microphone array size on the accuracy of the VAH. Five microphone arrays of different sizes were simulated and used to synthesize individual HRTFs. After a brief review of the applied methods, the objective and perceptual evaluation of synthesis with different arrays sizes will be discussed.

HIGH SPATIAL RESOLUTION LEAST-SQUARES FILTER-AND-SUM BEAMFORMER

The desired directivity pattern $D(f,\Theta_k)$ of, e.g., an individual HRTF at either the left or right ear as a function of frequency f and direction Θ_k can be synthesized with the VAH as a filter and sum beamformer. Considering the $N \times 1$ steering vector $\mathbf{d}(f,\Theta_k)$ which describes the frequency and direction dependent transfer function between the source at direction Θ_k and the N microphones of the microphone array, the synthesized directivity pattern $H(f,\Theta_k)$ of the VAH is defined as

$$H(f,\Theta) = \mathbf{w}^H(f)\mathbf{d}(f,\Theta) \qquad \text{(Eq. 1)}$$

The $N \times 1$ vector $\mathbf{w}(f) = [w_1(f), w_2(f), ..., w_N(f)]^T$ contains the complex-valued filter coefficients for each microphone which can be derived by minimizing a narrowband least-squares cost function J_{LS}, defined as the sum over P directions of the squared absolute differences between the synthesized and desired directivity pattern, i.e.,

$$J_{LS}(\mathbf{w}(f)) = \sum_{k=1}^{P} |H(f,\Theta_k) - D(f,\Theta_k)|^2 \qquad \text{(Eq. 2)}$$

In order to increase the robustness of the system against deviations in microphone positions or characteristics (c.f. Rasumow *et al.*, 2011; Doclo *et al.*, 2003), Rasumow *et al.* (2016) derived a closed form solution for the minimization of the least-squares cost function subject to some regularization constraints, however at the cost of a general loss of accuracy. With the aim of maintaining the accuracy of synthesis for a high number of directions Fallahi *et al.* (2017) suggested a constrained optimization approach, minimizing the least-squares cost function for directions Θ_k , $k = 1,2,...,P$ subject to constraints set on the mean white noise gain ($\mathrm{WNG_m}$, Rasumow *et al.*, 2016) and spectral distortion (*SD*) at synthesis directions θ_k , $k = 1,2,...,p$ by setting an upper and lower limit, L_{up} and L_{low}, for the synthesis error at each one of these directions, i.e., for all k

$$L_{\mathrm{low}} \leq SD(f,\theta_k) = 10\lg \frac{|\mathbf{w}^H(f)\mathbf{d}(f,\theta_k)|^2}{|D(f,\theta_k)|^2} \mathrm{dB} \leq L_{\mathrm{up}} \qquad \text{(Eq. 3)}$$

The minimization of J_{LS} subject to inequality constraints described above was solved by applying the interior-point method, using the results of the closed form solutions by Rasumow *et al.* (2016) as the initial values for this iterative optimization algorithm.

INFLUENCE OF MICROPHONE ARRAY SIZE ON THE VAH SYNTHESES

The main goal of the current study was to investigate the effect of the array size on the synthesis accuracy. Adopting the topology of the microphone array shown in Fig. 1 (20 cm×20 cm planar array with 24 microphones, c.f. Rasumow *et al.*, 2011), the original array as well as downsized copies of it, namely 50%-, 37.5%- and 25%-size arrays were simulated. In addition, a combination of the 50%-size and 25%-size arrays was considered as well (named as ‘Mix.’ in the following), by taking the positions of the eight outermost microphones of the 50%-size array and the innermost positions of the 25%-size array for the rest. Using the constrained optimization approach described above, a set of individually measured horizontal HRTFs with 5° azimuthal resolution were synthesized with these arrays. L_{low} and L_{up} were set to -1.5 dB and 0.5 dB respectively, leading to a maximum range of interaural level difference (ILD) deviation of 2 dB at each of the 72 synthesized directions.

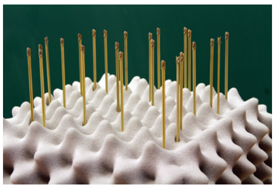


Fig. 1: Virtual artificial head: planar microphone array with 24 microphones (Rasumow *et al.*, 2016).

The results for spectral deviation at the left ear and the resulting ILD deviation are shown in Fig. 2. As can be seen, the constraints set on spectral error could not be met for all directions, especially not at high frequencies. For a given microphone array with a fixed size this could be due to more spatial details contained in the HRTF directivity patterns at higher frequencies as well as aliasing effects. For a smaller array, although the complexity of the HRTF’s directivity pattern at the contralateral side (e.g., 270° for the left ear) could still be a challenge, the aliasing effects for the ipsilateral side could be shifted to higher frequencies, as can be seen in the simulation results of smaller arrays at higher frequencies in comparison to larger arrays (Fig. 2, left). At the same time, however, a reduced array size corresponds to an increased wavelength relative to the array size which leads to the widening of the synthesized directivity pattern (c.f. Ward *et al.*, 2001). At low frequencies this might not introduce a major problem since the HRTFs have a mostly omnidirectional directivity pattern. But in the mid-frequency range of about 1 to 4 kHz, the directivity pattern starts to get more complicated while the wavelength might still be large enough to prevent the beamformer from reaching sufficient damping at the contralateral side, leading to the increased spectral distortions and ILD deviations of smaller arrays in the mid-frequency range. The question now arises whether the resulting spectral distortions are perceptually relevant and which array size should be preferred.

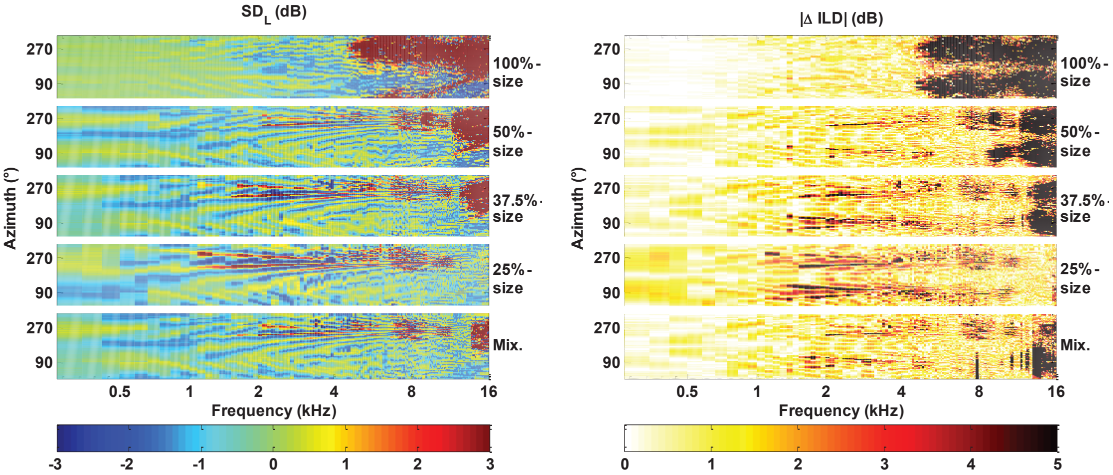


Fig. 2: Left: Spectral distortion (left ear). Right: ILD deviation between original and synthesized HRTFs with different array sizes.

EXPERIMENTAL PROCEDURE

In order to evaluate the perceptual quality of different array sizes, a listening experiment was performed. Prior to the listening test, individual horizontal HRTFs of 10 subjects were acquired with a 5° azimuthal resolution in a non-anechoic room (c.f. Koehler *et al.*, 2014), followed by individual measurements of the headphone transfer functions (HPTFs). The measured HRTFs were then smoothed both in frequency and spatial domain (c.f. Rasumow *et al.*, 2014) before being synthesized with different simulated microphone arrays. Three short bursts of pink noise with a spectral content of $300\text{Hz} \leq f \leq 16000\text{Hz}$ were chosen as the test signal. Each noise burst lasted $\frac{1}{3}$ s with 1-ms onset-offset ramps followed by $\frac{1}{6}$ s of silence. This test signal was then convolved with the individually measured and synthesized HRTFs and filtered individually with the inverse individual HPTF and presented via headphones.

Ten subjects, five of them with extensive experience with binaural psychoacoustical experiments, participated in the experiment. Participants were instructed to rate the binaural signals generated with synthesized HRTFs of different array sizes with respect to the reference (binaural reproduction with original HRTFs). Subjects performed the ratings in three separate sessions for three different aspects: spectral coloration, localization, and overall performance, giving their ratings on a continuous scale between bad, poor, fair, good, and excellent. In order to limit the total number of repetitions of the experiment to a feasible amount, five directions for the virtual source

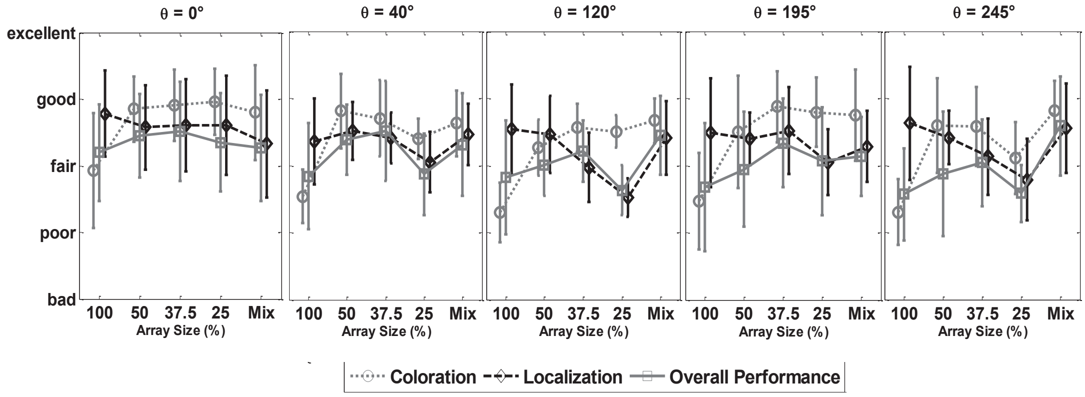


Fig. 3: Mean evaluations over all 10 subjects regarding different array sizes (x-axis) for different aspects and source directions.

were chosen, including two cases with the largest and two cases with small variations in the resulting ILD deviation caused by different array sizes (θ = 120°, 245° and θ = 40°, 195°), and the frontal direction (θ = 0°). Each direction appeared three times for each given aspect. The directions were presented in a randomized order.

PERCEPTUAL EVALUATION – RESULTS AND DISCUSSION

The results of the perceptual evaluations over all subjects with regard to different aspects and source direction are shown in Fig. 3 as mean and standard deviation across subjects. As can be seen, the results vary not only with different array sizes and source directions but also with subjects which can be due to different internal scales used by subjects or individual differences in the HRTFs.

According to the perceptual results, the mean evaluations of spectral coloration improved generally for a smaller array. This effect could be noticed at all of the five tested directions. Objective results (see Fig. 2 for one participant as an example) had shown that for a reduced-size array spectral distortions at the contralateral side start to get prominent in the mid-frequency range of 1 to 4 kHz, whereas at this frequency range the spectral distortion for the largest array (100%-size array) remained within the allowed range (defined in Eq. 3), however increased drastically above ca. 4 kHz. Considering the two frequency ranges $1\,\text{kHz} \leq f \leq 4\,\text{kHz}$ and $4\,\text{kHz} \leq f \leq 16\,\text{kHz}$, the ratings for spectral coloration of all participants vs. the absolute spectral distortion averaged across frequency for these two frequency ranges are shown in Fig. 4a at θ = 245° as an example. It seems that the increased spectral distortion of the smaller arrays at mid-frequency range did not influence the ratings on spectral coloration (Fig. 4a, top). Moreover, the prominent increased spectral distortion of the largest array at frequencies above 4 kHz coincided with the generally reduced coloration ratings for this array (Fig. 4a, bottom).

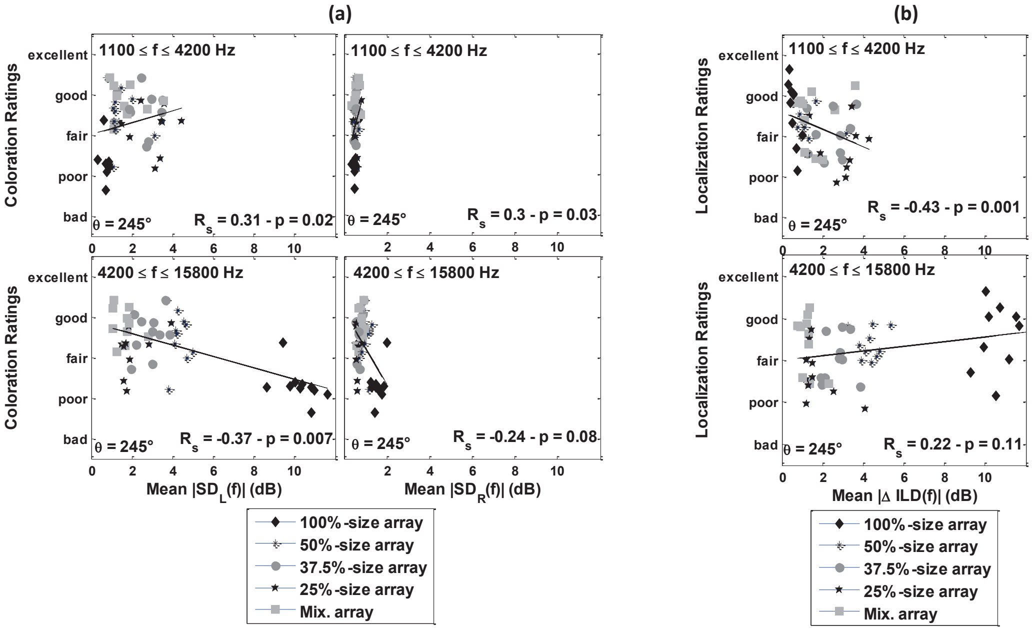


Fig. 4: (a) Evaluations on spectral coloration vs. mean spectral distortion at the left and right ears. (b) Evaluations on localization vs. mean ILD deviations. $\theta = 245°$, $1 \leq f \leq 4$ kHz (upper row) and $4 \leq f \leq 16$ kHz (lower row). Mean $|.(f)|$ = average over the depicted frequency range.

Smaller array sizes led in average to decreased localization ratings especially at $\theta = 120°$ and $245°$. The increased ILD deviation in the mid-frequency range was apparently more relevant for the decline in the ratings for smaller arrays (Fig. 4b, top), whereas the effect of ILD deviations in the frequency range $4\,\text{kHz} \leq f \leq 16\,\text{kHz}$ seemed not to be as important for the localization (Fig. 4b, bottom).

The mean evaluations on overall performance lay almost for all cases at the lower edge of coloration and localization ratings (Fig. 3). This could indicate that the overall perception of the synthesis depended both on coloration as well as on localization cues. In other words, the accuracy of synthesis should be preserved both for spectral coloration and localization cues. The ratings on overall performance show a general increase towards arrays with the middle-range size (37.5%-size array or the Mix-array) confirming the compromise between localization artifacts of smaller arrays and coloration artifacts of larger arrays.

In order to analyse whether at least one of the microphone arrays for a fixed direction and perceptual aspect led to significantly different evaluations the Friedman test was applied. The p-values for the given aspect and source direction are listed in Table 1. Considering the Bonferroni correction for the 3 repetitions of each direction, the p-values of conditions indicating a significant effect of the array size ($p \leq \frac{0.05}{3}$) are

	$\theta = 0°$	$\theta = 40°$	$\theta = 120°$	$\theta = 195°$	$\theta = 245°$
Spectral Coloration	**0.0068**	**0.0113**	**0.0003**	**0.0029**	**0.0002**
Localization	0.4435	0.0192	**0.0025**	**0.0146**	**0.0001**
Overall Performance	0.06917	0.0976	**0.0011**	0.1108	**0.0157**

Table 1: p-values according to Friedman test for different source directions and perceptual aspects. p-values indicating a significant effect of array size ($p \leq \frac{0.05}{3} = 0.0167$) are depicted as bold numbers.

shown as bold numbers. According to the results, array size seemed to have a significant effect on the coloration ratings at all of the five considered directions due to the difference in ratings for the 100%-size array compared to the other arrays. Different array sizes seemed to affect the evaluations regarding localization mostly at θ = 120°, 195°, and 245°. The effect was significant due to decreased ratings given to the 25%-size array. The effect of array size on the evaluation of overall performance was significant at θ = 120° and 245°, due to either different ratings given to the 100%-size array or 25%-size array, compared to the other arrays. This confirms that participants chose the spectral and localization cues differently as the critical cue for giving overall ratings.

CONCLUSION

In this study the effect of microphone array size on the accuracy of HRTF synthesis with a virtual artificial head was investigated. Simulation results for five different array sizes (planar arrays with 24 microphones, approximately quadratic with side lengths ranging from 20 cm to 5 cm) indicated that there are noticeable differences in the resulting monaural and binaural features (spectral distortion and ILD deviation) between original and synthesized HRTFs for different array sizes. While spectral distortions especially at the ipsilateral side could be shifted to higher frequencies by choosing a smaller array, spectral distortion increased in the mid-frequency range of $1\,\text{kHz} \leq f \leq 4\,\text{kHz}$ at the contralateral side for smaller arrays, leading to increased ILD deviations at these frequencies. Furthermore, experimental results showed that the array size had a significant effect on the perceived spectral coloration and source localization. In particular, large spectral distortions introduced by the largest array at frequencies above 4 kHz affected the perceived spectral coloration. Contralateral spectral distortions at the mid-frequency range appearing for smaller arrays did not affect the perceived coloration, but led to decreased localization ratings. These ratings presumably resulted from ILD deviations at these frequencies. The overall evaluation of different arrays sizes confirmed the importance of accuracy both with respect to spectral and localization cues. In general, the overall ratings were the highest for microphone arrays of mid-range size (37.5%-size array or the 'Mix' combination of 50%- and 25%-size arrays) since the deficiencies of larger and smaller arrays could be balanced for these array sizes. A further investigation should analyze the interaural phase differences resulting from different array sizes, both regarding their perceptual relevance and also regarding an effective incorporation of phase contraints for the minimization of the cost function J_{LS}.

ACKNOWLEDGMENTS

This project was funded by Bundesministerium für Bildung und Forschung under grant no. 03FH021IX5.

REFERENCES

Doclo, S., and Moonen, M. (**2003**). "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics," IEEE T. Signal Proces., **51**, 2511-2526. doi: 10.1109/TSP.2003.816885

Fallahi, M., Hansen, M., Doclo, S., van de Par, S., Mellert, V., Püschel, D., and Blau, M. (**2017**). "High spatial resolution binaural sound reproduction using a virtual artificial head," Fortschritte der Akustik, DAGA 2017, Kiel, Germany, pp. 1061-1064.

Köhler, S., Blau, M., van de Par, S., and Rasumow, E. (**2014**). "Simultane Messung mehrerer HRTFs in nicht reflexionsarmer Umgebung," Fortschritte der Akustik, DAGA 2014, Oldenburg, Germany, pp. 202-203.

Rasumow, E., Blau, M., Hansen, M., Doclo, S., van de Par, S. Mellert, V., and Püschel, D. (**2011**). "Robustness of virtual artificial head topologies with respect to microphone positioning errors," Proc. Forum Acusticum, Aalborg, pp. 2251-2256.

Rasumow, E., Blau, M., Hansen, M., van de Par, S., Doclo, S., Mellert, V., and Püschel, D. (**2014**). "Smoothing individual head-related transfer functions in the frequency and spatial domains," J. Acoust. Soc. Am., **135**, 2012-2025. doi: 10.1121/1.4867372

Rasumow, E., Hansen, M., van de Par, S., Püschel, D., Mellert, V., Doclo, S., and Blau, M. (**2016**). "Regularization approaches for synthesizing HRTF directivity patterns," IEEE T. Audio Speech, **24**, 215-225. doi: 10.1109/TASLP.2015.2504874

Rasumow, E., Blau, M., Doclo, S., van de Par, S., Hansen, M., Püschel, D., and Mellert, V.(**2017**). "Perceptual evaluation of individualized binaural reproduction using a virtual artificial head," J. Audio Eng. Soc., **65**, 448-459. doi: 10.17743/jaes.2017.0012

Ward, D.B., Kennedy, R.A., and Williamson, R.C. (**2001**). "Constant directivity beamforming," in *Microphone Arrays, Signal Processing Techniques and Applications*, Eds. M. Brandstein and D. Ward (Berlin Heidelberg: Springer-Verlag), pp. 3-18.

Evaluation of noise reduction in digital hearing aids in situations with multiple signal sources

MARLITT FRENZ[1,*], ALFRED MERTINS[2], AND HENDRIK HUSSTEDT[1]

[1] *German Institute of Hearing Aids, Lübeck, Germany*

[2] *Institute for Signal Processing, Universität zu Lübeck, Lübeck, Germany*

Features of modern hearing aids such as the digital noise reduction and directional microphones can enhance speech. One way to evaluate the effect of these features is to measure the increase of the signal-to-noise ratio (SNR). To this end, the method of Hagerman and Olofsson is often used. However, only two signals can be distinguished with this method, e.g., speech and noise. Since many realistic situations include more than two signals, an extension of the method of Hagerman and Olofsson for an arbitrary number of signals is introduced. To proof the concept, this extended method is applied to a setup with 9 different signals presented by 8 speakers. This study considers a separation of speech and noise for 8 signal sources. All speakers are positioned around a hearing aid on a circle with a radial distance of $r = 1\,\mathrm{m}$ and an angular distance of 45° between 0° and 360°. Speech is presented from 0° and noise from all 8 directions $(0^\circ, 45^\circ, ..., 360^\circ)$. With this setup, a state-of-the-art hearing aid is analysed for different settings where the digital noise reduction and/or the directional microphones are turned on or off. As a result, the SNR for all directions can be investigated individually. This demonstrates the practicability of the extended method.

INTRODUCTION

Speech intelligibility in noisy situations decreases depending on the characteristics of speech and noise such as the frequency spectrum and the signal-to-noise-ratio (SNR). As a result, the listening effort for hearing impaired people increases, and leads to a faster exhaustion than for normal hearing people (Holube *et al.*, 2005).

To increase the SNR, the distance between speaker and listener can be reduced or the listening environment can be changed by the listener. Also using hearing aids can lead to a higher speech intelligibility within the same environment. Hearing aids with adaptive features such as digital noise reduction and microphone directionality enhance the SNR. Microphone directionality enhances the spatial SNR and digital noise reduction analyses the signal temporally or spectrally (Chung, 2004). This leads to an increase of speech intelligibility (Bentler, 2005; Brons *et al.*, 2014).

To objectively evaluate speech enhancement in hearing aids, there exists several approaches such as computing the modulation transfer function (Holube *et al.*, 2005),

[*] Corresponding author: m.frenz@dhi-online.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

performing a percentile analysis (Harries, 2010) or using the method of Hagerman & Olofsson (Hagerman and Olofsson, 2004).

Among these three examples, the method of Hagerman & Olofsson can be used to separate the speech and the noise signal to directly calculate the SNR. However, two signals can be separated only.

To investigate the speech enhancement of hearing aids in a more realistic listening environment, more than two signals from various directions should be considered. This study introduces an extended version of the method of Hagerman and Olofsson. With this version, an arbitrary number of N signals can be considered, Therefore, a maximum number of N directions can be evaluated.

To proof the concept, this extended method is applied to a setup with 9 different signals presented by 8 speakers. Speech is presented from 0°, and noise from all 8 directions $(0^\circ, 45^\circ, ..., 315^\circ)$. With this setup, a state-of-the-art hearing aid is analysed for different settings where the digital noise reduction and/or the directional microphones are turned on or off.

This paper is organised as follows: First, the method of Hagerman & Olofsson is shortly introduced, and then the extension is presented. Next, the measurement setup is explained, and the results are discussed. Finally, the results are summarized and a conclusion is given.

HAGERMAN & OLOFSSON METHOD

Two superpositions $a_{\text{in},1}(t), a_{\text{in},2}(t)$ of two signals $v_1(t), v_2(t)$ are used, in which one superposition takes a phase delay of 180° for $v_2(t)$ into account, where

$$a_{\text{in},1}(t) = v_1(t) + v_2(t), \qquad \text{(Eq. 1)}$$

$$a_{\text{in},2}(t) = v_1(t) - v_2(t). \qquad \text{(Eq. 2)}$$

Both superpositions are presented to a system, e.g., a hearing aid, successively. $a_{\text{in},1}(t)$ and $a_{\text{in},2}(t)$ are modified by the system, Therefore, two output signals $a_{\text{out},1}(t), a_{\text{out},2}(t)$ are produced.

$v'_1(t)$ and $v'_2(t)$ are calculated out of the output signals with

$$v'_1(t) = \frac{1}{2}(a_{\text{out},1}(t) + a_{\text{out},2}(t)), \qquad \text{(Eq. 3)}$$

$$v'_2(t) = \frac{1}{2}(a_{\text{out},1}(t) - a_{\text{out},2}(t)). \qquad \text{(Eq. 4)}$$

$v'_1(t), v'_2(t)$ are the input signals processed by the hearing aid. Therefore, they are similar to $v_1(t), v_2(t)$.

Ricketts (2000) showed that traditional test environments with a single noise source located at 180° azimuth cannot be used to accurately predict directional benefit in comparison to other tests with more than one noise sources or real-world environments. Therefore, it is necessary to include more than two competing sound sources, e.g., one speech source and more than one noise sources.

EXTENDED METHOD

An arbitrary number of N signals is given, e.g., $v_1(t), v_2(t), ..., v_N(t)$. The number of input signals equals the number of measurement rounds. By that, N times a superposition of all signals is built. The phase of one signal is inverted within one input signal. Therefore, N input signals are built with

$$\mathbf{a}_{\text{in}}(t) = \begin{pmatrix} a_{\text{in},1}(t) \\ a_{\text{in},2}(t) \\ \vdots \\ a_{\text{in},N}(t) \end{pmatrix} = \begin{bmatrix} -1 & 1 & \dots & 1 \\ 1 & -1 & \dots & 1 \\ \vdots & 1 & \ddots & \vdots \\ 1 & \dots & 1 & -1 \end{bmatrix} \begin{pmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_N(t) \end{pmatrix} = \mathbf{A}\mathbf{v}(t). \qquad \text{(Eq. 5)}$$

A is the system matrix and of type NxN. Its i-th column changes the phase of signal $v_i(t)$ in one input signal and its j-th row defines the input signal $a_{\text{in},j}(t)$ at the hearing aid input. The system Matrix shall only change the phase and not weight signals. Therefore, its values are only -1 or 1 respectively.

The rank of **A** is N for all $N > 2$, Therefore, the inverse of **A** exists with

$$\mathbf{A}^{-1} = \begin{bmatrix} -\frac{N-3}{2(N-2)} & \frac{1}{2(N-2)} & \dots & \frac{1}{2(N-2)} \\ \frac{1}{2(N-2)} & -\frac{N-3}{2(N-2)} & \dots & \frac{1}{2(N-2)} \\ \vdots & \frac{1}{2(N-2)} & \ddots & \vdots \\ \frac{1}{2(N-2)} & \dots & \frac{1}{2(N-2)} & -\frac{N-3}{2(N-2)} \end{bmatrix} \qquad \text{(Eq. 6)}$$

The condition of the system matrix **A** for all $N > 3$ is $\varphi(\mathbf{A}) = \frac{N-2}{2}$. Therefore, as the impact of measurement tolerances increases, the more signals are used for a setup.

At the output of the hearing aid, all signals can be reconstructed with

$$\begin{pmatrix} v'_1(t) \\ v'_2(t) \\ \vdots \\ v'_N(t) \end{pmatrix} = A^{-1} \begin{pmatrix} a_{\text{out},1}(t) \\ a_{\text{out},2}(t) \\ \vdots \\ a_{\text{out},N}(t) \end{pmatrix}. \qquad \text{(Eq. 7)}$$

With this extension, various complex listening situations with multiple noise and/or speech signals from multiple directions can be analysed. After the separation of all signals $(v_1(t), ..., v_N(t))$, it is possible to compute the absolute SNR between two or multiple signals. Moreover, the SNR at the output can be compared with the SNR at the input. A positive value indicates an enhancement of the desired signal, e.g., speech. In this way the benefit of hearing aid features can be analysed in complex listening situations.

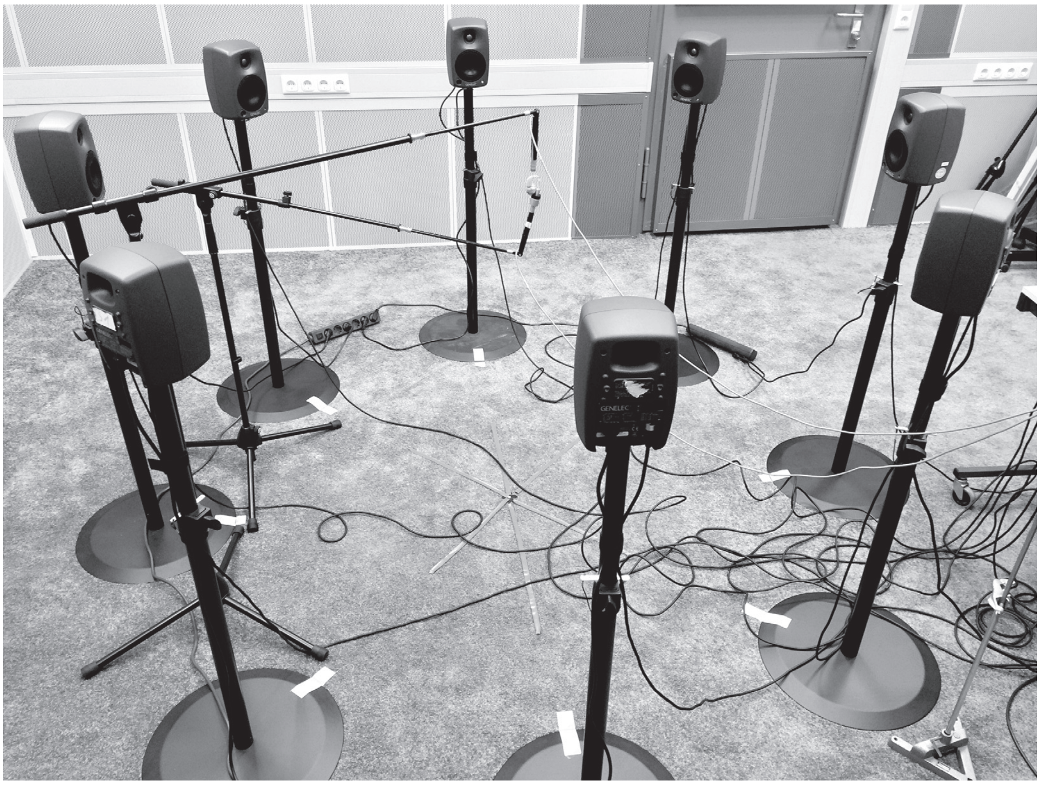


Fig. 1: The measurement setup consists of 8 loudspeakers, one reference microphone and one state-of-the-art hearing aid (BTE) connected to an ear simulator. The BTE is facing the loudspeaker in 0°. Not shown is the RME fireface 800 soundcard and the PC.

MEASUREMENT SETUP

The measurement setup consists of 8 speakers, which are equally distributed on a circle around the hearing aid. The radius of this circle is 1 m and the angular distance between the speakers is 45° (see Fig. 1). In this study, a speech signal, such as the International Speech Test Signal (ISTS), is presented from an angle of 0° and 8 different noise signals are presented from all 8 directions so that N equals 9. The noise signals are incoherent and built out of the ISTS so that the long term average spectrum is equal to the ISTS. The ISTS is presented with 65 dB SPL and an overall SNR of $+5$ dB is chosen. Thus, the individual level for each noise signal is 51 dB. The hardware of the setup consists of 8 GENELEC speakers of type 8020, a RME Fireface 800 soundcard, a Bruel & Kjaer (B&K) ear simulator according to IEC 60318-4, a reference microphone from B&K of type 4190, and a PC. All measurements as well as the signal analysis are performed with Matlab version 2017a.

The measurement signals $a_{\text{in},1}(t),...,a_{\text{in},9}(t)$ for each loudspeaker are set in a vector with $x(t) = (a_{\text{in},1}(t), 0_{5\,\text{s}}(t), a_{\text{in},2}(t),,...,0_{5\,\text{s}}(t), a_{\text{in},9}(t))^T$, Therefore, an output vector $y(t)$ is recorded. With the superposition of each input signal, a separation of the output signals is possible. With the inverse of the system matrix the signals $v'_1(t),...,v'_9(t)$ are reconstructed.

As transient effects take place within the first 15 seconds, the mean power is calculated in a time window from 15 s to 60 s. The SNR between the ISTS and each of the 8 noise signals is computed separately. Therefore, a spatial analysis of the SNR is possible.

To check the measurement setup, the extended method is applied to the signals of the reference microphone. The polarplot in Fig. 2 shows that the desired SNR of $+5\,\text{dB}$ is measured with an accuracy from $-0.4\,\text{dB}$ to $+1.3\,\text{dB}$. The SNR indicates random behavior independent to a hearing aid setting or the direction. Thus, the results indicate negligible modification of the given SNR for each measurement round.

For the measurements, a state-of-the-art behind the ear (BTE) hearing aid is used. The gain of the device is adjusted by simulating an auditory threshold of type N3 as defined in **?**. For this setup the maximum pressure output, the compression ratio as well as adaptive features such as feedback reduction, wind control are deactivated. As parameters, noise reduction and a fixed microphone directionality are investigated. Therefore, 4 test settings are evaluated:

1. noise reduction off & omnidirectional microphone settings,
2. noise reduction on & omnidirectional microphone settings,
3. noise reduction off & directional microphone settings, and
4. noise reduction on & directional microphone settings.

RESULTS AND DISCUSSION

The output SNR in relation to the input SNR is calculated for each sound source, so that the 4 hearing aid settings are evaluated independently. Figure 3 shows the results for all 4 hearing aid settings as a polar plot. The curves are linearly interpolated between the measurement points. A negative SNR shown in the polar plot indicates an improvement of speech intelligibility. The test setting with a deactivated noise reduction and an omnidirectional microphone setting indicates no modification of the SNR for all directions. This result indicates a good reproduction of the defined SNR for the proposed extended method. The result for an activated noise reduction and an omnidirectional microphone setting shows a negative SNR independent to the direction (see dashed and dotted line in Fig. 3). This is expected due to a working noise reduction in the temporal or frequency domain (Chung, 2004).

Furthermore, the test setting with a deactivated noise reduction and a directional microphone setting indicates a negative SNR dependent to the direction (see continuous

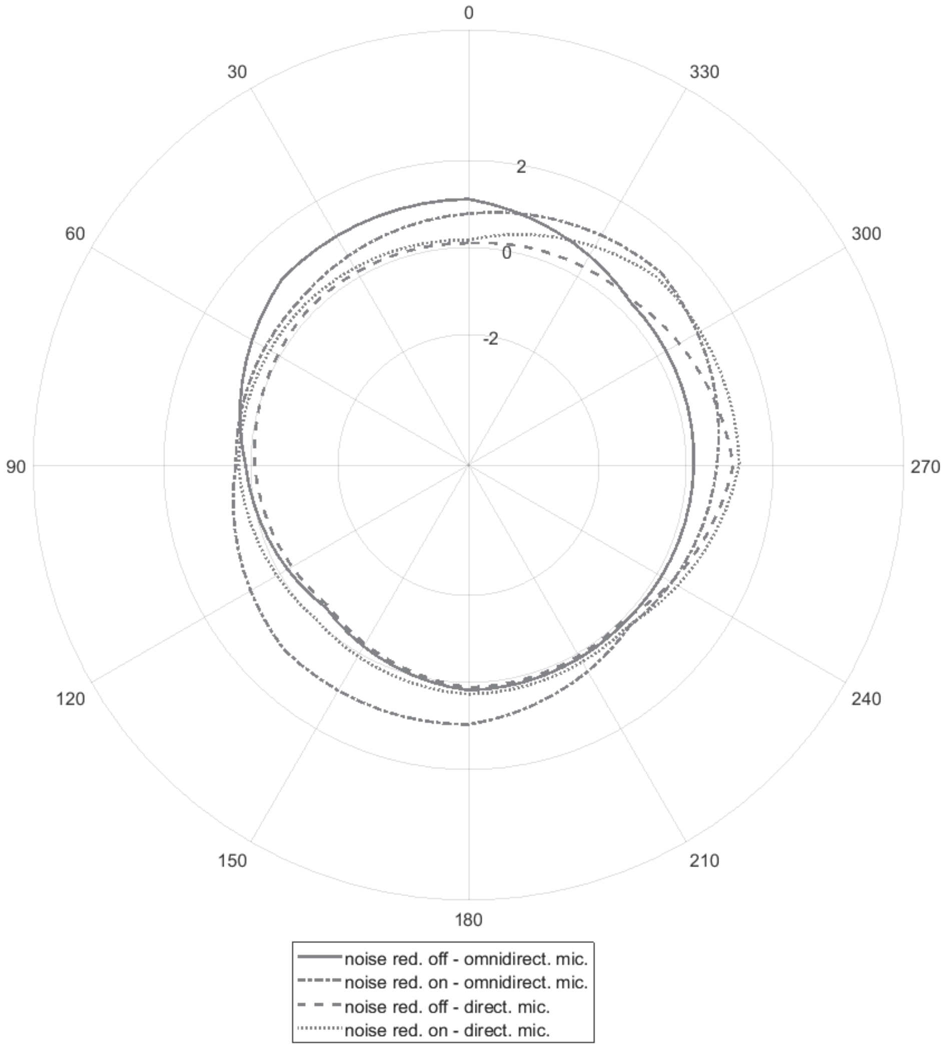


Fig. 2: Output SNR of the reference microphone in relation to the input SNR of the given signal. Each line represents the result of one of the four hearing aid settings. Its allocations are shown in the legend.

line). As the SNR is hardly modificated in 0° azimuth, the lowest SNR can be found in 180° azimuth. This indicates a subcardioid polar pattern, which can typically be found in hearing aids (Bentler, 2005; Chung, 2004).

In addition, an activated noise reduction as well as a directional microphone characteristic show a maximum amount of SNR reduction in 180°. These findings also support Ricketts, who stated in 2000, that a spatial analysis of features should be investigated with more than two sound sources (Ricketts, 2000).

The results in Fig. 3 indicate the proof of the concept of the extended method.

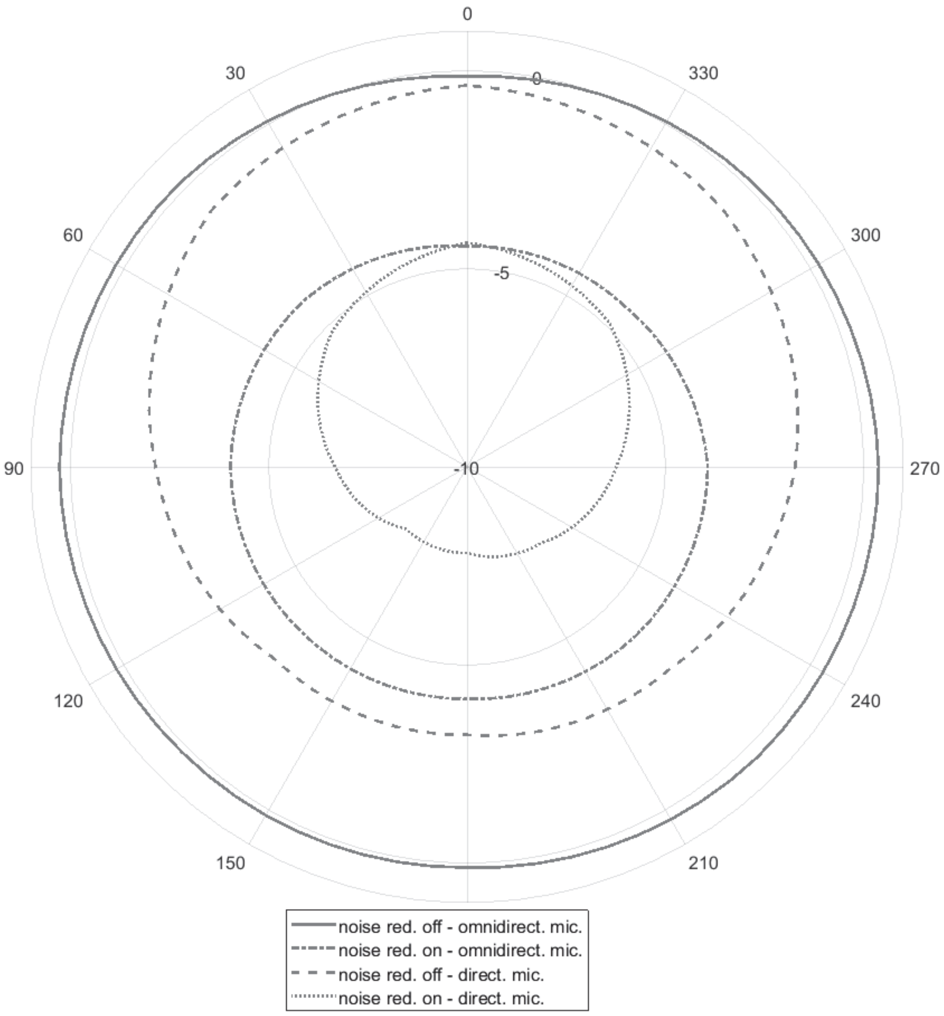


Fig. 3: Output SNR of the hearing aid in relation to the input SNR measured at the reference microphone. The directional microphones of the hearing aid are orientated in 0° direction. Each line represents the result of one of the four hearing aid settings. Its allocations are shown in the legend.

CONCLUSIONS

This paper presents an extended method of the concept presented by Hagerman & Olofsson in 2004. Hagerman and Olofsson introduced a method in which two signals are superpositioned at the input of system and can be reconstructed at the output of a system. As only two signals can be distinguished in this concept, a maximum number of two sound sources can be evaluated. This paper presents an extended method, in which an arbitrary number of N signals can be distinguished. Within the concept, a system matrix is introduced, which describes the phase of each signal for every superpositioned input signal. An inverse of the system matrix can be calculated for more than two signals. Therefore, the system matrix is used to reconstruct the signals at the output of the system. A measurement setup was designed to proof the concept. The results show an enhancement of the SNR independent to the direction for an activated noise reduction and an omnidirectional microphone setting. Also an enhancement of the SNR for a fixed microphone directionality dependent on the direction is measured. A maximum amount of SNR enhancement can be found in 180° for a test setting with an activated noise reduction and a fixed microphone directionality. These test results demonstrate the practicability of the extension of the method by Hagerman and Olofsson.

REFERENCES

Bentler, R.A. (**2005**) "Effectiveness of directional microphones and noise reduction schemes in hearing aids: a systematic review of the evidence," J. Am. Acad. Audiol., **16**, 473-484. doi: 10.1177/1084713806289514

Brons, I., Houben, R., and Dreschler, W.A. (**2014**) "Effects of noise reduction on speech intelligibility, perceived listening effort, and personal preference in hearing-impaired listeners," Trends Amplif., **18**, 2331-2165. doi: 10.1177/2331216514553924

Chung, K. (**2004**) "Challenges and recent developments in hearing aids: Part I. Speech understanding in noise, microphone technologies and noise reduction algorithms," Trends Amplif., **8**, 83-124. doi: 10.1177/108471380400800302

Hagerman, B., and Olofsson, A. (**2004**). "A method to measure the effect of noise reduction algorithms using simultaneous speech and noise," Acta Acust. United Ac., **90**, 356-361.

Harries, T. (**2010**). "Untersuchung der Perzentilanalyse und Messungen von Funktionselementen nicht linearer Hörgeräte mittels dieser Methode," Undergraduate thesis.

Holube, I., Fredelake, S., and Hansen, M. (**2005**). "Subjective and objective evaluation methods of complex hearing aids," Proceedings of the 8th International Congress on Audiology, Heidelberg, Germany.

Ricketts, T. (**2000**) "Impact of noise source configuration on directional hearing aid benefit and performance", Ear Hearing, **21**, 194-205.

Effects of slow- and fast-acting compression on hearing-impaired listeners' consonant-vowel identification in interrupted noise

BORYS KOWALEWSKI,[1,*] JOHANNES ZAAR[1], MICHAL FERECZKOWSKI[1],
EWEN N. MACDONALD[1], OLAF STRELCYK[2], TOBIAS MAY[1], AND TORSTEN DAU[1]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Sonova U.S. Corporate Services, Warrenville, IL, USA*

There is conflicting evidence about the relative benefit of slow- and fast-acting compression for speech intelligibility. It has been hypothesized that fast-acting compression improves audibility at low signal-to-noise ratios (SNRs) but may distort the speech envelope at higher SNRs. The present study investigated the effects of compression with nearly instantaneous attack time but either fast (10 ms) or slow (500 ms) release times on consonant identification in hearing-impaired listeners. Consonant-vowel speech tokens were presented at several presentation levels in two conditions: in the presence of interrupted noise and in quiet (with the compressor "shadow-controlled" by the corresponding mixture of speech and noise). These conditions were chosen to disentangle the effects of consonant audibility and noise-induced forward masking on speech intelligibility. A small but systematic intelligibility benefit of fast-acting compression was found in both the quiet and the noisy conditions for the lower speech levels. No negative effects of fast-acting compression were observed when the speech level exceeded the level of the noise. These findings suggest that fast-acting compression provides an audibility benefit in fluctuating interferers as compared to slow-acting compression, while not substantially affecting the perception of consonants at higher SNRs.

INTRODUCTION

It is widely accepted that due to the limited dynamic range of levels perceived by hearing-impaired (HI) listeners, some sort of level-dependent amplification is required to compensate for hearing loss. The majority of modern hearing aids apply dynamic-range compression (see Souza, 2002, and Edwards, 2004, for reviews). In such systems, the gain is determined based on one or more level-estimation circuit(s), characterized by attack and release time constants. The most commonly used attack times have values below 10 ms (Jenstad and Souza, 2005) in order to quickly reduce the gain in response to loud sounds. However, the optimal speed of gain recovery, i.e.,

*Corresponding author: bokowal@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

the release time, is still a subject of discussion. Shorter release times allow more gain to be applied to the low-intensity speech components (e.g., consonants) that follow other, high-intensity components (e.g., vowels) or noise bursts. This increased gain can potentially improve audibility and reduce the amount of forward masking, which in turn might lead to an improved speech recognition performance in HI listeners (Souza and Bishop, 1999; Edwards, 2002; Desloge *et al.*, 2010; Jenstad and Souza, 2005). On the other hand, with a very short release time, the gain follows the fast fluctuations of the signal, effectively reducing the temporal contrast. The temporal characteristics of the speech signal provide important cues for speech intelligibility, especially for HI listeners (Souza *et al.*, 2015). Temporal envelope distortion introduced by fast-acting amplification might therefore lead to a decrement in recognition performance. It is possible that optimal performance would be achieved if the time constants were adapted dynamically according to the current signal-to-noise ratio (SNR). For example, May *et al.* (2017) proposed a blind broadband-SNR estimator (based on the speech and noise power spectrum density), which could be applied in hearing aids. However, the relation between the optimal release time and SNR in connection to speech intelligibility is not yet known.

In the present study, it is hypothesized that potential negative effects of short release times will be more pronounced at higher SNRs, where audibility and masking are less of a concern and the compression is driven mostly by the speech signal. On the other hand, the additional gain applied by the fast-acting system is expected to provide an increasing benefit as the SNR decreases. To test these ideas, stimuli were designed to maximize the effects of compression release time. The noise consisted of high-intensity bursts, separated by silent gaps and had very sharp onsets and offsets. Short consonant-vowel (CV) tokens were used and listeners were asked to report the initial consonant – a speech component that typically has a low intensity. The temporal onset of the CV token relative to the noise was controlled and chosen based on a previous study (Zaar *et al.*, 2017). A wide range of SNRs and compression release times were tested in order to capture the potential interaction between the two factors.

METHODS

Listeners

Twelve young, normal-hearing (NH) listeners aged between 19 and 26 years (average age: 21.7 years) completed the task in the unaided condition. They all had pure-tone thresholds lower than 20 dB HL in the 250 to 8000 Hz range. The aided conditions were completed by nine older HI listeners aged between 66 and 77 years (average age: 71.4 years). Their hearing losses ranged from mild to moderately-severe losses and were most prominent at the high frequencies.

Stimuli

The target speech consisted of 15 consonant-vowel (CV) tokens: */bi, di, fi, gi, hi, ji, ki, li, mi, ni, pi, si, ʃi, ti, vi/* spoken by one male and one female talker (30 utterances in total), used previously by Zaar and Dau (2015). Four presentation levels were used:

45, 55, 65, and 75 dB SPL. In the aided conditions, these were the levels at the *input* to the amplification system. In each condition, each utterance was presented five times to the listeners.

The noise was composed of five 100-ms long bursts, separated by 100-ms silent gaps (corresponding to a 5-Hz repetition rate). White noise was chosen as a carrier in order to maximize masking of high-frequency consonants. The sound pressure level was 65 dB, defined as the level of the noise bursts at the input to the hearing aid simulator. The onset of the CV token was positioned 25 ms into the silent gap after the third noise burst, as shown schematically in Fig. 1. The instantaneous SNR was therefore infinite. The broadband SNR values are still reported for consistency with previous literature. They are defined as the difference between the sound pressure level of the token and the preceding noise burst.

Thirty noise waveforms (one per utterance) were pre-generated and stored as wav-files. Each utterance was always presented with the same noise recording. This was done in order to limit the across-repetition variability due to the random fluctuations in the Gaussian noise carrier, whilst preventing noise-learning effects that could occur if only one noise-waveform was used for all utterances (see Zaar and Dau, 2015).

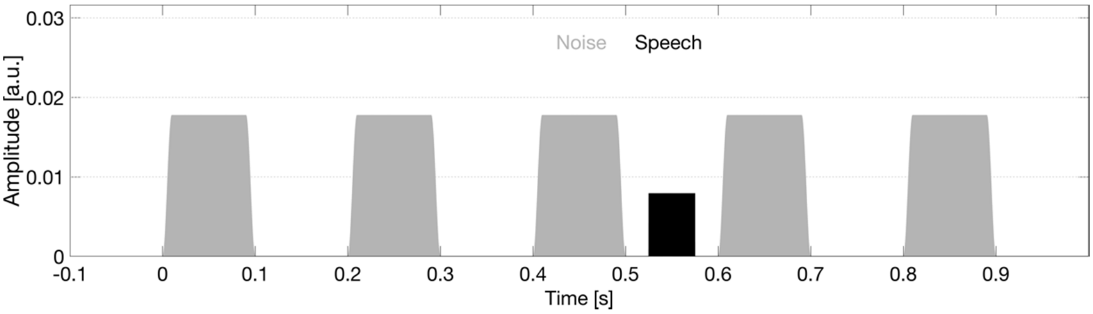


Fig. 1: A schematic representation of the stimulus time-course.

Amplification

For the HI listeners, the stimuli were pre-processed using a hearing-aid simulator with eight independent compression channels, implemented in MATLAB. The insertion gain was applied to the signals presented monaurally over Sennheiser HD650 headphones. The gain was frequency-dependent, based on the NAL-NL2 target (using the *Slow* setting, which yields more aggressive amplification, cf. Keidser *et al.*, 2011) for the N2 audiogram (Bisgaard *et al.*, 2010). This audiogram was chosen because it was most representative of the participants' hearing losses. Thus, compression ratios were the same for all listeners. However, in order to maximize audibility for each participant, the linear part of the gain (gain applied to stimuli below the compression threshold) was determined based on the individual audiogram.

The compression thresholds (kneepoints) were also frequency-dependent and calibrated such that each channel went into compression when the level of a broadband (white-noise) input exceeded 50 dB SPL. The attack time (of the level-detector circuit, or the so-called *RC time constants*, Kates, 1993) was always 5 ms. The release time depended on the amplification condition. It was 10 ms in the *fast* compression condition and 500 ms in the *slow* condition. The third condition was *linear*, which used the same maximum gain values but a compression ratio of 1:1 (i.e., no compression) and null attack and release times. It thus simulated an "idealized" hearing aid that never applies compression and provides the maximum possible amplification. Such high gain is unrealistic for high-intensity inputs, as it would be excessively loud. Thus, this condition served as a baseline for the behavior of compression systems, but only for lower-intensity speech inputs – 45 and 65 dB SPL in quiet (see "Experimental conditions" below).

In all conditions, the level-detection circuit of the compression and the resulting gain were driven by the mixture of speech and noise. Thus, the gain applied to the clean speech in the quiet condition was not controlled by the clean speech signal but rather *shadow-controlled* by the mixture. This setup allowed the investigation of the effects of the gain fluctuations (resulting from the presence of the interrupted noise) on the CV token without actually presenting the interferer to the listeners' ears.

Experimental conditions

The NH listeners were tested unaided while HI listeners were always presented with amplified stimuli. Slow and fast compression were tested in all conditions, while linear amplification was tested only in a limited number of conditions. In quiet, the compressed stimuli were always shadow-filtered with the corresponding mixture of speech and noise. An overview of all experimental conditions is shown in Table 1.

	Speech input level (dB SPL)	NH	HI		
		Unaided	Linear	Slow	Fast
Quiet (shadow-filtered)	45	x	x	x	x
	65	x	x	x	x
	75	-	-	x	x
Noise	45	x	-	x	x
	55	x	-	x	x
	65	x	-	x	x
	75	x	-	x	x

Table 1: Summary of experimental conditions: speech-noise configurations and amplification used.

RESULTS

For the quiet and noisy data sets, separate linear mixed-effects models were used with the fixed factors *speech level* and *amplification type* and random factor *listener.* Backwards elimination of non-significant effects was performed (Kuznetsova *et al.*, 2015) and the final model was used to establish significance between the results obtained with each *amplification type* at each *speech level.*

The distribution of the model residuals for the data in quiet deviated from normal (it was "light-tailed"). Therefore, these data were RAU-transformed before further analysis. The transformation was not necessary for the data in noise (the distribution of residuals was much closer to normal), so only the non-transformed data are reported for consistency.

Consonant recognition in quiet

The average consonant recognition rates for the stimuli presented in quiet are shown in Fig. 2. It can be observed that the unaided NH listeners achieved recognition rates close to 100% at both speech input levels, whereas the aided HI listeners performed much worse in all conditions and achieved maximum recognition rates of about 87% at 75 dB SPL. Significant differences were found between all amplification types for the lowest speech input level (45 dB SPL). The best recognition rate of 55% was achieved with linear amplification (that provided maximum possible gain), followed by fast (46%) and slow compression (34%).

For the 65 dB SPL speech input, no significant differences between amplification types were observed. Between 65 and 75 dB SPL, there was a slight increase in performance with slow compression but no significant change with fast compression (possibly due to ceiling effects). At 75 dB SPL, slow compression yielded, on average, slightly higher recognition rates than fast compression, but the difference was not significant.

Consonant recognition in noise

The recognition rates in noise are shown in Fig. 3. NH listeners achieved recognition rates of 95% for speech levels of 65 and 75 dB SPL (corresponding to SNRs of 0 and +10 dB). The rate decreased to 73% at 45 dB SPL (SNR −20 dB). Aided HI listeners achieved the highest recognition rate of 77% at 75 dB SPL. For speech input levels of 45, 55, and 65 dB SPL, the recognition rates observed with fast compression were 7-9% higher than with slow compression, with all differences being statistically significant ($p < 0.001$). At 75 dB SPL, slow compression yielded slightly higher recognition performance than fast, but the difference was not statistically significant.

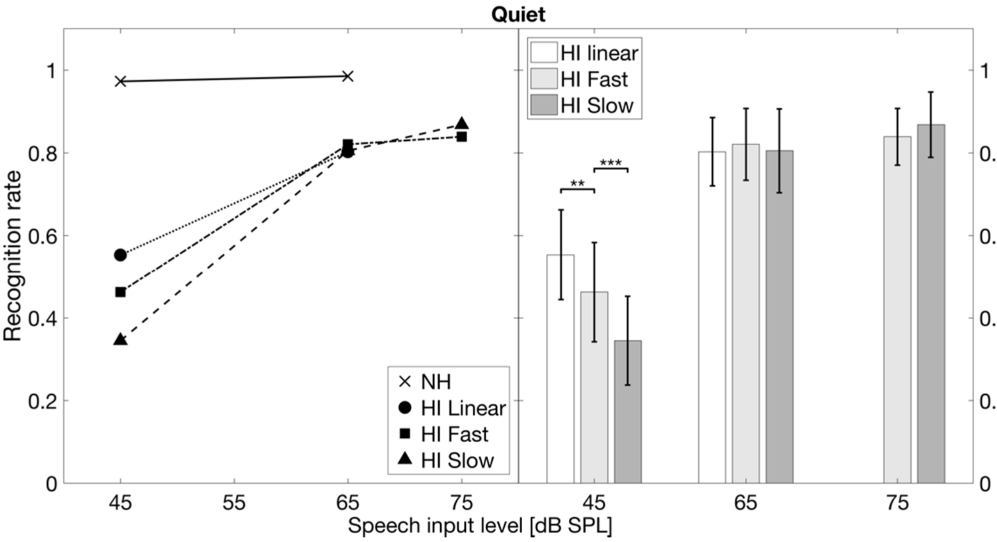


Fig. 2: Averaged consonant recognition rates for speech tokens in quiet, “shadow-controlled” by the mixture of speech and noise. Left: normal-hearing (NH) unaided and hearing-impaired (HI) aided with three types of amplification. Right: Only the HI data replotted. The error bars indicate +/- one standard deviation. The significance levels are: ** 0.01, *** 0.001.

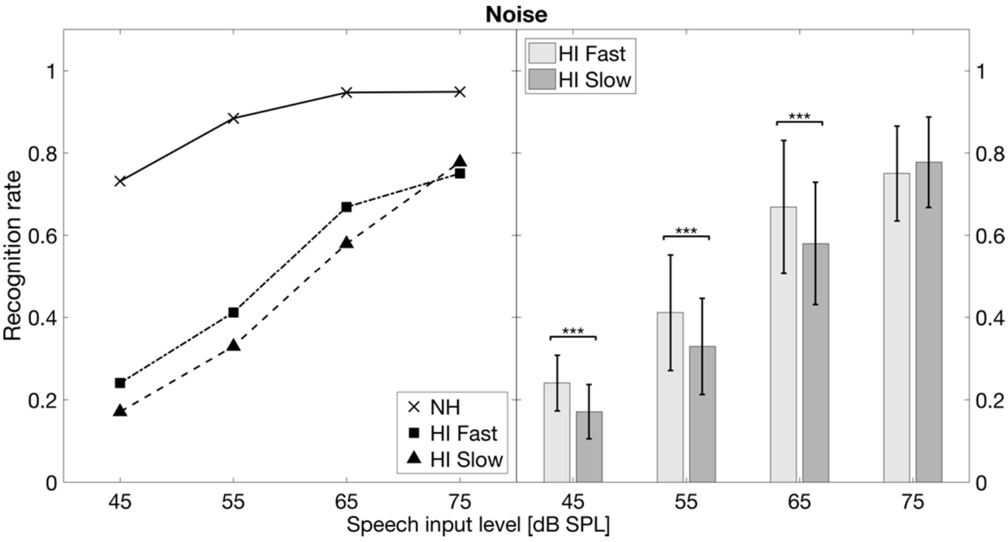


Fig. 3 The same as Fig 2. but presented in 5-Hz interrupted Gaussian noise (noise level: 65 dB SPL).

DISCUSSION

In the quiet condition, the consonant recognition rates at low speech input levels strongly depended on the amplification type. The best performance was obtained with linear amplification and fast compression, which provide higher gain and thus better audibility than slow compression. On the other hand, slow compression induced a small increase in performance between 65 and 75 dB SPL, while in the case of fast compression a ceiling effect was observed. Moreover, slow compression seemed to lead to a better average performance at 75 dB SPL, but the difference was small and not statistically significant.

In noise, fast-acting compression led to higher recognition rates for speech levels of up to 65 dB SPL, corresponding to a broadband SNR of 0 dB. Similar to the results in quiet, the performance at 75 dB SPL (+10 dB SNR) tended to be better with slow compression, but the effect was small and not significant. Overall, there is thus no statistically significant evidence for negative effects of fast compression (e.g., due to temporal envelope distortion) on consonant recognition performance at higher speech input levels. However, possible ceiling effects in the HI listeners' data may be a confounding factor here.

Effects on the recovery from forward masking

In quiet, the relative benefit of fast vs. slow compression decreased from 12% at 45 dB SPL to 2% at 65 dB SPL (SNRs −20 and 0 dB). In noise, the benefit increased from 7 to 9% between speech input levels of 45 and 65 dB SPL. As the compressor was shadow-controlled by the mixture of speech and noise when being applied to speech in quiet, it behaved identically in both conditions such that the only difference between the quiet and noise conditions was the presence of the noise. Therefore, an explanation for the above observation may be that the higher gain provided to the speech token by fast-acting compression improved the recovery from the noise-induced forward masking, at least at SNRs close to 0 dB (i.e., speech levels close to 65 dB SPL).

CONCLUSIONS AND OUTLOOK

A small but systematic benefit of fast-acting compression was found both in the quiet and the noisy conditions for speech levels below 65 dB SPL (0 dB SNR in noise). Despite potentially detrimental speech envelope distortions, no significant negative effects of fast-acting compression were observed when the speech level exceeded the level of the noise. These findings suggest that fast-acting compression provides an audibility benefit and, possibly, an improved recovery from forward masking in fluctuating interferers as compared to slow-acting compression, while not substantially compromising the perception of short CV tokens at higher SNRs.

It is yet to be investigated whether these effects persist in more realistic conditions, i.e., with longer speech stimuli (multi-syllable words, sentences) in fluctuating interferers with softer onsets/offsets.

The findings from this study and prospective future studies may help design SNR-dependent amplification strategies and individualized hearing-aid fitting strategies.

The potential use of blind SNR estimation for hearing-aid applications has been investigated in May *et al.* (2017). The output of such an estimator could be used to dynamically manipulate compression parameters in real-time and will be the subject of future investigations.

REFERENCES

Bisgaard, N., Vlaming, M., and Dahlquis, M. (**2010**). "Standard audiograms for the IEC 60118-15 measurement procedure," Trends Amplif., **14**, 113-120, doi: 10.1177/1084713810379609

Desloge, J.G., Reed, C.M., Braida, L.D., Perez, Z.D., and Delhorne, L.A. (**2010**). "Speech reception by listeners with real and simulated hearing impairment: effects of continuous and interrupted noise," J. Acoust. Soc. Am., **128**, 342-359, doi: 10.1121/1.3436522.

Edwards, B. (**2002**). "Signal processing, hearing aid design and the psychoacoustic Turing test," IEEE ICASSP. doi: 10.1109/ICASSP.2002.5745533

Edwards, B. (**2004**). "Hearing aids and hearing Impairment," in *Speech Processing in the Auditory System*, Edited by R.R. Fay and S. Greenberg (Springer, New York), pp. 339-421. doi: 10.1007/b97399

Jenstad, L., and Souza, P. (**2005**). "Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility," J. Speech Lang. Hear. Res., **48**, 651-667, doi: 10.1044/1092-4388(2005/045)

Kates, J. (**1993**). "Optimal estimation of hearing-aid compression parameters," J. Acoust. Soc. Am., **94**, 1-12.

Keidser, G., Dillon, H., Flax,. M, Ching, T., and Brewer, S. (**2011**). "The NAL-NL2 prescription procedure," Aud. Res., **1**, 88-90. doi: 10.4081/audiores.2011.e24

Kuznetsova, A., Brockhoff P.B., and Bojesen Christensen, R.H. (**2015**). "Package 'lmerTest'", *R package version* 2.0.

May, T., Kowalewski, B., Fereczkowski, M., and MacDonald E.N. (**2017**). "Assessment of broadband SNR estimation for hearing-aid applications," Proceedings of IEEE ICASSP, 231-235. doi: 10.1109/ICASSP.2017.7952152

Souza, P., and Bishop, R.D. (**1999**). "Improving speech audibility with wide dynamic range compression in listeners with severe sensorineural loss," Ear Hearing, **20**, 461-470. doi: 10.1097/AUD.0b013e3181aec5bc

Souza, P. (**2002**). "Effects of compression on speech acoustics, intelligibility and sound quality," Trends Amplif., **6**, 131-165. doi: 10.1177/108471380200600402.

Souza, P.E., Wright, R.A., Blackburn, M.C., Tatman, R., and Gallun, F.J. (**2015**). "Individual sensitivity to spectral and temporal cues in listeners with hearing impairment," J. Speech Lang. Hear. Res., **58**, 520-534. doi: 10.1044/2015_JSLHR-H-14-0138.

Zaar, J., and Dau, T. (**2015**). "Sources of variability in consonant perception of normal-hearing listeners", J. Acoust. Soc. Am., **138**, 1253-1267. doi: 10.1121/1.4928142.

Zaar, J., Kowalewski, B., and Dau, T. (**2017**) "Effects of non-stationary noise on consonant identification," Poster presented at the International Symposium on Auditory and Audiological Research, Nyborg, Denmark.

Speech intelligibility in dual task with hearing aids and adaptive digital wireless microphone technology

MATTHIAS LATZEL[1], KIRSTEN C. WAGENER[2], MATTHIAS VORMANN[2],
AND HANSE MÜLDER[3]

[1] *Phonak AG, Stäfa, Switzerland*

[2] *Hörzentrum Oldenburg, Oldenburg, Germany*

[3] *Phonak Communications AG, Murten, Switzerland*

Remote microphones (RMs) have been developed to support hearing aid users to understand distant talkers. A drawback of these systems is the deteriorated speech intelligibility in the near-field, as the hearing aids need to be in omnidirectional mode in combination with these RMs. This has changed with the introduction of a new hearing-aid technology developed specifically to support the user in the near-field when using a RM, by enabling directional microphones of the hearing aid. To verify the performance of this novel system, speech intelligibility tests were conducted using a dual-task paradigm. *Primary task:* Sentences of the female Oldenburg Matrix Test were presented continuously. The task of the subject was to mark the recognized name on a tablet. *Secondary task:* A speech recognition test with meaningful sentences (Göttinger Sentence Test, male voice) was carried out with the task to repeat the sentences. The primary-task stimuli were presented from a loudspeaker in the far-field and the secondary-task stimuli from a loudspeaker in the near-field (and vice versa), within a surrounding loudspeaker array playing restaurant noise. Results of 15 hearing-impaired subjects showed that the directional hearing-aid microphone delivered superior performance compared to the omni microphone. Benefits of the RM were confirmed for both primary and secondary tasks. For a higher ecological validity, the data were analyzed considering both tasks simultaneously. This analysis showed a positive effect of the directional hearing aid microphone.

INTRODUCTION

One of the most common problems that individuals with hearing loss face is to follow conversations in complex listening environments. Listening is often difficult when there is excessive background noise, reverberation, and large distances between the target signal and the individual with a hearing loss. This is also seen when the hearing loss is at least partly compensated by hearing aids. In order to overcome these three main factors, individuals with hearing loss require a better signal-to-noise (SNR) ratio than those with normal hearing (Baquis, 2014).

*Corresponding author: matthias.latzel@phonak.com

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

To address this need, modern hearing aids (HAs) include directional microphones that have been shown to increase speech understanding in noise (Dillon, 2012).

While directional microphones provide measureable benefit, they have their limitations. For example, a 4-5 dB SNR benefit can be achieved with directional microphones, but up to 25 dB SNR (depending on degree of hearing loss) may be needed to help individuals with hearing loss (Baquis, 2014). Additionally, directional microphones are primarily effective when used in the near-field, approximately 1.5 meters from the target signal (Kim and Kim, 2014).

Individuals who need additional SNR improvement beyond the potential of directional microphones may therefore consider utilizing remote microphones (RMs). Using RMs, the distance between the target signal and the microphone and thus the amount of background noise and reverberation can be significantly reduced. RMs are intended for far-field use and have historically been realized using frequency modulation (FM) transmission, where the FM radio transmitter is coupled with a microphone that the talker wears. The microphone signal is directly transmitted to the listener's hearing aid or cochlear implant via a (miniature) radio receiver using direct audio input or telecoil. These systems have shown significant benefit for both hearing-aid users (Anderson and Goldstein, 2004) and cochlear-implant users (Wolfe *et al.*, 2012). Traditional FM systems were generally configured as either fixed analog or adaptive analog systems.

Digital adaptive wireless systems are able to provide higher SNR improvements than traditional analogue FM systems, resulting in significantly better speech intelligibility of up to 35% in (high-level) noise (Wolfe *et al.*, 2013; Thibodeau, 2014). Therefore, the hearing-aid industry started to develop digital transmission systems described as adaptive digital wireless microphone technology. The present study investigates the benefit of an adaptive digital wireless technology ("Phonak Roger") for hearing aid users in adverse listening environments.

Historically, digital hearing-aid technology utilized two analog to digital converters forcing a single microphone mode (omni-directional) when using a RM. This led to a decrease in speech intelligibility in noise in the near field when speech was simultaneously presented to the RM and transmitted to the hearing aid.

In order to solve this problem, Phonak introduces a new solution/technology utilizing three analog-to-digital converters in the input stage of the hearing device, allowing for directional microphone settings to be used in conjunction with RMs.

Several studies have examined the use of RMs in combination with omni-directional hearing-aid microphones versus directional hearing-aid microphones in children either with static or with adaptive behavior. In a recent study Jones and Rakita (2016) used Phonak Sky V hearing aids. Speech was either presented from a loudspeaker simulating a peer talker *or* from a second loudspeaker simulating a class mate from behind *but not* a simultaneous presentation of both talkers. They found that children performed better on speech recognition tests in noise when using Roger plus hearing aid in directional microphone mode compared to Roger plus hearing aid in omni-directional mode by up to 25%.

Previous research showed the performance of RM with omni versus directional hearing aid microphones in either near-field or far-field target signals. However, it has not been investigated yet what happens when the target signal switches between the near field and far field. It is not uncommon for a listener to change their auditory focus in a given situation. For example, during a wedding reception the listener would like to listen to both, the official speech and comments from the people next to him/her. The present study aims to reproduce this type of adverse listening environment where the target signal changes from being close to the hearing aid wearer to being further away.

To that end, a dual-task paradigm was employed, which consists of two parallel speech intelligibility tasks and was developed to assess the interaction of target signals in near field and far field for hearing aid users with RMs.

METHODS

Subjects

Fifteen experienced hearing aid users with a severe sensorineural hearing loss (mean 4HFA (Roeser, 1996) of the better ear was 62.8 dB HL with a standard deviation of 6.1 dB HL) took part in the study. All subjects were inexperienced users of RM technologies. Subject ages ranged from 63 to 83 years with a mean age of 72 years (4 female, 11 male).

Hearing devices and test conditions

All subjects were bilaterally fitted with Phonak Naída V90 SP hearing aids (HAs). The initial setting was based on the subjects' audiograms and the fitting rule "Adaptive Phonak Digital" (Latzel 2013). The default acoustical coupling suggested by the fitting software Phonak Target 4.1 was selected. Fine tuning of the hearing aids (without RM) was allowed during an acclimatization period. The final settings were verified using real ear measurements.

During the laboratory measurement the subjects additionally received a RM ("Phonak Roger Pen") that was connected to the hearing aids ("Phonak Naída V90 SP") via receivers ("Phonak Roger 18").

Three different hearing aid conditions were defined:

P1: RM plus Hearing aids in omni-directional microphone setting
P2: RM plus Hearing aids in directional microphone setting
P3: Hearing aids alone without RM, binaural microphone setting ("StereoZoom")

Dual-task paradigm

In the primary task, sentences from the Oldenburg Sentence Test, spoken by a female speaker (OLSA, Wagener *et al.*, 2014), where presented continuously at a constant SNR. The subjects were asked to identify the name within each sentence from a list of 10 alternatives presented via a tablet PC.

A secondary task was performed simultaneously using the Göttingen Sentence Test presented via a male talker (GÖSA, Kollmeier and Wesselkamp, 1997) at a selected constant SNR. The subject was instructed to repeat all recognizable words. Based on word scoring, speech intelligibility in percent was determined.

Both tasks were simultaneously performed in a diffuse cafeteria noise scenario (L_{eq}=62 dB SPL, measured at the position where the RM was placed right in front of the far-field loudspeaker and at the position of the subject (see Fig. 1).

There were two set-up conditions:

(1) The primary task was presented in the far field from a loudspeaker at a distance of 6.4 m. The secondary task was presented from a loudspeaker in the near field (1.4 m).
(2) The primary task was presented from a loudspeaker in the near field (1.4 m) and the secondary task from a loudspeaker in the far field (6.4 m).

The presentation level of the primary task was kept constant at 65 dB SPL for the the primary task in the near field and at 70 dB SPL in the far field. These presentation levels assured audibility of the OLSA sentences for all subjects.

The loudspeaker set up is illustrated in Fig. 1.

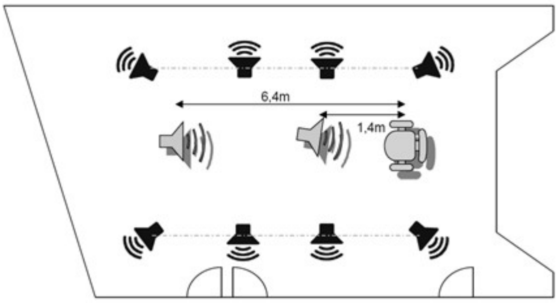


Fig. 1: Schematic display of the set-up used for the dual-task paradigm.

Training

To avoid training effects a training of the dual task was performed during each session prior to the measurements. During the training, the test hearing aids were used without RM (target signals only in near-field). Additionally, the speech presentation level for the secondary task (GÖSA) was determined individually and was used for all measurement conditions for said subject for far- and near-field presentation. The constant presentation level of the secondary task was individually determined from the GÖSA SRT result, measured adaptively, plus 3 dB. This resulted in GÖSA speech presentation levels ranging from 59 to 71.3 dB SPL across subjects.

After the training, the dual-task measurements were performed in the three hearing-aid conditions and two set-up conditions described above.

RESULTS AND DISCUSSION

Dual-task paradigm: Single performance

The left panel of Fig. 2 shows the results of the primary task in terms of the percentage of correctly identified names. The hearing aid conditions are called P1 to P3 and the set-up conditions are indicated by the loudspeaker that the primary-task stimuli were presented from: "far" or "near". An ANOVA of repeated measures revealed a significant main effect of hearing aid condition in the far-field [F(4,14) = 66.606, p < 0,001] and the post-hoc analysis showed a significant advantage of RM (p < 0.001) after Bonferroni correction (bfc.). This indicates that both programs (P1, P2) with RM active provided better performance than the HA alone (P3). This confirms earlier findings (Anderson and Goldstein, 2004; Wolfe *et al.*, 2012). In the near field, P1 and P2 showed no statistically significant differences with regard to primary-task performance. This finding does not suggest an advantage of P2 (directional microphone mode of HA) over P1 (omnidirectional microphone mode of HA) in both near-field and far-field conditions.

In the right panel of Figure 2 the results of the secondary task are illustrated in terms of percentage of correctly identified GÖSA words . The notation is similar to the left panel of Figure 2, except for "far" and "near" denoting the loudspeaker that the secondary-task stimuli were presented from. An ANOVA of repeated measures revealed a significant main effect of hearing-aid condition in the near-field [F(4,14) = 189.408, p < 0,001] and the post-hoc analysis showed a significant disadvantage of RM in the near-field (p < 0.001, bfc.). This indicates that the input from the RM was overlapping with the input of the HA microphone, resulting in poorer performance in the near-field. Without the RM, less interfering information was apparently provided to the listener in the near-field task. Results show P3 to markedly outperform the other hearing aid conditions regarding speech intelligibility in a noisy environment in the near-field. Furthermore, these results support the binaural beamformer (StereoZoom) which was active in P3, providing excellent performance in a noisy environment in the near-field (also noted by Appleton & König, 2014). Additionally, the post-hoc analysis revealed a significant advantage of the directional microphone (P2) over the omni-directional microphone (P1) in the near field (p<0.05, bfc.). In the far-field, no statistically significant difference between P1 and P2 was found, and therefore a general conclusion could not be established for the secondary task. This leads us to conclude that only the analysis of the common performance (primary and secondary task) is able to reflect the benefit of the different test conditions in near-field and far-field.

Dual-task paradigm: Common performance

The motivation of this study was to determine speech perception performance in a listening situation where the target signal changes from being close to the hearing-aid wearer to being further away. Dual-task costs were determined in order to calculate the common performance of both tasks within the dual-task paradigm.

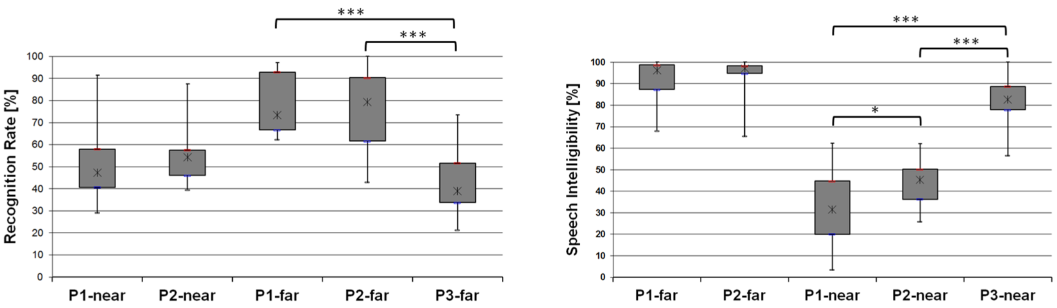


Fig. 2: *Left panel:* Recognition rate of names in continuous OLSA sentences (primary task) for hearing aid conditions P1, P2 and P3. Near: primary task was presented in the near-field; Far: primary task was presented in the far-field. *Right panel:* Speech intelligibility of GÖSA sentences (secondary task) for the hearing aid conditions P1, P2 and P3. Near: secondary task was presented in the near-field; Far: secondary task was presented in the far-field. Values are displayed as boxplots with median, minimum, maximum, 25th and 75th percentiles. *: Denotes statistically significant difference (p<0.05) ***: Denotes statistically significant differences (p<0.001). *Note:* P3 was not measured in the set-up condition 2, as from an ethical point of view it was not justifiable to conduct the secondary task from far-field as it would end up in a speech perception of about 0%. So the P3 hearing aid condition is only shown for the primary task in far-field and secondary task in near-field.

The main concept of dual-task costs is to measure the change of performance of the primary task due to the additional cognitive load of the secondary task (and vice versa). Most likely the performance of each task drops when performing both tasks at the same time compared to the case when every task is carried out alone. In the following, dual-task costs are calculated using the “probit” (probability units) transformation according to Oberauer *et al.* (2004): The differences in speech recognition (in percent correct) for doing every task in the single condition (data not shown here) compared to the dual task condition (Fig. 2) are calculated and expressed as the corresponding z-scores of a standard normal distribution. The probit values for both tasks are summed up afterwards to account for the common performance change on the primary and secondary task.

The dual-task costs are visualized in Fig. 3. A 2x2 factorial ANOVA of repeated measures revealed a significant main effect of hearing-aid condition [F(1,14) = 56.282, *p* < 0,000] and of set-up condition [F(1,14) = 4.6153, *p* = 0,49]. No statistically significant effect of the interaction of hearing aid condition & set-up condition could be found. *Hearing aid condition:* This result indicates that the directional microphone increases speech intelligibility regardless of whether the talker is near or far. It shows that the directional microphone not only improves speech perception in the near field but also for a distant speaker transmitted to the hearing aid via RM. In this case, the directional microphone acts as an additional means of noise suppression. *Set-up condition:* The results indicate that the different levels of difficulty for the tasks are influenced by the source position. When transmitting GÖSA speech material via RM to the HA, the performance is much better than when it is received with the microphones of the HA (see also Fig. 2, right panel). This may be due to the default mixing factor at the input stage of the HA, which is set to 10 dB amplification of the RM signal versus the HA microphone signal due to regulations for using a remote microphone/hearing aid system in school (Johnson *et al.*, 2011). The missing interaction effect for both hearing aid condition and set-up condition supports the extra benefit provided by the directional microphone regardless of where both tasks are presented from. P2 (directional microphones) showed to be beneficial both in the far-field and the near-field in terms of common performance.

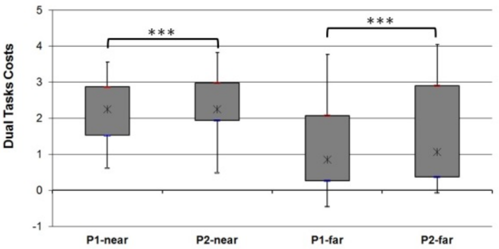


Fig. 3: Dual-task costs in probit (probability units) for hearing-aid conditions P1 and P2. Near: primary task was presented in the near field; Far: primary task was presented in the far-field. Values are displayed as boxplots with median, minimum, maximum, 25th and 75th percentiles (higher values represent better performance, thus less dual-task costs). ***: Denotes statistically significant difference ($p < 0.001$).

CONCLUSIONS

The described dual-task paradigm is an effective tool for testing the interaction of simultaneous input signals both in near field and in far field when using a hearing-aid in combination with a remote microphone. The set-up that has been used in this experiment is: (1) able to identify the advantages and disadvantages of a remote microphone/hearing-aid combination. The results confirm that the novel system (hearing aid with directional microphone in connection with a remote microphone) provides better common speech understanding in near and far-field. It can be expected that this set up could be optimized in terms of the mixing factor (amplification of RM input versus HA microphone input), particularly when used as a hearing solution for adults; (2) sensitive to differences between omnidirectional and directional microphone settings.

When analysing data of a dual task it is necessary to consider the common performance of both tasks. Calculating the “costs” of how much the performance of each single task drops when executing both tasks simultaneously has been shown to be a suitable way to derive the common performance.

REFERENCES

Anderson, K.L., and Goldstein, H. (**2004**). "Speech perception benefits of FM and infrared devices to children with hearing aids in a typical classroom," Lang. Speech Hear. Ser., **35**, 169-184, doi: 10.1044/0161-1461(2004/017)

Appleton, J., and König, G,.(**2014**). "Improvement in speech intelligibility and subjective benefit with binaural beamformer technology" Hearing Review, **21**, 40-42.

Baquis, D. (**2014**). *Assistive Listening Devices.* Retrieved from National Institute of the Deaf.

Dillon, H. (**2012**). *Hearing aids – A Comprehensive Text.* New York: Boomerang Press and Thieme.

Johnson, C.D., Anderson, V., Boothroyd, A., Eiten, L., Gabbard, S.A., Lewis, D., and Thibodeau, L. (**2011**). *Remote Microphone HearingAssistance Technologies for Children and Youth from Birth to 21 Years.* American Academy of Audiology Clinical Practice Guidelines.

Jones, C., and Rakita L. (**2016**). "A powerful noise-fighting duo: Roger and Phonak directionality," Phonak Field Study News. http://www.phonakpro.com/com/b2b/en/evidence.html

Kim, J.S., and Kim, C.H. (**2014**). "A review of assistive listening device and digital wireless technology for hearing instruments" Korean J. Audiol., **18**, 105-111. doi: 10.7874/kja.2014.18.3.105

Kollmeier, B., and Wesselkamp, M. (**1997**). "Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment," J. Acoust. Soc. Am., **102**, 2412-2421. doi: 10.1121/1.419624

Latzel, M. (**2013**). "Compendium 4 - Adaptive Phonak Digital (APD)", Phonak Compendium. http://www.phonakpro.com/com/b2b/en/evidence.html

Oberauer, K., Lange, E., and Engle, R.W. (**2004**). "Working memory capacity and resistance to interference," J. Mem. Lang., **51**, 80-96. doi: 10.1016/j.jml.2004.03.003

Roeser, R.J. (**1996**). *Roeser's Audiology Desk Reference.* New York, Stuttgart: Thieme, pp. 171.

Thibodeau, L. (**2014**). "Comparison of speech recognition with adaptive digital and FM wireless technology by listeners who use hearing aids," Am. J. Audiol., **23**, 201-210. doi: 10.1044/2014_AJA-13-0065

Wagener, K.C., Hochmuth, S., Ahrlich, M, Zokoll, M.A., and Kollmeier, B. (**2014**). "Der weibliche Oldenburger Satztest," 17th annual conference of the DGA, Oldenburg, CD-Rom.

Wolfe, J., Schafer, E.C., Parkinson, A., John, A.B. Hudson, M., Wheeler, J., and Mucci, A. (**2012**). "Effects of input processing and type of personal FM systems on speech recognition performance of adults with cochlear implants," Ear Hearing, **34**, 52-62. doi: 10.1097/AUD.0b013e3182611982

Wolfe, J., Morais, M., Schafer, E., Mills, E., Mülder, H.E., Goldbeck, F., Marquis, F., John, A., Hudson, M., Peters, B.R., and Lianos, L. (**2013**). "Evaluation of speech recognition of cochlear implant recipients using a personal digital adaptive radio frequency system," J. Am. Acad. Audiol., **24**, 714-724. doi: 10.3766/jaaa.14099.

Influence of multi-microphone signal enhancement algorithms on auditory movement detection in acoustically complex situations

MICHA LUNDBECK[1,2], LAURA HARTOG[1,2], GISO GRIMM[1,2], VOLKER HOHMANN[1,2], LARS BRAMSLØW[3], AND TOBIAS NEHER[1,4]

[1] *Medizinische Physik and Cluster of Excellence Hearing4all, Oldenburg University, Oldenburg, Germany*

[2] *HörTech gGmbH, Oldenburg, Germany*

[3] *Eriksholm Research Centre, Snekkersten, Denmark*

[4] *Institute of Clinical Research, University of Southern Denmark, Odense, Denmark*

The influence of hearing aid (HA) signal processing on the perception of spatially dynamic sounds has not been systematically investigated so far. Previously, we observed that interfering sounds impaired the detectability of left-right source movements and reverberation that of near-far source movements for elderly hearing-impaired (EHI) listeners (Lundbeck *et al.*, 2017). Here, we explored potential ways of improving these deficits with HAs. To that end, we carried out acoustic analyses to examine the impact of two beamforming algorithms and a binaural coherence-based noise reduction scheme on the cues underlying movement perception. While binaural cues remained mostly unchanged, there were greater monaural spectral changes and increases in signal-to-noise ratio and direct-to-reverberant sound ratio as a result of the applied processing. Based on these findings, we conducted a listening test with 20 EHI listeners. That is, we performed aided measurements of movement detectability in two acoustic scenarios. For both movement dimensions, we found that the applied processing could partly restore source movement detection in the presence of reverberation and interfering sounds.

INTRODUCTION

Listeners with sensorineural hearing loss exhibit considerable difficulties in complex acoustic environments. Hearing aids (HAs) can help by restoring audibility and by improving the signal-to-noise ratio (SNR). This can improve speech reception in noise, but it may also compromise spatial hearing abilities including movement perception. To start addressing this possibility, we recently conducted a study where we observed that for elderly hearing-impaired (EHI) listeners interfering sounds impaired the detectability of left-right source movements, and reverberation that of

*Corresponding author: micha.lundbeck@uni-oldenburg.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

near-far source movements (Lundbeck *et al.*, 2017). These results raise the question of how to compensate these deficits with HAs. In the current study, we therefore investigated the influence of different multi-microphone signal enhancement algorithms on source movement detection in acoustically complex situations. To that end, we used a higher-order Ambisonics-based system for simulating complex sound scenes together with a computer simulation of bilateral multi-microphone HAs. To start with, we investigated the influence of different multi-microphone signal enhancement algorithms on acoustic measures that are presumed to be related to movement perception. We then evaluated the most promising HA settings in a listening test to explore the potential of improving source movement detection with HAs. In summary, the current study had the following aims:

1. To identify HA settings that can enhance acoustic cues that are presumed to underlie left-right (L-R) and near-far (N-F) source movement detection;
2. To evaluate the most promising HA settings for improving L-R and N-F source movement detection with a group of EHI listeners.

METHODS

Experimental setup

We simulated a complex acoustic environment using a toolbox for creating dynamic virtual environments (TASCARpro version 0.128; Grimm *et al.*, 2015). We configured our setup such that it produced 48 virtual loudspeaker signals in the horizontal plane with a spatial resolution of 7.5°. The virtual listener was seated at the center of the loudspeaker array. As the aim of this study was to include different HA algorithms, we generated multi-microphone signals by convolving the loudspeaker signals with binaural room impulse responses from the database of Thiemann and van de Par (2015) for the corresponding directions.

Stimuli

We made use of five different environmental sounds. For the target, we used a broadband noise-like fountain signal (S1; at 0° azimuth and 1 m distance re. the listener in the reference position). As interfering sounds, we used recordings of ringing bells, bleating goats, pouring water and humming bees (S2-S5: at ±45° and ±90° and 1 m distance each). We presented the target sound (S1) at 65 dB SPL (nominal) and the other sounds (S2-S5) at 62 dB SPL (nominal) each, as measured under reverberant conditions at the position of the virtual listener. The duration of each sound was 3.1 s.

HA signal processing

We used the Master Hearing Aid (MHA) research platform (Grimm *et al.*, 2006) for simulating five HA settings: *unproc*, *dir*, *coh*, *dircoh*, and *beam*. The *unproc* condition corresponded to a pair of omnidirectional microphones that we simulated using the front microphones of two behind-the-ear (BTE) devices without any additional processing. The *dir* condition corresponded to a pair of static forward-facing cardioid microphones (e.g., Dillon, 2012), which were realized based on the front and rear microphone signals of each BTE device. We then spectrally equalized the output signals to ensure that the frontal target signals sounded highly similar across the unproc and dir settings. The *coh* condition corresponded to a binaural noise reduction scheme for attenuating incoherent signal segments (Grimm *et al.*, 2009). The gains applied to the left and right channels were always the same, so that interaural level and time differences (ILDs and ITDs) were unaffected, while incoherent sounds (as caused by early reflections and late reverberation, for example) were attenuated. The *dircoh* condition consisted of the serial combination of the dir and coh settings. The *beam* setting corresponded to a bilateral beamforming algorithm with a post-filter for binaural cue preservation (Rohdenburg *et al.*, 2007). For the current study, we used six input signals (three per side) and the front BTE microphone signals as reference signals for the binaural post-filter. We then also spectrally equalized the output signal so that the frontal target signals sounded similar across the unproc and beam settings. In the following, we will concentrate on the dircoh and beam settings, as they showed the clearest effects relative to the unproc setting.

Technical measurements

General setup and procedure

For the technical measurements, we generated stimuli based on the median L-R and N-F detection thresholds of the EHI listeners tested previously (Lundbeck *et al.*, 2017). Specifically, we generated stimuli where the target signal moved 28° in the L-R direction or 1.5 m in the N-F direction re. the reference position. The signal processing chain used for the acoustical analyses is shown in Fig. 1.

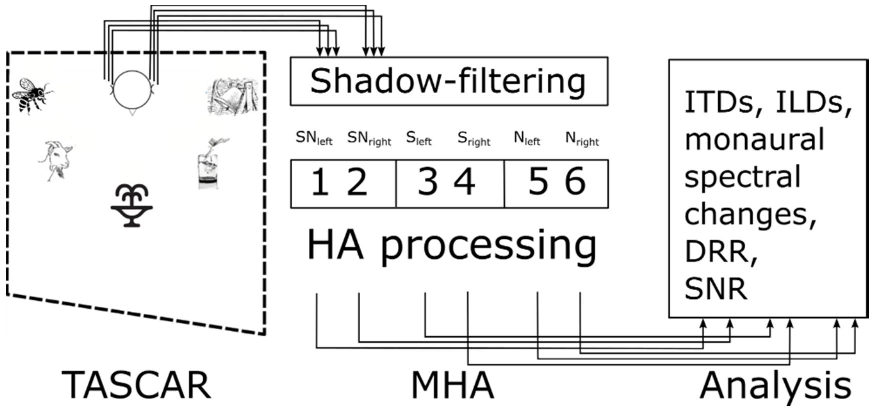

Shadow-filtering
SNleft SNright Sleft Sright Nleft Nright
1 2 3 4 5 6
HA processing
ITDs, ILDs, monaural spectral changes, DRR, SNR
TASCAR
MHA
Analysis

Fig. 1: Signal processing chain used for the acoustical analyses. Following the generation of the stimuli using TASCAR (left) and the shadow-filtering in the MHA (middle), different output channels (1-6) were analyzed using different measures (right). SN_{left}, SN_{right} = Left and right channels of the signal mixture; S_{left}, S_{right} = Left and right channels of the target signal; N_{left}, N_{right} = Left and right channels of the interfering signals.

We equipped the virtual listener with two BTE devices with up to three microphones each. We then processed the microphone signals with the MHA. We used the so-called shadow-filtering method to apply the processing computed for the signal mixture separately to the target and interferers. Depending on the measure of interest (see below), we then analyzed different output signals. To reveal short-time changes in the chosen measures, we used a 100-ms analysis window with 50% overlap.

Monaural spectral changes

To analyze the influence of the different HA settings on monaural spectral cues, we applied a spectral coloration measure of Moore and Tan (2004). We always analyzed the stimulus channel ipsilateral to the movement direction and referenced it to the stationary equivalent of the same stimulus. In this way, we measured relative monaural spectral changes due to the source movement and the HA settings.

Signal-to-noise ratio (SNR) changes

For estimating the SNR, we used the separate target and interferer signals (see middle panel of Fig. 1, channels 3+4 and 5+6, respectively). We then calculated the short-term level ratio between the target and the interferers at either the ipsilateral side (L-R dimension) or averaged across the two sides (N-F dimension).

Direct-to-reverberant sound ratio (DRR) changes

For the stimuli moving along the N-F dimension, we estimated short-term changes in the DRR. To that end, we created two stimuli per condition: one with and one without reverberation. We then subtracted the anechoic stimulus (comprising the direct sound only) from the reverberant stimulus and fed the direct and reverberant sound separately into the MHA. By comparing the DRR at the input and output of the MHA, we could estimate DRR changes due to the applied HA processing.

Perceptual measurements

Participants

For the perceptual measurements, we used 20 EHI listeners aged 63-80 yr (mean: 72.4 yr). Fifteen of them had bilateral HA experience of at least 2 yr. All participants had symmetric, sloping mild-to-moderate sensorineural hearing losses. We divided the participants into two groups according to their performance on a target detection task (see below). The mean pure-tone average hearing loss calculated across 0.5, 1, 2 and 4 kHz and both ears (PTA4) differed significantly across the two groups (group 1: 58 dB HL; group 2: 47 dB HL; $p < 0.001$), whereas age did not (group 1: 75 yr; group 2: 70 yr; $p > 0.1$).

General setup and procedure

To investigate the perceptual consequences of the tested HA settings, we carried out a listening test with 20 EHI listeners. Initially, we assessed each listener’s ability to detect the target signal in the presence of the four interferers. For the participants who could not consistently detect the target signal in the presence of the interferers (group 1; $N = 9$), we performed the movement detection threshold measurements without the interferers. For the other participants (group 2; $N = 11$), we performed the measurements with all five signals.

The listening test was carried out under reverberant conditions ($T_{60} \approx 0.8$ sec). Stimulus presentation was via a 24-bit RME (Haimhausen, Germany) Hammerfall DSP 9632 soundcard, a Tucker-Davis Technologies (Alachua, USA) HB7 headphone buffer and a pair of Sennheiser (Wennebostel, Germany) HDA200 headphones. For the psychoacoustic measurements, we used the "psylab" toolbox (Hansen, 2006). To ensure adequate audibility for each participant, we spectrally shaped all stimuli in accordance with the "National Acoustics Laboratories–Revised-Profound" (NAL-RP) fitting rule (Dillon, 2012).

Source movement detection thresholds

We presented stimuli with moving target sounds on half of the trials and stimuli with static target sounds (at the reference position) on the other trials. For the angular measurements, we randomized the direction of movement (towards the left or right). For the radial measurements, we always simulated a withdrawing (N-F) movement. In this way, the starting position of the target sound source was the same in all conditions (0°, 1 m re. the listener). To control the extent of the movement, we varied the velocity (in °/s or m/s) in the adaptive procedure. For the adaptive procedure, we used the single-interval-adjustment-matrix procedure of Kaernbach (1990) to ensure unbiased measurements. A run was terminated after 12 reversals, and the first four reversals were discarded from the analyses. Before the actual measurements, each participant completed two training runs (one with *unproc* and one with *beam*).

We estimated the detection thresholds by taking the arithmetic mean of the last eight reversal points. In this manner, we quantified the smallest displacement (in ° or m) of the target source that the participants could perceive within the 2.3 s over which the movements occurred. In the following, we will refer to these thresholds as the minimum audible movement angle (MAMA) or distance (MAMD) thresholds. We performed the L-R and N-F measurements in separate blocks. Within each block, we tested the various conditions in randomized order. After 1-2 weeks, we conducted retest measurements. In total, we measured six detection thresholds per movement dimension (L-R and N-F) and listener (and thus 240 thresholds in total). According to Kolmogorov-Smirnov's test, all datasets fulfilled the requirement for normality (all $p > 0.05$). We therefore used parametric statistics to analyze our data. Whenever appropriate, we corrected for violations of sphericity using the Greenhouse-Geisser correction.

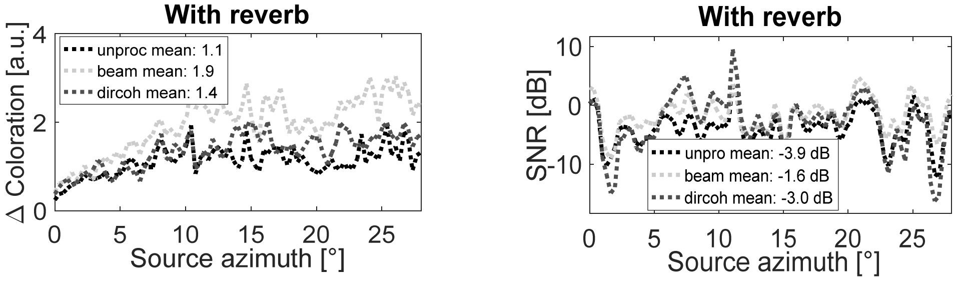


Fig. 2: Left panel: Monaural spectral coloration re. a static stimulus subjected to the same processing for unproc (black), beam (light gray) and dircoh (dark gray) as a function of source azimuth. Right panel: SNR for unproc (black), beam (light gray) and dircoh (dark gray) as a function of source azimuth. Legends show mean values for the different HA settings calculated over the whole stimulus duration.

RESULTS

Technical measurements

L-R dimension

Concerning the L-R dimension, the changes in the measures of interest that we observed were generally as expected. Regarding the monaural spectral changes, our analyses revealed that the beam and dircoh settings both increased this measure, suggesting that they are suited for improving source movement detectability. The left panel of Fig. 2 shows the resultant spectral coloration relative to the static condition in the presence of the four interferers and reverberation.

The right panel depicts the SNR caused by the three HA settings over the course of the target source movement in the presence of the four interferers. It is noticeable that the SNR varied substantially over the course of the source movement. This was because of the spectro-temporal fluctuations inherent to the environmental sounds that we used. Concerning the influence of dircoh and beam, beam increased the SNR more relative to unproc.

N-F dimension

Concerning the N-F dimension, the changes in the chosen measures were generally as expected (data not shown). The DRR generally decreased with increasing source distance, irrespective of the HA setting. Furthermore, the beam and especially the dircoh setting increased the DRR. The same was essentially true for the monaural spectral coloration, suggesting that monaural spectral cues may provide salient information about source movements. Regarding the SNR improvement relative to unproc, beam and especially dircoh led to clear increases.

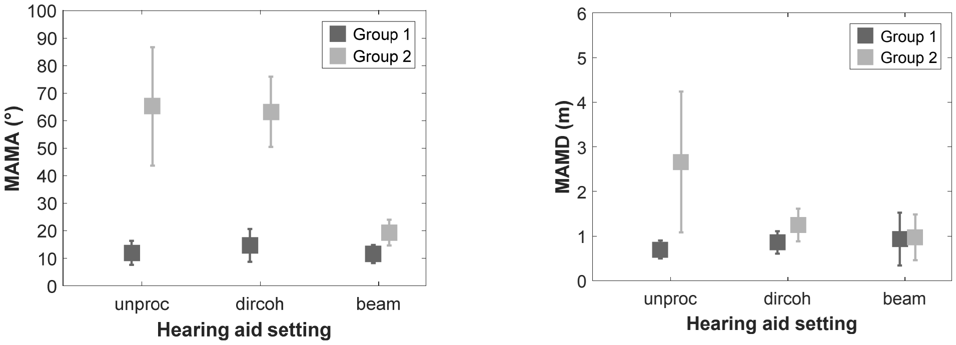


Fig. 3: Means and standard deviations of the MAMA (left) and MAMD (right) thresholds for the two groups and three HA settings.

Perceptual measurements

L-R dimension

Figure 3 (left panel) shows means and standard deviations of the MAMA thresholds for the two groups and three HA settings. For group 1, the thresholds varied little across HA settings and listeners. For group 2, the thresholds were much higher with unproc and dircoh than with beam. Furthermore, unproc was characterized by the largest spread and beam by the smallest spread.

To test for statistical differences among the three HA settings, we conducted two analyses of variance (ANOVA), that is, one per group with the within-subject factor HA setting (unproc, dircoh, beam). For group 1, we found no effect of HA setting [$F(2,16) = 2.5$, $p = 0.14$]. For group 2, the effect of HA setting was highly significant [$F(2,20) = 38.1$, $p < 0.0001$]. A series of planned contrasts showed that the beam setting differed significantly from both unproc and dircoh (both $p < 0.001$).

N-F dimension

Figure 3 (right panel) shows means and standard deviations of the MAMD thresholds for the two groups and three HA settings. As can be seen, group 1 obtained thresholds of around 1 m or lower in all conditions. In other words, the different HA settings did not appear to affect their performance. In contrast, for group 2 there was a clear influence of HA setting on movement detectability. To test for statistical differences among the three HA settings, we conducted an ANOVA per group with HA setting (unproc, dircoh, beam) as within-subject factor. For group 1, the effect of HA setting was not significant [$F(2,14) = 1.8$, $p = 0.2$], while for group 2 it was strongly significant [$F2,18 = 13.6$, $p < 0.001$]. A series of planned contrasts showed that the beam and dircoh settings differed significantly from unproc (both $p < 0.05$) and also from each other ($p < 0.01$).

Summary

The current study, which we conducted based on a setup for simulating complex virtual environments, showed that selected multi-microphone signal enhancement algorithms can enhance acoustical cues presumed to underlie source movement perception. Furthermore, the subsequent listening test showed substantial improvements in source movement detectability for a group of EHI listeners in complex scenarios with reverberation and interfering signals. In view of the fact that our study focused on one particular spatial dimension (i.e., source movement detection), it is of interest to extend this research to other aspects of spatial awareness perception and to head-worn HAs in future studies.

ACKNOWLEDGEMENTS

This research was funded by the Oticon Foundation and the DFG Cluster of Excellence EXC 1077/1 "Hearing4all".

REFERENCES

Dillon, H. (**2012**). *Hearing aids / Harvey Dillon* (Boomerang Press; Thieme, Sydney, New York).

Grimm, G., Herzke, T., Berg, D., and Hohmann, V. (**2006**). "The master hearing aid: a PC-based platform for algorithm development and evaluation," Acta Acust United Ac., **92**, 618-628.

Grimm, G., Hohmann, V., and Kollmeier, B. (**2009**). "Increase and subjective evaluation of feedback stability in hearing aids by a binaural coherence-based noise reduction scheme," IEEE T. Audio Speech, **17**, 1408-1419.

Grimm, G., Luberadzka, J., Herzke, T., and Hohmann, V. (**2015**). "Toolbox for acoustic scene creation and rendering (TASCAR)-Render methods and research applications," Proceedings of the Linux Audio Conference, Mainz.

Hansen, M. (**2006**). "Lehre und Ausbildung in Psychoakustik mit psylab: Freie Software fur psychoakustische Experimente," Fortschritte der Akustik, **32**, 591.

Kaernbach, C. (**1990**). "A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing," J. Acoust. Soc. Am., **88**, 2645-2655.

Lundbeck, M., Grimm, G., Hohmann, V., Laugesen, S., and Neher, T. (**2017**). "Sensitivity to angular and radial source movements as a function of acoustic complexity in normal and impaired hearing," Trends Hear., **21**, 2331216517717152.

Moore, B.C., and Tan, C.-T. (**2004**). "Development and validation of a method for predicting the perceived naturalness of sounds subjected to spectral distortion," J. Audio Eng. Soc., **52**, 900-914.

Rohdenburg, T., Hohmann, V., and Kollmeier, B. (**2007**). "Robustness analysis of binaural hearing aid beamformer algorithms by means of objective perceptual quality measures," in *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop on* (IEEE), pp. 315-318.

Thiemann, J., and van de Par, S. (**2015**). "Multiple model high-spatial resolution HRTF measurements," Proc. DAGA 2015.

Influence of a remote microphone on localization with hearing aids

JOHAN G. SELBY[1,2], ADAM WEISSER[2], AND EWEN N. MACDONALD[1,*]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *GN Hearing A/S, Ballerup, Denmark*

When used with hearing aids (HA), the addition of a remote microphone (RM) may alter the spatial perception of the listener. First, the RM signal is presented diotically from the HAs. Second, the processing in the HA often delays the RM signal relative to the HA microphone signals. Finally, the level of the RM signal is independent of the distance from the RM to HA. The present study investigated localization performance of 15 normal-hearing and 9 hearing-impaired listeners under conditions simulating the use of an RM with a behind the ear (BTE) HA. Minimum audible angle discrimination around an average angle of 45° was measured for three sets of relative gains and seven sets of relative delays for a total 21 conditions. In addition, a condition with just the simulated BTE HA signals was tested. Overall, for both groups, minimum audible angle discrimination was best when the relative RM gain was small (-3 and -6 dB) and the delay was approximately 10-20 ms. Under these conditions, localization performance approached the level obtained in the BTE HA only condition.

INTRODUCTION

Listening in background noise and/or reverberant environments can be particularly difficult for hearing-impaired (HI) listeners. In some situations, it is possible for an HI listener to use a remote microphone (RM) positioned close to the talker and have the signal streamed wirelessly to his or her HA (Ross and Giolas, 1971). Relative to the microphones in the HA, the RM receives a better quality signal from talker (i.e., higher signal-to-noise ratio and/or direct-to-reverberant ratio of the talker), which can improve intelligibility (Hawkins, 1984; Nábělek and Donahue, 1986; Boothroyd, 2004) and possibly reduce listening effort. However, the RM signal is mono and mixed diotically with bilateral HAs. If presented on its own, the RM signal should result in the perceived location of the talker being internalized in the head of the HI listener. Thus, the use of an RM may improve intelligibility but may decrease an HI listener's ability to localize the talker.

For cases where a single RM is in use (e.g., teacher in a classroom, listening to a speaker in an auditorium, etc.) localization may be only a minor issue as the HI

[*] Corresponding author: emcd@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. © The Authors. ISBN: 978-87-990013-6-1.

listener may not need to switch his/her focus away from the RM signal. However, in situations where multiple RMs are in use or multiple talkers are sharing the same RM, judging the location of the different talkers via acoustic cues may be advantageous in keeping up with the conversation.

There are two parameters when mixing the RM signal in the HA: the relative gain and the relative delay of the RM signal. Guidelines for audiologists regarding the relative gain fitting for children have been established and suggest a goal of "Transparency", by setting the gain of the RM to equal that from HA when presented with a 65 dB SPL signal (American Academy of Audiology, 2011). However, to our knowledge, no guidelines for RM delay have been established. Instead, the delay is usually determined by the the digital communication protocol used to wirelessly stream the RM signal to the HA.

The goal of the present study was to investigate the effects of relative gain and delay on spatial perception by comparing minimum audible angle (MAA) thresholds for a source with an incident angle of 45° in conditions that simulated using an RM with an ideal behind-the-ear (BTE) hearing aid.

METHODS

Spatialized stimuli were created through acoustic recordings in an anechoic room using two head and torso simulators (HATS, Brüel & Kjær 4128C) with omnidirectional microphones, which were positioned in locations corresponding to an RM and two BTE HAs. The recorded stimuli were then used in the listening test and were presented via headphones in a sound booth.

Stimuli

Recordings of 15 short Danish sentences spoken by a female talker were played back via a "talking" HATS in an anechoic room. These sentences were taken from the corpus created for Sørensen *et al.* (2017).

The speech signals were played back by the "talking" HATS, which was equipped with a mouth simulator. Both the "talking" and "listening" HATS were placed at an equal height with a distance of 97 cm from mouth to mouth and 118 cm from ear to ear (see Fig 1). The listening HATS was placed on a Brüel & Kjær Type 9640 turntable in order to record different angles between the talking and listening HATS.

To simulate an RM, a DPA 4060 omnidirectional microphone was positioned on the chest of the talking HATS 20 cm from the center of the mouth opening. Two other DPA 4060 microphones were placed at the top of the pinnae of each ear to simulate the microphone positions of BTE HAs.

For each of the 15 sentences, recordings were made with the listening HATS angled from 0 to 90° in 1° steps.

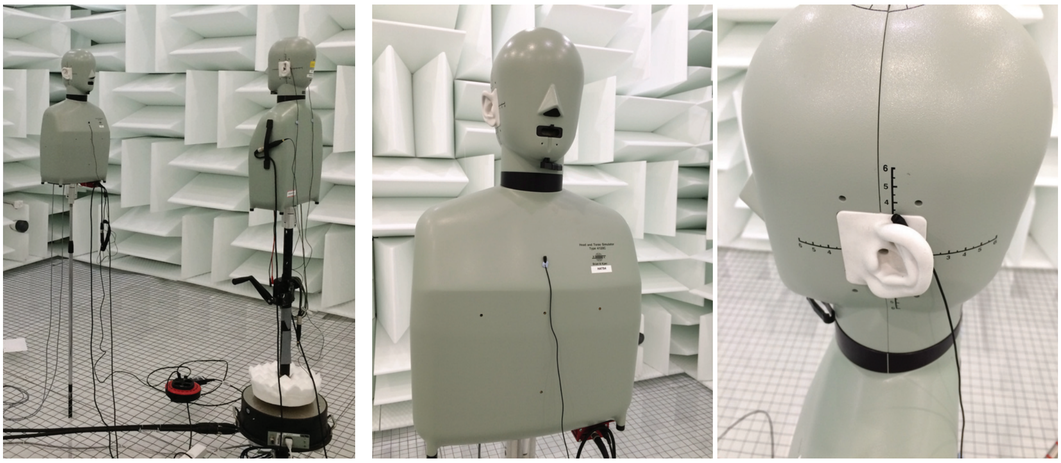


Fig. 1: Set-up for recording acoustic stimuli (left panel). Position of remote mic on “talking” HATS (middle panel). Position of microphones on “listening” HATS (right panel)

Participants

15 normal-hearing (NH) and 9 HI listeners participated in the study. All NH listeners had air-conduction audiometric thresholds below 25 dB between 125 Hz and 4 kHz. The average audiogram of the HI listeners is plotted in Fig. 2.

Procedure

A 3-interval 3-alternative forced-choice paradigm using a 1-up 2-down adaptive rule was used to estimate MAA. Between each run of the adaptive procedure, the target angle (i.e., the direction used in two of three intervals) was roved uniformly between 40–50°. At the start of each run, the initial angular difference was always as large as possible. The initial step size was 8° and the step size was halved after each reversal until the minimum step size of 1° was reached. Runs were terminated after four reversals and the threshold was estimated as the mean of the angular difference between target and presented angular value at the last two reversals. Sentences were randomly chosen between trials but remained the same within each 3-interval triplet.

For each presentation, two audio files were loaded, one with the HA signal and one with the RM signal. According to the current state of the test, a gain of either 0, -3 or -6 dB was applied to the RM signal and a delay of either 0, 10, 20, 40, 60, 80 or 100 ms was applied to the RM signal. After applying gain and delay to the signals, they were mixed into one stereo file, which was then presented to the listener via headphones.

A further control condition, in which only the HA recordings were presented, was conducted to estimate the baseline MAA performance with microphones in BTE HA positions.

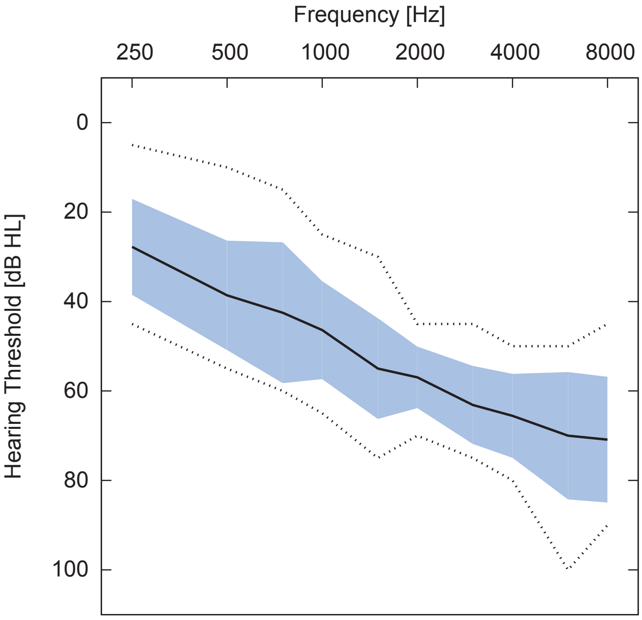


Fig. 2: Audiometric thresholds of the hearing-impaired listeners. The solid black line indicates the mean hearing threshold, the shaded region indicates one standard deviation, and the dotted lines indicate minimum and maximum measured thresholds.

To compensate for reduced audibility, the mixed stimuli was further amplified using the linear Cambridge Formula (Moore and Glasberg, 1998) for each individual HI listener.

No correction was applied to compensate for the acoustic delay between RM and HA microphones, which was approximately 2.5 ms. Thus, in the 0 ms condition, RM signals preceded HA signals.

RESULTS

The average MAA as a function of relative delay for the three relative gains is plotted in Fig 3. The lower dotted line in each panel indicates the MAA threshold for the control condition when no RM signal was mixed with the HA recordings. Thus, this estimates the minimum MAA listeners could achieve when using ideal BTE HAs. As expected, NH listeners exhibited lower MAA overall than the HI listeners.

For the normal hearing listeners, two overall trends were observed: (1) MAA thresholds decreased as the RM gain was decreased; (2) MAA thresholds were minimized when the relative delay was between 10–20 ms. The same trends were observed in the HI listeners but the size of the effects were smaller.

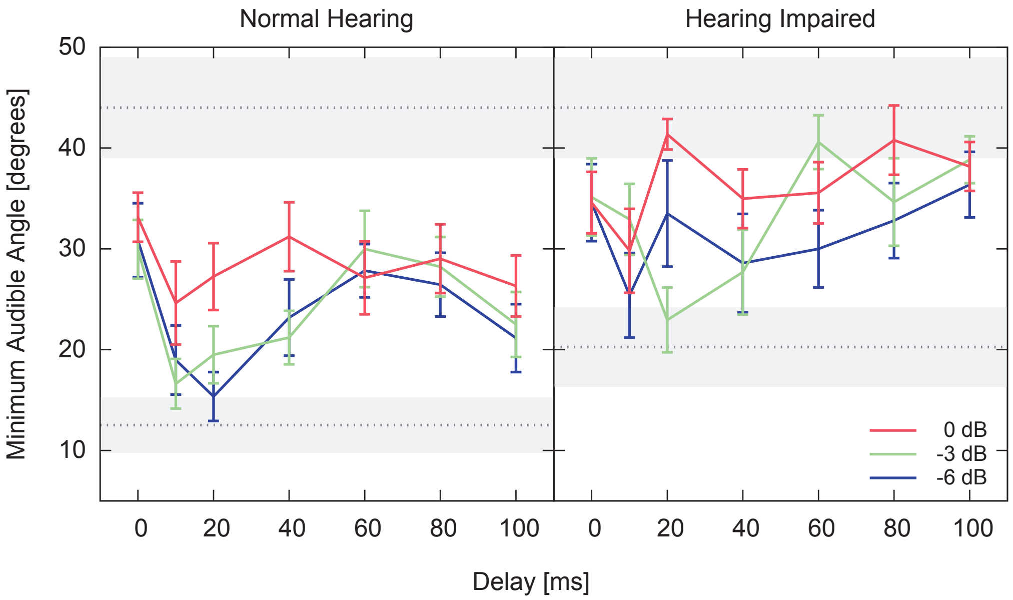


Fig. 3: Average MAA as a function of relative delay for three different relative gains of RM vs HA microphone signals for normal-hearing (left panel) and hearing-impaired listeners (right panel). The lower dotted line and shaded area indicate the average MAA for the HA only condition. The upper dotted lines indicate chance performance. The bars and shaded areas indicate standard error.

These observations were confirmed by the results from a repeated measures ANOVA with gain and delay as within-group and hearing status as between-group factors. Statistically significant main effects of gain [$F(2,46) = 9.002$, $p = 0.01$] and delay [$F(6,138) = 7.637$, $p < 0.001$] and hearing status [$F(1,23) = 304.827$, $p < 0.001$] were observed.

DISCUSSION

For the NH listeners, MAA was smallest when RM gain was reduced and the RM delay was approximately 10–20 ms. Indeed, for relative gains of -3 and -6 dB, NH listeners exhibited MAA thresholds that were similar to those they achieved with only the microphone signals from the simulated BTE positions. A similar pattern of results was obtained from the HI group of listeners. However, their overall MAA thresholds were higher and the size of the effects were smaller.

For speech, previous studies have found echo thresholds ranging from 30–50 ms (Lochner and Burger, 1958; Haas, 1951). Thus, a relative delay of the RM signal of 10–20 ms should result in a fused percept of a single talker and the precedence effect suppresses the incongruent spatial information presented by the RM signal.

Some of the largest MAA thresholds were obtained in the 0-ms delay conditions. However, no correction was applied to compensate for the acoustic delay between RM and HA microphone positions. As a result, in the 0-ms condition, RM signals preceded HA signals by approximately 2.5 ms. Thus, it is not surprising that larger MAA thresholds were observed in these conditions as the spatial cues available from the HA microphones were likely suppressed by the precedence effect.

Overall, the present study suggests that to achieve optimal spatial perception, the signal from the RM microphone should be mixed with both little gain as possible and a relative delay of 10–20 ms (after compensating for the difference in acoustic delay between the RM and HA). However, there are some potential issues with this advice.

First, the reason for employing an RM is to improve speech intelligibility and/or reduce listening effort by providing the HI listener with a higher quality speech signal (i.e., one with a higher SNR and/or direct-to-reverberant energy ratio). Thus, reducing the gain of the RM reduces its potential benefit for speech intelligibility and listening effort.

Second, in current devices, the delay caused by mixing the RM signal is heavily influenced by the communication protocols used in the digital transmission of the signal from RM to HA. As the actual delay of each device is proprietary, it is difficult to compare our results to the delays that are present in products that are currently available on the market. That said, common digital transmission protocols, such as Bluetooth, can result in delays of 100 ms or more. For use cases involving multiple RMs, much lower delays are likely desirable.

Assuming a sufficiently low-latency transmission protocol was employed, achieving the suggested delay target requires the HA to estimate and compensate for the acoustic delay differences between RM and HA microphones. While this presents a technical challenge, other research into improving selective attention in HI listeners assumes similar requirements (e.g., Favre-Felix *et al.*, 2017).

In the present study, the HA signals were simulated using omnidirectional microphones positioned at the pinnae. Thus, the signals presented replicated an “ideal” HA (i.e., full bandwidth) and did not include standard HA signal processing that might affect spatial perception (e.g., directional microphones/beamforming, compression, and noise-reduction). Further, the signals were recorded in anechoic conditions. Thus, MAA thresholds in more realistic conditions are likely to be larger.

CONCLUSION

Based on the results from the present study, the detrimental effects of an RM on localization can be minimized by targeting a relative delay of 10–20 ms between the RM and HA signals. Further, the gain of the RM should be reduced as much as is possible while still maintaining its beneficial effects on speech intelligibility.

ACKNOWLEDGEMENTS

This experiment was conducted as a bachelor project by the first author in collaboration with GN ReSound. The authors would like to thank Andreas Schousboe from GN ReSound A/S for his help in facilitating the project.

REFERENCES

American Academy of Audiology Clinical Practice Guidelines (**2011**). “Remote microphone hearing assistance technologies for children and youth from birth to 21 years.”

Boothroyd, A. (**2004**). Hearing aid accessories for adults: The remote FM microphone. Ear Hearing, **25**, 22-33. doi: 10.1097/01.AUD.0000111260.46595.EC

Favre-Felix, A., Graversen, C., Dau, T., and Lunner, T. (**2017**). “Steering of audio input in hearing aids by eye gaze through in-ear electrodes,” in *Adaptive Processes in Hearing*, Proc. ISAAR, **6**, in press.

Haas, H. (**1951**). “On the influence of a single echo on the intelligibility of speech,” Acustica, **1**, 48-58.

Hawkins, D.B. (**1984**). “Comparisons of speech recognition in noise by mildly-to-moderately hearing-impaired children using hearing aids and FM systems,” J. Speech Hear. Disord., **49**, 409-418. doi: 10.1044/jshd.4904.409

Lochner, J.P.A., and Burger, J.F. (**1958**). “The subjective masking of short time delayed echoes by their primary sounds and their contributions to the intelligibility of speech,” Acustica, **8**, 1-10.

Moore, B.C.J., and Glasberg, B.R. (**1998**). “Use of loudness model for hearing-aid fitting. I. Linear hearing aids,” Brit. J. Aud., **32**, 317-335. doi: 10.3109/03005364000000083

Nábělek, A.K., and Donahue, A.M. (**1986**). “Comparison of amplification systems in an auditorium,” J. Acoust. Soc. Am., **79**, 2078-2082. doi: 10.1121/1.393167

Ross, M., and Giolas, T.G. (**1971**). “Effect of three classroom listening conditions on speech intelligibility,” Am. Ann. Deaf, **116**, 580-584.

Sørensen, A.J., Weisser, A., and MacDonald, E.N. (**2017**). “Preliminary investigation of the categorization of gaps and overlaps in turn-taking interactions: Effects of noise and hearing loss,” in *Adaptive Processes in Hearing*, Proc. ISAAR, **6**, in press.

List of authors