Individual Hearing Loss – Characterization, Modelling, Compensation Strategies

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

www.isaar.eu

ISAAR is organized by the Danavox Jubilee Foundation
and supported by GN ReSound

Members of the board of the Danavox Jubilee Foundation (2015)

ISBN 978-87-990013-5-4 EAN 9788799001347

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

Layout: Graphic Mind
Printed by Centertryk A/S 2016

Preface

The 5th International Symposium on Audiological and Auditory Research (ISAAR) was held at Hotel Nyborg Strand in Nyborg, Denmark, from August 26 to 28, 2015. Two-hundred colleagues from all over the world participated; 30 talks and 53 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 characterization, modelling, and compensation of individual hearing loss. Different perspectives were presented and discussed, including individual differences in impaired auditory perception; genetics of hearing loss; supra-threshold deficits and neural degeneration in the presence of normal hearing thresholds; modelling of individual hearing loss; as well as novel hearing rehabilitation and compensation strategies in state-of-the-art hearing instruments.

The goal of the symposium was to gain insights from current research in different areas and disciplines within hearing science and to relate some of the findings across these disciplines. The programme was comprised of the following sections: characterization of individual differences in hearing loss; genetics of hearing loss; "hidden hearing loss" and neural degeneration in "normal" hearing; modelling of individual hearing impairment; and hearing rehabilitation with hearing aids and cochlear implants. 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 ReSound for the economic 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 Lene Jørgensen and her group at GN ReSound for their preparation of all 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 2015

Scientific

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

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

Lisbeth Tranebjærg, University of Copenhagen, Copenhagen, Denmark

Ture Andersen, Odense University Hospital, Odense, Denmark

Sébastien Santurette, 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, Denmark

About ISAAR

The "International Symposium on Auditory and Audiological Research" is formerly known as the "Danavox Symposium". The 2015 edition was the 26th symposium in the series and the 5th symposium under the ISAAR name, adopted in 2007. The Danavox Jubilee Foundation was established in 1968 on the occasion of the 25th anniversary of GN Danavox. The aim of the foundation is to support and encourage audiological research and development.

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

A list of proceedings from previous symposia may be found at the ISAAR website: www.isaar.eu – 'Previous Symposia'. All contributions from previous symposia can be found, searched, and downloaded from the GN ReSound Audiological Library: www.audiological-library.gnresound.dk.

Contents

I: Characterizing individual differences in hearing loss

II: Hidden hearing loss: Neural degeneration in "normal" hearing

III: Modelling individual hearing impairment

Characterizing individual differences: Audiometric phenotypes of age-related hearing loss

JUDY R. DUBNO*

Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, Charleston, SC, USA

Metabolic presbyacusis, or the degeneration of the cochlear lateral wall and decline of the endocochlear potential, largely accounts for age-related threshold elevations observed in laboratory animals raised in quiet and may underlie the characteristic audiogram of older humans. The "audiometric phenotype" associated with metabolic presbyacusis differs from audiograms associated with sensory losses resulting from ototoxic drug and noise exposures. Evidence supporting metabolic and sensory phenotypes in audiograms from older adults can be derived from demographic information (age, gender), environmental exposures (noise and ototoxic drug histories), and stability or changes in audiometric phenotypes as individuals age. When confirmed with biological markers and longitudinal analyses, well-defined audiometric phenotypes of human age-related hearing loss can contribute to explanations of individual differences in auditory function for older adults.

INTRODUCTION

Naturally occurring age-related changes to the auditory periphery in older adults combine with damaging effects of a lifetime of environmental exposures and disease processes. Subsequent anatomic, physiologic, and neurochemical deficits result in reduced detection for low-level signals (hearing loss) and impaired suprathreshold auditory function, including complex signal processing and speech understanding. As such, the aging auditory periphery delivers degraded signal representations for processing by the central auditory pathways and cortex. At the same time, older adults may be increasingly affected by changes in cognitive abilities, including declines in working memory, executive function, attention, and processing speed; reduced ability to suppress irrelevant information; and inadequate compensation strategies. Taken together, these effects may impose increased cognitive demands on an aging brain with already limited resources and loss of inhibition. Thus, multiple risk factors (aging, noise, drugs, disease, infections, comorbid conditions) and multiple sources of pathology in the auditory system (hair cells and lateral wall of the cochlea, auditory nerve, central auditory pathways, cortex) contribute to large individual differences. These complex and interactive effects throughout the aging auditory system highlight the critical need for evidence to (1) allocate declines to each risk factor, especially aging, (2) explain individual differences, (3) identify

*Corresponding author: dubnojr@musc.edu

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

promising targets for intervention, and (4) develop strategies to prevent or delay the onset of age-related changes.

SOURCES OF AGE-RELATED PATHOLOGY

Lateral wall and stria vascularis

The cochlear lateral wall is responsible for production and maintenance of the endocochlear potential (EP), which is a positive voltage of 80-100 mV present in the endolymph of the scala media and serves as the battery that provides voltage to the outer hair cells (OHCs), or cochlear amplifier. Laboratory animals raised in quiet (e.g., "quiet-aged" gerbils) demonstrate a systematic degeneration of the lateral wall and reduced EP, which deprives the cochlear amplifier of its essential power supply (Schulte and Schmiedt, 1992; Schmiedt, 1996; Gratton *et al*., 1997). These changes (1) reduce cochlear amplifier gain in the lower frequencies by as much as 20 dB and in the higher frequencies by as much as 60 dB, and (2) reduce but maintain cochlear nonlinearities, such as compression and otoacoustic emissions (OAEs). Although OHCs are preserved, age-related reduction in EP results in changes in OHC function (Schmiedt *et al*., 1990; Schmiedt, 1996). The frequency-specific neural threshold loss of quiet-aged gerbils measured with compound action potentials (CAP) is associated with EP loss and is not associated with OHC loss, and defines the gradually sloping audiogram of older gerbils (Schmiedt *et al*., 2002; Lang *et al*., 2003; Lang *et al*., 2010; Mills *et al*., 2004; Schmiedt, 2010).

Outer hair cells and cochlear amplifier

With environmental exposures from ototoxic drug or excess noise exposure, sensory and non-sensory cell loss result in threshold shifts of ~50-70 dB, loss of the cochlear amplifier, and loss of cochlear nonlinearities (such as absent OAEs). These characteristics are not seen in quiet-aged gerbils (Mills *et al*., 1990; Schmiedt *et al*., 1990; Tarnowski *et al*., 1991).

Primary auditory neurons

Quiet-aged gerbils also demonstrate primary neural degeneration, which is not related to sensory cell loss. The spiral ganglion cells are reduced in size and number along the entire cochlear duct, and there is selective loss or inactivity of low spontaneous-rate auditory nerve fibers (Hellstrom and Schmiedt, 1990; Schmiedt *et al*., 1996; Schulte *et al*., 1996; Suryadevara *et al*., 2001; Lang *et al*., 2002; Mills *et al*., 2006). These results are consistent with evidence from human archival temporal bones, which show spiral ganglion cells declining with age, even without hair cell loss (Otte *et al*., 1978; Makary *et al*., 2011). In addition to primary neural degeneration in aging animal models and humans, the early noise trauma mouse model (Kujawa and Liberman, 2009) also shows a loss of hair cell synapses and terminals, loss of spiral ganglion neurons, and selective loss of low spontaneous-rate auditory nerve fibers. Of importance is that this "neural presbyacusis" (1) can occur without threshold elevation (i.e., with a normal audiogram), (2) affects neural coding at high signal levels (i.e., shallow CAP amplitude-intensity or input-output functions) and in noise, and affects suprathreshold auditory behavior, all of which can put older adults at greater disadvantage. Moreover, these neural declines, along with deterioration of the lateral wall and hair cell dysfunction, represent additional sources of individual differences.

ARE AGE-RELATED PATHOLOGIES SEEN IN HUMAN AUDIOGRAMS?

The observance of a characteristic audiogram of older laboratory animals raised in quiet (a mild, flat hearing loss at lower frequencies coupled with a gradually sloping hearing loss at higher frequencies) led to the question of whether metabolic presbyacusis also defines the gradually sloping audiogram of older humans. That is, we were interested in determining if age-related conditions of cochlear and neural pathology, as described earlier from animal models, can be consistently observed in human audiograms of older adults. To answer this question, we first developed schematic boundaries for 5 audiometric phenotypes, based on 5 hypothesized conditions: older-normal, pre-metabolic, metabolic, sensory, and metabolic+sensory (Schmiedt, 2010; Dubno *et al*., 2013; see Fig. 1). The combined metabolic+sensory phenotype is consistent with the notion that, in contrast to quiet-aged gerbils, audiograms of older adults likely reflect the effects of environmental exposures (noise, drug) combined with age-related declines in the auditory periphery unrelated to these exposures. Next, we searched initial audiograms stored in the MUSC longitudinal human subject database for “exemplars”, (best examples) of each phenotype. Of 1,728 initial audiograms (obtained at enrollment in the longitudinal study), 22% were identified by expert raters as exemplars with no knowledge of subject demographics.

Validation of audiometric phenotypes

To validate this approach, we predicted the phenotypes of the exemplar audiograms using three machine learning tools, Support Vector Machines, Random Forests, and nonlinear Quadratic Discriminant Analysis (QDA). Each of the machine learning tools classified the audiograms by comparing pure-tone thresholds to a prior distribution and finding the maximum probability. Nonlinear QDA was selected because covariances across frequency are not equivalent and the exemplar phenotypes (thresholds as a function of frequency) are nonlinear. Each of the three machine learning tools replicated expert judgements with a similarly high degree of accuracy (93.2% for QDA). Given that the results for the three procedures were comparable, QDA was selected as the procedure for future analyses based on its ability to capture the nonlinearities in the audiogram and because it is a widely understood method as compared to the other two methods. Also at this time, the decision was made to eliminate the pre-metabolic phenotype because only a small number of audiograms (3%) were classified. An additional concern was that pre-metabolic is not a distinct phenotype, but an early stage of the metabolic phenotype (i.e., a transition from older-normal to metabolic). An automated classifier as used here provides a means to study new samples to further replicate and validate our results, with the long-term goal of evaluating genetic and biological mechanisms of

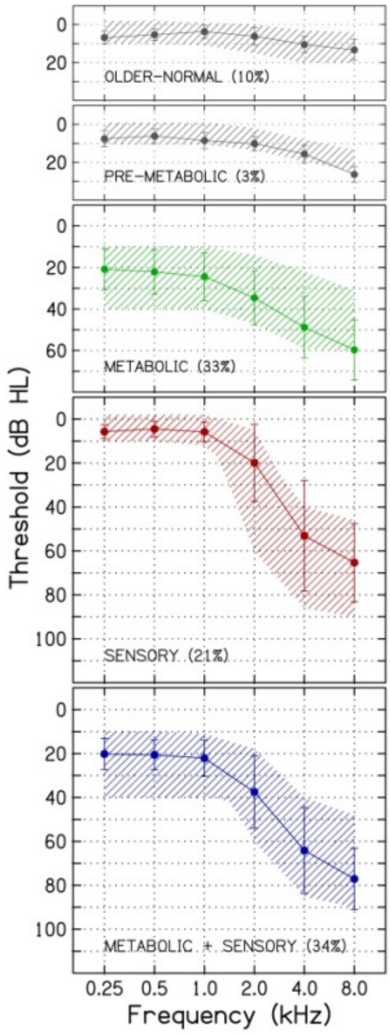


Fig. 1: Schematic boundaries of five phenotypes of age-related hearing loss (shaded regions). Symbols and error bars are mean thresholds (±1 standard error) of exemplar audiograms. See text for additional details.

age-related hearing loss in humans. Until that occurs, the phenotypic classifications, and their hypothesized underlying mechanisms based on animal models of metabolic and sensory loss (as described earlier), should be considered putative in nature (Dubno *et al.*, 2013).

How well do phenotypes correspond to predicted demographics?

We further assessed the accuracy of the classifier by determining how well the audiograms assigned to the phenotypes were consistent with the predefined schematic boundaries and corresponded to predicted demographics, such as age, gender, and noise exposure history. Individual estimates of noise history were obtained from a 7-item self-report questionnaire on occupational and non-occupational noise exposures – see Lee *et al.* (2005) and Dubno *et al.* (2013) for additional results. Results showed that, on average, individuals with older-normal phenotypes were youngest, whereas individuals with metabolic phenotypes were oldest (consistent with EP declining with increasing age). Individuals with sensory phenotypes (sensory and metabolic+sensory) were primarily male, whereas older-normal and metabolic phenotypes were primarily female. Sensory phenotypes were more likely to have positive noise exposure histories, whereas older-normal and metabolic phenotypes were less likely to have positive noise histories. Next, we validated the approach by classifying non-exemplar initial audiograms according to the four phenotypes (N=1,379). QDA classifications showed high consistency of threshold, age, gender, and noise histories within groups (Dubno *et al*., 2013). Thus, using cross-sectional data, classifications of audiometric phenotypes were consistent with expert judgements, and revealed that individuals with audiograms classified as metabolic phenotypes were older, predominately female, and had negative noise exposure histories, consistent threshold elevations resulting from a declining EP.

Using longitudinal data from the MUSC human subject database to assess stability of audiometric phenotypes over time

Using longitudinal data from the MUSC human subject database, we determined the likelihood of metabolic phenotypes increasing with age. The human subject database currently contains data from ~1,500 participants (~450 active participants), of which 69% are age 60 and older, 60% are female, and nearly 30% are racial/ethnic minorities. The database contains more than 20,000 audiograms (more than 10,000 lab visits × 2 ears). Participants of all ages are recruited from the Charleston area, including local audiology and otolaryngology clinics, assisted living facilities, senior centers, and health fairs. Participants must be 18 and older, in good general health to be able to visit the laboratory multiple times, and no evidence of conductive hearing loss, active otologic disease, or significant cognitive decline. There is no restriction on amount of hearing loss, but hearing abilities must be good enough to provide measurable results on a majority of the test battery. Measures are repeated yearly or every 2-3 years for longitudinal data.

Audiometric measures include hearing for pure tones, including extended high frequencies, ability to understand speech in quiet and in noise, otoacoustic emissions, upward and downward spread of masking, middle ear function, and auditory brainstem responses (e.g., Lee *et al*., 2005; Dubno *et al*., 1997; 2008). Study participants provide oral or written responses to self-report questionnaires on medical history, prescription and over-the-counter drugs, noise history, hearing-aid history, hearing handicap, tinnitus, smoking, and handedness. A cognitive battery includes tests of attention, working memory, processing speed, and perceived workload. Brain imaging is obtained on a subset of participants while they are listening to and understanding low-pass filtered speech or speech in background noise. Each participant has an otologic exam and provides blood for clinical chemistries and to extract DNA for whole exome sequencing. Finally, participants are offered the opportunity to donate their temporal bones for future structural-functional analyses.

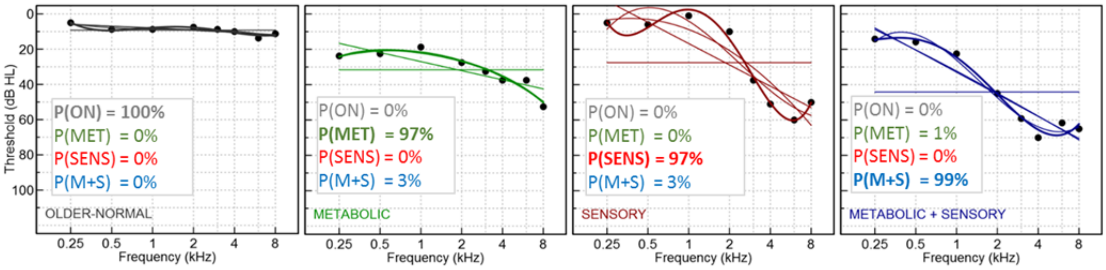


Fig. 2: Examples of audiometric phenotypes classified based on fitted curve parameters. Each filled circle is a pure-tone threshold and lines represent fits using 1-5 parameters. The legend in each panel includes the probabilities assigned to each classification. *Notes*: ON=older-normal; MET=metabolic; SENS=sensory; M+S=metabolic+sensory.

For this next phase of phenotype classifications, a new procedure was introduced, whereby phenotypes were classified based on fitted curve parameters, and then selected as before based on the maximum probability (Fig. 2). With 5 parameters, a cross-validated accuracy of 94.4% was obtained using the curve-based approach and smoothed audiograms from “clustered” time points. Laboratory visits (and pure-tone thresholds and other measurements) occur in clusters of several visits within a short time-frame (less than one year), which are then repeated every 2-3 years. Therefore, longitudinal data were defined as 2 or more clustered time points, where a cluster is 3 or more audiograms within one year. This resulted in ~7,700 audiograms averaged into 1,826 clusters from 686 ears (ranging in age from 50-93). Participants with missing data were excluded.

Phenotypes were found to be stable over time for a majority of ears (64%). In addition, a majority of right/left ears had the same phenotype (71%) and most ears matched across all time points (89%). Nevertheless, although a majority of individual ears maintained the same phenotype with increasing age, pure-tone thresholds increased (as demonstrated by longitudinal changes in thresholds obtained from serial audiograms). These increases in thresholds varied with phenotype and gender, with thresholds for metabolic phenotypes showing greater declines with increasing age.

For the 36% of ears with phenotypes that changed with increasing age, unique patterns were observed. Older-normal phenotypes transitioned to each of the other three phenotypes with approximately equal probability. Metabolic phenotypes transitioned primarily to metabolic + sensory phenotype, as did sensory phenotypes (Fig. 3). Nearly all metabolic+sensory phenotypes transitioned to the metabolic phenotype; probabilities for the initial and final phenotypes were typically less than 1.0, indicating that these audiograms may have been in an intermediate stage.

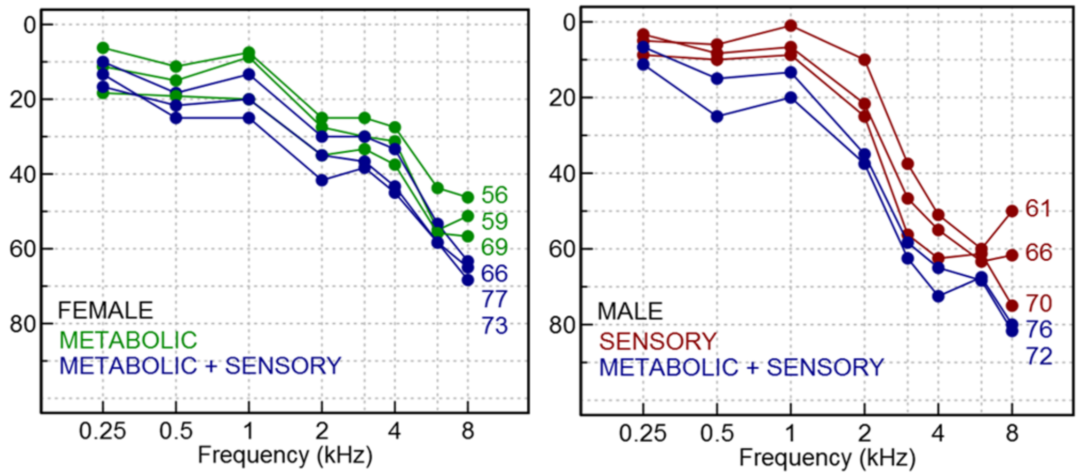


Fig. 3: Serial audiograms from two study participants illustrating phenotypes changing with age. In both cases, the transition was to the metabolic+sensory phenotype, but the initial phenotype was metabolic in one case (female, left panel) and sensory in the other case (male, right panel). Ages at the times of the measured audiograms are indicated on the right side of each panel.

Do ears with metabolic phenotypes increase with age?

An analysis of the numbers and percentages of ears that changed phenotypes with increasing age indicated that of those that changed, most changed to metabolic phenotypes. Specifically, 50.8% of ears transitioned to the metabolic + sensory phenotype and 23.6% of ears transitioned to the metabolic phenotype (Fig. 4). Moreover, these transitions to metabolic phenotypes occurred at older ages than transitions to other phenotypes. Finally, transitions to metabolic phenotypes were much more likely in females than males (76-85% were female), except for the transition from sensory to the metabolic+sensory phenotype, for which 70% were males.

Ongoing analyses designed to provide additional validation include assessing phenotypes with additional measures of auditory function measured longitudinally (including OAEs and speech recognition) and confirming with biological markers (genetics and otopathology from human temporal bones). For genetic analyses, audiometric phenotypes provide a framework beyond classifying older adults as either “affected” (hearing impaired) or “non-affected” (normal hearing). Currently, our approach is to search for genetic associations and structural variations in genes related to metabolic vs. non-metabolic phenotypes, which may also explain individual differences. Following that, we will initiate studies to determine the pathological and potential functional consequence of genetic variations as they relate to phenotypes of age-related hearing loss, largely through studies of human temporal bones. Such information can also drive the development of mouse models with specific mutations. Future studies will apply this phenotypic approach to understanding neural presbyacusis.

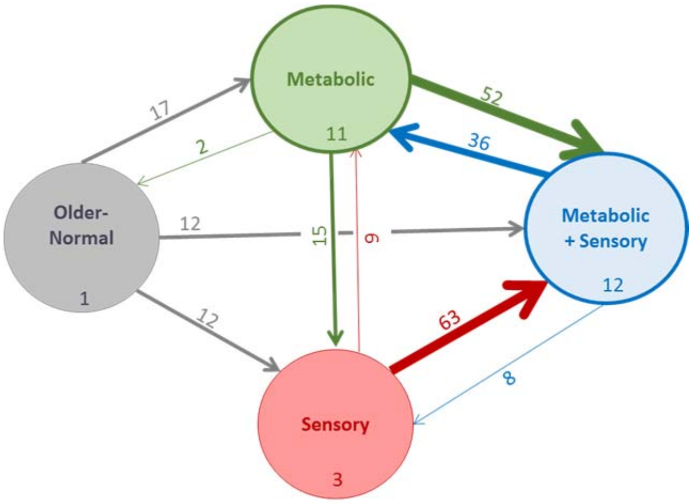


Fig. 4: Schematic illustrating transitions of ears from initial to final phenotypes. Thickness of arrows and numbers adjacent to arrows correspond to number of ears by initial and final phenotype. Thickness of borders around each circle and numbers within each circle indicate the number of ears that changed phenotype but ultimately returned to their initial phenotype. See text for additional details.

SUMMARY AND CONCLUSIONS

In summary, audiograms from middle age to older adults are consistent with predictions from animal findings associated with sensory and strial pathology. Audiograms appear to contain information about distinct presbyacusis phenotypes and also reflect large individual differences. A machine learning algorithm was trained to classify audiograms based on expert ratings (animal models) and fitted curve parameters. Classifications were consistent with phenotypic predictions based on thresholds, age, gender, and noise history. Analysis of longitudinal data showed a majority with stable phenotypes over time, even while hearing loss was increasing. The remainder showed changes in phenotypes with increasing age, with the most common change to metabolic phenotypes. Changes in phenotype differed with age and gender, also consistent with metabolic presbyacusis increasing with age. In conclusion, audiometric phenotypes are consistent with the view of age-related hearing loss as a metabolic disorder rather than a sensory disorder.

ACKNOWLEDGEMENTS

This project is supported by a grant from the National Institutes of Health/National Institute on Deafness and Other Communication Disorders (NIH/NIDCD). Significant contributions from colleagues Kenneth I. Vaden, Mark A. Eckert, Lois J. Matthews, Richard A. Schmiedt, Bradley A. Schulte, and Jayne B. Ahlstrom are gratefully acknowledged.

REFERENCES

Dubno, J.R., Lee, F.S., Matthews, L.J., and Mills, J.H. (**1997**) "Age-related and gender-related changes in monaural speech recognition," J. Speech Lang. Hear. Res., **40**, 444-452.

Dubno, J.R., Lee, F.S., Matthews, L.J., Ahlstrom, J.B., Horwitz, A.R., and Mills, J.H. (**2008**). "Longitudinal changes in speech recognition in older persons," J. Acoust. Soc. Am., **123**, 462-475.

Dubno, J.R., Eckert, M.A., Lee, F.S., Matthews, L.J., and Schmiedt, R.A. (**2013**). "Classifying phenotypes of age-related hearing loss from human audiograms," J. Assoc. Res. Otolaryngol., **14**, 687-701.

Gratton, M.A., Schulte, B.A., and Smythe, N.M. (**1997**). "Quantification of stria vascularis and strial capillary areas in young and old gerbils raised in quiet," Hear. Res., **114**, 1-9.

Hellstrom, L.I. and Schmiedt, R.A. (**1990**). "Compound action potential input/output functions in young and quiet-aged gerbils," Hear. Res., **50**, 163-174.

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.

Lang H, Schulte, B.A., and Schmiedt, R.A. (**2002**). "Endocochlear potentials and compound action potential recovery: functions in the C57BL/6J mouse," Hear. Res., **172**, 118-126.

Lang, H., Schulte, B.A., and Schmiedt, R.A. (**2003**). "Effects of chronic furosemide treatment and age on cell division in the adult gerbil inner ear," J. Assoc. Res. Otolaryngol., **4**, 164-175.

Lang, H., Jyothi, V., Smythe, N.M., Dubno, J.R., Schulte, B.A., and Schmiedt, R.A. (**2010**). "Chronic reduction of endocochlear potential reduces auditory nerve activity: further confirmation of an animal model of metabolic presbyacusis," J. Assoc. Res. Otolaryngol., **11**, 419-434.

Lee, F.S., Matthews, L.J., Dubno, J.R., and Mills, J.H. (**2005**) "Longitudinal study of pure-tone thresholds in older persons," Ear. Hear., **26**, 1-11.

Makary, C.A., Shin, J., Kujawa, S.G., Liberman, M.C., and Merchant, S.N. (**2011**). "Age-related primary cochlear neuronal degeneration in human temporal bones," J. Assoc. Res. Otolaryngol., **12**, 711-717.

Mills, J.H., Schmiedt, R.A., and Kulish, L.F. (**1990**). "Age-related changes in auditory potentials of Mongolian gerbil," Hear. Res. **46**, 201–210.

Mills, D.M. and Schmiedt, R.A. (**2004**). "Metabolic presbycusis: Differential changes in auditory brainstem and otoacoustic emission responses with chronic furosemide application in the gerbil," J. Assoc. Res. Otolaryngol., **5**, 1-10.

Mills, J.H., Schmiedt, R.A., Schulte, B.A., and Dubno, J.R. (**2006**). "Age-related hearing loss: A loss of voltage, not hair cells," Semin. Hear., **27**, 228-236.

Otte, J., Schuknecht, H.F., and Kerr, A.G. (**1978**). "Ganglion cell populations in normal and pathological human cochleae. Implications for cochlear implantation," Laryngoscope, **88**, 1231-1246.

Schmiedt, R.A., Mills, J.H., and Adams, J.C. (**1990**). "Tuning and suppression in auditory nerve fibers of aged gerbils raised in quiet or noise," Hear. Res., **45**, 221-236.

Schmiedt, R.A. (**1996**). "Effects of aging on potassium homeostasis and the endocochlear potential in the gerbil cochlea," Hear. Res., **102**, 125-132.

Schmiedt, R.A., Mills, J.H., and Boettcher, F.A. (**1996**). "Age-related loss of activity of auditory-nerve fibers," J. Neurophysiol., **76**, 2799-2803.

Schmiedt, R.A., Lang, H., Okamura, H.O., and Schulte, B.A. (**2002**). "Effects of furosemide applied chronically to the round window: a model of metabolic presbyacusis," J. Neurosci., **22**, 9643-9650.

Schmiedt, R.A. (**2010**). "The physiology of cochlear presbyacusis," in *The aging auditory system: Perceptual characterization and neural bases of presbyacusis*, Eds. S. Gordon-Salant, R.D. Frisina, A.N. Popper, and R.R. Fay (Springer, New York), pp. 9-38.

Schulte, B.A. and Schmiedt, R.A. (**1992**). "Lateral wall Na,K-ATPase and endocochlear potentials decline with age in quiet-reared gerbils," Hear. Res., **61**, 35-46.

Schulte, B.A, Gratton, M.A., Smythe, N., and Lee, F.S. (**1996**). "Morphometric analysis of spiral ganglion neurons in young and old gerbils raised in quiet," Assoc. Res. Otolaryngol., **19**, 160.

Suryadevara, A.C., Schulte. B.A., Schmiedt, R.A., and Slepecky, N.B. (**2001**). "Auditory nerve fibers in young and quiet-aged gerbil: Morphometric correlations with endocochlear potential," Hear. Res., **161**, 45-53.

Tarnowski, B.I., Schmiedt, R.A., Hellstrom, L.I., Lee, F.S., and Adams, J.C. (**1991**). "Age-related changes in cochleas of Mongolian gerbils," Hear. Res., **54**, 123-134.

Beyond the audiogram: Influence of supra-threshold deficits associated with hearing loss and age on speech intelligibility

AGNÈS C. LÉGER[1,*], CHRISTIAN LORENZI[2], AND BRIAN C. J. MOORE[3]

[1] *School of Psychological Sciences, University of Manchester, Manchester, UK*

[2] *Laboratoire des Systèmes Perceptifs, UMR CNRS 8248, Département d'Etudes Cognitives, Institut d'Etudes de la Cognition, École Normale Supérieure, Paris, France*

[3] *Department of Experimental Psychology, University of Cambridge, Cambridge, UK*

Sensorineural hearing loss and greater age are associated with poor speech intelligibility, especially in the presence of background sounds. The extent to which this is due to reduced audibility or to supra-threshold deficits is still debated. The influence of supra-threshold deficits on intelligibility was investigated for normal-hearing (NH) and hearing-impaired (HI) listeners with high-frequency losses by limiting the effect of audibility. The HI listeners were generally older than the NH listeners. Speech identification was measured using nonsense speech signals filtered into low- and mid-frequency regions, where pure-tone sensitivity was near normal for both groups. The older HI listeners showed mild to severe intelligibility deficits for speech presented in quiet and in various backgrounds (noise or speech). Overall, these results suggest that speech intelligibility can be strongly influenced by supra-threshold auditory deficits.

INTRODUCTION

Both sensorineural hearing loss and greater age are associated with poorer-than-normal speech intelligibility (for reviews, see George *et al.*, 2006; Moore, 2007; Rhebergen *et al.*, 2010a; 2010b), especially for speech presented in background sounds. Some authors have suggested that the problems arise primarily from reduced audibility (e.g., Desloge *et al.*, 2010; Humes *et al.*, 1987; Lee and Humes, 1993; Zurek and Delhorne, 1987), i.e., from the fact that parts of the speech cannot be heard at all. Others have suggested that the problems arise not only from reduced audibility, but also from supra-threshold deficits that lead to perceived distortion or lack of clarity of the speech signal (e.g., Dreschler and Plomp, 1980; 1985; Glasberg and Moore, 1989; Plomp, 1978; 1986), i.e., from a reduced ability to discriminate the acoustic features of the speech, despite it being audible. The studies reviewed here aimed at investigating specifically the influence of supra-threshold deficits on the intelligibility of speech.

*Corresponding author: agnes.leger@manchester.ac.uk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

Many studies have attempted to tease apart the contribution of reduced audibility and supra-threshold auditory deficits to speech intelligibility, especially for speech in complex backgrounds (e.g., Bernstein and Grant, 2009; Christiansen and Dau, 2012; Léger *et al.*, 2012b; Rhebergen *et al.*, 2006; Strelcyk and Dau, 2009). In many studies, reduced audibility (as estimated using audiometric thresholds) was not sufficient to explain the deficits of the hearing-impaired (HI) and/or elderly listeners (e.g., Bernstein and Grant, 2009; Dubno *et al.*, 2002; Füllgrabe *et al.*, 2015; Grose *et al.*, 2009; Hopkins and Moore, 2011; Horwitz *et al.*, 2002; Humes, 2002; Lorenzi *et al.*, 2006; Neher *et al.*, 2012; Sheft *et al.*, 2012; Summers *et al.*, 2013). Several supra-threshold deficits have been identified, including reduced frequency selectivity and reduced temporal processing (especially processing of the temporal fine structure [TFS] of the signal); see Moore (2007, 2014) for reviews. However, in some other studies, audibility has been suggested to fully explain the deficits of the HI listeners (e.g., Desloge *et al.*, 2010; Phatak and Grant, 2012). Thus, it is still unclear to what extent supra-threshold deficits contribute to the speech intelligibility deficits of the elderly and/or HI listeners. The goal of the studies reviewed here was to estimate the influence of supra-threshold deficits while controlling for the effect of audibility, therefore disentangling those two factors.

To control for, or at least reduce the influence of audibility and level differences of the stimuli for normal-hearing (NH) and HI listeners, speech intelligibility was compared for stimuli filtered into frequency regions where the audiometric thresholds were normal or near-normal for both groups. The results of previous studies using this approach (e.g., Horwitz *et al.*, 2002; Strelcyk and Dau, 2009) suggested that HI listeners with a high-frequency hearing loss could have speech processing deficits at lower frequencies. Several studies (Léger *et al.*, 2012b; 2012c; 2014) conducted using this approach are reviewed here. Note that the HI listeners were often older than the NH listeners; the effects of age are considered in the analyses that follow.

METHODS

Listeners

Listeners were informed about the goals of the studies and provided written consent before their participation. All studies were approved by French Regional Ethics Committee. Listeners were native French speakers and had no history of cognitive impairment or psychiatric disorders. A total of 112 listeners were tested in the studies reported here. Listeners were classified as NH or HI, based on their audiometric thresholds. Individual and mean audiometric thresholds are shown in Fig. 1.

A total of 63 NH listeners were tested. They had normal (≤20 dB HL) audiometric thresholds for octave-spaced frequencies between 0.125 and 8 kHz, except for 5 older listeners with audiometric thresholds of 25 dB HL at 6 and/or 8 kHz. The NH listeners were aged 20 to 61 years (mean=33 years, median 25 years, SD=13 years).

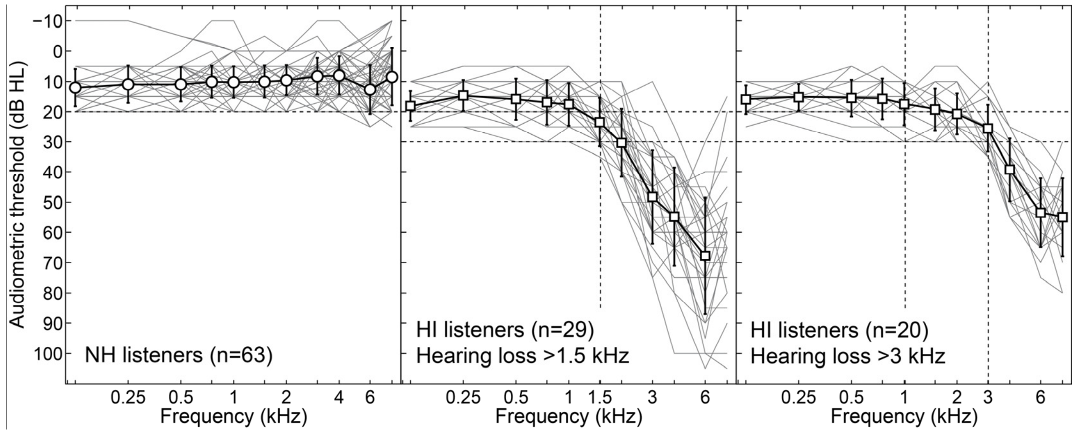


Fig. 1: Audiometric threshold (in dB HL) as a function of frequency for the NH listeners (left panel) and HI listeners. The HI listeners had near-normal audiometric thresholds up to either 1.5 kHz (middle panel) or 3 kHz (right panel), and a hearing loss above that frequency. In each panel, the grey lines show individual audiograms, and the thick black line shows the average audiogram (error bars: standard deviation, SD). The horizontal dotted lines show the limits of normality (20 dB HL) and near-normality (30 dB HL) for audiometric thresholds. The vertical dotted lines show the limits of some of the frequency regions of test for the HI listeners (see text).

A total of 49 HI listeners were tested. They had normal (≤20 dB HL) or near-normal (≤30 dB HL) audiometric thresholds for octave frequencies between 0.125 kHz and a cutoff frequency N_f, and a moderate to severe hearing loss at higher frequencies. The value of N_f was 1.5 kHz for 29 listeners and 3 kHz for the remaining 20 listeners. All losses were of sensorineural origin, as confirmed by the absence of air-bone gaps in the audiometric thresholds. The HI listeners were aged 20 to 76 years (mean=59 years, median=60 years, SD=13 years). An analysis of variance (ANOVA) unsurprisingly confirmed that the HI listeners were older than the NH listeners [$F(1,110)=104$, $p<0.001$]. An ANOVA was conducted on the pure-tone-averages in the low-frequency regions (up to 1.5 and 3 kHz), later referred to as PTA-f. Despite attempts at matching audiometric thresholds, the HI listeners had higher PTA-f than the NH listeners [$F(1,110)=79$, $p<0.001$]. On average, there was an 8-dB difference in PTA-f between the NH (mean=10 dB HL, SD=4 dB) and HI (mean=18 dB HL, SD=5 dB) listeners.

Speech materials

The methods used to measure speech intelligibility were similar across the studies reviewed here (Léger *et al.*, 2012b; 2012c; 2014). The reader is referred to those studies for details. Intelligibility was measured for speech signals filtered into three frequency regions. For the "*low-frequency region*", signals were low-pass filtered at 1.5 kHz. For the "*mid-frequency region*", signals were band-pass filtered between 1 and 3 kHz. For the "*low+mid-frequency region*", signals were low-pass filtered at

3 kHz. To prevent off-frequency listening, the filtered speech signals were always presented with a speech-shaped noise (SSN) filtered into the frequency region(s) outside of the region of test (e.g., above 1.5 kHz for speech low-pass filtered into the low-frequency region). This off-frequency noise was presented at a signal-to-noise ratio (SNR) of +12 dB. The HI listeners with a hearing loss above 1.5 kHz were tested only using the low-frequency region.

The speech signals were 48 Vowel-Consonant-Vowel (VCV) stimuli, each spoken twice by a female and a male native French speaker. Each set was composed of 16 consonants combined with three vowels. The four sets of VCVs (male and female speakers, two repetitions each) were used to generate a speech-shaped-noise (SSN). Note that in Léger *et al.* (2014), listeners were tested with the VCVs spoken by the male speaker only, the remaining VCVs being used as maskers (see below).

The filtered speech signals were presented at 65 dB SPL, except for HI listeners whose PTA-f was above 20 dB HL, in which case a frequency-independent gain equal to half the PTA was applied to (attempt to) restore audibility.

Background stimuli

The filtered speech signals were presented either in quiet (apart from the noise designed to limit off-frequency listening), or in an unmodulated or a modulated background. The unmodulated background was a SSN. The characteristics of the modulated backgrounds are described below. All listeners were tested with speech presented in quiet and in the unmodulated background; which listeners were tested with the various modulated backgrounds is reported below. Backgrounds were presented at fixed SNRs of −6, −3, and 0 dB. Backgrounds were filtered into the same frequency region as the speech signals they were presented with.

There were three types of modulated background. *"Backgrounds modulated in amplitude"*: a SSN was modulated in amplitude using an 8-Hz rectangular wave (modulation depth of 100%, random starting phase). Two duty cycles (DC, the percentage of time for which the masker was at full amplitude) were used to assess the effect of the duration of the temporal dips: 25% ("long dips") and 50% ("short dips"). "*Backgrounds modulated in spectrum"*: a SSN was passed through 32 non-overlapping gammatone filters each with a bandwidth of 1 ERB_N (equivalent rectangular bandwidth of the auditory filter for young listeners with normal hearing, Glasberg and Moore, 1990), and the outputs of the filters were multiplied by zero or 1 to introduce a spectral modulation. For "narrow dips", the pattern was 1, 0, 1, 0, 1, 0, 1, …, for "medium dips" the pattern was 1, 1, 0, 0, 1, 1, 0, 0, …, and for "wide dips" the pattern was 1, 0, 0, 0, 1, 0, 0, 0, 1, …. The value assigned to the lowest filter was randomised, thereby randomising the phase of the modulation. "*Speech*": VCVs were spoken by the female speaker (one VCV was chosen randomly for each trial). The fundamental frequency (f0) of the interfering speech was processed to assess the effect of the f0 separation between the target (male) and the interfering (female) speaker: the f0 separations were about 1 octave ("large f0 separation"), 3 semitones ("medium f0 separation"), and 1 semitone ("small f0 separation").

Procedure and analyses

In all studies, speech intelligibility was assessed by measuring consonant identification. Listeners were first tested with unfiltered VCVs in quiet, to familiarise them with the task; all listeners achieved scores of 80% correct or above. They were then tested with background sounds in a semi-random order (see each study for details).

Scores were converted into rationalized arcsine units (RAU; Studebaker, 1985) to make the data more suitable for ANOVAs. Because all listeners were not tested in the same frequency regions and background conditions, analyses on the results obtained by all listeners were not conducted here; the effects discussed below are supported by the analyses conducted for each study separately. Within each study and for each frequency region, ANOVAs were conducted on the scores with factors group (NH and HI) and condition (see papers for details). The influence of PTA-f and age was generally assessed using correlation analyses.

RESULTS AND DISCUSSION

Speech intelligibility scores

Consonant identification scores (hereafter referred to as “scores”) are shown in Fig. 2. Scores are shown in each background condition, for NH and HI listeners (slight offset between the two groups, shown in black and white, respectively), and for all frequency regions (slight offset between the regions, and coded by the symbols). Conclusions were similar for the different SNRs tested; therefore, in Fig. 2, individual scores were averaged over different SNRs.

The older HI listeners had poorer scores than the NH listeners in most conditions, despite the fact that both groups were tested in frequency regions of normal or near-normal audibility. These findings support the hypothesis that supra-threshold deficits can lead to speech intelligibility deficits, even in the absence of a reduction in audibility. This is consistent with several studies in which speech intelligibility deficits were reported under conditions where audibility was normal or near-normal, for HI and/or elderly listeners (e.g., Füllgrabe *et al.*, 2015; Grose *et al.*, 2009; Horwitz *et al.*, 2002; Lorenzi *et al.*, 2009; Strelcyk and Dau, 2009).

There was large variability in the scores of the HI listeners in all conditions, with deficits ranging from mild to severe (up to ~60 RAU, relative to the average for the NH listeners). This confirms that HI listeners with similar audiograms can present with a wide range of speech intelligibility deficits. Large deficits were observed even for HI listeners with near-normal audiometric thresholds up to 3 kHz (see, for example, the results for the two HI listeners with the lowest scores in quiet in Fig. 2 – for the mid-frequency region). Thus, a clinically normal or near-normal audiogram up to 3 kHz does not ensure that speech intelligibility is normal.

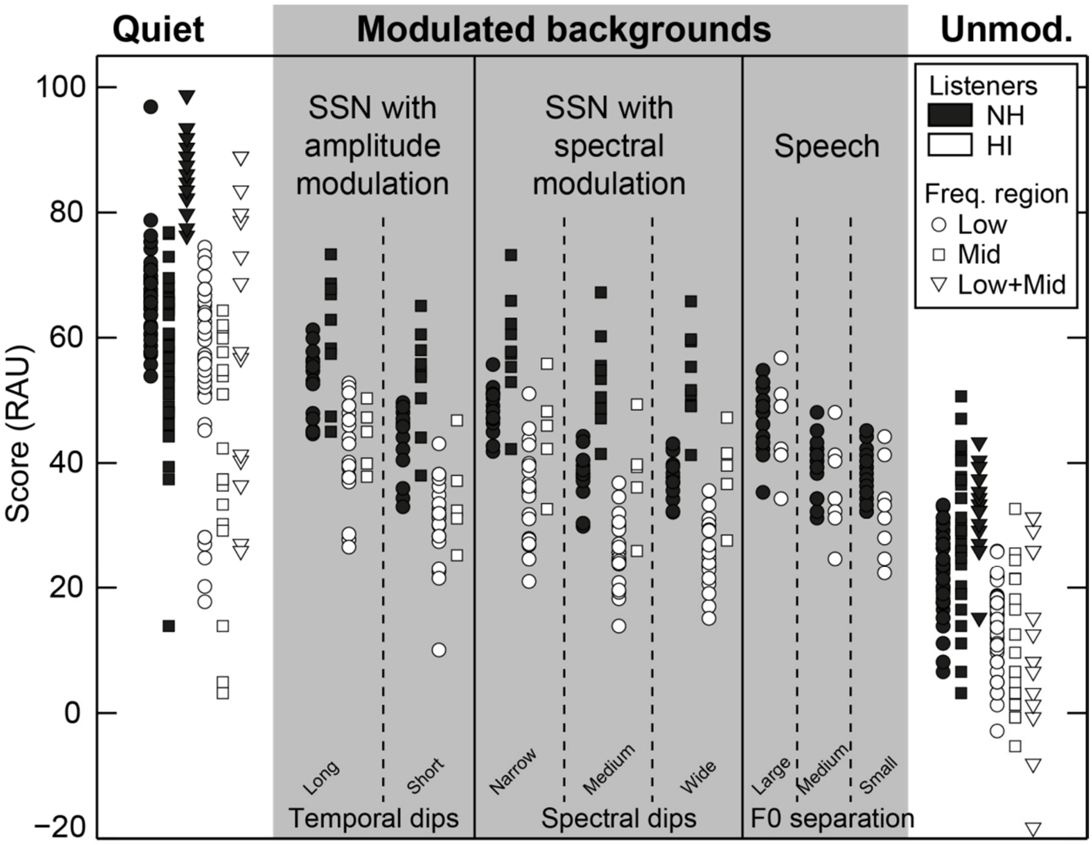


Fig. 2: Scores (in RAU) for the different background conditions: quiet, modulated backgrounds, or unmodulated backgrounds (“unmod.”). The three different types of modulated backgrounds are specified at the top of the grey area, and for each type of modulated background, the different conditions (sizes of the dips or f0 separation) are identified at the bottom. Scores are shown in black for NH listeners and in white for HI listeners, and the symbols show the region of test (see legend). Within each condition (separated by the vertical lines), the results are slightly offset between listener groups and region of test.

The deficits of the HI listeners were generally larger for speech presented in background sounds than in quiet (see Fig. 2). This was true for all backgrounds except speech, for which the deficits of the HI listeners were mild. However, this might be due to the small sample sizes; see discussion in Léger *et al.* (2014). For the noise backgrounds, the deficits of the HI listeners were similar across different types of backgrounds. Notably, the differences in scores between unmodulated and modulated noises did not differ significantly for the NH and HI listeners tested by Léger *et al.* (2012b). In other words, the HI listeners did not show reduced “masking release” (or “release from modulation masking”; Stone *et al.*, 2011). This is at odds with studies suggesting that hearing loss and/or age can reduce masking release (for a review, see

Léger *et al.*, 2012d). It may be the case that suprathreshold deficits were less severe for the HI listeners tested here than for listeners showing moderate/severe hearing losses in the tested frequency regions, and that this difference explains the discrepancies in the masking release deficits. Taken together, those results suggest that in frequency regions of near-normal audiometric thresholds, supra-threshold deficits can lead to speech intelligibility deficits that are larger for speech in background sounds than in quiet, but this reflects a global deficit, not related to the type of background.

Origin of the speech intelligibility deficits

As suggested earlier, the deficits of the HI listeners in frequency regions of near-normal audiometric thresholds may have been caused by supra-threshold deficits. It may be the case that the high-frequency hearing losses were associated with supra-thresholds deficits, including in the low-frequency region. However, the speech deficits of the HI listeners were generally not related to the severity of their high-frequency hearing loss (that is, there was generally no significant correlation between scores and the PTA in the high-frequency regions of hearing loss). It may also be the case that the supra-threshold deficits were associated with age, given that the HI listeners were generally older than the NH listeners. The differences between the groups may also be the consequence of differences in PTA-f in the frequency region of test. These non-mutually exclusive hypotheses are discussed below.

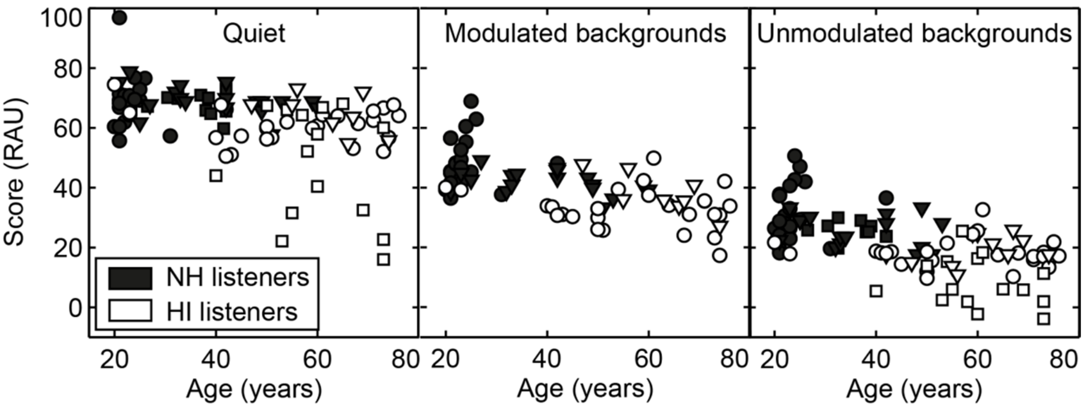


Fig. 3: Scores (in RAU) as a function of age (in years), averaged within the following background conditions: quiet, modulated backgrounds, unmodulated backgrounds. Scores for NH and HI listeners are shown in black and white, respectively. For each panel, scores were averaged within all frequency regions, background conditions and SNRs tested for a given listener. The symbols show in which study a given listener was tested: circles for Léger *et al.* (2012b; low- and mid-frequency regions), squares for Léger *et al.* (2012c; all frequency regions) and triangles for Léger *et al.* (2014; low-frequency region).

Greater age has been shown to have a deleterious effect on speech intelligibility (e.g., Arehart *et al.*, 2011; Dubno *et al.*, 2002; Grose *et al.*, 2009; Vongpaisal and Pichora-Fuller, 2007), as well as on many supra-threshold auditory abilities (e.g., Füllgrabe, 2013; Füllgrabe *et al.*, 2015; Harris *et al.*, 2008; He *et al.*, 2007; 2008; Strelcyk and Dau, 2009). As illustrated in Fig. 3, there was a global relationship between scores and age. However, the effects of age and hearing loss were generally confounded in the studies reviewed here (see Fig. 3: oldest listeners tend to be HI listeners, and vice versa; see Methods section). Analyses of the effect of age and the relationship between age and scores were carried out in all the studies reviewed here, but the conclusions were inconsistent across studies. Therefore, the effect of age on those results remains unclear. It could be the case that the speech intelligibility deficits of the HI listeners resulted largely from supra-threshold auditory deficits caused by factors associated with aging.

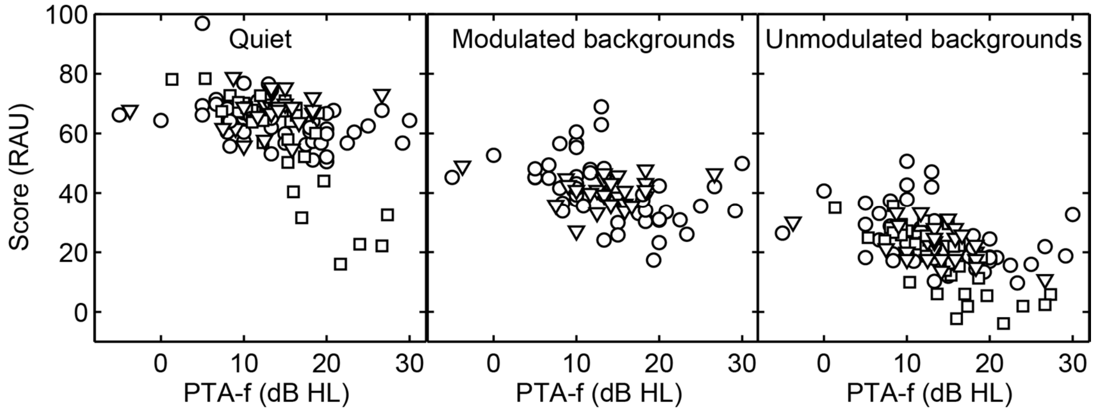


Fig. 4: Scores (in RAU) as a function of PTA-f (in dB HL) averaged within the following background conditions: quiet, modulated backgrounds, unmodulated backgrounds. Otherwise as Fig. 3.

The influence of audibility was assumed to be limited in the studies reviewed here, since all listeners were tested in frequency regions of near-normal audiometric thresholds and the speech was amplified for HI listeners with PTA-f above 20 dB HL. However, it remains unclear whether slightly increased audiometric thresholds in the tested frequency region were related to the speech identification deficits demonstrated by the HI listeners. As illustrated in Fig. 4, there generally was a relationship between PTA-f and speech intelligibility (see papers for details). This might indicate an influence of audibility. To assess whether the results were influenced by differences in the audibility of the target speech across listeners, extended speech intelligibility index (ESII; Rhebergen *et al.*, 2006; Rhebergen and Versfeld, 2005) values were computed by Léger *et al.* (2012b) for each listener. The ESII values were not correlated with the scores for either the NH or HI listeners, suggesting that the deficits demonstrated by the HI listeners were not due to small audibility differences.

However, the ESII may not give accurate predictions of intelligibility for those conditions. Therefore, the contribution of small audibility differences remains uncertain. It is possible, indeed likely, that the correlation between scores and PTA-f values occurred because higher audiometric thresholds are associated with larger supra-threshold deficits, which could reduce speech intelligibility without any influence of audibility.

Nature of the supra-threshold deficits

There are several candidate supra-thresholds deficits that might have affected the older HI listeners. Frequency selectivity can be slightly reduced for HI listeners at frequencies where the audiometric thresholds are normal or near-normal (for a review, see Léger *et al.*, 2012a). A simulation study (Léger *et al.*, 2012a) using spectral smearing with NH listeners, suggested that a slight reduction of frequency selectivity could lead to small intelligibility deficits in similar testing conditions. Furthermore, measurement of otoacoustic emissions suggested that outer hair cell functioning was related to the speech intelligibility deficits of the HI listeners tested by Léger *et al.* (2012c). However, slightly reduced frequency selectivity could not entirely explain the deficits demonstrated by the HI listeners (Léger *et al.*, 2012a). This is consistent with the result of Strelcyk and Dau (2009), who did not find any correlation between frequency selectivity and speech intelligibility for stimuli filtered into frequency regions of normal audibility. Therefore, it may be the case that other supra-threshold deficits are involved, for example in the processing of TFS cues. Indeed, both age and hearing loss have been shown to be associated with poorer TFS processing, which can contribute to intelligibility deficits for speech in background sounds (for a review, see Moore, 2014). However, in the studies reviewed here, there was no evidence in favour of or against an influence of impaired TFS processing on the deficits demonstrated by the (elderly) HI listeners for speech presented in background sounds. Therefore, the potential contribution of reduced TFS processing remains unclear.

It may also be the case that central factors played a role in the deficits of the HI listeners. Nonsense syllables were used in all studies to minimise the role of higher-level cognitive and linguistic processing. However, an influence of higher-level factors cannot be ruled out.

CONCLUSIONS

To control for the influence of reduced audibility, speech identification was measured using nonsense speech signals filtered into low- and mid-frequency regions, where pure-tone sensitivity was near normal for both (younger) NH and (older) HI listeners. The older HI listeners showed mild to severe intelligibility deficits for speech presented in quiet and in various backgrounds (noise or speech). Overall, these results suggest that speech intelligibility can be strongly influenced by supra-threshold auditory deficits associated with hearing loss and/or age, in the absence of reduced audibility.

REFERENCES

Arehart, K.H., Souza, P.E., Muralimanohar, R.K., and Miller, C.W. (**2011**). "Effects of age on concurrent vowel perception in acoustic and simulated electroacoustic hearing," J. Speech Lang. Hear. Res., **54**, 190-210.

Bernstein, J.G. and Grant, K. W. (**2009**). "Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am., **125**, 3358-3372.

Christiansen, C. and Dau, T. (**2012**). "Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise," J. Acoust. Soc. Am., **132**, 1655-1666.

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.

Dreschler, W.A. and Plomp, R. (**1980**). "Relation between psychophysical data and speech perception for hearing-impaired subjects I," J. Acoust. Soc. Am., **68**, 1608-1615.

Dreschler, W.A. and Plomp, R. (**1985**). "Relations between psychophysical data and speech perception for hearing-impaired subjects II," J. Acoust. Soc. Am., **78**, 1261-1270.

Dubno, J.R., Horwitz, A.R., and Ahlstrom, J.B. (**2002**). "Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing," J. Acoust. Soc. Am., **111**, 2897-2907.

Füllgrabe, C. (**2013**). "Age-dependent changes in temporal-fine-structure processing in the absence of peripheral hearing loss," Am. J. Audiol., **22**, 313-315.

Füllgrabe, C., Moore, B.C.J., and Stone, M.A. (**2015**). "Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition," Front. Aging Neurosci., **6**, 347.

George, E.L., Festen, J.M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am., **120**, 2295-2311.

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.

Glasberg, B.R. and Moore, B.C.J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res., **47**, 103-138.

Grose, J.H., Mamo, S.K., and Hall, J.W. (**2009**). "Age effects in temporal envelope processing: speech unmasking and auditory steady state responses," Ear Hearing, **30**, 568-575.

Harris, K.C., Mills, J.H., He, N.-J., and Dubno, J.R. (**2008**). "Age-related differences in sensitivity to small changes in frequency assessed with cortical evoked potentials," Hear. Res., **243**, 47-56.

He, N.J., Mills, J.H., and Dubno, J.R. (**2007**). “Frequency modulation detection: effects of age, psychophysical method, and modulation waveform,” J. Acoust. Soc. Am., **122**, 467-477.

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.

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.

Horwitz, A.R., Dubno, J.R., and Ahlstrom, J.B. (**2002**). “Recognition of low-pass-filtered consonants in noise with normal and impaired high-frequency hearing,” J. Acoust. Soc. Am., **111**, 409-416.

Humes, L.E., Dirks, D.D., Bell, T.S., and Kincaid, G.E. (**1987**). “Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers,” J. Acoust. Soc. Am., **81**, 765-773.

Humes, L.E. (**2002**). “Factors underlying the speech-recognition performance of elderly hearing-aid wearers,” J. Acoust. Soc. Am., **112**, 1112-1132.

Lee, L.W. and Humes, L. E. (**1993**). “Evaluating a speech-reception threshold model for hearing-impaired listeners,” J. Acoust. Soc. Am., **93**, 2879-2885.

Léger, A.C., Moore, B.C.J., Gnansia, D., and Lorenzi, C. (**2012a**). “Effect of spectral smearing on temporal and spectral masking release in low- and mid-frequency regions,” J. Acoust. Soc. Am., **131**, 4114-4123.

Léger, A.C., Moore, B.C.J., and Lorenzi, C. (**2012b**). “Temporal and spectral masking release in low- and mid-frequency regions for normal-hearing and hearing-impaired listeners,” J. Acoust. Soc. Am., **131**, 1502-1514.

Léger, A.C., Moore, B.C.J., and Lorenzi, C. (**2012c**). “Abnormal speech processing in frequency regions where absolute thresholds are normal for listeners with high-frequency hearing loss,” Hear. Res., **294**, 95-103.

Léger, A.C., Moore, B.C.J., and Lorenzi, C. (**2012d**). “A review of speech masking release for hearing-impaired listeners with near-normal perception of speech in unmodulated noise maskers,” In T. Dau, M.L. Jepsen, J.-C. Cristensen-Dalsgaard, and T. Poulsen (Eds.), Proc. ISAAR 2011, *Speech Perception and Auditory Disorders*, Danavox Jubilee Foundation, Denmark, pp. 159-166.

Léger, A.C., Ives, D.T., and Lorenzi, C. (**2014**). “Abnormal intelligibility of speech in competing speech and in noise in a frequency region where audiometric thresholds are near-normal for hearing-impaired listeners,” Hear. Res., **316**, 102-109.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B.C.J. (**2006**). “Speech perception problems of the hearing impaired reflect inability to use temporal fine structure,” Proc. Natl. Acad. Sci. U. S. A., **103**, 18866-18869.

Lorenzi, C., Debruille, L., Garnier, S., Fleuriot, P., and Moore, B.C.J. (**2009**). “Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal (L),” J. Acoust. Soc. Am., **125**, 27-30.

Moore, B.C.J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*, Wiley-Interscience, Chichester, England, 332 pages.

Moore, B.C.J. (**2014**). *Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss*, World Scientific, Hackensack, NJ, 196 pages.

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.

Phatak, S.A. and Grant, K.W. (**2012**). "Phoneme recognition in modulated maskers by normal-hearing and aided hearing-impaired listeners," J. Acoust. Soc. Am., **132**, 1646-1654.

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

Plomp, R. (**1986**). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," J. Speech Hear. Res., **29**, 146-154.

Rhebergen, K.S. and Versfeld, N.J. (**2005**). "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am., **117**, 2181-2192.

Rhebergen, K.S., Versfeld, N.J., and Dreschler, W.A. (**2006**). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," J. Acoust. Soc. Am., **120**, 3988-3997.

Rhebergen, K.S., Lyzenga, J., Dreschler, W.A., and Festen, J.M. (**2010a**). "Modeling speech intelligibility in quiet and noise in listeners with normal and impaired hearing," J. Acoust. Soc. Am., **127**, 1570-1583.

Rhebergen, K.S., Versfeld, N.J., de Laat, J.A.P.M., and Dreschler, W.A. (**2010b**). "Modelling the speech reception threshold in non-stationary noise in hearing-impaired listeners as a function of level," Int. J. Audiol., **49**, 856-865.

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

Stone, M.A., Füllgrabe, C., Mackinnon, R.C., and Moore, B.C.J. (**2011**). "The importance for speech intelligibility of random fluctuations in 'steady' background noise," J. Acoust. Soc. Am., **130**, 2874-2881.

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.

Studebaker, G.A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res., **28**, 455-462.

Summers, V., Makashay, M.J., Theodoroff, S.M., and Leek, M.R. (**2013**). "Suprathreshold auditory processing and speech perception in noise: hearing-impaired and normal-hearing listeners," J. Am. Acad. Audiol., **24**, 274-292.

Vongpaisal, T. and Pichora-Fuller, M.K. (**2007**). "Effect of age on F0 difference limen and concurrent vowel identification," J. Speech Lang. Hear. Res., **50**, 1139-1156.

Zurek, P.M. and Delhorne, L.A. (**1987**). "Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment," J. Acoust. Soc. Am., **82**, 1548-1559.

Characterizing individual differences in frequency coding: Implications for hidden hearing loss

ANDREW J. OXENHAM[1,2,*]

[1] *Department of Psychology, University of Minnesota – Twin Cities, Minneapolis, MN, USA*

[2] *Department of Otolaryngology, University of Minnesota – Twin Cities, Minneapolis, MN, USA*

A long-standing debate in hearing research has focused on whether frequency is coded in the peripheral auditory system via phase-locked timing information in the auditory nerve (temporal code), or via tonotopic information based on the firing rates of auditory-nerve fibers tuned to different frequencies (rate-place code). Because frequency discrimination is generally much more accurate than intensity discrimination, it has been thought that frequency is likely to be coded via a temporal code, whereas intensity is represented via a rate code. However, direct empirical tests of this assumption have produced mixed results. This paper reviews a way in which the coding of both frequency and intensity might be reconciled within a single mechanism, and then uses an approach based on simple signal detection theory to predict the effects of a loss auditory-nerve synapses (synaptopathy) on some basic psychoacoustic phenomena, such as detection thresholds, frequency discrimination, and intensity discrimination. The predictions provide a baseline with which to compare future empirical findings based on the perceptual consequences of synaptopathy, or "hidden hearing loss."

INTRODUCTION

The coding of frequency is critical to many aspects of auditory perception, such as speech perception, music perception, and auditory scene analysis. A long-standing question in auditory science is how frequency is coded in the peripheral auditory system. The two most common candidates involve a code based on the tonotopic representation of frequency along the cochlea's basilar membrane, leading to differences in firing rate in auditory nerve fibers tuned to different characteristic frequencies (rate-place code), and a code based on the phase-locked timing of auditory nerve spikes (temporal code) (e.g, Siebert, 1970; Heinz *et al.*, 2001a).

In general, the information carried in the timing information is far greater than that carried in the rate-place information, assuming optimal processing of that information. Processing of timing information would require some neural mechanism that can precisely measure the time intervals between neural spikes with a resolution of microseconds and for delays as large as tens of milliseconds. Although evidence for

*Corresponding author: oxenham@umn.edu

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

neural coding accuracy down to microseconds has been found for the binaural system when processing interaural time differences (ITDs) (e.g., Yin and Chan, 1990; Brand *et al.*, 2002), similarly direct evidence has not been identified for the processing of frequency. There is, however, a body of more indirect evidence, pointing to a role for the temporal code. First, frequency discrimination becomes much worse at high frequencies, with difference limens (as a percentage of the reference frequency) increasing by about an order of magnitude between 2 and 8 kHz (e.g., Moore, 1973; for a review, see Micheyl *et al.*, 2012). This deterioration at high frequencies is difficult to explain based simply on a peripheral rate-place code, but may be explained in terms of the upper limits of phase-locking in a temporal code (Heinz *et al.*, 2001b). Second, studies have generally found little to no relationship between pure-tone frequency discrimination at low or high frequencies and frequency selectivity, suggesting that a rate-place code based on tonotopic representation is unlikely to limit performance (Tyler *et al.*, 1982; Moore and Peters, 1992). Third, detection thresholds for frequency modulation (FM) depend on modulation rate and carrier frequency in a way that is not found for amplitude modulation (AM). At low carrier frequencies (< 4 kHz) and slow modulation rates (< 5 Hz), listeners are generally very sensitive to FM, whereas at higher frequencies and/or at higher modulation rates, performance deteriorates. This pattern of results, along with other evidence from the interference of AM on FM detection, has led to the proposal that slow-rate FM at low carrier frequencies is coded via a timing code that is temporally sluggish (i.e., unresponsive to rapid changes in frequency), whereas fast-rate FM, or FM at high carrier frequencies, relies on an FM-to-AM transformation via the auditory filters (e.g., Moore and Sek, 1995).

Perhaps because of the apparent need for fine timing information to code frequency, it has been hypothesized that temporal fine structure and temporal envelope coding may be particularly affected by a form of hearing loss, termed “hidden hearing loss” (Schaette and McAlpine, 2011) or “synaptopathy” that results from a loss of synapses between the hair cells and auditory nerve fibers (e.g., Kujawa and Liberman, 2009). Several studies have now suggested a link between synaptopathy and certain behavioral deficits observed in temporal coding in the absence of traditional clinical hearing loss (Plack *et al.*, 2014; Bharadwaj *et al.*, 2015).

In this paper we review two recent studies, one empirical and one theoretical, that address the question of how frequency and intensity changes are coded. Finally, we present a simple analysis based on signal detection theory for predicted effects on signal detection, as well as frequency and intensity coding, of hidden hearing loss.

EMPRICAL TEST OF TEMPORALLY CODED SLOW FREQUENCY MODULATION

Whiteford and Oxenham (2015) carried out a correlational study involving 100 young normal-hearing listeners. They measured detection thresholds for FM, AM, dichotic FM (introducing dynamic ITD cues), and dichotic AM (with dynamic interaural level difference, ILD, cues), all with a carrier frequency of 500 Hz and a slow (1-Hz) or fast (20-Hz) modulation rate. In addition, frequency selectivity around 500 Hz was estimated using a forward-masking paradigm. The hypothesis was that slow-rate FM and ITD coding are both governed by the same temporal (phase-locking) code, and so should be correlated, whereas fast-rate FM is determined by FM-to-AM translation, and so should be correlated with the threshold predicted from fast-rate AM thresholds combined with the measure of frequency selectivity. Whiteford and Oxenham (2015) found a reasonable correlation (around r = 0.5) between essentially all measures of modulation detection, slow and fast, FM and AM, and dichotic and diotic. Although the correlation between slow FM and dichotic FM thresholds was consistent with the hypothesis, the fact that the correlations were similar for all modulation-detection tasks was not. In addition, the measure of frequency selectivity was not correlated with either fast or slow FM, even when AM sensitivity was accounted for. In other words, the results provided no support for the idea that slow FM is coded differently from other forms of modulation.

Whiteford and Oxenham's (2015) negative result may be because thresholds are not limited by peripheral sensory factors, such as auditory-nerve coding, but are instead limited by higher-level (e.g., cortical) sensory or cognitive factors. Alternatively, similar peripheral mechanisms may limit both FM and AM perception at both low and high modulation rates, leading to the common source of variance. This common variance may reflect a common neural code, or it may simply reflect a common mode of transmission; for instance, damage to the auditory nerve would result in poorer transmission of both rate-place and timing codes. A next step for this line of investigation is to study correlations using a more diverse population of subjects, to study the effects of ageing and the effects of hearing loss. For instance, it has been suggested that ageing results in a selective deficit in temporal fine structure processing (Moore *et al.*, 2012). If so, then stronger correlations between diotic and dichotic slow-rate FM detection thresholds might be observed in a population that had a wider age range. Similarly, cochlear hearing loss due to dysfunction of the outer hair cells leads to a loss of sensitivity and poorer frequency selectivity (e.g., Moore *et al.*, 1999). Therefore, including subjects with a range of cochlear hearing losses may result in a clearer correlation between fast-rate FM detection thresholds and estimated frequency selectivity.

The next section reviews one possible way in which AM (fluctuations in intensity) and FM (fluctuations in frequency) might be coded similarly, and yet remain consistent with the finding that frequency coding appears more accurate than intensity coding.

A COMMON CODE FOR FREQUENCY AND INTENSITY?

Even if a temporal code is admitted for representing frequency at the level of the auditory periphery, it is unlikely that such a code survives the transformations between the cochlea and primary auditory cortex. Instead, by the time the processing reaches auditory cortex, any timing information extracted from the temporal fine structure of tones has probably been transformed into some form of rate-based population code (e.g., Wang *et al.*, 2008). This leaves a potential problem: Frequency difference limens (FDLs), as well as FM detection thresholds at slow rates and low carrier frequencies, are generally much smaller than would be predicted by a just-detectable change in excitation pattern, based on measured intensity difference limens or AM detection thresholds at similar rates and carrier frequencies (Glasberg and Moore, 1986; Lacher-Fougere and Demany, 1998). If both intensity and frequency are coded by a rate-place in auditory cortex, then how can the apparent discrepancy between the accuracy of intensity discrimination and frequency discrimination be resolved?

Micheyl *et al.* (2013) recently proposed a solution to this apparent discrepancy. Their solution was based on the potential for correlations between the responses of neurons to the same stimulus, even in the absence of stimulus variability. This so-called "noise correlation" (Cohen and Kohn, 2011) generally decreases the benefit of pooling information across neurons. For instance, consider the case where an increment in the intensity of a stimulus is to be detected via a change in the firing rate of a population of neurons. The sensitivity of a single neuron is given by difference in mean firing rate in response to the baseline and the incremented stimuli (M_{R2}-M_{R1}), divided by the standard deviation (σ, i.e., the trial-to-trial variability of the neural response). This provides a measure of sensitivity, d', for each individual neuron: $d' = (M_{R2}-M_{R1})/\sigma$.

Assuming independence between all neurons, the optimal decision rule is to combine the information from across all N neurons (e.g., Green *et al.*, 1959):

$$d'_{TOT}=\sqrt{\sum_{i=1}^{N} d'^2_i} \qquad \text{Eq. (1)}$$

So, for instance, doubling the number of independent neurons leads to an increase in d' of a factor of $\sqrt{2}$, or about 1.4. However, if the neurons are all completely correlated (noise correlation coefficient = 1), then no benefit is derived from combining the information from multiple neurons, as the total information is the same as the information from just a single neuron. Therefore, as the degree of correlation increases from 0 to 1, the increase in sensitivity as a function of N decreases from a factor of $\sqrt{N}$ to 1 (no change).

When the task involves detecting a change in frequency, however, the situation is different. Now, a noise correlation can in some cases *improve* performance. For instance, consider two neurons with characteristic frequencies (CFs) on either side of the test-tone frequency. When the frequency of the tone is increased, the response of the neuron with the higher CF will increase, whereas the response of the neuron with the lower CF will decrease. Thus, an optimal combination of information will involve some form of *subtraction* of the two responses, as opposed to the *addition* that would be required in the intensity-discrimination condition. When responses are added, noise correlation increases in the internal noise; when responses are subtracted, any noise correlation can be potentially subtracted out and hence eliminated. Thus, in the case of frequency discrimination, noise correlation may improve performance. This difference between frequency and intensity coding is illustrated in Fig. 1, which shows the responses of two sample neurons with some degree of correlation. The spike rate of one unit (*j*) is plotted as a function of the spike rate of another unit (*i*). If the spiking rate of the units were uncorrelated, the distributions would be circles; the oval distributions show that there exists a positive correlation (perfect correlation would be represented by a straight line along the major diagonal). Panel A illustrates the case of frequency discrimination, where unit *j* has a CF higher than the test frequencies, and unit *i* has a CF lower than the test frequency. When the stimulus frequency is increased from the reference (blue) to the higher frequency (red), the average firing rate of *j* increases, whereas the average firing rate of *i* decreases. In this situation, the fact that the firing-rate distributions are oval means less overlap (and hence better discriminability) between the two joint distributions than would be the case with independent firing rates. The opposite is true for the case of intensity discrimination (Panel B). Here the oval distributions lead to more overlap (and hence worse discriminability) than would be the case with independent firing rates. Using this kind of approach, Micheyl *et al.* (2013) showed that the same model could account for human performance in both intensity and frequency discrimination, using the same rate-place neural coding, by assuming a degree of correlation that was within the range of those observed in auditory cortical recordings. The work thus shows that it is not necessary to assume different neural codes to account for human frequency and intensity discrimination abilities.

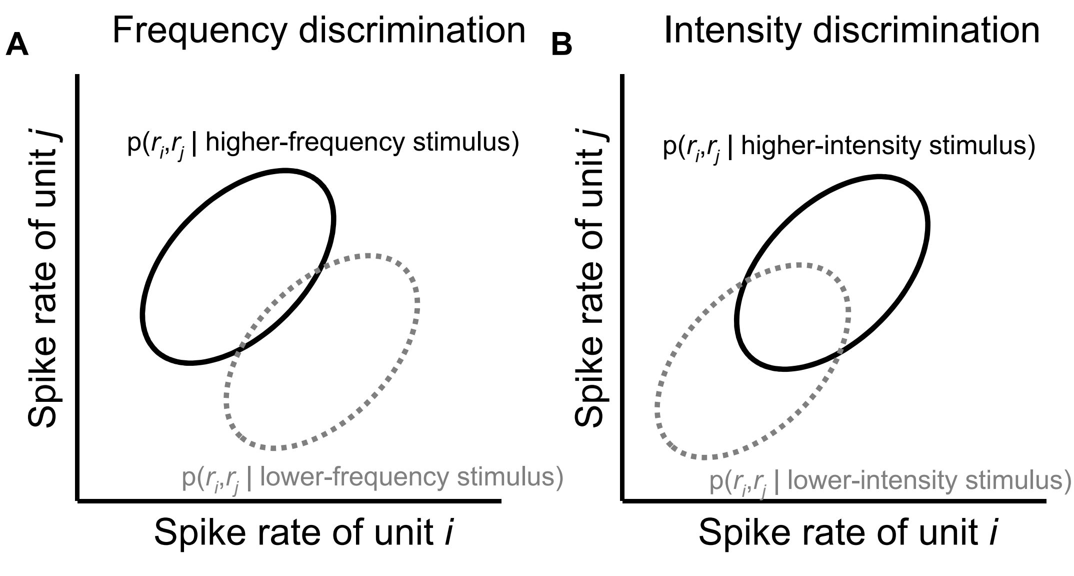


Fig. 1: Schematic diagram of the response distributions of two sample neurons (*i* and *j*) to illustrate the effects of noise correlation between neurons in frequency and intensity discrimination task. Redrawn from Micheyl *et al.* (2013).

PREDICTED EFFECTS OF HIDDEN HEARING LOSS

As the recent work of Liberman, Kujawa, and colleagues has shown (e.g., Kujawa and Liberman, 2009; Sergeyenko *et al.*, 2013; Fernandez *et al.*, 2015), noise exposure that causes only a temporary shift in thresholds (measured behaviorally and neurally) can nevertheless lead to permanent loss of synapses (of 50% or more) between the inner hair cells and auditory nerve fibers. This synaptopathy, has been termed "hidden hearing loss" (Schaette and McAlpine, 2011; Plack *et al.*, 2014), because it would not be detected by a traditional audiogram.

The reasons why absolute thresholds remain unaffected by hidden hearing loss are not completely clear. One possibility is that the synaptic loss seems to be concentrated in fibers with low spontaneous firing rates and high thresholds (Furman *et al.*, 2013), meaning that the high-spontaneous-rate fibers with low thresholds, which are presumably responsible for detecting low-intensity sounds, are less affected. There has been some speculation as to what perceptual abilities might be most affected by hidden hearing loss, including poorer temporal processing (similar to that found in people with auditory neuropathy or dys-synchrony), deficits in processing supra-threshold sounds, particularly at higher sound levels, and understanding speech in noise (Plack *et al.*, 2014; Bharadwaj *et al.*, 2015).

At this point it may be useful to generate some basic expectations regarding performance in perceptual tasks, based on signal detection theory (Green and Swets, 1966), along with some highly simplified assumptions concerning peripheral auditory processing. The analysis below follows in the tradition of Viemeister (1988), who calculated the number of auditory nerve fibers required to achieve human levels of intensity discrimination, based on the response properties of single neurons.

Model assumptions

In estimating the effect of losing synapses (and hence functionally losing auditory nerve fibers), the simplest assumptions are that: 1) the response of each auditory nerve fiber is independent from the responses of the others, and 2) the information from all the auditory nerve fibers is optimally combined. In this case, the sensitivity of the system is described by the d'_{TOT} shown in Eq. 1, where d'_i is the sensitivity of an individual auditory nerve fiber, *i*. For this initial analysis, a further simplifying assumption is that all auditory nerve fibers carry equal information or, equivalently, a loss of functional auditory nerve fibers affects the entire population proportionally.

Predictions for detecting a signal in quiet or in noise

Many studies have shown that the sensitivity to a signal in noise or quiet is proportional to the signal intensity, for a given signal duration and frequency (e.g., Green *et al.*, 1959; Hicks and Buus, 2000). For instance, a doubling in sensitivity should lead to a halving in the sound intensity, or a 3-dB decrease in level, required for detection threshold. Taking our simplified assumptions along with Eq. 1, we can see that decrease in the number of functional auditory nerve fibers by a factor *F* will lead to a decrease in the overall d'_{TOT} by a factor $\sqrt{F}$. In other words a 50% (factor of

2) loss in auditory nerve fibers will lead to a reduction in sensitivity by a factor of √2. Because d' and intensity are proportional, a √2 decrease in d' implies a √2 increase in the intensity required to achieve threshold. This translates into a 1.5-dB increase in threshold. In other words, the model predicts that a 50% loss of fibers would lead to only a 1.5-dB change in threshold – one that is probably not measurable with standard audiometric equipment. Similarly, a dramatic 90% loss of fibers would still only predict a 5-dB increase in thresholds in quiet or in noise. The relationship between predicted threshold change (where a negative number implies a loss of sensitivity or increase in threshold) and proportional loss of synapses is shown in Fig. 2 for losses between 0 and 99% of synapses.

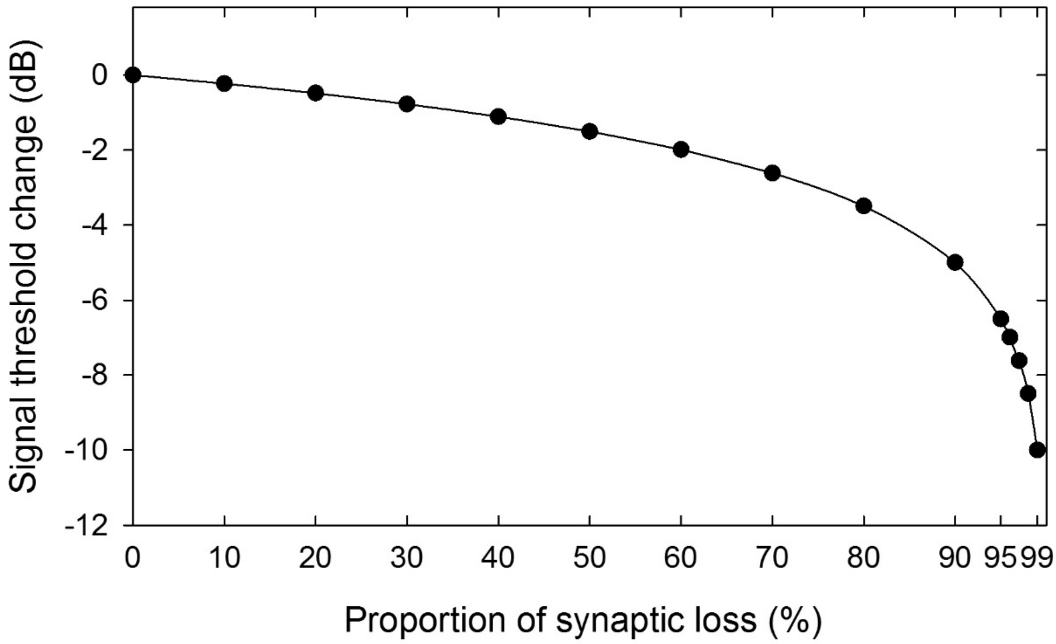


Fig. 2: Illustration of the predicted change in threshold, as a function of the proportion of lost synapses. Negative numbers imply a worsening, or increase, in threshold.As shown, even a 99% loss of synapses results in only a 10-dB change in threshold.

Predictions for auditory discrimination of frequency or interaural time differences

Similar predictions can be derived for any auditory task where the simplifying assumptions are reasonable and where the relationship between d' and the relevant stimulus parameter is known. For frequency discrimination, d' is generally proportional to the difference in frequency, Δf (e.g., Dai and Micheyl, 2011). Thus, by the same logic as outlined above, any decrease in d' due to loss of fibers would result in a proportional increase in the Δf at a given threshold. For instance, a 50% loss of fibers would result in a predicted decrease in d' of √2, and so frequency-discrimination thresholds should increase by the same amount. Although a change in threshold from, say, 1% to 1.4% might be measurable within an individual subject, the large individual differences observed in normal-hearing listeners (e.g., Whiteford and Oxenham, 2015) would make it difficult to distinguish from other factors in the general population.

For the discrimination of intensity differences, d' has been found to be roughly proportional to the change in level (in dB), ΔL (Buus and Florentine, 1991; Buus *et al.*, 1995). Thus, according to our simplified model, a 50% loss in synapses would be predicted to produce a factor of √2 increase in the just-noticeable difference (JND). For instance, a JND of 1 dB would increase to 1.4 dB, which again would be barely measurable. It would take a more dramatic loss of 75% of synapses to double the JND to 2 dB.

The detection of interaural time differences (ITDs) is one psychoacoustic measure that almost certainly depends on auditory-nerve phase locking. Here again, d' is proportional to the ITD, so that a 50% reduction in fibers is predicted to lead to an increase in the threshold ITD by a factor of √2.

Predicting the effects on more complex tasks, such as speech understanding in noise, will take a more detailed approach. However, signal-detection-based approaches have been applied to the problem of speech understanding (e.g., Musch and Buus, 2001a; 2001b; Micheyl and Oxenham, 2012), so such approaches could likely be used to predict how speech intelligibility would be predicted to change in the face of auditory synaptopathy.

Model limitations

The predictions of the perceptual consequences of synaptopathy from the model outlined above are, of course, dependent on the model assumptions. All assumptions are highly simplified, and some are more justifiable than others, as outlined below.

The first assumption is that the responses from individual auditory-nerve fibers are independent. Based on available data, this assumption seems reasonable (in contrast to auditory cortical responses described in the previous section). However, if some correlation is assumed between neurons then the predicted effect of a loss of fibers becomes even smaller; as the assumed correlation increases from 0 to 1, the predicted change in d' decreases from a factor of √F to no change.

The second assumption is that all fibers carry equal information. This is clearly not the case. For instance, at low intensities, most coding will be done by high-spontaneous-rate fibers, and fibers with low characteristic frequencies will have little influence on the coding of high-frequency sounds. In terms of high- vs. low-spontaneous-rate fibers, if synaptopathy does selectively affect low-spontaneous-rate fibers, then it may selectively and disproportionately impair processing at higher sound levels.

The third assumption is that the statistical distributions can be considered Gaussian and continuous. This assumption may fail in the cases where small numbers of neurons are involved and/or where the responses are more discrete in nature. For instance, if a brainstem neuron requires coincident input from two auditory-nerve fibers, then it will fail completely if just one of the fibers is no longer active.

Overall, the model should be treated as a very rough first approximation, but it nonetheless provides some insights into why a dramatic loss of fibers may result in behavioral changes that are barely measurable. More sophisticated and realistic models will likely provide an important tool in our quest to better understand the nature and consequences of different forms of damage to the human auditory system.

CONCLUSIONS

This paper reviewed two recent studies that investigated the possible neural codes underlying frequency and intensity coding in the auditory system. The first empirical study failed to find evidence that phase locking mediates the coding of slow-rate frequency modulation at low carrier frequencies (Whiteford and Oxenham, 2015). The second theoretical study showed how human performance in both frequency and intensity discrimination could be explained using a single rate-place code, if some degree of correlation between the responses of neighboring neurons is assumed. Regardless of the neural code used for frequency and intensity, decreasing the number of fibers carrying information, via synaptopathy or hidden hearing loss, will result in decreased performance. The final part of the paper outlined predictions of a highly simplified model based on signal detection theory that showed how a dramatic loss of auditory nerve fibers may only result in small, and in some cases unmeasurable, decreases in behavioral performance. Such modeling can be used as a 'baseline' with which to make specific predictions regarding the perceptual consequences of hidden hearing loss.

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.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). "Precise inhibition is essential for microsecond interaural time difference coding," Nature, **417**, 543-547.

Buus, S. and Florentine, M. (**1991**). "Psychometric functions for level discrimination," J. Acoust. Soc. Am., **90**, 1371-1380.

Buus, S., Florentine, M., and Zwicker, T. (**1995**). "Psychometric functions for level discrimination in cochlearly impaired and normal listeners with equivalent-threshold masking," J. Acoust. Soc. Am., **98**, 853-861.

Cohen, M.R. and Kohn, A. (**2011**). "Measuring and interpreting neuronal correlations," Nat. Neurosci., **14**, 811-819.

Dai, H. and Micheyl, C. (**2011**). "Psychometric functions for pure-tone frequency discrimination," J. Acoust. Soc. Am., **130**, 263-272.

Fernandez, K.A., Jeffers, P.W., Lall, K., Liberman, M.C., and Kujawa, S.G. (**2015**). "Aging after noise exposure: acceleration of cochlear synaptopathy in "recovered" ears," J. Neurosci., **35**, 7509-7520.

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.

Glasberg, B. and Moore, B.C.J. (**1986**). "The relationship between frequency selectivity and frequency discrimination for subjects with unilateral and bilateral cochlear impairments," in *Auditory Frequency Selectivity*. Eds. B.C.J. Moore and R.D. Patterson (Plenum, New York, NY), pp. 407-417.

Green, D.M., McKey, M.J., and Licklider, J.C.R. (**1959**). "Detection of a pulsed sinusoid in noise as a function of frequency," J. Acoust. Soc. Am., **31**, 1446-1452.

Green, D.M. and Swets, J.A. (**1966**). *Signal Detection Theory and Psychophysics* (Krieger, New York).

Heinz, M.G., Colburn, H.S., and Carney, L.H. (**2001a**). "Evaluating auditory performance limits: I. one-parameter discrimination using a computational model for the auditory nerve," Neural Comput., **13**, 2273-2316.

Heinz, M.G., Colburn, H.S., and Carney, L.H. (**2001b**). "Rate and timing cues associated with the cochlear amplifier: level discrimination based on monaural cross-frequency coincidence detection," J. Acoust. Soc. Am., **110**, 2065-2084.

Hicks, M.L. and Buus, S. (**2000**). "Efficient across-frequency integration: evidence from psychometric functions," J. Acoust. Soc. Am., **107**, 3333-3342.

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.

Lacher-Fougere, S. and Demany, L. (**1998**). "Modulation detection by normal and hearing-impaired listeners," Audiology, **37**, 109-121.

Micheyl, C. and Oxenham, A.J. (**2012**). "Comparing models of the combined-stimulation advantage for speech recognition," J. Acoust. Soc. Am., **131**, 3970-3980.

Micheyl, C., Xiao, L., and Oxenham, A.J. (**2012**). "Characterizing the dependence of pure-tone frequency difference limens on frequency, duration, and level," Hear. Res., **292**, 1-13.

Micheyl, C., Schrater, P.R., and Oxenham, A.J. (**2013**). "Auditory frequency and intensity discrimination explained using a cortical population rate code," PLoS Comput. Biol., **9**, e1003336.

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

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 Sek, A. (**1995**). "Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation)," J. Acoust. Soc. Am., **97**, 2468-2478.

Moore, B.C.J., Vickers, D.A., Plack, C.J., and Oxenham, A.J. (**1999**). “Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism,” J. Acoust. Soc. Am., **106**, 2761-2778.
Moore, B.C.J., Vickers, D.A., and Mehta, A. (**2012**). “The effects of age on temporal fine structure sensitivity in monaural and binaural conditions,” Int. J. Audiol., **51**, 715-721.
Musch, H. and Buus, S. (**2001a**). “Using statistical decision theory to predict speech intelligibility. I. Model structure,” J. Acoust. Soc. Am., **109**, 2896-2909.
Musch, H. and Buus, S. (**2001b**). “Using statistical decision theory to predict speech intelligibility. II,” J. Acoust. Soc. Am., **109**, 2910-2920.
Plack, C.J., Barker, D., and Prendergast, G. (**2014**). “Perceptual consequences of “hidden” hearing loss,” Trends Hear., **18**.
Schaette, R. and McAlpine, D. (**2011**). “Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model,” J. Neurosci., **31**, 13452-13457.
Sergeyenko, Y., Lall, K., Liberman, M.C., and Kujawa, S.G. (**2013**). “Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline,” J. Neurosci., **33**, 13686-13694.
Siebert, W.M. (**1970**). “Frequency discrimination in the auditory system: place or periodicity mechanisms,” Proc. IEEE, **58**, 723-730.
Tyler, R.S., Wood, E.J., and Fernandes, M.A. (**1982**). “Frequency resolution and hearing loss,” Brit. J. Audiol., **16**, 45-63.
Viemeister, N.F. (**1988**). “Psychophysical aspects of auditory intensity coding,” in *Auditory function*. Eds. G.M. Edelman, W.E. Gall, and W.A. Cowan (Wiley, New York).
Wang, X., Lu, T., Bendor, D., and Bartlett, E. (**2008**). “Neural coding of temporal information in auditory thalamus and cortex,” Neuroscience, **157**, 484-494.
Whiteford, K.L. and Oxenham, A.J. (**2015**). “Using individual differences to test the role of temporal and place cues in coding frequency modulation,” J. Acoust. Soc. Am., **138**, 3093-3104.
Yin, T.C.T. and Chan, J.C.K. (**1990**). “Interaural time sensitivity in medial superior olive of cat,” J. Neurophysiol., **64**, 465-488.

Relations between auditory brainstem response and threshold metrics in normal and hearing-impaired listeners

SARAH VERHULST*, ANOOP JAGADEESH, MANFRED MAUERMANN, AND FRAUKE ERNST

Medizinische Physik and Cluster of Excellence "Hearing4all", Dept. of Medical Physics and Acoustics, Oldenburg University, Oldenburg, Germany

Auditory brainstem responses (ABRs) offer a potential tool to diagnose auditory-nerve deficits in listeners with normal hearing thresholds as abnormalities in the amplitude of this population response may result from a loss in the number of auditory-nerve fibers contributing to this response. However, little is known about how cochlear gain loss interacts with auditory-nerve deficits to impact ABRs. We measured level-dependent changes in click-ABR wave-I and V in listeners with normal and elevated thresholds to study which measures are dominated by cochlear gain loss. ABR wave-V latency-vs-intensity functions correlated well to the distortion-product otoacoustic emission threshold and this relation was also observed for the slope of supra-threshold ABR wave-I level growth in listeners with thresholds above 20 dB SPL. ABR wave-I and wave-V growth were not related to each other, demanding caution when using ABR wave-V growth or level as a direct measure for auditory-nerve deficits.

INTRODUCTION

Auditory brainstem responses (ABRs) have regained popularity in the diagnostics of subcomponents of peripheral hearing loss. As the ABR is easily recorded in humans and its wave peaks result from population responses at different ascending processing stages along the auditory pathway, it can be used to isolate auditory-nerve (AN) deficits (i.e., cochlear neuropathy) due to noise exposure or ageing. Particularly, in subjects with normal auditory thresholds, the ABR wave-I level is reduced when the number of auditory nerve (AN) fibers synapsing onto the inner-hair cell is reduced (Kujawa and Liberman, 2009; Sergeyenko *et al.*, 2013; Furman *et al.*, 2013). The ABR wave-I contains information about many AN fibers firing synchronously to transient stimuli and its level reduction can thus occur while correlates of outer-hair-cell health such as otoacoustic emissions are normal.

While in animal physiology the ABR wave-I is strong, humans have a weak wave-I compared to the wave-V that is thought to be generated by medial-superior-olive (MSO) primary cells projecting on to the lateralis lemniscus and inferior colliculus (Melcher and Kiang, 1993). It is currently unclear whether auditory-nerve deficits impact wave-V in similar ways as wave-I since it has been suggested that homeostatic

*Corresponding author: sarah.verhulst@uni-oldenburg.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

mechanisms can undo effects of the ABR wave-I reduction in tinnitus patients (Schaette and McAlpine, 2011) and that greater wave-V/I level ratios have been associated with hyperacusis and tinnitus (Hickox and Liberman, 2014; Gu *et al.*, 2012). Although hearing diagnostics have mostly focused on wave-V in humans, it may not directly reflect auditory-nerve deficits that occur at more peripheral processing stages. Another confounding factor in using ABRs is that they are an output measure that is influenced by both auditory-nerve *and* hair-cell deficits. While cochlear neuropathy studies have so far focused on listeners with normal thresholds, it is not clear how outer-hair-cell-loss-related cochlear gain loss impacts the ABR wave-I and wave-V. Because in clinical practice, one would ideally use one measure that can differentially diagnose subcomponents of hearing loss in listeners with mixtures of pathologies, it is important to study how different hearing deficits interact and impact the ABR.

The present study addresses this topic by reporting click-ABR wave-I and wave-V levels and latencies in listeners with normal and elevated hearing thresholds. The ABR measures were correlated to distortion-product otoacoustic emission (DPOAE) thresholds as an objective correlate of hearing threshold to test whether cochlear gain mechanisms are the dominant factor accounting for the ABR results. Lastly, it was tested whether one measure for cochlear neuropathy – ABR wave-I growth (Furman *et al.*, 2013) – follows the same trend in listeners with normal and elevated hearing thresholds.

METHODS

Audiograms, ABRs, and DPOAEs were measured in 37 subjects who were divided into two groups. The *normal-threshold* group consisted of 23 participants (mean age = 26.8 years) who were ensured to have hearing thresholds below 15 dB HL in the octave frequencies between 250 and 4000 Hz (mean threshold at 4 kHz = 3.4 dB). The *elevated-threshold* group consisted of 14 participants (mean age=64.4 years) who had a minimum of 20 dB of hearing loss at and above 4 kHz in the better ear (mean threshold at 4 kHz = 26.4 dB). The subjects had mild to moderate hearing losses and measureable DPOAEs. All study participants signed an informed consent according to the ethical review board of the University of Oldenburg.

Instrumentation: Sounds were presented using ER-2 insert earphones attached to a TDT-HB7 headphone driver and a Fireface UCX sound card. All stimuli were generated in Matlab and calibrated using a B&K type 4157 ear simulator and sound level meter. OAEs were recorded using the OLAMP software and an ER10B+ microphone. ABRs were recorded using a 32-channel Biosemi EEG amplifier and a custom built triggerbox, and analysed using the ANLFFR and Matlab software.

Click-ABR: 100-μs condensation clicks (0-1-0) were presented monaurally to the better ear at a rate of 33.3 Hz with a 10% jitter on the recording window duration. 7000 clicks were presented at peak-equivalent sound pressure levels (peSPL) of 70, 80, 90, and 100 dB. For each stimulus level, the raw EEG from the Cz channel was referenced to the mean of the reference electrodes placed on the earlobes, and filtered from 70 Hz to 2000 Hz. ABR waveforms were epoched from −10 ms to 20 ms, baseline corrected, and averaged. ABR peak latency and the peak-to-peak amplitudes were determined for wave-I and V. ABR latency was reported to the start of the stimulus and no compensation for the (fixed) recording delay of the sound delivery system was applied. The ABR peaks were hand picked by two independent observers. If the peak-to-peak levels were less than 2 dB apart results were averaged, else the data-point was discarded. Similarly, for ABR latency measurement points were discarded when readings were more than 0.3 ms apart. Slopes of ABR level and latency across the 30-dB stimulus level range were calculated using a linear fit across the data points corresponding to the four stimulus levels. Group statistics on the intensity curves of ABR latency and level were calculated using a *t*-test and correlation statistics were obtained using linear regression.

DPOAEs: DPOAEs were measured for a fixed f_2/f_1 ratio of 1.2 and primary levels were either chosen according to the Neely-level paradigm (half of the participants) or the Scissors paradigm (other half; Kummer *et al.*, 1998). Because the two level paradigms yield similar growth functions at low stimulus levels (Neely *et al.*, 2005), this methodological difference is not expected to influence the derived DPOAE thresholds substantially. The primary frequencies were exponentially swept up (2s/octave) over a 1/3 octave range around the geometric mean of 4 kHz at a constant frequency using a sweep method (Long *et al.*, 2008). Using a sufficiently sharp least squared fit filter (here ca. 2.2 Hz), the distortion component was extracted from the DPOAE recording. This distortion component is generated around the characteristic site of f_2 and thus predominantly provides information about the f_2 site without being influenced by DPOAE fine structure (Mauermann and Kollmeier, 2004). Growth functions were computed as the average over 34 distortion-source DPOAE functions across the measured frequency range and a matched cubic function:

$$L_{DP} = a + \left(\frac{1}{q}(L_2\text{-}b)\right)^3$$

with parameters a, b, and q fitted to the data points. DPOAE thresholds were determined as the level of L_2 at which the extrapolated fitting curve reached a level of −25 dB SPL (~0 Pa).

RESULTS

Figure 1 shows ABR intensity functions of ABR wave-V latency (A) and ABR wave-V and I level (B and C) for the normal-threshold and elevated-threshold group.

Whereas the wave-V latencies of the two groups are not significantly different at 100 dB peSPL (p=0.23), the latency difference between groups becomes significantly greater as stimulus level is reduced (p<0.01). Specifically, the listeners with elevated hearing thresholds exhibit steeper latency-vs-level slopes (p<0.001), due to overall increased wave-V latencies at the lower stimulus levels. Even though increased ABR wave-V latencies for listeners with elevated hearing thresholds are somewhat at odds with linear filter theory that predicts shorter local basilar-membrane impulse responses for wider auditory filters, another study has similarly reported increased

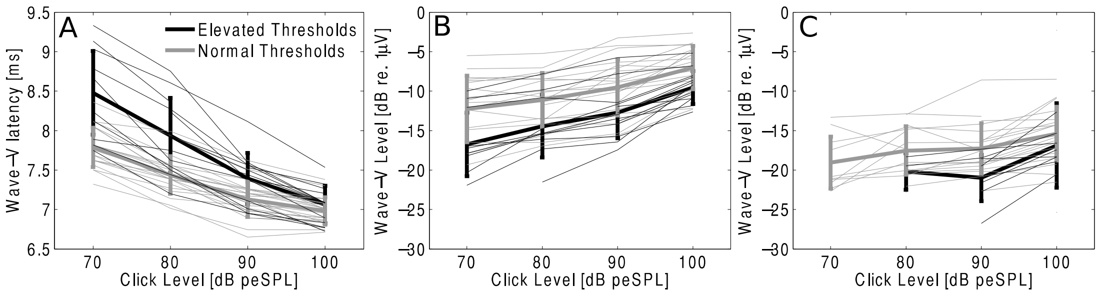


Fig. 1: ABR wave-V latency (A), level (B), and ABR wave-I level (C) for the normal (gray) and elevated-hearing threshold (black) group. Both the mean results +/- 1 standard deviation (thick lines) and individual (thin lines) results are shown.

wave-V latencies for 2-kHz derived band ABRs at moderate intensities (60-70 dB peSPL) in listeners with sloping hearing losses (Strelcyk *et al.*, 2009). A study with few participants also reported steeper click ABR wave-V latency slopes for those listeners with sloping audiograms (Gorga *et al.*, 1985).

Overall, ABR wave-V levels (panel B) were higher for the normal-threshold group (p<0.05), and the growth of ABR wave-V levels across all 4 intensities was steeper for the elevated threshold group (p<0.05). However, when only considering the highest two stimulus levels (90 and 100 dB peSPL), ABR wave-V was not significantly steeper for the elevated threshold group than for the normal-threshold group (p=0.15). A similar trend was observed for the ABR wave-I level growth functions (panel C), which did not show significant level growth differences across groups (90-100 dB peSPL; p=0.13).

Whereas cochlear neuropathy studies report shallower ABR wave-I growth in normal-threshold subjects with auditory nerve-deficits (Furman *et al.*, 2013), the results for the ABR wave-V seem to indicate that cochlear gain loss might in stead steepen the ABR growth function (across 4 intensities). Because it is not clear whether this relation also holds true for the measured ABR wave-I, Fig. 2 studies the relation between cochlear gain loss and the ABR slope metrics in more detail.

Figure 2 demonstrates that DPOAE thresholds at 4 kHz did not significantly correlate to ABR wave-V (panel B; p=0.08) and wave-I level growth (panel C; p=0.5) when only considering the highest two stimulation levels (90-100 dB peSPL). Differently, for participants with DPOAE thresholds above 20 dB SPL, Fig. 2C shows significantly steeper ABR wave-I growth functions, whereas this relation is missing for listeners with thresholds below 20 dB. With respect to the neuropathy hypothesis that predicts shallower ABR wave-I growth functions for normal-threshold listeners with auditory-nerve deficits, it could be that the absence of a relation between the DPOAE threshold and ABR wave-I growth in the normal-threshold group can be explained by cochlear neuropathy (or other) effects. However, the steeper wave-I growth functions found for elevated-threshold listeners are in clear contrast to the

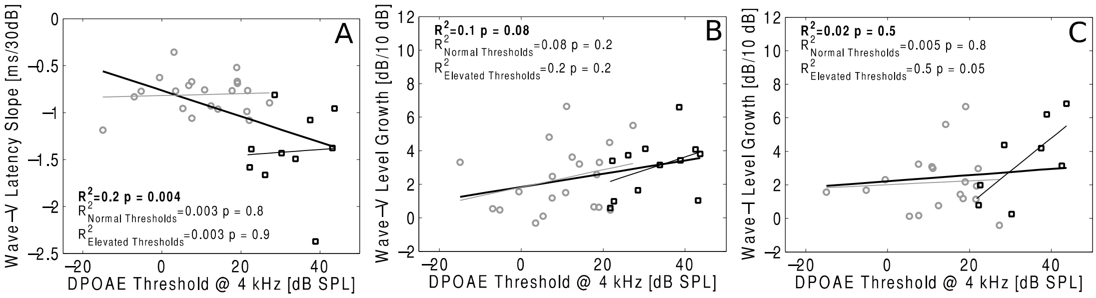


Fig. 2: Relation between the DPOAE threshold at 4 kHz and the ABR wave-V latency slope (A), wave-V level slope (B), and wave-I level slope (C) for the normal (gray) and elevated (black) hearing threshold group. The latency slope was calculated for levels between 70 and 100 dB peSPL, and the wave-I and V level slopes were calculated between the 90 and 100 dB peSPL levels.

neuropathy hypothesis and demonstrate that cochlear gain losses can steepen the ABR growth function even though it can be assumed that cochlear neuropathy occurs before thresholds are elevated (Sergeyenko *et al.*, 2013). Even though more compelling (and physiological) evidence is required, the steep wave-I level growth functions could result from cochlear gain loss being the dominant effect in determining ABR wave-I level growth when both gain and auditory-nerve deficits are present. Lastly, ABR wave-I and wave-V level growth did not correlate in individual listeners (p=0.7) demanding caution when using wave-V level or growth metrics to diagnose auditory-nerve deficits as level-dependent properties of processing centers between the auditory nerve and inferior colliculus might contribute to the ABR wave-V level.

Lastly, Fig. 2A shows a significant relation between the ABR wave-V latency slope and the DPOAE threshold at 4 kHz (p<0.01) yielding steeper slopes for listeners with elevated thresholds. Because the ABR latency slope is a relative metric within a specific listener, and not related to the amplitude of the ABR that can be reduced because of cochlear neuropathy (Sergeyenko *et al.*, 2013; Furman *et al.*, 2013), it might potentially be a differential diagnostic tool of cochlear gain loss. Further, the observed correlation between wave-I and wave-V latency-vs-intensity slopes in individual listeners (p<0.05) supports the view that the cochlear gain loss influence on cochlear excitation that steepens the wave-I latency-vs-intensity slope is still present at the level of the ABR wave-V. However, to prove that ABR wave-V or wave-I latency-vs-intensity curves are fully independent from cochlear neuropathy, it needs to be demonstrated that the onset latency characteristics of the different AN fiber types (low vs high-spontaneous rate) do not significantly influence ABR wave-V latency-vs-intensity curves. There is currently no explicit proof. However, a physiological study that shows a relatively small contribution of low-spontaneous rate AN fibers to the onset peak of the supra-threshold population compound action potential in gerbils (Bourien *et al.*, 2014) does support the limited role of low-spontaneous rate AN fibers to the population onset response.

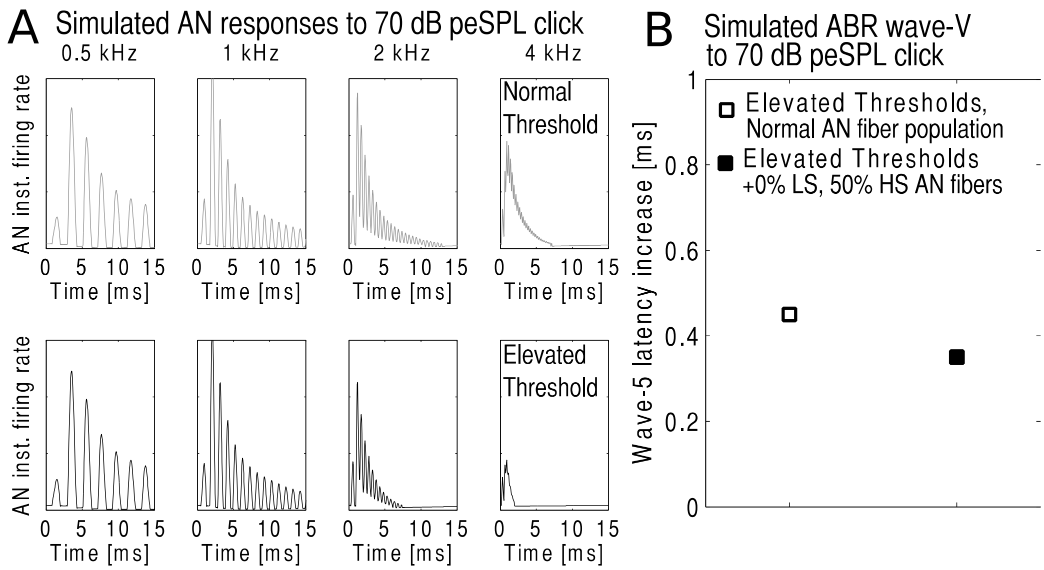


Fig. 3: A: Simulated single-unit auditory-nerve responses to a 70 dB peSPL click at 4 CFs for a normal-hearing model (top) and model with a sloping hearing loss (bottom). Note the reduced contribution of the 2 and 4 kHz channels in the model with elevated thresholds. B: ABR wave-V latency change in the elevated threshold model compared to the normal-hearing model in response to a 70 dB-peSPL click.

DISCUSSION

In this section, the relationship between the ABR wave-V latency slope and cochlear gain loss is further investigated by studying how cochlear filter changes at local basilar-membrane locations can yield population responses with increased ABR wave latencies for listeners with elevated DPOAE thresholds. For this purpose, the functional ABR model by Verhulst *et al.* (2015) was adopted in which cochlear gain loss and auditory-nerve fiber loss can be manipulated on a frequency-dependent basis. ABR wave-V latency was evaluated for a 70 dB peSPL click in three models: (i) a normal-hearing model with normal cochlear filter tuning characteristics (Shera *et al.*, 2010) and a normal auditory-nerve fiber population (70% high, 15% medium and low spontaneous-rate fibers), (ii) a model with a normal AN fiber population, but with a sloping cochlear gain loss (0 dB until 1kHz; 30 dB at 8 kHz), and (iii) a model with a sloping cochlear gain loss and a reduced AN fiber population (loss of 100% medium/low and 50% high spontaneous-rate fibers).

Figure 3A shows that local AN firing responses (summed across all available AN fibers and types for each CF) for the elevated-threshold model (bottom) have reduced in amplitude and exhibit earlier peak latencies at those frequency where a cochlear gain loss was introduced (2 and 4 kHz). This observation stems from the shorter duration and lower amplitude basilar-membrane impulse responses as cochlear gain is reduced and local cochlear filters widen. However, when summing up all energy in individual CF channels to yield the population response wave-V, Fig. 3B demonstrates that even though local BM impulse responses had shorter peak-latencies, the overall wave-V latency is increased by 0.4 ms. The increased latency of the wave-V response can be explained by the higher dominance of the longer-latency low-frequency channels to the population response when a sloping high-frequency cochlear gain loss is introduced.

In support, the slope of the audiometric hearing loss also has also experimentally been shown to influence to the latency-vs-intensity characteristics of the ABR (Gorga *et al.*, 1985). Additional simulations confirm this relation to the audiogram shape, as a flat hearing loss configuration yields an overall ABR wave-V latency decrease because shorter basilar-membrane impulse response peak-latencies occur in all frequency channels contributing to the population response (Verhulst *et al.*, 2013).

Figure 3B further demonstrates that the impact of cochlear gain loss on the ABR wave-V latency outweighs that of the loss of medium and low spontaneous-rate fibers. Low-spontaneous rate AN fibers generally fire with delayed onset peaks (Bourien et al., 2014) and seem to have a small effect on the population response latency in the present simulations. An explanation for their small contribution to population responses has been empirically explained by the large jitter in first-spike-latency for low- compared to high-spontaneous rate AN fibers (Bourien et al., 2014).

CONCLUSION

Overall, the ABR wave-V latency slopes showed a good correlation to the DPOAE threshold in listeners with normal and elevated hearing thresholds. Taken together with the simulations that show that cochlear neuropathy only has a small effect on click-ABR latency, this metric may form an auditory brainstem correlate of cochlear gain loss. Additionally, it was found that supra-threshold ABR wave-I growth was related to the DPOAE threshold in listeners with elevated thresholds, demanding caution when using this metric as an indicator for cochlear neuropathy in listeners with mixed pathologies. However, it remains possible that the wave-I growth function to narrow-band tone-pip stimuli (Furman *et al.*, 2013) – as opposed to the clicks adopted here – are more sensitive to neuropathy in listeners with elevated thresholds. Lastly, click-ABR wave-V and wave-I growth characteristics were different in individual listeners, complicating a direct and straightforward interpretation of wave-V levels in terms of auditory-nerve deficits.

ACKNOWLEDGMENTS

Work supported by DFG Cluster of Excellence EXC 1077/1 “Hearing4all”.

REFERENCES

Bourien, J., Tang, Y., Batrel, C., Huet, A., Lenoir, M., Ladrech, S., Desmadryl, G., Nouvian, R., Puel, J.L., and Wang, J. **(2014)**. “Contribution of auditory nerve fibres to compound action potential of the auditory nerve,” J. Neurophys., **12**, 1025-1039.

Furman, A.C., Kujawa, S.G., and Liberman, M.C. **(2013)**. "Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates," J. Neurophys., **110**, 577-586.

Gorga, M.P., Worthington, D.W., Reiland, J.K., Beauchaine, K.A., and Goldar, D.E. **(1985)**. "Electrophysiological techniques in audiology and otology – some comparisons between auditory brain-stem response thresholds, latencies and the pure-tone audiogram," Ear. Hearing, **6**, 105-112.

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

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

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.

Kummer, P., Janssen T., and Arnold, W. **(1998)**. "The level and growth behavior of the 2 f1−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.

Mauermann, M. and Kollmeier, B. **(2004)**. "Distortion product otoacoustic emission (DPOAE) input/output functions and the influence of the second DPOAE source," J. Acoust. Soc. Am., **116**, 2199-2212.

Melcher, J.R. and Kiang, N. **(1996)**. "Generators of the brainstem auditory evoked potential in cat III: identified cell populations," Hear. Res., **93**, 52-71.

Neely, S.T., Johnson, T.A., and Gorga, M.P. **(2005)**. "Distortion-product otoacoustic emission measured with continuously varying stimulus level," J. Acoust. Soc. Am., **117**, 1248-1259.

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

Sergeyenko, Y., Lall, K., Liberman, M.C., and Kujawa, S.G. **(2013)**. "Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline," J. Neurosci., **33**, 13686-13694.

Shera, C.A., Guinan, J.J.J., and Oxenham, A.J. **(2010)**. "Otoacoustic estimation of cochlear tuning: Validation in the chinchilla," J. Assoc. Res. Otolaryngol., **11**, 343-365.

Strelcyk, O., Christoforidis, D., and Dau, T. **(2009)**. "Relation between derived-band auditory brainstem response latencies and behavioral frequency selectivity," J. Acoust. Soc. Am. **124**, 1878-1888.

Verhulst, S., Bharadwaj, H., Mehraei G., and Shinn-Cunningham, B.G. **(2013)**. "Understanding hearing impairment through model predictions of brainstem responses," Proc. Meetings Acoust., **19**, 050182.

Verhulst, S., Bharadwaj, H., Mehraei, G., Shera, C.A., and Shinn-Cunningham, B.G. **(2015)**. "Functional modeling of the human auditory brainstem response to broadband stimulation," J. Acoust. Soc. Am., **138**, 1637-1659.

Is cochlear gain reduction related to speech-in-babble performance?

KRISTINA DEROY MILVAE*, JOSHUA M. ALEXANDER,
AND ELIZABETH A. STRICKLAND

Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN, USA

Noisy settings are difficult listening environments. With some effort, individuals with normal hearing are able to overcome this difficulty when perceiving speech, but the auditory mechanisms that help accomplish this are not well understood. One proposed mechanism is the medial olivocochlear reflex (MOCR), which reduces cochlear gain in response to sound. It is theorized that the MOCR could improve intelligibility by applying more gain reduction to the noise than to the speech, thereby enhancing the internal signal-to-noise ratio. To test this hypothesized relationship, the following measures were obtained from listeners with clinically normal hearing. Cochlear gain reduction was estimated psychoacoustically using a forward masking task. Speech-in-noise recognition was assessed using the QuickSIN test (Etymotic Research), which generates an estimate of the speech reception threshold (SRT) in background babble. Results were surprising because large reductions in cochlear gain were associated with large SRTs, which was the opposite of the hypothesized relationship. In addition, there was a large range for both cochlear gain reduction and SRT across listeners, with many individuals falling outside of the normal SRT range despite having normal hearing thresholds.

INTRODUCTION

We are able to navigate the world around us using sensorineural systems that give us a sense of touch, sight, smell, taste, and sound. These sensory systems work by detecting changes in our environment, such as the sound of a friend's voice above the noise of a restaurant. To detect the friend's voice, it would be helpful for our auditory system to have a differential response to the varying speech relative to the ongoing background noise. It is known that one function of cochlear outer hair cells is to provide gain to basilar membrane motion for low-level acoustic stimulation. If relatively less gain is applied to the steady noise, then acoustic changes associated with the speech can be detected more easily. One possible mechanism to accomplish this is the medial olivocochlear reflex (MOCR).

*Corresponding author: kderoy@purdue.edu

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

The MOCR is a bilateral reflex in the auditory system involving the inner ear and brainstem pathways. Once activated by acoustic stimulation along some place on the basilar membrane, the MOCR acts to reduce the outer hair cell gain at that place (Cooper and Guinan, 2006). This reflex takes about 25 ms to fully activate, making it a sluggish feature of the auditory system (Backus and Guinan, 2006; James *et al.*, 2005).

One hypothesis for the role of the MOCR is that it improves perception in noise. Auditory nerve fibers are able to better respond to changes in a signal embedded in noise when the MOCR is activated (Kawase *et al.*, 1993; Winslow and Sachs, 1987). In addition, the MOCR may improve the signal-to-noise ratio (SNR) for speech in noise, as shown in modelling studies (Ghitza, 1988; Messing *et al.*, 2009).

The relationship between physiological estimates of MOCR gain reduction strength (using contralateral suppression of otoacoustic emissions [OAEs]) and speech-in-noise performance has been measured in correlational studies. Results have been mixed. Some studies have found a positive correlation (Bidelman and Bhagat, 2015; de Boer and Thornton, 2008; Giraud *et al.*, 1997; Kumar and Vanaja, 2004), but another found a negative correlation (de Boer *et al.*, 2012). In addition, some work has found no correlation between the two measures (Wagner *et al.*, 2008). The reason for this variety of findings is not yet clear.

An alternative measure of cochlear gain reduction, likely related to MOCR activity, can be estimated using psychoacoustic measures (Krull and Strickland, 2008; Roverud and Strickland, 2010; Strickland, 2001; Yasin *et al.*, 2014). There are some advantages to the use of behavioral measures over OAEs. Behavioral measures allow for quantification of cochlear gain reduction in terms that may help us better understand functional consequences for perception. In addition, measures of cochlear gain reduction from ipsilateral stimulation can be easily measured with this technique, so that ipsilateral gain reduction can be compared to ipsilateral speech-in-noise performance. Studies investigating psychoacoustic measures of ipsilateral gain reduction have primarily investigated cochlear gain reduction at the 4-kHz place. However, because speech is a broadband signal and has more energy at lower frequencies, it is important to consider MOCR function at lower frequencies as well. With contralateral acoustic stimulation, psychoacoustic evidence of cochlear gain reduction has been found at frequencies as low as 500 Hz (Aguilar *et al.*, 2013). The present study will estimate cochlear gain reduction at both 2 and 4 kHz to improve our understanding of ipsilateral cochlear gain reduction across frequency.

We hypothesize that participants with relatively larger gain reduction estimates will perform better on a speech-in-noise task. This study builds on previous work in that we measure ipsilateral gain reduction at a frequency that is more relevant to speech perception than that traditionally measured. In addition, the perceptual measure of gain reduction is compared to performance on the QuickSIN (Etymotic Research), thereby allowing us to investigate the relationship between psychoacoustic measures of ipsilateral cochlear gain reduction and speech-in-noise performance.

METHOD

Participants

Twenty young adults (7 male, 13 female) between the ages of 18 and 28 years (median: 20 years) completed this experiment in exchange for modest monetary compensation. All participants reported English as a first language. Participants passed a hearing screening of 15 dB HL at 0.25, 0.5, 1, 2, 4, and 8 kHz in both ears. The hearing screening was completed in a sound-treated booth. One additional participant did not pass the hearing screening and testing was discontinued.

Stimuli and procedure

Speech-in-babble performance. Speech understanding in noise was measured using the QuickSIN (Speech-in-Noise) Test (Etymotic Research; Killion *et al.*, 2004). Participants listened to a recording of a woman's voice and background speech babble at various SNRs. Sentences were presented from 0-25 dB SNR in 5-dB steps and descending order for each list of six sentences (easiest to most difficult condition). Participants responded by repeating the target sentence at each SNR. A practice list was used to familiarize participants with the task. Next, four test lists (lists 1-4) were used. Sentences were scored according to the test instructions and were based on the number of keywords repeated correctly.

The QuickSIN measures speech reception threshold (SRT), which is the SNR required for 50%-correct performance. SRTs above +4 dB (the normative range for SNR loss plus the 2 dB reference for listeners with normal hearing) are considered outside the normal range (Killion *et al.*, 2004). The QuickSIN was presented at 70 dB HL to each participant's right ear via ER-3A insert earphones using a CD player routed to an audiometer (GSI-61).

Estimate of cochlear gain reduction. Estimates of cochlear gain reduction were measured in the right ear. The signal was a 2-kHz, 10-ms tone (5-ms $\cos^2$ ramps) or a 4-kHz, 6-ms tone (3-ms $\cos^2$ ramps). These durations were chosen to keep the signals as short as possible with minimal frequency spread. Participant thresholds for the tone alone were compared with those for the same tone preceded by a 50-ms, 60-dB SPL broadband noise precursor and a 20-ms silent gap. The precursor bandwidth was 0.25-8 kHz, and 5-ms $\cos^2$ ramping was used at onset and offset. High-pass noise was presented during each precursor interval to limit off-frequency listening (e.g., Nelson *et al.*, 2001). The noise began 50 ms before the first stimulus and ended 50 ms after the signal (5-ms $\cos^2$ ramps), and was presented at a spectrum level 50 dB below the signal level. The frequency content of the high-pass noise ranged from 1.2 times the signal frequency to 10 kHz.

This paradigm is based on the one used by Roverud and Strickland (2010), with silence replacing the off-frequency masker in an effort to isolate masking due to cochlear gain reduction from masking due to excitation. Previous research has provided evidence that preceding stimulation in this temporal paradigm is more consistent with cochlear gain reduction than temporal integration of sound (Jennings *et al.*, 2009; Roverud and Strickland, 2014).

Stimuli were generated using a custom Matlab (2012a, The Math Works, Natick, MA) program and delivered by a Lynx II XLR sound card. The sounds were passed through a headphone buffer (TDT HB6) and then delivered to insert earphones (ER-2). Adaptive tracking (Levitt, 1971) was implemented in the computer program to approximate the 70.7% correct threshold on the psychometric function, using a rule that increases the intensity of the signal after one incorrect response and decreases the intensity of the signal after two correct responses. Step sizes began at 5 dB and then decreased to 2 dB after the fourth reversal. The program continued testing until 12 reversals were completed. Threshold was defined as the average of the levels of the final 8 reversals.

Participants were instructed that they would hear three intervals. The task was to identify the interval containing the signal for each set of stimuli. Four thresholds were measured for each of the four conditions. Adaptive runs with a reversal point standard deviation greater than 5 dB were discarded, and additional runs were completed to obtain four estimates of threshold for each condition. However, due to a programming error, only 3 threshold estimates were obtained for one of the forward masking conditions in 4 participant data sets.

The order in which the conditions were completed always began with a signal-alone condition and ended with a signal-and-precursor condition. In addition, same-frequency conditions had no more than 1 condition separating them in time. This ordering was used to ensure that participants were familiarized with the signal before completing the forward masking task, and resulted in 4 groups of 5 participants who completed the task in the same order.

RESULTS

QuickSIN results from each list tested were averaged to estimate the SRT of each participant. All scores fell in the normal/near normal to mild SNR loss range as indicated by the scoring guide.

Estimates of cochlear gain reduction were calculated by subtracting the average threshold for the signal alone condition from the average threshold for the broadband noise condition for each signal frequency. One outlier with a gain reduction estimate at 2 kHz that was greater than 2 times the interquartile range (Tukey's criteria) was excluded from further analysis. A within-subjects ANOVA [$F(1,18) = 9.66$, $p = 0.006$] revealed that gain reduction was significantly greater at 4 kHz ($M = 11.61$, $SD = 4.73$) than at 2 kHz ($M = 8.17$, $SD = 3.22$).

A similar relationship was seen between SRTs and estimates of gain reduction at 2 and 4 kHz. Participants with better speech-in-noise performance (lower SRT) had smaller gain reduction estimates than those with poorer speech-in-noise performance. In fact, a linear relationship was found between these two variables for 2 kHz [$r(17) = 0.70$, $p = 0.001$], excluding one outlier. However, the relationship was not statistically significant at 4 kHz [$r(18) = 0.32$, $p = 0.174$].

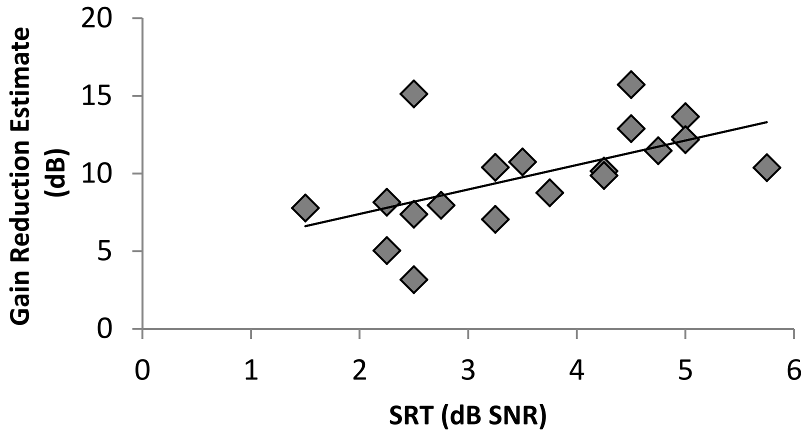


Fig. 1: Observed relationship between SRT and gain reduction estimates averaged across 2 and 4 kHz for 19 participants (one outlier excluded).

Figure 1 demonstrates the correlation between gain reduction estimates and speech-in-noise performance when gain reduction estimates were averaged across the two frequencies [$r(17) = 0.57$, $p = 0.010$]. One participant's data were again excluded according to Tukey's criteria for outliers.

DISCUSSION

The correlation showed that participants with better speech-in-noise performance had smaller gain reduction estimates than those with poorer speech-in-noise performance. The correlation was stronger when gain reduction was estimated at 2 kHz than when gain reduction was estimated at 4 kHz. This relationship between speech-in-noise performance and cochlear gain reduction is the opposite of that hypothesized.

This counterintuitive finding is similar to that found by de Boer *et al.* (2012), who used a different technique in an attempt to examine the same relationship. They used a consonant identification-in-noise task and compared those results to a reduction in OAE amplitude with contralateral stimulation. De Boer and colleagues found that participants with large contralateral suppression of OAEs performed poorer on the speech-in-noise task. They reasoned that the demand on attention is different between the two measures. It has been shown that the MOCR is under some attentional control (Delano *et al.*, 2007; Maison *et al.*, 2001). Since OAEs do not require the participant's attention, de Boer and colleagues (2012) hypothesized that differences in attentional control across the conditions could possibly explain their counterintuitive findings.

In our study, however, participants were actively engaged in both the measure of MOCR strength and the speech-in-noise task, which suggests that the attentional control explanation does not explain the observed relationship. Alternatively, perhaps there is something about the measure that explains this relationship. It is unlikely that activation of the MOCR leads to poorer perception in noise, given physiological data that suggests the opposite (Kawase *et al.*, 1993). Behavioral data also suggests that the MOCR improves perception in noise. For example, May *et al.* (2004) found that cats performed much better on a localization task in noise with their olivocochlear neurons intact.

The measure of speech perception in noise in this study involved an estimate of the SNR where performance was 50% correct. Because of this, each participant's SRT represented threshold performance at different SNRs. Although this is a valid way to measure a decrement in speech-in-noise performance, perhaps measurement at different SNRs is not the best choice to examine the relationship between MOCR strength and speech-in-noise performance. Research has shown that this method can confound data when an effect is SNR-dependent (Bernstein, 2012).

The MOCR may improve performance at certain SNRs and hinder performance at other SNRs. Kumar and Vanaja (2004) found that contralateral acoustic stimulation improved speech perception for ipsilateral SNRs of +10 and +15 dB, but not +20 dB. In hearing aid research, a parallel is the action of wide dynamic range compression (WDRC). When WDRC is activated, gain is provided by the hearing aid to the input sound, increasing the level presented to the ear. As the level of the input sound to the hearing aid rises, the hearing aid decreases the amount of gain provided. This variable gain has similarities to that provided by the outer hair cells in the inner ear. Research has shown that WDRC progressively decreases positive SNRs, especially for fast-acting multichannel compression and steady background noise (Alexander and Masterson, 2015). This body of research inspires the idea that a more systematic approach to measurement of speech perception in noise is preferable. By measuring performance at several SNRs, it will be possible to see if the relationship between ipsilateral cochlear gain reduction and speech-in-noise performance changes with SNR.

It is also possible that bilateral stimulation is needed to observe a beneficial relationship between cochlear gain reduction and speech-in-noise performance. The MOCR is, after all, a bilateral reflex. Natural listening situations such as cocktail parties, where MOCR activity could be beneficial, are situations where both ears are involved in listening to a target. There is possible interplay between cochlear feedback and localization cues.

The results of this experiment also bring to light individual differences. All participants passed a hearing screening at 15 dB HL, yet there was a range of both SRT and gain reduction estimates for these individuals. In the case of the SRT measurements, many participants with hearing thresholds in the normal range had SRTs outside of the normal range.

This study is the first to examine ipsilateral cochlear gain reduction with psychoacoustic methods at 2 kHz. This frequency may be more relevant to speech perception than 4 kHz, which is the frequency most frequently examined. This experiment is a first step in connecting psychoacoustic observations of cochlear gain reduction to speech perception, by showing that cochlear gain reduction is observed at a frequency with higher importance for speech intelligibility (Fletcher and Galt, 1950). This study is also the first to relate a psychoacoustic measure of cochlear gain reduction to speech-in-noise performance, allowing a comparison between two conditions where the ipsilateral MOCR pathway may be activated.

REFERENCES

Alexander, J.M. and Masterson, K.M. (**2015**). "Effects of WDRC release time and number of channels on output SNR and speech recognition," Ear Hearing, **36**, e35-e49.

Aguilar, E., Eustaquio-Martin, A., and Lopez-Poveda, E.A. (**2013**). "Contralateral efferent reflex effects on threshold and suprathreshold psychoacoustical tuning curves at low and high frequencies," J. Assoc. Res. Otolaryngol. **14**, 341-357.

Backus, B.C. and Guinan, J.J., Jr. (**2006**). "Time-course of the human medial olivocochlear reflex," J. Acoust. Soc. Am., **119**, 2889-2904.

Bernstein, J.G.W. (**2012**). "Controlling signal-to-noise ratio effects in the measurement of speech intelligibility in fluctuating maskers," Proc. ISAAR 2011: *Speech Perception and Auditory Disorders*. 3rd International Symposium on Auditory and Audiological Research, Nyborg, Denmark. Eds. T. Dau, M.L. Jepsen, T. Poulsen, and J.C. Dalsgaard. ISBN 978-87-990013-3-0 (The Danavox Jubilee Foundation, Copenhagen), pp. 33-44.

Bidelman, G.M. and Bhagat, S.P. (**2015**). "Right-ear advantage drives the link between olivocochlear efferent 'antimasking' and speech-in-noise listening benefits," NeuroReport, **26**, 483-487.

Cooper, N.P. and Guinan, J.J., Jr. (**2006**). "Efferent-mediated control of basilar membrane motion," J. Physiol., **576**, 49-54.

de Boer, J. and Thornton, A.R.D. (**2008**). "Neural correlates of perceptual learning in the auditory brainstem: Efferent activity predicts and reflects improvement at a speech-in-noise discrimination task," J. Neurosci., **28**, 4929-4937.

de Boer, J., Thornton, A.R.D., and Krumbholz, K. (**2012**). "What is the role of the medial olivocochlear system in speech-in-noise processing?" J. Neurophysiol., **107**, 1301-1312.

Delano, P.H., Elgueda, D., Hamame, C.M., and Robles, L. (**2007**). "Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas," The J. Neurosci., **27**, 4146-4153.

Fletcher, H. and Galt, R.H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am., **22**, 89-151.

Ghitza, O. (**1988**). "Auditory neural feedback as a basis for speech processing," ICASSP, **88**, 91-94.

Giraud, A.L., Garnier, S., Micheyl, C., Lina, G., Chays, A., and Chery-Croze, S. (**1997**). "Auditory efferents involved in speech-in-noise intelligibility," NeuroReport, **8**, 1779-1783.

James, A.L., Harrison, R.V., Pienkowski, M., Dajani, H.R., and Mount, R.J. (**2005**). "Dynamics of real time DPOAE contralateral suppression in chinchillas and humans," Int. J. Audiol., **44**, 118-129.

Jennings, S.G., Strickland, E.A., and Heinz, M.G. (**2009**). “Precursor effects on behavioral estimates of frequency selectivity and gain in forward masking,” J. Acoust. Soc. Am., **125**, 2172-2181.

Kawase, T., Delgutte, B., and Liberman, M.C. (**1993**). “Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones,” J. Neurophysiol., **70**, 2533-2549.

Killion, M.C., Niquette, P.A., Gudmundsen, G.I., Revit, L.J., and Banerjee, S. (**2004**). “Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners,” J. Acoust. Soc. Am., **116**, 2395-2405.

Krull, V. and Strickland, E.A. (**2008**). “The effect of a precursor on growth of forward masking,” J. Acoust. Soc. Am., **123**, 4352-4357.

Kumar, U.A. and Vanaja, C.S. (**2004**). “Functioning of olivocochlear bundle and speech perception in noise,” Ear Hearing, **25**, 142-146.

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

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

May, B.J., Budelis, J., and Niparko, J.K. (**2004**). “Behavioral studies of the olivocochlear efferent system – Learning to listen in noise,” Arch. Otolaryngol., **130**, 660-664.

Messing, D.P., Delhorne, L., Bruckert, E., Braida, L.D., and Ghitza, O. (**2009**). “A non-linear efferent-inspired model of the auditory system; matching human confusions in stationary noise,” Speech Commun., **51**, 668-683.

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.

Roverud, E. and Strickland, E.A. (**2010**). “The time course of cochlear gain reduction measured using a more efficient psychophysical technique,” J. Acoust. Soc. Am., **128**, 1203-1214.

Roverud, E. and Strickland, E.A. (**2014**). “Accounting for nonmonotonic precursor duration effects with gain reduction in the temporal window model,” J. Acoust. Soc. Am., **135**, 1321-1334.

Strickland, E.A. (**2001**). “The relationship between frequency selectivity and overshoot,” J. Acoust. Soc. Am., **109**, 2062-2073.

Wagner, W., Frey, K., Heppelmann, G., Plontke, S.K., and Zenner, H.-P. (**2008**). “Speech-in-noise intelligibility does not correlate with efferent olivocochlear reflex in humans with normal hearing,” Acta Oto-Laryngol., **128**, 53-60.

Winslow, R.L. and Sachs, M.B. (**1987**). “Effect of electrical stimulation of the crossed olivocochlear bundle on auditory nerve response to tones in noise,” J. Neurophysiol., **57**, 1002-1021.

Yasin, I., Drga, V., and Plack, C.J. (**2014**). “Effect of human auditory efferent feedback on cochlear gain and compression,” J. Neurosci., **34**, 15319-15326.

Effects of cochlear compression and frequency selectivity on pitch discrimination of complex tones with unresolved harmonics

FEDERICA BIANCHI*, MICHAL FERECZKOWSKI, JOHANNES ZAAR, SÉBASTIEN SANTURETTE, AND TORSTEN DAU

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

Physiological studies have shown that noise-induced sensorineural hearing loss (SNHL) enhances the amplitude of envelope coding in auditory-nerve fibers. As pitch coding of unresolved complex tones is assumed to rely on temporal envelope coding mechanisms, this study investigated pitch-discrimination performance in listeners with SNHL. Pitch-discrimination thresholds were obtained in 14 normal-hearing (NH) and 10 hearing-impaired (HI) listeners for sine-phase (SP) and random-phase (RP) unresolved complex tones. The HI listeners performed, on average, similarly as the NH listeners in the SP condition and worse than NH listeners in the RP condition. Cochlear compression and auditory filter bandwidths were estimated in the same listeners. A significant correlation was found between the reduction of cochlear compression and the difference between RP and SP pitch-discrimination thresholds. The effects of degraded frequency selectivity and loss of compression were considered in a model as potential factors in envelope enhancement. The model revealed that a broadening of the auditory filters led to an increase of the modulation depth in the SP condition, while it did not have any effect for the RP condition. Overall, these findings suggest that both reduced cochlear compression and auditory filter broadening alter the envelope representation of unresolved complex tones, leading to changes in pitch-discrimination performance.

INTRODUCTION

Sensorineural hearing loss (SNHL) is commonly associated with reduced frequency selectivity and a reduced ability to extract temporal fine structure information (Moore et al., 2006; Hopkins and Moore, 2007; Strelcyk and Dau, 2009). However, recent physiological studies in animals showed that noise-induced SNHL increases the temporal precision and the amplitude of envelope coding in single auditory-nerve fibers (Kale and Heinz, 2010; Henry et al., 2014). These findings were ascribed to a variety of factors, such as broader auditory filters, a reduction of cochlear compression due to outer hair cell damage and altered auditory-nerve response temporal dynamics. Thus, while fine spectro-temporal cues are disrupted,

*Corresponding author: fbia@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

temporal envelope cues may be enhanced and the relative importance of spectral and temporal cues for pitch processing may be altered in listeners with SNHL. Although it is commonly reported that hearing-impaired (HI) listeners have disrupted abilities in pitch discrimination of complex tones (Hoekstra and Ritsma, 1977; Arehart, 1994; Bernstein and Oxenham, 2006), a more precise examination of these findings suggests that the performance of HI listeners is not always disrupted as compared to NH listeners. In fact, some studies showed a similar performance of HI *vs.* NH listeners in pitch discrimination of complex tones with unresolved harmonics (Arehart, 1994; Bernstein and Oxenham, 2006). Since the broadening of auditory filters in HI listeners leads to an increased amount of unresolved harmonics as compared to NH listeners, it seems plausible that HI listeners rely more on the temporal information conveyed by the unresolved complex tones than on the fine spectro-temporal information conveyed by the resolved complexes. It is still unclear whether the altered importance of temporal *vs.* spectral cues for pitch discrimination may be additionally due to the suggested enhancement of temporal envelope coding with SNHL (Kale and Heinz, 2010; Henry *et al.*, 2014).

The aim of the present behavioural study was to clarify (i) whether human listeners with SNHL show an enhanced pitch discrimination performance for unresolved complexes, and (ii) if this enhancement is related to the broadening of auditory filters and/or to the reduction of cochlear compression. Pitch discrimination of complex tones was investigated behaviourally as a function of the fundamental frequency (F_0) in NH listeners and listeners with SNHL. Additionally, auditory filter bandwidths and cochlear compression were estimated in the same listeners to assess how SNHL was related to pitch discrimination performance. Finally, a simplified peripheral model was used to predict how filter broadening and cochlear compression affected the envelope representation of unresolved complex tones.

METHOD

Listeners

Fourteen NH listeners and ten HI listeners participated in this study. All NH listeners had hearing thresholds of less than 20 dB hearing level (HL) at all audiometric frequencies between 125 Hz and 8 kHz. The HI listeners had hearing thresholds between 30 and 60 dB HL at the audiometric frequencies between 1 and 4 kHz.

Pitch discrimination of complex tones

A three-alternative forced-choice (3-AFC) paradigm was used in combination with a weighted up-down method (Kaernbach, 1991) to measure the 75% point on the psychometric function. For each trial, two intervals contained a reference complex tone with a fixed F_0 and one interval contained a deviant complex tone with a larger F_0. The listeners' task was to select the interval containing the tone with the highest pitch. Before the actual test, the listeners performed three repetitions of training. The final value of F_0DL was calculated from the mean of three repetitions.

All signals consisted of 300-ms complex tones embedded in threshold equalizing noise (TEN; Moore *et al.*, 2000). The level of each component of the complex tone was set at 12.5 dB sensation level (SL) relative to the mean pure tone thresholds (at 1.5, 2, and 3 kHz) in TEN. For the NH listeners, the level of the TEN was set to 55 dB SPL per equivalent rectangular bandwidth (ERB; Glasberg and Moore, 1990) to mask the combination tones. For the HI listeners, pure-tone detection in quiet was performed at 1.5, 2, and 3 kHz and the level of the TEN was set at the maximum threshold measured in this range. The complex tones were created by summing harmonic components either in sine phase (SP) or random phase (RP) to vary the envelope peakiness. All HI listeners carried out the SP and RP conditions, whereas only 9 out of the 14 NH listeners completed the measurements for both conditions. Conditions of varying resolvability were achieved by bandpass filtering the complexes between 1.5 and 3.5 kHz, with 50 dB/octave slopes.

Auditory-filter bandwidth estimation

The auditory-filter bandwidth at 2 kHz was estimated from the temporal modulation transfer functions (TMTFs) in the 10 HI listeners. A 3-AFC paradigm, in combination with a weighted up-down rule, was used to measure modulation detection thresholds at the 75% point of the psychometric function. For each trial, two intervals contained a 300-ms pure tone at 2 kHz and one interval contained a sinusoidally amplitude-modulated 2-kHz tone at modulation frequencies (f_ms) between 25 and 1500 Hz. For each listener, the auditory-filter bandwidth was estimated at the f_m leading to a modulation threshold that was 9.5 dB below the maximum point of the TMTF. This point was selected since it led to an estimated filter bandwidth of 325 Hz at 2 kHz for NH listeners, which corresponds to the mean equivalent rectangular bandwidth (ERB) estimated via the notched-noise method by Bernstein and Oxenham (2006).

Cochlear compression estimation

Masker thresholds were measured in nine out of the 10 HI listeners as a function of the temporal gap between a 16-ms probe at 2 kHz and a 200-ms masker, either “on-frequency” at 2 kHz or “off-frequency” at 0.6 of the probe frequency (Fereczkowski, 2015). The on-frequency and off-frequency masker thresholds were paired to form a set of basilar membrane input/output (BM I/O) points (Nelson et al., 2001). A two-section function was fitted to the set of points to approximate the listener’s BM I/O function and the inverse slope of the shallow section was taken as an estimate of the compression ratio at 2 kHz (CR).

RESULTS

Figure 1 depicts the mean pitch-discrimination thresholds for NH listeners (black solid symbols), as well as the individual thresholds for HI listeners (open symbols), for the SP condition (left panel) and the RP condition (right panel). The thresholds for both conditions showed similar trends for the NH listeners, whereby F_0DLs decreased with increasing F_0. A one-way ANOVA confirmed a significant effect of

F_0 for both conditions [SP: $F(8,117) = 10, p < 0.001$; RP: $F(8,72) = 12.6, p < 0.001$]. The current findings are in agreement with previously reported pitch-discrimination thresholds (Bernstein and Oxenham, 2006), where the improvement in performance with increasing F_0 was thought to reflect the progressive increase of the resolvability of the harmonics. The grey shaded area in Fig. 1 depicts the two conditions (at F_0s of 100 and 125 Hz) for which the harmonics are considered to be completely unresolved – i.e., no significant effect of F_0 between the mean thresholds of NH listeners [SP: $F(1,26) = 0.05, p = 0.833$; RP: $F(1,16) = 0.69, p = 0.420$].

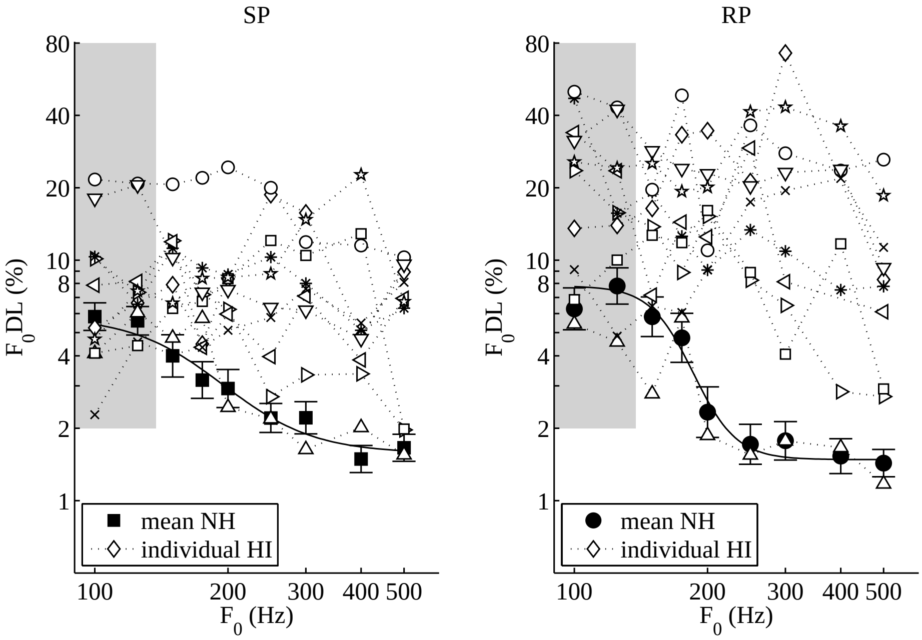


Fig. 1: Pitch-discrimination thresholds for the SP condition (left panel) and RP condition (right panel). The solid symbols depict the mean results for 14 NH listeners (left panel) and 9 NH listeners (right panel). The open symbols depict the individual results for the 10 HI listeners. Error bars depict the standard error of the mean. The grey-shaded region highlights the unresolved conditions.

The performance for the 10 HI listeners was generally worse than for the NH listeners, whereby the mean threshold across HI listeners differed significantly from the mean threshold of the NH listeners [SP: $F(1,16) = 26.21, p < 0.001$; RP: $F(1,16) = 33.93, p < 0.001$]. However, for the two unresolved conditions in the grey-shaded area there was no significant difference between the mean of the NH vs. the HI listeners for the SP condition [100 Hz: $F(1,22) = 0.6$, p = 0.446; 125 Hz: $F(1,22) = 2.63, p = 0.119$], while a post-hoc one-tailed t-test revealed significantly larger mean thresholds for the HI vs. the NH listeners for the RP condition [100 Hz: $p = 0.002$; 125 Hz: $p = 0.020$]. Overall, these findings revealed that HI listeners performed similarly as NH listeners in pitch discrimination of unresolved complex tones for the SP condition and worse than NH listeners for the RP condition.

In order to quantify and compare these changes in performance across participants, the ratio between the RP and SP threshold (F_0DL ratio) was calculated for the individual HI listeners and for the mean of the NH listeners. Figure 2 depicts the calculated F_0DL ratios as a function of the estimated cochlear compression (i.e., the slope of the BM I/O function, 1/CR; left panel) and filter bandwidth (right panel). Nine out of 10 HI listeners had F_0DL ratios larger than the upper boundary of the confidence interval for the ratio of the NH listeners (y-axis in the left and right panels). The increase of F_0DL ratios positively correlated with the estimated loss of cochlear compression for all listeners (left panel in Fig. 2) [$R^2 = 0.58$, $p = 0.011$]. Thus, the lower the residual cochlear compression, and thus CR, the larger was the difference in performance between RP-complex tones and SP-complex tones. Three HI listeners (asterisk, left-pointing triangle and star) showed the largest loss of cochlear compression and the largest F0DL ratio, while their filter bandwidths were similar to the average in the remaining HI listeners. Thus, for these three listeners, the loss of cochlear compression seemed to be the dominant factor increasing the F0DL ratio. No significant correlation was found between F_0DL ratio and auditory filter bandwidth (right panel in Fig. 2), although a significant positive correlation [$R^2 = 0.66$, $p = 0.015$] was reported when leaving out the three HI listeners. Overall, these findings suggest that both auditory filter broadening and loss of cochlear compression contribute in altering the pitch discrimination performance of the unresolved complexes, although the relative contribution of each factor remains unclear.

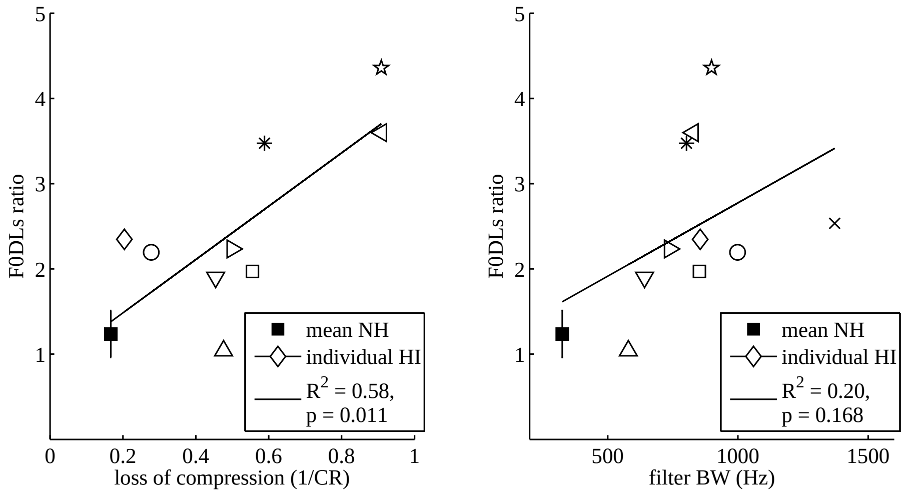


Fig. 2: F_0DL ratios as a function of the estimated loss of cochlear compression (left panel) and filter bandwidth (right panel). Solid symbols depict the mean results for NH listeners. The open symbols depict the individual results for HI listeners. Error bars depict the standard deviation of the mean (for the 9 NH listeners that measured both SP and RP conditions).

DISCUSSION

The hypothesis of the current study was that if the envelope representation is enhanced in listeners with SNHL (Kale and Heinz, 2010; Henry *et al.*, 2014), pitch cues for unresolved complex tones should also be enhanced if one assumes an envelope coding mechanism for pitch extraction of unresolved complexes. The pitch-discrimination thresholds measured in the present study revealed that the performance of the HI listeners for the unresolved conditions was similar to that of the NH listeners when the harmonics were added in SP (left panel in Fig. 1) and worse for the RP condition (right panel in Fig. 1). Although for most HI listeners pitch-discrimination performance was not better than for the NH listeners, these findings do not rule out an enhanced envelope representation following SNHL. In fact, other factors might be involved in limiting the behavioural performance of HI listeners (e.g., disrupted temporal fine structure cues, higher internal noise, other central limitations). Overall, these findings suggest that changes in the internal envelope representation occurred in listeners with SNHL.

The difference in performance between the RP and SP conditions (F_0DL ratio) was considered as an indicator of envelope coding. Correlations between F_0DL ratios and individual estimates of cochlear compression and filter bandwidth (Fig. 2) revealed that both cochlear compression reduction and filter broadening increased the F_0DL ratio. Although these two factors are known to be physiologically linked and dependent on outer-hair cell damage (Ruggero, 1992), the behavioural estimates of cochlear compression and filter bandwidth obtained in the present study did not show a one-to-one correspondence. Thus, a simplified peripheral model was used to qualitatively explain the relative effect of one factor versus the other on the envelope representation of the unresolved complex tones. SP and RP complexes were processed via a gammatone filter centred at 2 kHz, the output of which was processed by a broken stick non-linearity, as defined by Jepsen and Dau (2011). After envelope extraction, the modulation depth of the output signal was calculated for the SP and RP conditions, as well as their modulation depth ratio. Four different filter bandwidths were used (i.e., from 1 to 2.5 ERBs), as well as three levels of compression (NH compression, mild compression loss, and severe compression loss). The model parameters of the broken stick non-linearity were adjusted according to the fits of Jepsen and Dau (2011).

Figure 3 depicts the obtained modulation depth ratio between the SP and RP envelopes together with the F_0DL ratios calculated for the NH (solid symbol) and HI listeners (open symbols). The output of the model qualitatively predicted the trends in the data, whereby both compression reduction and filter broadening increased the modulation depth ratio. The larger the loss of cochlear compression (indicated by the different lines in Fig. 3), the larger was the effect of filter broadening on increasing the modulation depth ratio (i.e., the steeper the curve). Additionally, the model revealed that a broadening of the auditory filters led to an increase of the modulation depth (i.e., a peakier envelope) in the SP condition, since more components added up in phase, while it did not have any effect for the RP condition. The reduction of cochlear compression led to an increase of the modulation depth for both SP and RP conditions, although the envelope enhancement was larger for the SP condition. Thus, for the SP condition both filter broadening and loss of compression increased the envelope amplitude, whereas for the RP condition only loss of compression led to a moderate envelope enhancement.

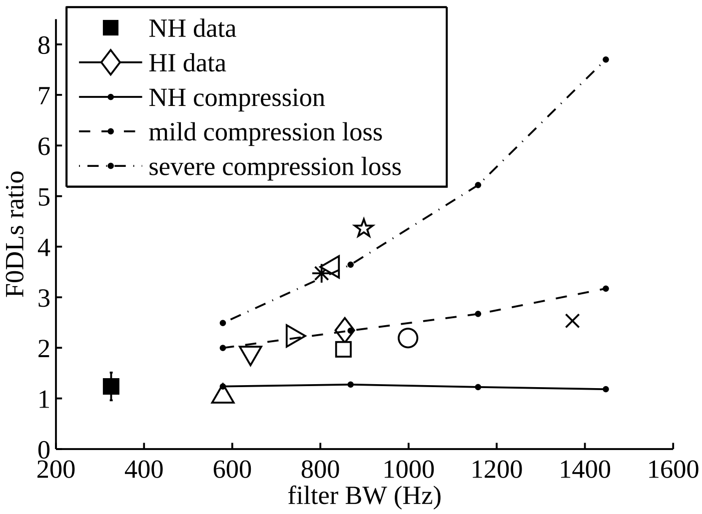


Fig. 3: Modelling results for 3 levels of residual compression (solid line: NH compression; dashed line: mild compression loss; dash-dot line: severe compression loss) and 4 levels of filter broadening, in comparison with the F_0DL ratios for NH listeners (solid square symbol) and for the HI listeners (open symbols).

CONCLUSION

Overall, the results of the pitch-discrimination experiment revealed that the performance of the HI listeners was, on average, similar to that of the NH listeners for the SP unresolved complex tones, and worse for the RP complexes. This difference in performance (F_0DL ratio) was significantly correlated with the decrease in residual cochlear compression. These findings suggest that changes in the internal envelope representation of unresolved complex tones occurred in listeners with SNHL, possibly as a result of their reduced compression, and altered their performance in pitch discrimination. Moreover, the outcomes of a simplified peripheral model revealed that both auditory filter broadening and loss of cochlear compression contributed to enhance the envelope peakiness of the unresolved complex tones, especially in the SP condition. Thus, the internal envelope representation of the unresolved complexes might be enhanced in listeners with SNHL for both the SP and the RP conditions, with the largest enhancement for the SP condition, while the behavioral performance of HI listeners could be affected by more central limitations.

REFERENCES

Arehart, K.H. (**1994**). “Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners,” J. Acoust. Soc. Am., **95**, 3574-3585.

Bernstein, J.G. and Oxenham, A.J. (**2006**). “The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level,” J. Acoust. Soc. Am., **120**, 3916-3928.

Fereczkowski, M. (**2015**). *Time-Efficient Behavioural Estimates of Cochlear Compression.* Doctoral thesis, Technical University of Denmark.

Glasberg, B. R. and Moore, B.C.J. (**1990**). “Derivation of auditory filter shapes from notched-noise data,” Hear. Res., **47**, 103-138.

Henry, K.S., Kale, S., and Heinz, M.G. (**2014**). “Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers,” Front. Sys. Neurosci., **8**, 1-10.

Hoekstra, A. and Ritsma, R.J. (**1977**). “Perceptive hearing loss and frequency selectivity,” in *Psychophysics and Physiology of Hearing.* Eds. Evans, E.F., Wilson, J.P. (London: Academic), pp. 263-271.

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.

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.

Kale, S. and Heinz, M.G. (**2010**). “Envelope coding in auditory nerve fibers following noise-induced hearing loss,” J. Assoc. Res. Otolaryngol., **11**, 657-673.

Kaernbach, C. (**1991**). “Simple adaptive testing with the weighted up-down method,” Percept. Psychophys., **49**, 227-229.

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

Moore, B.C.J., Glasberg, B.R., and Hopkins, K. (**2006**). “Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure information,” Hear. Res., **222**, 16-27.

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.

Ruggero, M.A. (**1992**). “Responses to sound of the basilar membrane of the mammalian cochlea,” Curr. Opin. Neurobiol., **2**, 449-456.

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.

Characterizing individual hearing loss using narrow-band loudness compensation

DIRK OETTING[1,2,*], JENS-E. APPELL[1], VOLKER HOHMANN[2], AND STEPHAN D. EWERT[2]

[1] *Project Group Hearing, Speech and Audio Technology of the Fraunhofer IDMT, Oldenburg, Germany*

[2] *Medizinische Physik and Cluster of Excellence Hearing4all, Universität Oldenburg, Oldenburg, Germany*

Loudness is one of the key factors related to overall satisfaction with hearing aids. Individual loudness functions can reliably be measured using categorical loudness scaling (CLS) without any training. Nevertheless, the use of loudness measurement like CLS is by far less common than use of audiometric thresholds to fit hearing aids, although loudness complaints are one of the most mentioned reasons for revisiting the hearing aid dispenser. A possible reason is that loudness measurements are typically conducted with monaural narrow-band signals while binaural broad-band signals as speech or environmental sounds are typical in daily life. This study investigated individual uncomfortable loudness levels (UCL) with a focus on monaural and binaural broad-band signals, as being more realistic compared to monaural narrow-band signals. Nine normal-hearing listeners served as a reference in this experiment. Six hearing-impaired listeners with similar audiograms were aided with a simulated hearing aid, adjusted to compensate the narrow-band loudness perception back to normal. As desired, monaural narrow-band UCLs were restored to normal, however large individual deviations of more than 30 dB were found for the binaural broad-band signal. Results suggest that broad-band and binaural loudness measurements add key information about the individual hearing loss beyond the audiogram.

INTRODUCTION

To compensate a hearing loss with multichannel dynamic compression, frequency- and level-dependent gains have to be adjusted to fit to the individual ear. The most common approach is to use audiogram-based fitting formulas, but still loudness complaints are one of the most mentioned reasons for revisiting the hearing aid dispenser (Jenstad *et al.*, 2003). The use of loudness measurements like categorical loudness scaling (CLS) is by far less common although individual supra-threshold information about the hearing loss can precisely be assessed (Brand and Hohmann, 2001). One reason might be that the typical monaural narrow-band test stimuli used in the CLS procedure are not suitable to describe the loudness perception of amplified binaural broad-band signals like speech as later processed by the hearing aid.

*Corresponding author: dirk.oetting@idmt.fraunhofer.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

However, so far no systematic measurements of binaural broad-band uncomfortable loudness levels (UCL) were conducted after hearing-impaired (HI) listeners were compensated for the monaural narrow-band loudness perception. In this study UCLs of signals with different bandwidth in monaural and binaural conditions were measured in HI listeners and compared with a normal-hearing (NH) group. The HI listeners had similar, typical age-related hearing losses and were aided with a simulated hearing aid that performed a static, frequency and level dependent amplification. The amplification was individually adjusted to restore the narrow-band loudness perception back to that of the NH control group.

METHODS

Nine younger NH (mean±std. age: 26.3±3.3 y) and six older HI (73.8±2.8 y) listeners participated in this study. All HI listeners had a high-frequency hearing loss and no self-reported tinnitus sensation. The HI listeners were selected to have similar hearing threshold levels as shown in Fig. 1. The PTA (500, 1k, 2k, and 4k) was between 30 and 44 dB HL.

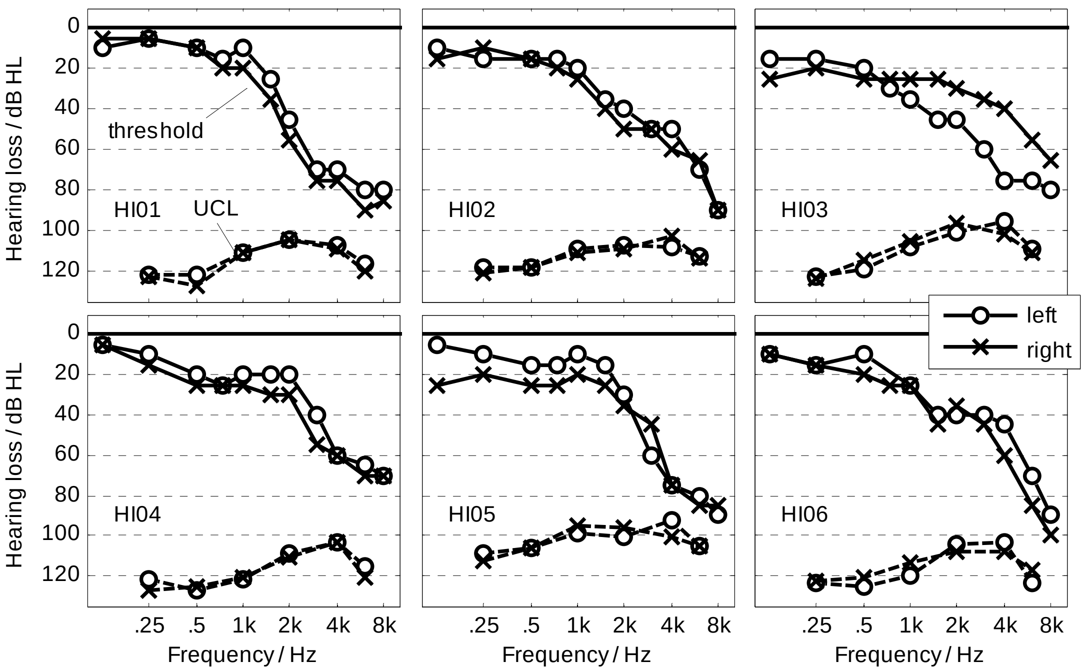


Fig. 1: Audiograms of the six HI listeners with high frequency hearing losses. Subjects were selected to have similar hearing threshold levels. The bottom lines in each panel show the uncomfortable loudness levels (UCL) corresponding to the level for “too loud” (50 CU) on the loudness function.

All subjects conducted the adaptive categorical loudness scaling procedure (ACALOS; Brand and Hohmann, 2002) with one-third octave signals (low-noise noise, LNN) at six center frequencies (250, 500, 1k, 2k, 4k, and 6k). Three repetitions on at least two different days were performed. The stimulus duration was 1000 ms. The uncomfortable loudness levels (UCL) for "too loud" (50 categorical units; CU) of the LNN signals were extracted and are included at the bottom of each panel in Fig. 1. The narrow-band loudness functions were used to provide each HI listeners with a narrow-band loudness compensating algorithm where the average loudness functions of the NH listeners served as the target loudness function for the gain calculations. The method for gain calculation is shown in Fig. 2. The channel levels of an input signal were determined in six channels having the same center frequency as the LNN signals as shown in Fig. 2a. The gain calculation for the 2-kHz channel for subject HI02 is shown in Fig. 2b. The narrow-band NH loudness corresponding to the channel level was determined (black vertical line in Fig. 2b) and the required gain to restore the narrow-band loudness back to normal was extracted (horizontal black line, 23.5 dB). The gain values at each center frequency were interpolated on a logarithmic frequency and on a logarithmic level axis and applied to the input signal as static gains in the frequency domain (Fig. 2c).

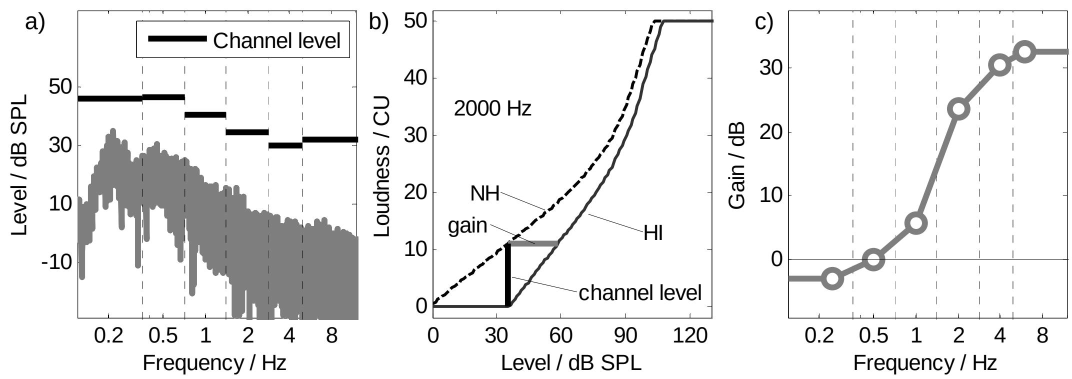


Fig. 2: Gain calculation for narrow-band loudness compensation using listener HI02 as example: a) Channel levels in six channels of the IFnoise at 50 dB SPL; b) gain values were extracted from the difference between the average NH loudness function (dashed) and the individual loudness function (solid); c) gains values were interpolated on a logarithmic frequency axis and applied in the frequency domain for the left and right ear independently.

UCLs were measured in NH and narrow-band loudness compensated HI listeners for different test signals. As test signals uniform exciting noise (UEN) with 1- and 5-Bark bandwidth and a speech-shaped noise (international female noise, IFnoise) were used. The Bark spectra of these signals are shown in Fig. 3a. Loudness scaling measurements to extract the UCL was conducted for monaural and binaural presentation. Three repetitions on at least two different days were performed. For data analysis, the differences of the UCL compared to the mean NH listeners (ΔUCL) were assessed as shown in Fig. 3b.

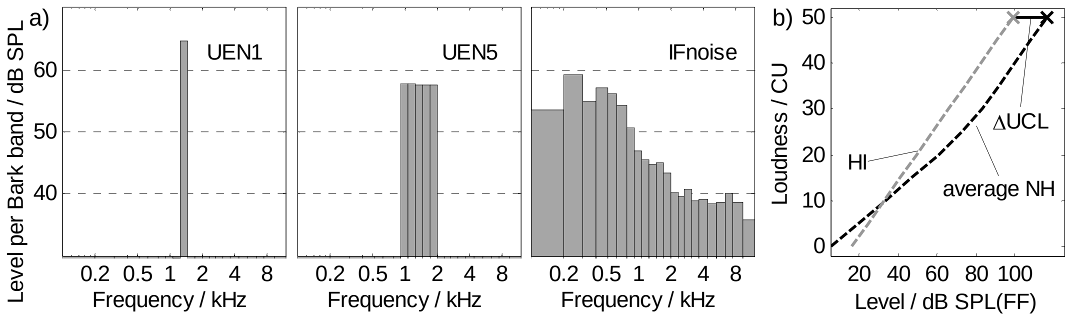


Fig. 3: a) Bark spectrum of the test signals; b) The difference of the UCL at 50 CU compared to the mean NH values (ΔUCL) was extracted and used for further data analysis.

All measurements were conducted with Sennheiser HDA200 headphones in a sound-proof both. Signals were presented using an RME Fireface UC at 44.1 kHz and a Tucker-Davis HB7 headphone driver. Headphones were calibrated using the B&K artificial ear 4153, B&K 0.5-inch microphone 4134, B&K microphone preamplifier 2669, and B&K measuring amplifier 2610. Signals were calibrated using the free-field equalization according to ISO 389 (2004). The maximum presentation level was 105 dB HL for the LNN signals and 105 dB SPL for the test signals (UEN1, UEN5, and IFnoise).

RESULTS

UCL differences are shown for each HI listener on a 2D map in Fig. 4 with the three test signals on the x-axis with increasing bandwidth (UEN1, UEN5, IFnoise) and the presentation mode (left, binaural, right) on the y-axis. The grayscale-coded map shows the difference between the UCL of the average NH listener and the individual measured UCL (ΔUCL). Measurement points are indicated by the white circles whereas all other pixels are interpolated to facilitate visual accessibility.

The scaling bar at the right side of the figure shows the absolute values of the grayscale. Light gray correspond to values around 0 dB, meaning that the compensated UCL is very similar compared to the average NH UCL. Dark gray indicates lower and lighter gray indicates higher UCLs compared to the average NH listeners. Each 5 dB step is indicated by black contour lines including figures of the absolute amount of ΔUCL.

The narrow-band UEN1 signal for the left and right condition results in light gray colors (top and bottom left corner of each panel) for all HI listeners. The restored monaural narrow-band UCLs were close to the average NH UCL. This confirmed that the gains from the narrow-band loudness compensation rule were appropriately set, at least around the center frequency of the UEN1 noise (1370 Hz). For most listeners, similar UCL values were also observed in the binaural condition for UEN1, but two listeners (HI01 and HI03) showed lower UCLs indicated by the darker gray towards the middle of the left edge of each 2D map. With increasing bandwidth of the test signals, the differences between listeners further increased. ΔUCL values for the binaural IFnoise condition (middle point of the right edge of each panel) were between −30 dB (HI01, HI03) and around 0 dB (HI05, HI06). Listeners HI02 and HI04 showed ΔUCL values around 0 dB in the for the monaural IFnoise condition, but ΔUCL values decreased to −10 and −15 dB in the binaural IFnoise condition.

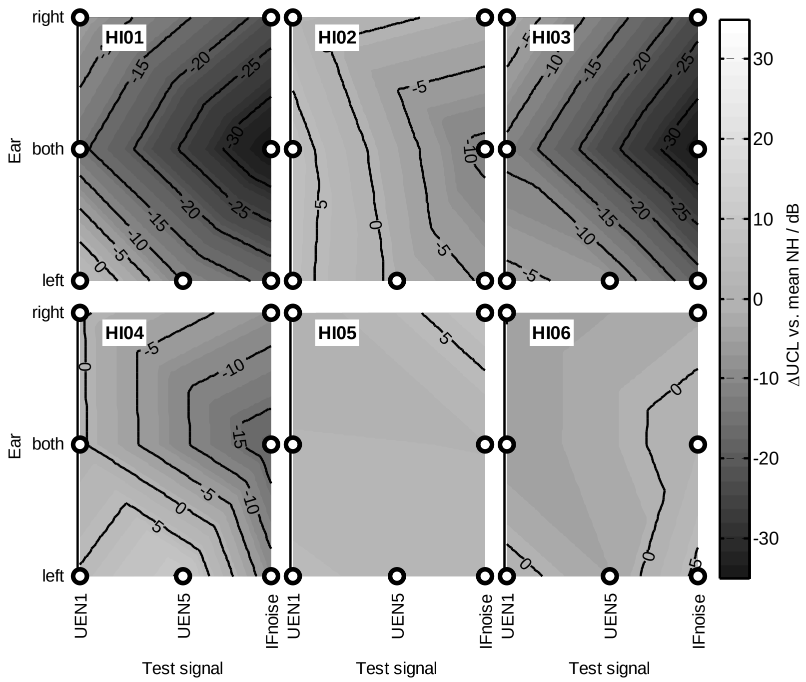


Fig. 4: Grayscale coded ΔUCL values indicate the level difference between individual UCL values after loudness compensation and the average NH UCL values. Large individual variabilities were observed: i) Similar values as for NH listeners were found in HI05 and HI06; ii) Lower UCL values for monaural and binaural broad-band signals were found in listeners HI01 and HI03; and iii) lower UCL values in the binaural broad-band case but not in the monaural case were observed in listeners HI02 and HI04.

Figure 5 shows the same 2D maps of ΔUCLs comparing individual NH listeners to the average NH listener. Overall, the ΔUCL values are within a ±10 dB range except for NH04 who showed up to 20 dB higher UCL values compared with the average NH listener.

Comparison of the results for the NH and the HI listeners indicates that individual variations were considerably higher in the narrow-band loudness compensated HI listeners than in the NH control group.

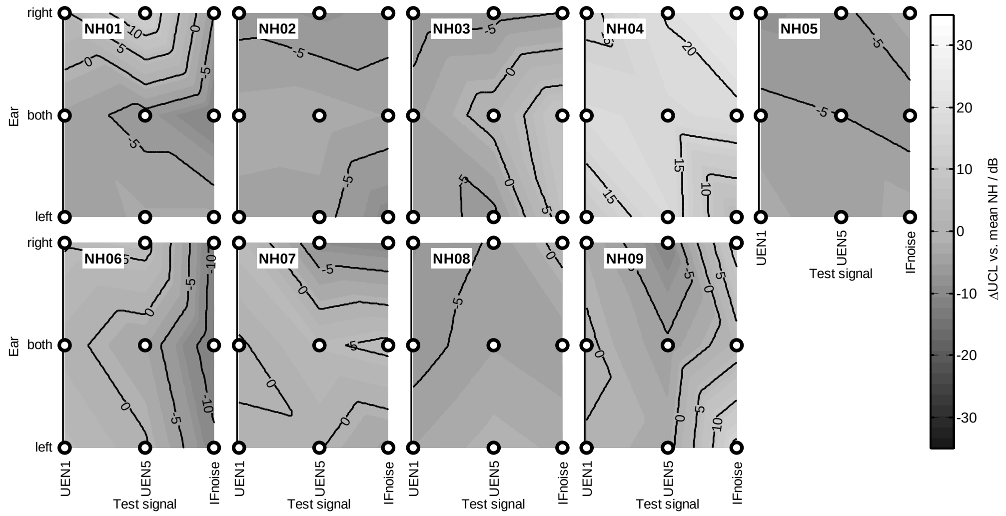


Fig. 5: Same as Fig. 4 but for the NH listeners. Individual variations were lower compared to the HI listeners.

DISCUSSION

After narrow-band, monaural loudness perception was compensated for in HI listeners, large individual variations in the uncomfortable loudness level (UCL) were observed for other types of signals. Especially for binaural broad-band test signals the UCL was lowered by up to 30 dB whereas other HI listeners showed totally normal UCL values.

Bentler and Pavlovic (1989) showed an increased amount of spectral loudness summation of tone complexes at the UCL in HI listeners compared to NH listeners. Furthermore, an increasing amount of individual variations was indicated by increased standard deviations compared to a NH group, but they tested only monaurally. Surprisingly, several subjects in the current data showed a decrease of more than 10 dB of the UCL value for the broad-band signal when comparing the monaural with the binaural presentation. This means that gains which were adjusted for the correct loudness perception in the left and right ear, separately, can be too high for loudness compensation if they are used in a bilateral presentation mode. Increased loudness sensitivity was found by Smeds *et al.* (2006), where hearing aid gains were adjusted according to NAL-NL1 which should led to normal or lower-than-normal loudness (Byrne *et al.*, 2001). The aided HI listeners rated the loudness higher than the NH listeners for broadband binaural signals with medium to high input levels. These observations are in line with the current data. Furthermore, Smeds *et al.* (2006) already speculated that there might be a problem with the underlying loudness model in NAL-NL1. The underlying loudness model is a monaural loudness model (Moore and Glasberg, 1997) which cannot account for an altered binaural summation in HI listeners. Keidser *et al.* (2012) mentioned that about 45% of the subjects preferred lower gains than prescribed by NAL-NL1. The successor fitting rule NAL-NL2 includes the empirical insights and therefore further reduces the prescribed gains. These gain adjustments might be more suitable for normal loudness of binaural broad-band signals, but do not consider the individual variations of binaural broad-band UCLs as found in the current data. Because of the similar hearing thresholds of the HI listeners the prescribed gains would be quite similar by fitting formulas based on the hearing threshold.

The current binaural broad-band UCL measurements might contain valuable information for hearing-aid fitting or the diagnosis of the underlying pathology. Until now, no binaural broad-band UCL measurement is included in standard clinical protocols, e.g., to determine the remaining dynamic range for broad-band binaural signals. Considering the observed large individual variability in the six subjects, it is obvious that no listener-independent correction factor for binaural presentation could be determined.

A possible reason for the increased loudness perception might be an increased central gain of the auditory system as reported for NH listeners with tinnitus by Schaette and McAlpine (2011). They measured brainstem responses in NH subjects with tinnitus compared to a NH control group and found reduced auditory-brainstem-response wave I (evoked from auditory nerve) in the tinnitus groups whereas there was no difference in wave V (evoked from inferior colliculus) between both groups which indicates an increased central gain in tinnitus patients. Qiu *et al.* (2000) found an increased auditory cortex potential in chinchillas after inner hair cell loss although the compound action potential elicit by the auditory nerve was reduced.

It remains unclear how such a potential central gain mechanism in HI listeners is realized in the auditory pathway as the increased gains based on the observed UCL differences in the HI listeners can be quite different between narrow- and broad-band signals but also for monaural and binaural presentation.

AKNOWLEDGMENTS

This work was supported by the DFG Cluster of Excellence EXC 1077/1 “Hearing4all”.

REFERENCES

Bentler, R.A. and Pavlovic, C.V. (**1989**). “Comparison of discomfort levels obtained with pure tones and multitone complexes,” J. Acoust. Soc. Am., **86**, 126-132.

Brand, T. and Hohmann, V. (**2001**). “Effect of hearing loss, centre frequency, and bandwidth on the shape of loudness functions in categorical loudness scaling,” Int. J. Audiol., **40**, 92-103.

Brand, T. and Hohmann, V. (**2002**). “An adaptive procedure for categorical loudness scaling,” J. Acoust. Soc. Am., **112**, 1597-1604.

Byrne, D., Dillon, H., Ching, T., and Katsch, R. (**2001**). “The NAL-NL1 procedure for fitting non-linear hearing aids: Characteristics and comparisons with other procedures,” J. Am. Acad. Audiol., **12**, 37-51.

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

Keidser, G., Dillon, H., Carter, L., and O'Brien, A. (**2012**). "NAL-NL2 empirical adjustments," Trends Amplif., **16**, 211-223.

Moore, B.C.J. and Glasberg, B.R. (**1997**). "A model of loudness perception applied to cochlear hearing loss," Audit. Neurosci., **3**, 289-311.

Qiu, C., Salvi, R., Ding, D., and Burkard, R. (**2000**). "Inner hair cell loss leads to enhanced response amplitudes in auditory cortex of unanesthetized chinchillas: evidence for increased system gain," Hear. Res., **139**, 153-171.

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

Smeds, K., Keidser, G., Zakis, J., Dillon, H., Leijon, A., Grant, F., Convery, E., and Brew, C. (**2006**). "Preferred overall loudness. I: Sound field presentation in the laboratory," Int. J. Audiol., **45**, 2-11.

Evaluation of a clinical auditory profile in hearing-aid candidates

NICOLINE THORUP[1,*], SÉBASTIEN SANTURETTE[2], SØREN JØRGENSEN[2], ERIK KJÆRBØL[1], TORSTEN DAU[2], AND MORTEN FRIIS[1,3]

[1] *Department of Otorhinolaryngology and Audiology, Rigshospitalet, Denmark*

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

[3] *University of Copenhagen, Copenhagen, Denmark*

Hearing-impaired (HI) listeners often complain about communicating in the presence of background noise, although audibility may be restored by a hearing-aid (HA). The audiogram typically forms the basis for HA fitting, such that people with similar audiograms are given the same prescription by default. However, this does not necessary lead to the same HA benefit. This study aimed at identifying clinically relevant tests that may be informative in addition to the audiogram and relate more directly to HA benefit. Twenty-nine HI listeners performed fast tests of loudness perception, spectral and temporal resolution, binaural hearing, speech intelligibility in stationary and fluctuating noise, and a working-memory test. Six weeks after HA fitting they answered the International Outcome Inventory – Hearing Aid evaluation. The HI group was homogeneous based on the audiogram, but only one test was correlated to pure-tone hearing thresholds. Moreover, HI listeners who took the least advantage from fluctuations in background noise in terms of speech intelligibility experienced greater HA benefit. Further analysis of whether specific outcomes are directly related to speech intelligibility in fluctuating noise could be relevant for concrete HA fitting applications.

INTRODUCTION

It has been estimated that 30% of Danish hearing-aid (HA) users found listening situations to improve only moderately, a little bit, or not at all after HA prescription (Jørgensen, 2009), suggesting inadequate HA treatment. Pure-tone audiometry typically forms the basis for administering and fitting HA devices. This implies that people with similar audiograms are given the same HA prescription by default. However, patients with the same audiometric profile may experience differences in HA benefit.

Although audibility may be restored by a HA, users often complain about communicating in the presence of background noise. Previous studies have shown

*Corresponding author: nicolinethorup13@gmail.com

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

that the audiogram correlates well with speech intelligibility in quiet but poorly with speech intelligibility in noise (Festen and Plomp, 1983; Glasberg and Moore, 1989). Moreover, hearing-impaired (HI) listeners with normal or near-normal pure-tone hearing thresholds at low frequencies may show speech identification deficits when the speech spectrum is limited to the regions of normal or near-normal hearing (Léger *et al.*, 2012). Speech intelligibility in noise has also been found to correlate with temporal fine-structure (TFS) processing abilities reflected by, e.g., frequency discrimination (Festen and Plomp, 1983; Papakonstantinou *et al.*, 2011), and TFS processing deficits can be present despite near-normal thresholds (Strelcyk and Dau, 2009). The evaluation of a test battery covering different hearing domains, hearing disability, listening effort, and cognitive function recently showed that HI listeners can suffer from auditory deficits that do not necessarily correlate with the audiogram but may be detectable in clinically-applicable tests (van Esch *et al.*, 2013).

Despite compelling evidence that the audiogram alone is insufficient to characterize hearing loss, it remains unclear which additional properties of hearing function should be assessed in the clinic to provide adequate HA rehabilitation. The aim of this study was to evaluate whether a clinical auditory profile including different psychoacoustics tests and a cognitive test adds relevant information to the audiogram. The auditory domains of interest were: spectral and temporal resolution, TFS processing, and speech perception in noise. Another aim was to evaluate HA benefit in relation to the auditory profile to investigate if specific test outcomes relate to HA benefit.

METHODS

Listeners

Twenty-nine HI listeners with sensorineural high-frequency hearing loss (age 52-80 years, mean 68.4, 13 female, 8 new and 21 experienced HA users) participated. The inclusion criteria were based on predefined audiometric categories (Bisgaard *et al.*, 2010). At low frequencies, the categories “mild” to “moderate” hearing loss were included. At high frequencies, the categories “mild” to “moderate/severe” hearing loss were included. A maximal deviation from these categories of +/- 5 dB at each frequency was allowed, except at 250 Hz and 500 Hz where no lower limits were defined. All HI listeners had bilateral HA therapy and were native Danish speakers. Listeners were excluded if they suffered from asymmetry > 15 dB hearing level (HL) at any frequency, or asymmetry in speech discrimination (DS) > 20%, or if they suffered from conductive hearing loss.

Experimental set-up

All measurements were conducted via a PC in a double-walled soundproof booth. The stimuli were generated in MATLAB and presented via a Fireface UCX sound card connected to Sennheiser HDA200 headphones. Calibration was performed using a B&K 2636 measuring amplifier and a B&K 4153 artificial ear simulator. 128-tap linear phase FIR equalization filters were applied to all broadband stimuli to flatten the headphone frequency response. For audiometric measurements

Interacoustics AC40 and AC440 connected to TDH39 headphones or Madsen Orbiter OB922 connected to HDA200 headphones were used. Insert earphones (EAR 3A) were used in listeners with a small auditory canal.

General procedure

For pre-examination, air and bone conduction pure-tone thresholds from 250-8000 Hz were measured. The test battery was always scheduled for another day than pre-examination and HA fitting. A standardized written and verbal introduction was given before each test and all tests contained a training run. The cognitive test was carried out before the psychoacoustic measurements, which sequence was randomized. All hearing tests were conducted without HA.

The test battery

A summary of all conducted tests and the corresponding outcome measures is given in Table 1. A brief description of all tests is given below.

Domain	**Test**	**Outcome**
Audibility	Pure-tone hearing thresholds	PTA_{low}: 0.25, 0.5, 1 kHz (dB HL) PTA_{high}: 2, 4, 6 kHz (dB HL)
Working memory	Reading span	Number of correct words
Spectral and temporal resolution	Combined spectral and temporal resolution test (F&T-test)	MR no gap vs. spectral gap (dB) MR no gap vs. temporal gap (dB)
Binaural TFS-processing	Interaural-phase-difference (IPD) detection	Upper frequency limit for IPD detection (Hz)
Speech perception in noise	Danish hearing-in-noise test (HINT)	MR stationary vs. fluctuating noise (dB)
Hearing-aid treatment evaluation	The “international-outcome-inventory – hearing-aid” (IOI-HA)	Score on introspection subscale Score on interaction subscale

Table 1: Tests included in the test battery and corresponding outcomes.

Reading span (RS). The reading span test was used to evaluate working memory storage and processing simultaneously (Lunner, 2003). The main task was to recall the first or the final word in a sequence of sentences. The remembered words were pronounced out loud and the test contributor registered the answers. The secondary task was to assess continuously if each sentence was correct or absurd. The participant responded by pressing the keyboard “F” (absurd) or “K” (correct) after each sentence. A total of 54 sentences (27 correct and 27 absurd) were presented. The outcome measure was the number of correctly recalled words (RS score).

Combined spectral and temporal resolution. Auditory spectral and temporal resolution were tested simultaneously using a modified version of the F&T test (Larsby and Arlinger, 1998). The task was to detect a pulsed tone at 500 Hz in the presence of broadband threshold-equalizing noise (TEN; Moore *et al.*, 2000) containing either no gap, a spectral gap, or a temporal gap. The tone length was 275 ms and the tone-pulse-interval 175 ms with a 50-ms ramp. The spectral gap was 3 equivalent rectangular bandwidths wide around the center frequency and the temporal gap around the pulsed tone was of 50-ms duration. The noise level was fixed at 55 dB SPL. The tone level was varied adaptively using a Békésy tracking method with a starting value of 70 dB SPL. Each condition was measured monaurally in the left and right ear. The sequence of noise conditions and ears was randomized. All conditions were measured twice. Masking releases (MRs) between the spectral-gap and no-gap (MR_{spec}) and temporal-gap and no-gap (MR_{temp}) conditions were calculated.

Interaural-phase-difference detection. Binaural TFS processing was evaluated by measuring the upper frequency limit for which an interaural phase difference (IPD) of 180° was detectable (Ross *et al.*, 2007), using a procedure similar to Santurette and Dau (2012). The task was to detect which of three stimulus intervals contained an IPD and thus sounded more spacious than the other two intervals with no IPD. The stimulus was a sinusoidal-amplitude-modulated pure-tone with a modulation rate at 40 Hz and modulation depth equal to 1. The presentation level was 35 dB sensation level defined from the pure-tone hearing-thresholds for each ear separately. The start frequency was 250 Hz. The frequency changed according to a 2-up 1-down rule. The frequency was changed in step-sizes of 1/2, 1/5, and 1/10 octave that decreased after each lower reversal. Two measurements were obtained for each listener.

Hearing-in-noise test. The speech reception threshold in noise (SRTn) was measured using the Danish hearing-in-noise test (HINT; Nielsen and Dau, 2011). The listener was asked to repeat the presented sentences and the answer was registered as “correct” or “false” by the test instructor. The noise was set at a fixed level of 65 dB SPL. The first sentence was presented at 0-dB speech-to-noise-ratio (SNR). The speech level was changed according to a 1-up 1-down rule. The SRTn was the mean of speech levels in the 15 last sentences minus the noise level. SRTn was measured in two different noise types: a stationary speech-shaped noise and a fluctuating background, the International Speech Test Signal (Holube *et al.*, 2010). Lists 1 and 2 from the Danish HINT sentences were used. Condition and list order were randomized. The masking release between SRTn in stationary and fluctuating noise (MR_{HINT}) was calculated.

International Outcome Inventory – Hearing Aid. To evaluate the benefit from the HA intervention the HI listeners answered the Danish IOI-HA (Jespersen *et al.*, 2006). The IOI-HA consists of 7 items and is divided into two subscales. One subscale evaluates the introspective aspects of the HA treatment and the other interaction with the surroundings. According to a new revision of the Danish translation, item 5 was omitted (Jespersen *et al.*, 2014). The greater the advantage a person has from the HA, the greater the score is in the IOI-HA evaluation.

RESULTS

The Pearson correlation coefficient was calculated between the outcomes of all tests and the low (0.25, 0.5, 1 kHz) and/or high (2, 4, 6 kHz) pure-tone average (PTA). For the F&T test, correlations were calculated between the masking releases (MRs) and the pure-tone hearing thresholds at 500 Hz after pooling the data from both ears. For the IPD detection and HINT tests, the average PTA from the right and left ear was used in the correlation analysis. Fisher's transformation was used to calculate the confidence interval (CI) for the correlation coefficient. Correlations between test outcomes and the IOI-HA subscales were obtained in the same way.

Correlations with the audiogram

Table 2 (upper rows) lists the correlations coefficient CIs between measures from each test and the PTA at high and low frequencies. Scatter plots of the individual outcomes for each test as a function of audibility are also given in Fig. 1. Only the MR_{HINT} was significantly correlated to PTA at low frequencies. Outcomes from the reading span, F&T test, and IPD detection test were not correlated to audibility.

Correlations with HA benefit

Table 2 (lower rows) lists the correlations coefficient CIs between IOI-HA and outcome measures from all tests. A negative significant correlation was found between the MR_{HINT} and the introspection subscale (also when controlled for PTA), indicating that HI listeners who took small advantage in fluctuating noise experienced a greater HA benefit. Neither audibility nor other test outcomes were correlated with HA benefit.

DISCUSSION

Comparison to earlier findings

In the present study an extended auditory profile was tested on a group of HA users. In the following, the results are compared to findings from previous studies.

		PTA_{low}	PTA_{high}	RS score	MR_{spec}	MR_{temp}	IPD	MR_{HINT}
Audio-gram	**PTA_{low}**			[-.38;.35]	[-.33;.27]	[-.53;.03]	[-.17;.56]	**[-.73;-.17]**
	p-value			*.94*	*.83*	*.08*	*.25*	***<.01***
	PTA_{high}			[-.34;.39]				[-.37;.36]
	p-value			*.90*				*.99*
IOI-HA	**Introspection**	[-.05;.64]	[-.37;.39]	[-.50;.25]	[-.27;.33]	[-.19;.41]	[-.51;.29]	**[-.79;-.26]**
	p-value	*.08*	*.94*	*.46*	*.83*	*.46*	*.54*	***<.01***
	Interaction	[-.28;.47]	[-.53;.20]	[-.56;.17]	[-.28;.34]	[-.33;.30]	[-.09;.64]	[-.45;.31]
	p-value	*.57*	*.34*	*.26*	*.83*	*.92*	*.12*	*.69*

Table 2: Confidence intervals and *p*-values for correlation coefficients between PTA and all tests (first 4 rows) and IOI-HA subscales and all tests (last 4 rows). PTA_{low} at 0.25, 0.5, 1 kHz; PTA_{high}: PTA at 2, 4, 6 kHz.

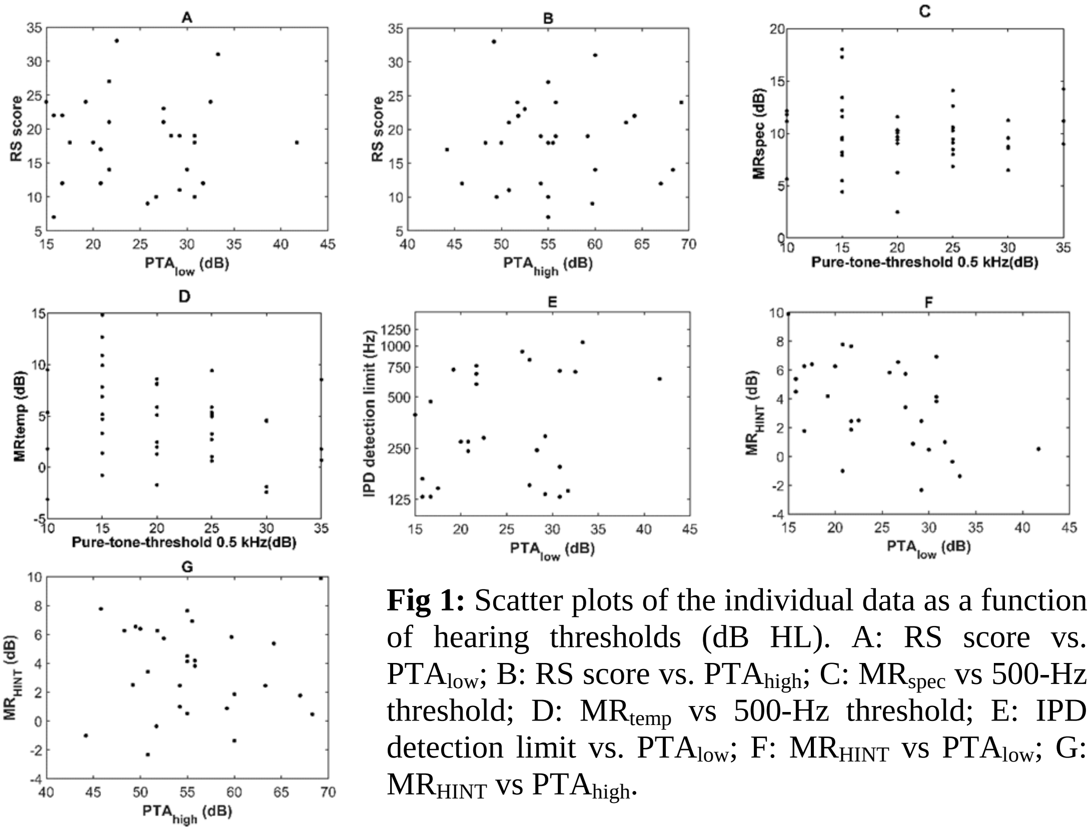


Fig 1: Scatter plots of the individual data as a function of hearing thresholds (dB HL). A: RS score vs. PTA_{low}; B: RS score vs. PTA_{high}; C: MR_{spec} vs 500-Hz threshold; D: MR_{temp} vs 500-Hz threshold; E: IPD detection limit vs. PTA_{low}; F: MR_{HINT} vs PTA_{low}; G: MR_{HINT} vs PTA_{high}.

Correlations with the audiogram. One earlier study found significant negative correlation between the reading span score and audibility (Lunner, 2003). However, in that study, both the RS score and hearing thresholds were also correlated with age, such that age could have been the determining factor. While no correlation of spectral and temporal masking releases in the F&T test with hearing thresholds was at first found (Larsby and Arlinger, 1998), a second study found a significant correlation (Larsby and Arlinger, 1999). The present study, which used a slightly different set-up aiming to make the test more independent of test frequency, showed no correlations with audibility. The present results are consistent with previous findings of absent correlation between low-frequency hearing thresholds and IPD detection thresholds in HI listeners (Santurette and Dau, 2012, Füllgrabe and Moore, 2014). The correlation between the Danish HINT and PTA was not previously examined. Here, the SRTn in stationary noise was correlated with PTA_{high} ($p<0.01$) and SRTn in fluctuating noise with PTA_{low} ($p<0.05$), and the MR between the two with PTA_{low}.

Correlations with HA benefit. Previous studies have investigated how audibility, demographic factors, HA type and fitting were related to the IOI-HA outcome. A positive correlation between hearing thresholds and items 1 and 4 and a negative correlation between hearing thresholds and item 6 were found in Jespersen *et al.* (2014). It was also found that more severe hearing impairment, previous HA experience, and bilateral fitting were significantly correlated to a higher score on the introspection subscale and a lower score on the interaction subscale (Jespersen *et al.*, 2006). In a recent study, no correlations were found between hearing thresholds, HA experience, and IOI-HA outcome (Brannstrom *et al.*, 2014). Predictions of IOI-HA outcome were also investigated by taking demographic factors and the audiogram into account. Only the DS of the better ear was found to predict 16% of the interaction sub-scale. Potential confounders could be age, poor audibility, and poorer SRT, because all these factors were related to the DS of the better ear. The IOI-HA measures HA satisfaction in general. Satisfaction in listening situations "conversation with one person", "in small group", "in larger groups", and "outdoors" were found to be important to receive a high IOI-HA outcome (Hickson *et al.*, 2010). This is consistent with the present finding that hearing in fluctuating noise is related to the IOI-HA outcome. Many factors influence IOI-outcome and the questionnaire may be too general to be directly related to specific psychoacoustic measurements in a clinical test battery. Moreover, etiological details of the hearing impairment such as family history and known genetic factors were not considered in the present study although they may play a role in differences in HA outcome in patients with similar audiometric profiles. These aspects would thus be relevant to consider in an extended hearing profile, as they may shed light on where damage is located along the auditory pathway.

Clinical feasibility

All tests were conducted in one session. A short training session and maximally two repetitions were performed. The set-up was comparable to a clinical setting. All participants were able to complete the test battery. The test set up was easily implementable as only a PC, headphones, and soundcard were needed. The duration time of the complete test-battery would have to be brought down for clinical implementation. However, only the F&T-test had a duration time above 20 minutes.

CONCLUSION

The tested auditory profile confirmed that HI listeners have difficulties in different hearing domains that are not predictable from their audiogram. The ability to make use of temporal fluctuations in background noise in terms of speech intelligibility was the only outcome measure directly related to subjective HA benefit. However, such a measure was also related to low-frequency audibility, although HA benefit was not. Further analysis of whether other specific outcomes are directly related to speech intelligibility in fluctuating noise could be relevant for concrete HA fitting applications. A large-scale evaluation of the test battery in relation to more objective measures of HA fitting and aided listening performance, as well as further reductions in testing time, are steps forward to select the key tests that would be beneficial in clinical hearing assessment.

REFERENCES

Bisgaard, N., Vlaming, M.S., and Dahlquist, M. (**2010**). "Standard audiograms for the IEC 60118-15 measurement procedure," Trends Amplif., **14**, 113-120.

Brannstrom, K.J., Lantz, J., Nielsen, L.H., and Olsen, S.O. (**2014**). "Prediction of IOI-HA scores using speech reception thresholds and speech discrimination scores in quiet," J. Am. Acad. Audiol., **25**, 154-163.

Festen, J.M. and Plomp, R. (**1983**). "Relations between auditory functions in impaired hearing," J. Acoust. Soc. Am., **73**, 652-662.

Füllgrabe, C., and Moore, B.C.J. (**2014**). "Effects of age and hearing loss on stream segregation based on interaural time differences," J. Acoust. Soc. Am., **136**, EL185-191.

Glasberg, B.R. and Moore, B.C.J. (**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.

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.

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.

Jespersen, C.T., Groth, M.B.J., and Hansen, T. (**2006**). "Danish reference data for the IOI," Proc. 21st Danavox Symposium 2005, *Hearing Aid Fitting*, Eds. A.N. Rasmussen, T. Poulsen, T. Andersen, and C.B. Larsen. ISBN 87-982422-0-2 (The Danavox Jubilee Foundation, Copenhagen), pp. 475-491.

Jespersen, C.T., Bille, M., and Legarth, J.V. (**2014**). "Psychometric properties of a revised Danish translation of the international outcome inventory for hearing aids (IOI-HA)," Int. J. Audiol., **53**, 302-308.

Jørgensen, S.L. (**2009**). "Kvalitetssikring på offentlige høreklinikker – Resultater fra tilfredshedsundersøgelse for høreapparatbrugere," DELTA Teknisk-Audiolog-isk Laboratorium, Odense, Denmark.

Larsby, B. and S. Arlinger (**1998**). "A method for evaluating temporal, spectral and combined temporal-spectral resolution of hearing," Scand. Audiol. **27**, 3-12.

Larsby, B. and S. Arlinger (**1999**). "Auditory temporal and spectral resolution in normal and impaired hearing," J. Am. Acad. Audiol. **10**, 198-210.

Léger, A.C., Moore, B.C.J., and Lorenzi, C. (**2012**). "Abnormal speech processing in frequency regions where absolute thresholds are normal for listeners with high-frequency hearing loss," Hear. Res., **294**, 95-103.

Lunner, T. (**2003**). "Cognitive function in relation to hearing aid use," Int. J. Audiol. **42**, S49-S58.

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

Nielsen, J.B., and Dau, T. (**2011**). "The Danish hearing in noise test," Int. J. Audiol., **50**, 202-208.

Papakonstantinou, A., Strelcyk, O., and Dau, T. (**2011**). "Relations between perceptual measures of temporal processing, auditory-evoked brainstem responses and speech intelligibility in noise," Hear. Res., **280**, 30-37.

Ross, B., Tremblay, K.L., and Picton, T.W. (**2007**). "Physiological detection of interaural phase differences," J. Acoust. Soc. Am., **121**, 1017-1027.

Santurette, S. and Dau, T. (**2012**). "Relating binaural pitch perception to the individual listener's auditory profile," J. Acoust. Soc. Am., **131**, 2968-2986.

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.

van Esch, T.E., Kollmeier, B., Vormann, M., Lyzenga, J., Houtgast, T., Hallgren, M., Larsby, B., Athalye, S.P., Lutman, M.E., and Dreschler, W.A. (**2013**). "Evaluation of the preliminary auditory profile test battery in an international multi-centre study," Int. J. Audiol., **52**, 305-321.

Relating $2f_1 - f_2$ distortion product otoacoustic emission and equivalent rectangular bandwidth

ANDERS T. CHRISTENSEN*, RODRIGO ORDOÑEZ AND DORTE HAMMERSHØI

Section for Signal and Information Processing, Department of Electronic Systems, Aalborg University, Aalborg, Denmark

To explore the extent of distortion product otoacoustic emission (DPOAE) toward low frequencies we measured in 21 normal-hearing human subjects its dependence on the ratio between evoking stimulus frequencies, f_1 and f_2, at $2f_1 - f_2$ distortion frequencies 88, 176, and 264 Hz. The "optimal" ratio evoking the largest DPOAE level is frequency dependent but well-guided by 1.52 equivalent rectangular bandwidth (ERB).

INTRODUCTION

Distortion-product otoacoustic emission (DPOAE) is the healthy ear's active response at distortion frequencies of two simultaneously-presented tones with frequencies f_1 and f_2 ($f_1 < f_2$) (Kemp, 1979). This two-tone stimulus evokes two traveling waves on the basilar membrane. Throughout the region excited by both tones (corresponding to the f_2 wave) distortion is generated, mostly at the $2f_1 - f_2$ frequency in humans.

DPOAE is thus, like typical measures of the frequency tuning of hearing, related to the excitation of the basilar membrane as controlled by varying the two-tone stimulus parameters. The DPOAE level for example is a bell shaped function of the frequency ratio f_2/f_1 and the "optimal" ratio is traditionally defined as that which on average evokes the largest DPOAE level.

Six systematic studies of the DPOAE level-ratio dependency consistently find an optimal ratio close to 1.22 (Christensen *et al.*, 2015a). Even though a slight increase in the optimal ratio is also consistently found as the $2f_1 - f_2$ decreases, a ratio fixed at 1.22 is standard in DPOAE measurements across frequency. With this ratio the average DPOAE level-to-noise ratio is below zero below a distortion frequency of about 0.5 kHz (Gorga *et al.*, 1993).

In the present study, the DPOAE level-ratio dependence was measured in 21 normal-hearing subjects at three $2f_1 - f_2$ frequencies: 88, 176, and 264 Hz. This is about an order of magnitude lower in frequency than typically measured and should help solidify the apparent frequency dependence of the optimal ratio.

METHODS

In 21 normal-hearing human subjects, the dependence of the $2f_1 - f_2$ DPOAE level on the stimulus frequency ratio f_2/f_1 was measured at $2f_1 - f_2$ frequencies of 88, 176,

*Corresponding author: atc@es.aau.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

and 264 Hz. The ratios measured are shown in Table 1. The stimulus levels were 65 and 55 dB sound pressure level (SPL) for f_1 and f_2, respectively, calibrated in a Brüel & Kjær 4157 ear simulator (IEC 60318-4:2010).

$2f_1 - f_2 = 88$ Hz	$2f_1 - f_2 = 176$ Hz	$2f_1 - f_2 = 264$ Hz
1.286 (9/7)	1.250 (5/4)	1.200 (6/5)
1.333 (4/3)	1.286 (9/7)	1.250 (5/4)
1.375 (11/8)	1.333 (4/3)	1.286 (9/7)
1.400 (7/5)	1.375 (11/8)	1.333 (4/3)
1.444 (13/9)	1.400 (7/5)	1.375 (11/8)
1.500 (3/2)	1.444 (13/9)	1.400 (7/5)
1.556 (14/9)	1.500 (3/2)	1.444 (13/9)

Table 1: Overview of tested stimulus parameters.

Low-frequency noise in the ear canal from breathing, blood circulation, etc., is usually filtered out electronically and the transducer sensitivities of commercial probe systems are generally tailored to OAE measurements above 0.5 kHz. Therefore, to condition measurements properly at low frequencies a custom probe system was built for use in the present study (Christensen *et al.*, 2015b), shown in Fig. 1.

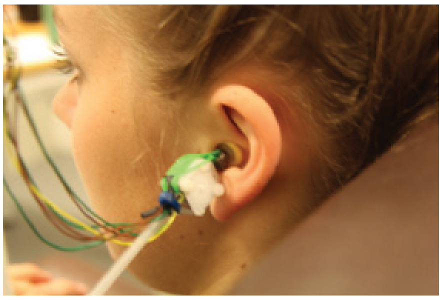


Fig. 1: Custom-made probe in a subject for DPOAE measurements at low frequencies. The $2f_1 - f_2$ frequency was held fixed at 88, 176, and 264 Hz as the stimulus ratio was varied to find the one evoking the largest DPOAE level.

Data were recorded with a frequency resolution of 1.46 Hz and the averaging duration was 95.7, 24.7, and 9.6 s at 88, 176, and 264 Hz, respectively. The duration of each measurement was usually about 10% longer because data were rejected if victim of either burst or slowly varying noise. The DPOAE level was calculated from the power in the fast Fourier transform (FFT) bin corresponding to the $2f_1 - f_2$ frequency and the noise level was calculated from power in the bins in the outer two thirds of an equivalent rectangular bandwidth (see Eq. 1) around the $2f_1 - f_2$ frequency.

RESULTS

Fig. 2 shows the DPOAE level-ratio dependence in four subjects.

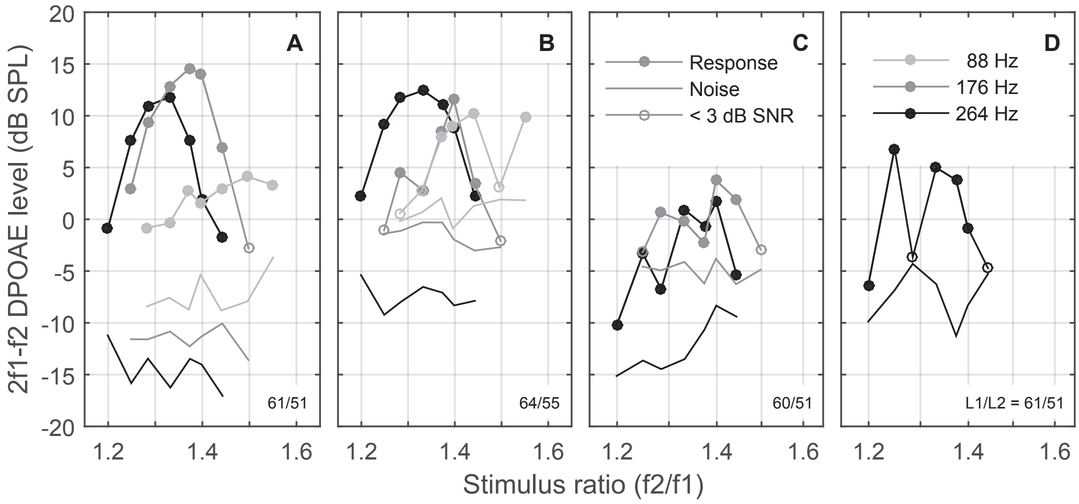


Fig. 2: Individual DPOAE levels as a function of the stimulus ratio f_2/f_1 at three different $2f_1 - f_2$ frequencies. Data points with signal-to-noise ratios less than 3 dB were rejected and the entire curve was excluded from the average if more than three points were rejected.

Aside from some irregularity, such as the dip in subject D, the bell shaped dependence known from mid and high frequencies also exists at low frequencies.

Fig. 3 shows the average DPOAE level-ratio dependence in subjects with enough measurements above the noise floor.

Eight, 15, and 20 out of 21 subjects had at least four out of 7 measured points with a signal-to-noise ratio better than 3 dB. The prevalence does not decrease toward low frequencies, as can be seen in Fig. 3, because the DPOAE level is generally lower. The prevalence decreases instead because the noise floor increases at lower frequencies, even though in this study the averaging duration was markedly increased as the $2f_1 - f_2$ frequency decreased, exemplified in Fig. 2.

The optimal ratio is not 1.22 as it is at higher frequencies. It is 1.46, 1.37, and 1.31 at 88, 176, and 264 Hz, respectively.

DISCUSSION

Equivalent rectangular bandwidth (ERB) is a measure of the bandwidth of behavioral tuning curves (Glasberg and Moore, 1990). Its empirical relation to the center frequency f [kHz] is given by

$$\mathrm{ERB}(f) = 24.7 \cdot (4.37f + 1) \quad [\mathrm{Hz}]. \tag{Eq. 1}$$

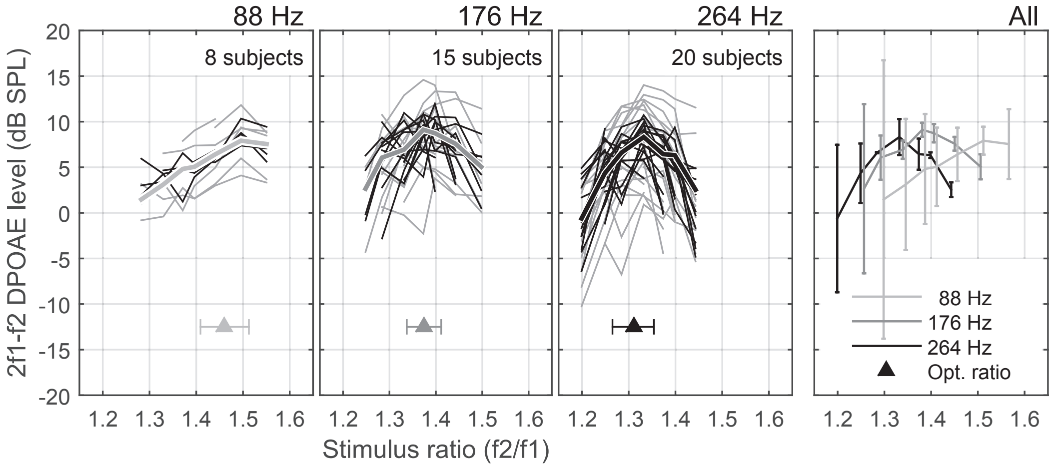


Fig. 3: Average DPOAE levels as a function of the stimulus ratio f_2/f_1 at three different $2f_1 - f_2$ frequencies. Individual data are plotted with thin lines behind the averages. The rightmost subfigure shows only the averages with error bars signifying one standard deviation. The optimal ratios shown as triangles are calculated as the average of the maximum in each individual curve.

The ERB can be related to the stimulus frequency separation in a DPOAE measurement by a scaling parameter λ (Christensen *et al.*, 2015a):

$$f_2 - f_1 = \lambda \cdot \mathrm{ERB}(f_2) \quad \text{[Hz], or} \qquad \text{(Eq. 2)}$$

$$f_1 - (2f_1 - f_2) = \frac{\lambda}{1 - 2 \cdot 24.7 \cdot 4.37} \cdot \mathrm{ERB}(2f_1 - f_2) \quad \text{[Hz].} \qquad \text{(Eq. 3)}$$

λ may then be fit by minimizing the squared difference to the data.

As summarized in Fig. 4, our results combined with the results of previous studies yield an optimal frequency separation well guided by 1.52 ERB. This shows that the optimal ratio is systematically dependent on frequency. It also suggests that the distinct places on the basilar membrane (Greenwood, 1990), excited maximally by the two stimulus tones, are secondary to the spread of excitation around those places in the generation of DPOAE. ERB is just one measure of that spread.

ENDNOTES

This is a preliminary report of an article submitted for publication in JARO.

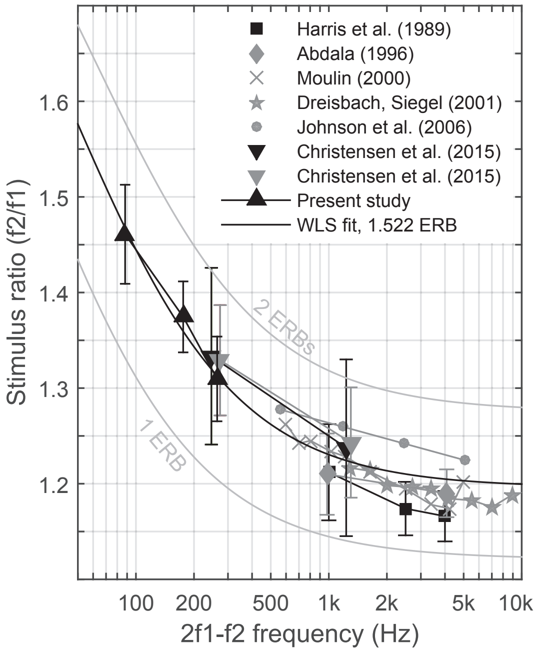


Fig. 4: Optimal ratios for DPOAE measurements as found by seven independent studies, including the present one (full references given in Christensen *et al.* (2015a)). The results comprise measurements in 98 individual subjects. Shown also is the least-squares fit of the ERB model to the data sets, weighted by the number of subjects they each represent.

REFERENCES

Christensen, A.T., Ordoñez, R., and Hammershøi, D. (**2015a**). "Stimulus ratio dependence of low-frequency distortion-product otoacoustic emissions in humans," J. Acoust. Soc. Am., **137**, 679-689.

Christensen, A.T., Ordoñez, R., and Hammershøi, D. (**2015b**). "Design of an Acoustic Probe to Measure Otoacoustic Emissions Below 0.5 kHz," Proc. 58th Audio Engineering Society Conference: *Music Induced Hearing Disorders*, June 2015.

Glasberg, B.R. and Moore, B.C.J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res., **47**, 103-138.

Gorga, M.P., Neely, S.T., Bergman, B., Beauchaine, K.L., Kaminski, J.R., Peters, J., and Jesteadt, W. (**1993**). "Otoacoustic emissions from normal-hearing and hearing-impaired subjects: Distortion product responses," J. Acoust. Soc. Am., **93**, 2050-2060.

Greenwood, D.D. (**1990**). "A cochlear frequency-position function for several species – 29 years later," J. Acoust. Soc. Am., **87**, 2592-2605.

Kemp, D.T. (**1979**). "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea," Arch. Otolaryngol. Head Neck Surg., **224**, 37-45.

Individual differences on an auditory-visual speech perception test for people with hearing loss

TATSUO NAKAGAWA*

College of Education and Human Sciences, Yokohama National University, Yokohama, Japan

Individual differences in auditory-visual speech perception in people with hearing loss were investigated using syllables, words, and sentences. The stimuli were presented in auditory-only, visual-only, and auditory-visual conditions for both congruent and incongruent conditions. In the congruent condition auditory speech stimuli were presented with their identical visual cues, and in the incongruent condition auditory stimuli were presented with conflicting visual cues. Nine young adults with varying degrees of hearing loss, fitted with hearing-aids or cochlear implants participated in the study. The relative increase in auditory-visual speech perception as measured by these tests resulting from the addition of visual cues to the auditory signal was calculated for each condition. The results showed that the subjects were better able to integrate both auditory and visual cues in the auditory-visual congruent condition. The auditory-visual gain in speech perception was less for the incongruent condition. The subjects showed significant individual differences in the amount of gain for different experimental conditions. These results suggest that auditory-visual integration of speech information does occur but that the degree of integration varies among the subjects. The speech stimuli showing the most auditory-visual integration are discussed in the text.

INTRODUCTION

It is well known that we depend on vision in addition to audition in daily speech communication, particularly in difficult listening situations such as in low signal levels and/or in high noise levels. The dependence on visual cues is greater for people with hearing loss. Several studies have shown that some people with hearing loss demonstrated visually biased responses to incongruent auditory-visual stimuli. These data were obtained with different auditory and visual syllables after cochlea implantation (Desai *et al.*, 2008; Rouger *et al.*, 2008). But recently other studies have failed to demonstrate this trend and concluded that the factors in the experimental design, such as subject's proficiency and informational content of the sensory channels, may have accounted for the different results (Schwartz, 2010; Huyse *et al.*, 2012). This study aims to clarify the characteristics of auditory-visual speech perception in college students with varying degrees of hearing loss and

*Corresponding author: t-nakagawa@ynu.ac.jp

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

different kinds of wearable listening devices. It is of particular interest to determine whether auditory visual integration is equivocally seen in different speech stimuli such as syllables, words and sentences, and if there is some difference in the perceived degree of integration between the stimuli, which stimuli are the most suitable for the clarification of auditory-visual integration

METHODS

Participants

Nine college students with hearing loss (7 women and 2 men, mean age=20.6 years, SD=0.5) participated in this study. All students communicated verbally with hearing people in daily life. Two of them had a unilateral cochlear implantation and the remaining 7 students had hearing aids bilaterally or monaurally. Table 1 provides a summary of the average hearing levels and the worn listening devices.

		S1	S2	S3	S4	S5	S6	S7	S8	S9
Right	**MHL**	91	89	55	109	75	50	80	100	109
Left		100	89	129	109	116	55	90	106	109
Right	**Devices**	HA	HA	HA	—	HA	HA	HA	HA	CI
Left		HA	HA	—	CI	—	HA	HA	HA	—

Table 1: Audiological characteristics of the participants. MHL: mean hearing level. Devices: worn listening devices, HA: hearing aid, CI: cochlear implant.

Stimuli

A female Japanese speaker was videotaped producing the speech stimuli. The twenty-one consonant-vowel syllables, twenty familiar words often used in daily school life, and twenty sentences which consist of 35 key words were used. The CV stimuli used in this study were as follows: /a, ki, shi, ta, ni, yo, ji, u, ku, su, ha, ba, ri, ba, o, te, mo, wa, to, ga, da/. A digital camera was set to record the speaker's face and shoulder. All recordings were made in a single walled sound treated room.

Original digitized videotaped stimuli were edited with the specially developed editing software. Auditory signals were digitized at a sampling rate of 24000 Hz and were equalized in level. The synchronization of audio-visual stimuli was measured. It was within 60 ms. For the auditory only (AO) condition, the visual image of the speaker was hidden by visual masking. For the visual only (VO) condition, the audio signal was turned off. The congruent auditory-visual (AV-C) stimuli consisted of digital audio-video files of the speaker saying and articulating the same speech stimuli. For incongruent auditory-visual (AV-I) condition, stimuli were created by combining audio files with non-corresponding video files and matching the onset times. More specifically, in the AV-I condition for syllable presentation, an auditory syllable was paired with a visual syllable whose vowel was the same as the original auditory syllable. In the AV-I condition for word presentation, an auditory word was paired with a visual word whose first vowel was the same as the original auditory word. In the AV-I condition for sentence presentation, the entire length of an auditory sentence was paired with a visual sentence whose entire length was the same as the original auditory sentence as much as possible.

Procedure

For all nine students stimuli were presented in four conditions: AO, VO, AV-C, and AV-I. All stimuli were presented randomly to the participants. Before testing, practice sessions were used to familiarize the subjects with the procedure. Participants were instructed to listen and/or watch each stimulus and repeat what they judged to have been said. The auditory stimuli were presented at 65 dB SPL.

RESULTS AND DISCUSSION

Figures 1 to 3 show the individual percentage correct scores of the 9 participants for CV syllables, words, and sentences, respectively. Except for one participant (S8), all of the other participants showed higher scores in the AO condition than in the VO condition. These subjects showed even higher scores in the AV-C condition than in the AV-I condition. From these data the participants seemed to use auditory-visual information effectively in integrating auditory and visual stimuli in speech perception but the degree of auditory-visual integration was different between subjects.

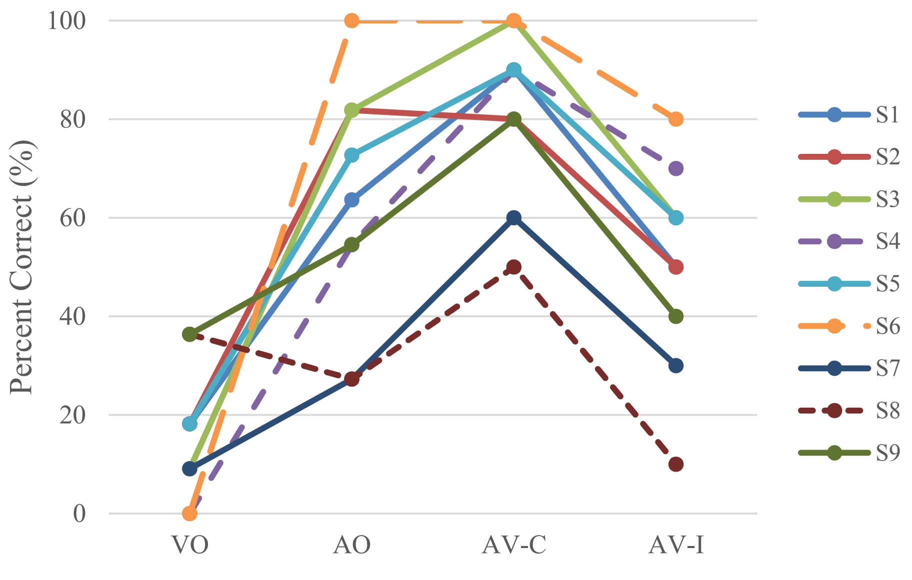


Fig. 1: Syllable identification score (%) for each condition. VO: visual only, AO: auditory only, AV-C: congruent auditory-visual, AV-I: incongruent auditory-visual conditions.

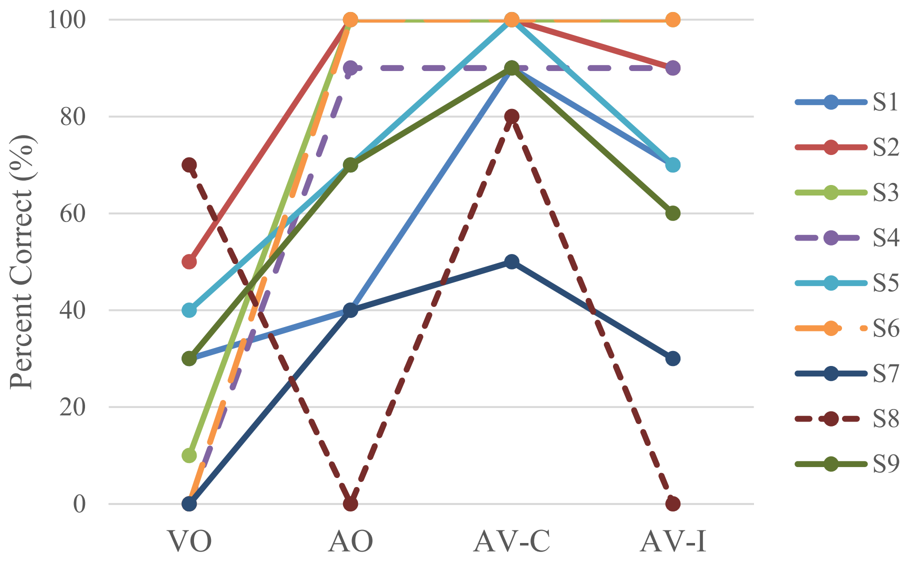


Fig. 2: Word identification score (%) for each condition. VO: visual only, AO: auditory only, AV-C: congruent auditory-visual, AV-I: incongruent auditory-visual conditions.

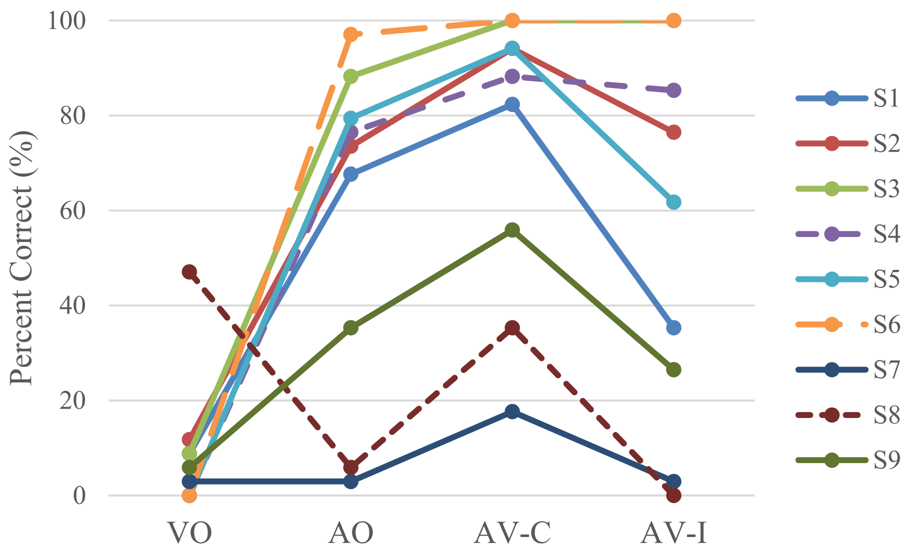


Fig. 3: Sentence identification score (%) for each condition. VO: visual only, AO: auditory only, AV-C congruent auditory-visual, AV-I: incongruent auditory-visual conditions.

Except for 2 participants (S4 and S6), all of the remaining participants showed increased percent correct scores for all speech stimuli in the AV-C condition than in the AV-I condition. For the two exceptional participants, percent correct scores for words and sentences were almost 100% and they appeared to have reached a ceiling effect. The context effects seem to be involved in the process of identification of the words and sentences.

On the other hand unlike other 8 participants, S8 got higher scores in the VO condition than in the AO condition and also got almost the same scores in the VO and AV-I conditions. When the scores between AV-C and AV-I conditions were compared, the former scores were higher than the later ones in syllables, words, and sentences. From these results S8 seems to be more dependent on vision in speech perception.

The relative benefit score (Grant and Seitz, 1998), defined as (AV-C−AO)/(1−AO) with AO and AV-C score expressed as percent correct, was calculated for each subject. Figures 4 to 6 show the individual relative benefit score of 9 participants for CV syllables, words, and sentences, respectively. When plotting relative benefit scores vs. AO performance, the data seemed to be distributed steadily in all three kinds of speech stimuli. But in Figs. 4 and 5 celling effects were seen for the highest performances in the AO condition. Some subjects got the same highest scores in the AO and AV-C conditions, such that their relative benefit scores were zero. But in Fig. 6 when key words in sentences were calculated, the data were scattered evenly. It thus seems that sentences were the most suitable stimuli for clarifying the AV integration.

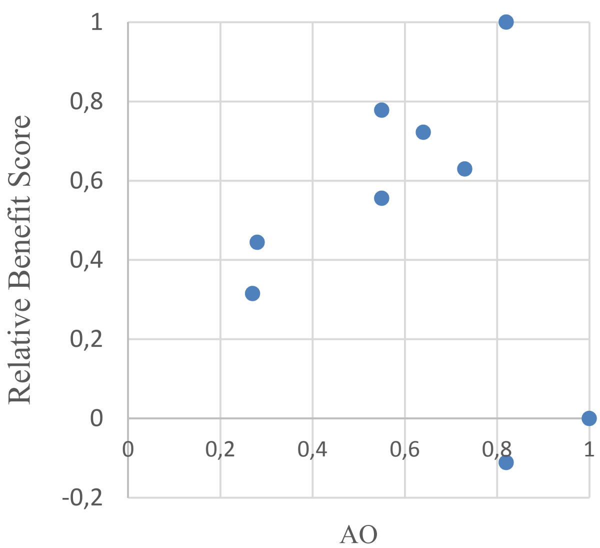


Fig. 4: Relative benefit score for syllables. The relative benefit score, defined as (AV-C−AO)/(1−AO) with AO and AV-C score expressed as percent correct was calculated for each subject. AO: auditory only, AV-C: congruent auditory-visual conditions.

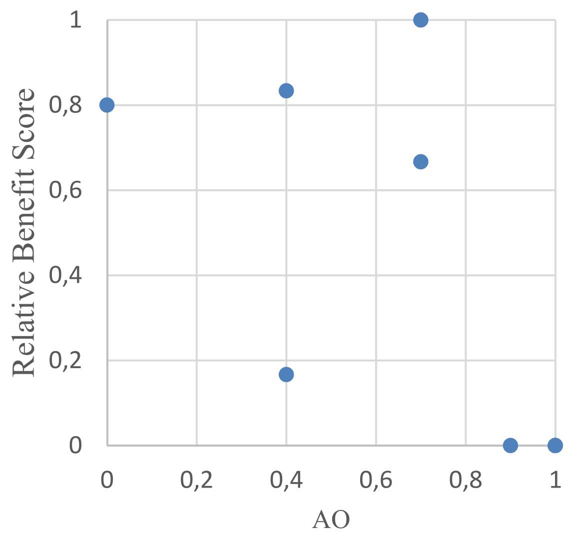


Fig. 5: Relative benefit score for words. The relative benefit score, defined as (AV-C－AO)/(1－AO) with AO and AV-C score expressed as percent correct was calculated for each subject. AO: auditory only, AV-C: congruent auditory-visual conditions.

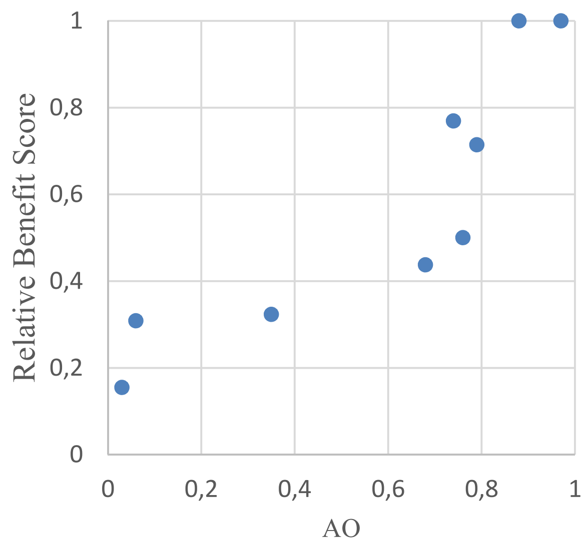


Fig. 6: Relative benefit score for sentences. The relative benefit score, defined as (AV-C－AO)/(1－AO) with AO and AV-C score expressed as percent correct was calculated for each subject. AO: auditory only, AV-C: congruent auditory-visual conditions.

ACKNOWLEDGEMENT

I wish to thank Harry Levitt and Sheila Moodie for helping me to prepare the manuscript.

REFERENCES

Desai, S., Stickney, G., and Zeng, F.G. (**2008**). “Auditory-visual speech perception in normal-hearing and cochlear-implant listeners,” J. Acoust Soc Am, **123**, 428-440.

Grant, K. and Seitz, P. (**1998**). “Measures of auditory-visual integration in nonsense syllables and sentences,” J. Acoust. Soc. Am., **104**, 2438-2450.

Huyse, A., Berthommier, F., and Leybaert, J. (**2012**). “Degradation of labial information modifies audiovisual speech perception in cochlea-implanted children,” Ear Hearing, **34**, 110-121.

Rouger, J., Fraysse, B., Deguine, O., and Barone, P. (**2008**). “McGurk effects in cochlear-implanted deaf subjects,” Brain Res., **1188**, 87-99.

Schwartz, J. (**2010**). “A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent,” J. Acoust. Soc. Am., **127**, 1584-1594.

Towards a diagnostic test for hidden hearing loss

Christopher J. Plack*, Garreth Prendergast, Karolina Kluk, Agnès Léger, Hannah Guest, and Kevin J. Munro

The University of Manchester, Manchester Academic Health Science Centre, Manchester, England

Cochlear synaptopathy (or "hidden hearing loss"), due to noise exposure or ageing, has been demonstrated in animal models using histological techniques. However, diagnosis of the condition in individual humans is problematic because of: (i) test reliability, and (ii) lack of a gold standard validation measure. Wave I of the transient-evoked auditory brainstem response (ABR) is a non-invasive electrophysiological measure of auditory nerve function, and has been validated in the animal models. However, in humans Wave I amplitude shows high variability both between and within individuals. The frequency-following response (FFR), a sustained evoked potential reflecting synchronous neural activity in the rostral brainstem, is potentially more robust than ABR wave I. However, the FFR is a measure of central activity, and may be dependent on individual differences in central processing. Psychophysical measures are also affected by inter-subject variability in central processing. Differential measures, in which the measure is compared, *within an individual*, between conditions that are affected differently by cochlear synaptopathy, may help to reduce inter-subject variability due to unrelated factors. There is also the issue of how the metric will be validated. Comparisons with animal models, computational modeling, auditory nerve imaging, and human temporal bone histology are all potential options for validation, but there are technical and practical hurdles, and difficulties in interpretation. Despite the obstacles, a diagnostic test for hidden hearing loss is a worthwhile goal, with important implications for clinical practice and health surveillance.

INTRODUCTION

Hearing ability is usually assessed using pure tone audiometry (Johnson, 1970), which measures the smallest detectable level of pure tones at a range of frequencies. The resulting audiogram is sensitive to dysfunction of the outer hair cells and, to a lesser extent, inner hair cells (IHCs) in the cochlea. However, it is becoming increasingly clear that the audiogram is not sensitive to some types of peripheral auditory dysfunction. In particular, results from rodent models suggest that noise exposure and/or aging, can cause permanent loss of synapses between the IHCs and auditory nerve fibers, *without permanently affecting sensitivity to quiet sounds* (Kujawa and Liberman, 2009; Sergeyenko *et al.*, 2013). The disconnected nerve

*Corresponding author: chris.plack@manchester.ac.uk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

fibers subsequently degenerate. This disorder has been variously termed “cochlear neuropathy”, “cochlear synaptopathy”, and popularly “hidden hearing loss” (Schaette and McAlpine, 2011), because the loss is not thought to be detectable using pure-tone audiometry. The loss seems to affect selectively the low spontaneous rate (SR) fibers that have high thresholds and are thought to be responsible for coding sound intensity at moderate-to-high levels (Furman *et al.*, 2013). This may explain why the loss does not affect sensitivity to quiet sounds.

Several research groups are currently trying to determine the extent to which hidden hearing loss is a contributor to hearing difficulties experienced by humans. There is evidence that listeners with a history of noise exposure but with normal audiograms have deficits in speech perception and temporal processing (Alvord, 1983; Kumar *et al.*, 2012). Similarly, the aging process may affect speech perception in noise even when there are no significant increases in audiometric threshold (Dubno *et al.*, 1984; Rajan and Cainer, 2008). An open question concerns the extent to which these deficits are a consequence of cochlear synaptopathy, or other types of dysfunction, for example, IHC dysfunction, or central neural dysfunction.

A major obstacle to the academic investigation of hidden hearing loss, and to the eventual incorporation of the research findings into clinical practice, is the absence of a reliable and validated diagnostic test for the disorder. In the animal models, selective immunostaining and confocal microscopy can be used to determine directly the loss of synapses. However, such invasive procedures are not possible in humans, at least pre-mortem. In this article we will consider non-invasive measures of hidden hearing loss, their potential as a diagnostic test, and the challenges faced in developing them to this stage. Table 1 provides a summary of the techniques that will be discussed.

Diagnostic technique	**Hypothesized effect of synaptopathy**	**Pros**	**Cons**
ABR	Reduction in wave I amplitude at high levels	Relatively direct measure of auditory nerve function	Highly variable in humans
FFR	Reduction in synchrony to amplitude modulation	Robust response; objective	Affected by variability in central processes
Behavioral	Increase in discrimination thresholds at high levels	Easy to measure	Affected by central processes; hypothesized effects are small

Table 1. A summary of potential diagnostic techniques for hidden hearing loss.

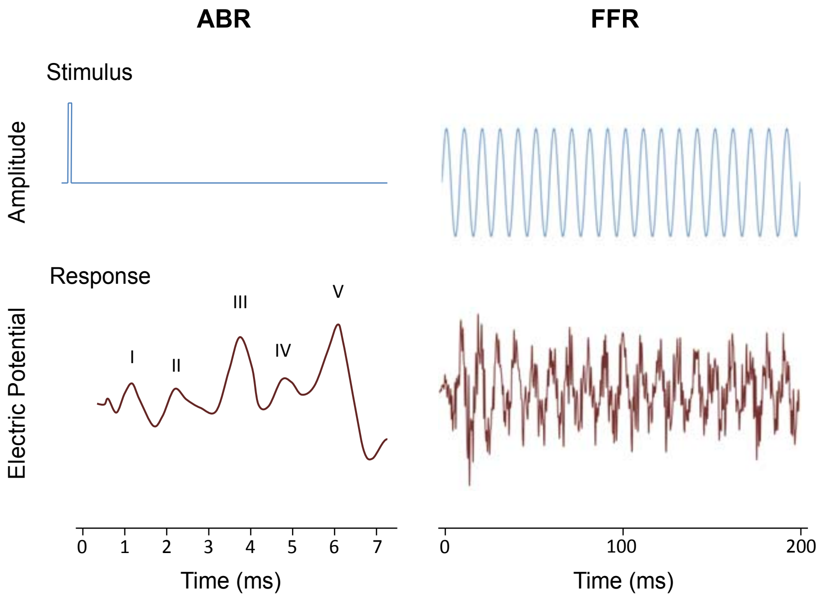


Fig. 1. An illustration of typical stimuli and recorded waveforms for two electrophysiological measures of auditory neural coding; the auditory brainstem response (ABR) and the frequency following response (FFR).

MEASURES OF HIDDEN HEARING LOSS IN HUMANS

The auditory brainstem response

The click-evoked electrophysiological auditory brainstem response (ABR, see Fig. 1) is a prime candidate for a measure of hidden hearing loss in humans. The ABR can be recorded in humans using electrodes placed on the scalp; typically an electrode is attached to a mastoid and to another location such as the contralateral mastoid, forehead, or vertex. The differential response to the two electrodes determines the recorded ABR. Wave I of the ABR reflects auditory nerve function, and in the rodent models has been shown to be sensitive to the effects of noise exposure (Kujawa and Liberman, 2009) and aging (Sergeyenko *et al.*, 2013). In these models, the amplitude of Wave I is reduced at moderate-to-high levels but not at low levels, consistent with a selective loss of low-SR fibers. Furthermore, Wave I amplitude correlates strongly with the proportion of intact synapses (Kujawa and Liberman, 2009; Sergeyenko *et al.*, 2013), which provides validation for the measure in rodents.

In humans the evidence is less compelling, but both aging (Konrad-Martin *et al.*, 2012) and, recently, noise exposure (Stamper and Johnson, 2015) have been shown to be associated with a reduction in ABR Wave I amplitude for high-level clicks, in the absence of, or controlling for, an increase in audiometric threshold. In addition, Wave I amplitude for high-level clicks is reduced in listeners with tinnitus even when the audiogram is normal (Schaette and McAlpine, 2011). It is suggested by

Schaette and McAlpine that loss of auditory nerve fibers may induce tinnitus due to a compensatory increase in central neural gain. However, there are some problems associated with the use of ABR Wave I as a diagnostic test of hidden hearing loss. First, unlike the rodent models in which the ABR can be measured accurately using sub-cutaneous electrodes, in humans ABR Wave I has a relatively low amplitude and shows high variability both between individuals and within individuals on repeated tests (Beattie, 1988; Lauter and Loomis, 1988). This variability may be the result of a number of factors unrelated to cochlear synaptopathy, including sex, head size, variations in tissue resistance, and variations in electrode placement (Schwartz and Berry, 1985). The use of intra-canal electrodes, including tympanic membrane electrodes, can increase the amplitude of Wave I, but may increase the variability (Stamper and Johnson, 2015). Hence, at present, while Wave I may be useful for demonstrating *group* differences in synaptopathy, between those noise exposed and those not for example, it is probably not useful for determining if an *individual* has hidden hearing loss.

Another issue is that the amplitude of Wave I in response to a broadband click is strongly influenced by activity in basal regions of the cochlea (Don and Eggermont, 1978). Even if the audiogram is normal over the clinical range, up to 4 kHz or 8 kHz say, hair cell loss in higher characteristic frequency regions may affect the amplitude of the response. Hence to identify synaptopathy, the results may have to be controlled for high-frequency audiometric thresholds, or, alternatively, the high-frequency region may be masked using high-pass noise during recording of the ABR to prevent the basal region contributing to the response (Don and Eggermont, 1978).

The frequency-following response

The frequency-following response (FFR) is a sustained auditory evoked potential, thought to reflect neural activity in the brainstem synchronized (phase locked) to the waveform of the stimulus (Krishnan, 2006; see Fig. 1). The FFR is particularly sensitive to amplitude modulation at modulation rates of a few hundred hertz, although it also reflects phase locking to temporal fine structure for frequencies up to about 1 kHz. Over recent years the FFR has become popular as a measure of auditory temporal coding. The FFR can be recorded using similar electrode montages to the ABR, and for lower frequencies at least, is a more robust measure than ABR Wave I, with most participants showing a clear response above the noise floor. Importantly, FFR amplitude can be measured objectively using a discrete Fourier transform of the response at the component frequency, whereas ABR Wave I measurement sometimes requires a subjective intervention to analyze the waveform and determine the peak location.

There is evidence that the amplitude of the FFR to both stimulus envelope and temporal fine structure decreases with increasing age even when controlling for absolute threshold (Clinard and Tremblay, 2013; Marmel *et al.*, 2013; Bones and Plack, 2015). The FFR is also predictive of behavioral performance on tasks such as frequency discrimination (Marmel *et al.*, 2013) and modulation discrimination (Bharadwaj *et al.*, 2015) for listeners with normal audiometric thresholds. There is also preliminary evidence that the FFR is reduced in noise-exposed ears for listeners with normal absolute thresholds (Plack *et al.*, 2014, see Fig. 2). These results suggest that the FFR may be sensitive to synaptopathy.

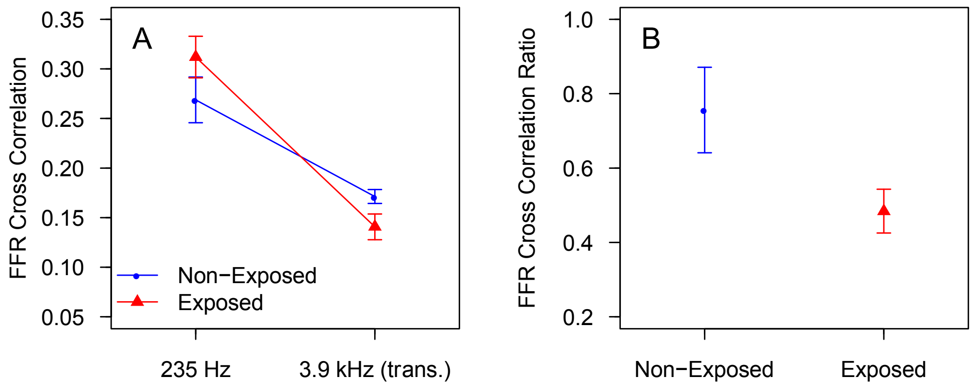


Fig. 2. Results from the conference presentation of Barker *et al.* (2014) reported by Plack *et al.* (2014). A: FFR synchrony to a 235-Hz pure tone and to a 235-Hz tone transposed to 3.9 kHz (i.e., a 3.9 kHz pure-tone carrier amplitude modulated at 235 Hz), for groups of listeners with (red triangles) and without (blue circles) a history of recreational noise exposure. For each stimulus, the dependent variable was the coefficient of correlation between the FFR and a 235-Hz pure tone. B: The ratios of the coefficients between the two frequencies (3.9 kHz : 235 Hz). Error bars are standard errors.

However, unlike ABR Wave I, the FFR is produced largely by generators in the brainstem, the largest component from the region of the inferior colliculus (Krishnan, 2006). Hence differences in central auditory processing may well contribute to individual differences in FFR amplitude. For example, it is known that musicians and tone language speakers have stronger FFRs for certain types of stimuli (Krishnan *et al.*, 2005; Wong *et al.*, 2007), likely due to experience-related plasticity. Aging affects central neural function (Konrad-Martin *et al.*, 2012), so an FFR deficit due to age could be a consequence of a combination of peripheral and central factors. Like the ABR, the FFR is also limited by between- and within-subject variability due to factors such as tissue resistance and electrode placement.

Behavioral measures

Behavioral measures, such as psychophysical thresholds, require a subjective response from the listener. Hence, they don't have the "objectivity" of electrophysiological measures, and may potentially depend on processing at all stages from the auditory periphery to the motor commands sent to the finger that presses the response key. As is the case for the FFR technique, there is the concern that performance may be influenced by central factors unrelated to synaptopathy. As well as purely auditory factors, these may include higher-level functions such as memory and attention. However, behavioral techniques have been shown to provide reliable measures of some aspects of peripheral function, in particular frequency selectivity and cochlear compression (Oxenham and Plack, 1997).

Reduction in the numbers of low-SR fibers might be expected to affect discrimination tasks at high sound levels. However, as pointed out by Oxenham and Heinz (personal communications) if considered in terms of signal detection theory, a 50% fiber loss (similar to that in the animal studies) would reduce the discrimination index, d-prime, by a factor of √2 only. This would result in a barely measurable increase in threshold, about 1 dB in the case of the intensity difference limen (Buus and Florentine, 1991), for example. Considering the between-subject variability in performance expected due to central factors, it is not clear that psychophysical measures have the necessary sensitivity to diagnose synaptopathy, *unless* almost all the synapses with low-SR fibers are lost in a given region of the cochlea.

There are little available data directly relating synaptopathy to behavioral performance. Tinnitus patients with normal hearing, who exhibit a reduction in ABR Wave I amplitude consistent with synaptopathy (Schaette and McAlpine, 2011), have elevated intensity discrimination thresholds (Epp *et al.*, 2012). Noise exposure and aging have been related to deficits in temporal processing tasks and speech discrimination in noise (Alvord, 1983; Dubno *et al.*, 1984; Kumar *et al.*, 2012).

MANAGING VARIABILITY

A common problem for measures of hidden hearing loss in humans is that of variability. *Within*-subject variability may be minimized for the electrophysiological techniques by using careful procedures, and ensuring electrode placements and impedances are tightly controlled. For psychophysical tests, practice and the use of a procedure that is easy to learn can ensure that performance is at asymptote (King *et al.*, 2013). An approach for minimizing both within- and between-subject variability is to use a differential measure, in which two measures are compared for each individual: one measure that is assumed to be affected by synaptopathy and one that isn't. Ideally, both measures should be affected equally by other sources of variability so that effectively this variability can be cancelled out or at least minimized. Such an approach may be effective for both electrophysiological and behavioral measures, and help to reduce or eliminate confounds due to central factors for the FFR and for the behavioral measures. There are two clear options for differential measures of synaptopathy; comparisons across frequency and comparisons across level.

Comparisons across frequency

One differential approach is to compare measures between a low-frequency region and a high-frequency region. It is generally reported that noise exposure causes most damage in higher frequency regions (around 4 kHz), hence the low-frequency measure can be used as a within-subject comparison. A preliminary study used this technique by comparing the FFR to a 235-Hz pure tone with that to a 235-Hz modulator imposed on a 4-kHz carrier (Plack *et al.*, 2014, see Fig. 2). The participants were audiogram matched. The noise-exposed group had no reduction in FFR amplitude to the low-frequency tone, but showed a reduction in the amplitude of the FFR to the envelope of the high-frequency stimulus. Furthermore, the difference between the groups was greater when the ratio of high-frequency to low-frequency responses was used as the measure.

For the ABR, filtered or masked clicks can be used to probe different frequency regions, and hence allow a cross-frequency comparison. For behavioral measures it is relatively simple for narrowband stimuli to compare performance in different frequency regions. Whenever narrowband stimuli are used, it may be advisable to include a broadband masking noise to ensure that high-SR fibers do not contribute to the response due to spread of excitation. However, a problem with using across-frequency comparisons is that it is not yet clear that synaptopathy only affects high-frequency regions.

Comparisons across level

An alternative is to rely on the finding that synaptopathy is selective for low-SR fibers, which have high thresholds and code intensity information at high levels, above the saturation level of the high-SR fibers. Hence evoked-response amplitude, and behavioral performance, should be selectivity impaired at high levels. By comparing the measure across different levels, it may be possible to isolate the effects of synaptopathy from other sources of variability. In the study of Schaette and McAlpine (2011) it was observed that the reduction in ABR Wave I was greater for the 100 dB pe SPL click than for the 90 dB pe SPL click. Bharadwaj *et al.* (2015) have taken a similar approach for their FFR measures, by measuring the FFR to a modulator imposed on a high-level carrier. They reasoned that the FFR for a low modulation depth would be determined primarily by the response of low-SR fibers, whereas the FFR for a high modulation depth would depend in part on the response of high-SR fibers, since the dips in the modulation would fall within their level range. Bharadwaj *et al.* (2015) showed that the slope of the function relating FFR strength to modulation depth correlated more strongly with behavioral modulation detection performance than did FFR strength in isolation.

THE PROBLEM OF VALIDATION

In the rodent models, validation of electrophysiological or psychophysical measures is possible because researchers can count synapses and nerve fibers post-mortem using histological techniques. While human temporal bones are available to researchers, and have been used to provide estimates of auditory nerve fiber loss due to aging (Makary *et al.*, 2011), it is not trivial to validate a test performed on a living human using a post-mortem measure! The problem essentially is that we currently lack a "gold-standard" measure of synaptopathy that can be used with a living human to validate the diagnostic test. We are hence confronted by the serious problem of being unable to confirm that our diagnostic test is measuring what we want it to. There are, however, a number of potential approaches to validation that may be productive.

Validation with animal models

One approach to validation is to assume that between-species differences are insignificant with regard to the diagnosis of synaptopathy, and validate the measure using animal models. For the ABR, for example, there is good evidence from comparisons with synapse counts that Wave I is a reliable measure of synaptopathy in animals with normal sensitivity to quiet sounds (Kujawa and Liberman, 2009; Sergeyenko *et al.*, 2013). The FFR could be validated in a similar way, and it should be possible to validate simple behavioral measures, such as psychophysical discrimination thresholds, in animals suited to behavioral tasks such as the chinchilla. These measures can then be compared with post-mortem synapse counts taken shortly after threshold measurement.

Computational modeling

There are now a number of computational models of the peripheral auditory system (e.g. Zilany *et al.*, 2009), based on animal and human data, that could be adapted to make predictions of the expected effects of synaptopathy on evoked potentials and behavioral performance. These results could help validate diagnostic tests based on these measures, to determine whether the pattern of results is consistent with the expected effects of synaptopathy. However, there are still too many uncertainties in these models to rely on them entirely, and these models of course cannot determine the actual synaptic loss for an individual. The utility of these models may lie in their use in conjunction with the animal data.

Auditory nerve imaging

Imaging techniques, in particular magnetic resonance imaging (MRI), have the potential to provide a direct measure of nerve fiber loss. At present it is not possible to image the auditory nerve non-invasively in humans with the resolution required to detect a proportional reduction in nerve fibers. However, it is conceivable that techniques such as diffusion tensor MRI may be refined to the point at which we can provide a direct estimate of the loss of fibers due to synaptopathy. Although such a measure may not itself be cost-effective or practical for routine use in the clinic as a diagnostic test, it could be used to validate a simpler clinical test, for example, by imaging a relatively small number of individuals with normal audiograms, with and without suspected hidden hearing loss.

Human temporal bone histology

Direct nerve fiber and synapse counts are certainly possible in humans post-mortem using donated temporal bones. The problem then is how to use this information to validate a test, without having to repeatedly perform that test on the individual until they die to account for changes in performance over time. Terminally ill patients may be one option if consent can be obtained, although these individuals are predominantly elderly and may have a number of hearing-related complications, including hair cell loss. Another option is to test young participants in the military, or other occupations with higher than average mortality, who have agreed to donate their temporal bones.

CONCLUSIONS

The discovery of cochlear synaptopathy, or hidden hearing loss, has potentially major implications for audiological practice, health surveillance, and noise exposure regulations. Investigations of the disorder in humans are hampered by the lack of a reliable diagnostic test. The amplitude of Wave I of the ABR is the most direct non-invasive measure of auditory nerve function in humans, but is limited by variability. The FFR and behavioral measures are less direct, and influenced by central factors, but may prove more reliable. Variability may be reduced by the use of differential measures, that compare performance across frequency or level for example, to isolate the effects of synaptopathy form other sources of variance. There is also the problem of test validation. It may be necessary to rely on animal data relating comparable electrophysiological and behavioral measures with direct histological measures, although it is conceivable that technological innovations in neuroimaging may allow a direct estimate of auditory nerve fiber loss in humans, permitting validation of a more clinically useable test.

ACKNOWLEDGMENTS

This content of this article was strongly influenced by discussions at the recent NIH/NIDCD Workshop on Noise-Induced Hearing Loss in Bethesda, MD. The authors are grateful to the workshop attendees, and in particular to the organizers: Janet Cyr, Amy Donohue, Sharon Kujawa, Charlie Liberman, and Rick Davis. Our work on hidden hearing loss is supported by grants from the MRC (MR/L003589/1) and Action on Hearing Loss.

REFERENCES

Alvord, L.S. (**1983**). “Cochlear dysfunction in "normal-hearing" patients with history of noise exposure,” Ear. Hearing, **4**, 247-250.

Barker, D., Hopkins, K., Baker, R., and Plack, C.J. (**2014**). “Detecting the early effects of noise exposure,” in *Midwinter Meeting of Association for Research in Otolaryngology* (San Diego).

Beattie, R.C. (**1988**). “Interaction of click polarity, stimulus level, and repetition rate on the auditory brainstem response,” Scand. Audiol., **17**, 99-109.

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.

Bones, O. and Plack, C. J. (**2015**). “Losing the music: aging affects the perception and subcortical neural representation of musical harmony,” J. Neurosci., **35**, 4071-4080.

Buus, S. and Florentine, M. (**1991**). "Psychometric functions for level discrimination," J. Acoust. Soc. Am., **90**, 1371-1380.
Clinard, C.G. and Tremblay, K.L. (**2013**). "Aging degrades the neural encoding of simple and complex sounds in the human brainstem," J. Am. Acad. Audiol., **24**, 590-599.
Don, M. and Eggermont, J.J. (**1978**). "Analysis of the click-evoked brainstem potentials in man unsing high-pass noise masking," J. Acoust. Soc. Am., **63**, 1084-1092.
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.
Epp, B., Hots, J., Verhey, J.L., and Schaette, R. (**2012**). "Increased intensity discrimination thresholds in tinnitus subjects with a normal audiogram," J. Acoust. Soc. Am., **132**, EL196-201.
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.
Johnson, E.W. (**1970**). "Tuning forks to audiometers and back again," Laryngoscope, **80**, 49-68.
King, A., Hopkins, K., and Plack, C.J. (**2013**). "Differences in short-term training for interaural phase difference discrimination between two different forced-choice paradigms," J. Acoust. Soc. Am., **134**, 2635-2638.
Konrad-Martin, D., Dille, M.F., McMillan, G., Griest, S., McDermott, D., Fausti, S. A., and Austin, D.F. (**2012**). "Age-related changes in the auditory brainstem response," J. Am. Acad. Audiol., **23**, 18-35.
Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (**2005**). "Encoding of pitch in the human brainstem is sensitive to language experience," Cogn. Brain Res., **25**, 161-168.
Krishnan, A. (**2006**). "Frequency-following response," in *Auditory evoked potentials: basic principles and clinical application*. Eds. R.F. Burkhard, M. Don, and J. Eggermont (Lipincott Williams and Wilkins, Philadelphia), pp. 313-333.
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.
Kumar, U.A., Ameenudin, S., and Sangamanatha, A.V. (**2012**). "Temporal and speech processing skills in normal hearing individuals exposed to occupational noise," Noise Health, **14**, 100-105.
Lauter, J.L. and Loomis, R.L. (**1988**). "Individual differences in auditory electric responses: comparisons of between-subject and within-subject variability. II. Amplitude of brainstem Vertex-positive peaks," Scand. Audiol., **17**, 87-92.
Makary, C.A., Shin, J., Kujawa, S.G., Liberman, M.C., and Merchant, S.N. (**2011**). "Age-related primary cochlear neuronal degeneration in human temporal bones," J. Assoc. Res. Otolaryngol., **12**, 711-717.
Marmel, F., Linley, D., Carlyon, R.P., Gockel, H.E., Hopkins, K., and Plack, C.J. (**2013**). "Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently," J. Assoc. Res. Otolaryngol., **14**, 757-766.

Oxenham, A.J. and Plack, C.J. (**1997**). “A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,” J. Acoust. Soc. Am., **101**, 3666-3675.

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

Rajan, R. and Cainer, K.E. (**2008**). “Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational maskers,” Neuroscience, **154**, 784-795.

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

Schwartz, D.M. and Berry, G.A. (**1985**). “Normative aspects of the ABR,” in *The auditory brainstem response*. Ed. J.T. Jacobson (Taylor and Francis, London), pp. 65-97.

Sergeyenko, Y., Lall, K., Liberman, M.C., and Kujawa, S.G. (**2013**). “Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline,” J. Neurosci., **33**, 13686-13694.

Stamper, G.C. and Johnson, T.A. (**2015**). “Auditory function in normal-hearing, noise-exposed human ears,” Ear. Hearing, **36**, 172-184.

Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., and Kraus, N. (**2007**). “Musical experience shapes human brainstem encoding of linguistic pitch patterns,” Nat. Neurosci., **10**, 420-422.

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

Evaluation of peripheral compression and auditory nerve fiber intensity coding using auditory steady-state responses

Gerard Encina-Llamas[1,*], James M. Harte[2], Torsten Dau[1], and Bastian Epp[1]

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

[2] *Interacoustics Research Unit, Kgs. Lyngby, Denmark*

The compressive nonlinearity of the auditory system is assumed to be an epiphenomenon of a healthy cochlea and, particularly, of outer-hair cell function. Another ability of the healthy auditory system is to enable communication in acoustical environments with high-level background noises. Evaluation of these properties provides information about the health state of the system. It has been shown that a loss of outer hair cells leads to a reduction in peripheral compression. It has also recently been shown in animal studies that noise over-exposure, producing temporary threshold shifts, can cause auditory nerve fiber (ANF) deafferentation in predominantly low-spontaneous rate (SR) fibers. In the present study, auditory steady-state response (ASSR) level growth functions were measured to evaluate the applicability of ASSR to assess compression and the ability to code intensity fluctuations at high stimulus levels. Level growth functions were measured in normal-hearing adults at stimulus levels ranging from 20 to 90 dB SPL. To evaluate compression, ASSR were measured for multiple carrier frequencies simultaneously. To evaluate intensity coding at high intensities, ASSR were measured using a single carrier frequency at four modulation depths between 25 and 100%. The data showed that ASSR level growth functions exhibited compression of about 0.25 dB/dB. For levels above 60 dB SPL, the slope showed higher variability for the different modulation depths across subjects than for lower levels. The results indicate that the slope of the ASSR level growth function can be used to estimate peripheral compression simultaneously at four frequencies below 60 dB SPL, while the slope above 60 dB SPL may provide information about the integrity of intensity coding of low-SR fibers.

INTRODUCTION

The integrity of the hearing system has traditionally been assessed through audiometry, where the minimum sound level (hearing threshold) of pure tones presented at different frequencies is measured. Patients showing hearing thresholds comparable to standardized normal hearing thresholds are categorized as being normal-hearing (NH)

*Corresponding author: encina@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

listeners, whereas those patients with elevated hearing thresholds are categorized as being hearing-impaired (HI) listeners. Nevertheless, there is emerging evidence that the complexity of the hearing system cannot be fully characterized by just evaluating sensitivity by measuring thresholds. Several clinical studies showed that about 5 to 10% of patients self-reporting hearing difficulties, particularly in noisy background situations, show pure-tone audiograms better than 20 dB HL (Saunders and Haggard, 1989; Kumar *et al.*, 2007; Hind *et al.*, 2011). Furthermore, there is strong evidence from physiological studies in animal models that profoundly damaged hearing systems do not produce permanent threshold shifts. In Lobarinas *et al.* (2013) an anti-cancer drug was injected in chinchillas to produce a selective loss of inner hair cells (IHC) while keeping the total integrity of outer hair cells (OHC). The data show only minor effects on behaviorally measured audiometric thresholds, even with a loss of IHC that exceeds 80%. There exist also other examples of damaged systems that are not detectable by an audiogram. It was reported that noise over-exposure causing a temporal threshold shift with threshold recovery within two weeks after the exposure produced a rapid and permanent loss of about 40-50% of auditory nerve fiber (ANF) synapses in mice (Kujawa and Liberman, 2009) and guinea pigs (Lin *et al.*, 2011), leading to a slow loss of ANF spiral ganglion cells (ANF cell bodies). Furman *et al.* (2013) showed that the loss of ANF synapses (deafferentation) after noise over-exposure is more selective to low- spontaneous rate (SR) ANF. Since this damage does not hamper sensitivity, but rather supra-threshold coding, this new form of hearing loss is known as *hidden hearing loss* because it cannot be detected by the currently available diagnostic metrics. Therefore, the development of novel methods able to evaluate the integrity and functionality of the human hearing system assessing supra-threshold processing is required.

The compressive nonlinearity of the peripheral auditory system is commonly assumed to be a result of healthy OHC function, and to be a good indicator of the system's integrity. Ruggero *et al.* (1997) showed that basilar membrane (BM) velocity grows linearly (slope of 1 dB/dB) when recorded as a function of sound level (dB SPL) in a dead cochlea. The BM input/output function grows compressively in an alive and healthy cochlea, where the healthy function of OHC generates a gain mechanism. In impaired systems, like listeners with a sensorineural hearing loss, a reduction of OHC leads to a reduction in compression and a loss of sensitivity. Since a reduction of sensitivity could also be caused by other mechanisms, like severe loss of IHC, the reduction in sensitivity does not necessarily imply a loss of OHC. Therefore, a method able to provide an estimate of peripheral compression in humans would be an excellent complement to the audiogram to characterize better the hearing system function at supra-threshold levels.

Intensity coding in the auditory nerve is done by different types of auditory nerve fibers (ANF). Typically, the types of afferent fibers that innervate an IHC have been divided according to their firing rate (number of spikes per second) in quiet (without sound stimulation). The fibers that produce more than 18 spikes/second are referred to as high-SR fibers, while the fibers producing less than 18 spikes per second are named low-SR fibers (Liberman, 1978). The high- and low-SR fibers show different rate-intensity functions (spike rate as a function of sound level). High-SR fibers have lower thresholds, their discharge rate increases with level at the lower stimulation levels range and it saturates at medium and high sound levels. On the contrary, low-SR fibers show higher thresholds and their discharge rate function grows with sound intensity (Winter *et al.*, 1990). Each of these types of fibers has a limited dynamic range of around 30 dB, such that high-SR fibers are already completely saturated at levels of 60 dB SPL. Since ANF are connected to a narrow region in the cochlea, noninvasive assessment of intensity coding would provide a good method to assess intensity coding in narrow frequency regions.

Auditory steady-state responses (ASSR) represent a well-studied objective measure of auditory function (see Picton *et al.*, 2003, for a review). ASSR are gross electroencephalography (EEG) potentials that follow the envelope of periodic acoustic stimuli. The most common acoustic stimulus used to record ASSR are sinusoidally amplitude modulated (SAM) tones. The use of SAM tones is very convenient because its envelope is a sinusoid defined by the modulation frequency (f_m). When the recorded ASSR is analyzed in the frequency domain, the energy at frequency f_m is a measure of the encoded stimulus envelope. Thus, increasing the amplitude of the SAM stimulus leads to a larger ASSR magnitude, and a reduction in the modulation depth (m) of the stimulus results in a smaller ASSR magnitude. One common clinical application of ASSR is the estimate of thresholds in non-responding listeners. It has been shown that ASSRs can be measured for multiple SAM tones simultaneously over a broad frequency range (Picton *et al.*, 2003). Since ASSRs encode the stimulus envelope at various intensities after peripheral processing, the measurement of ASSR at supra-threshold levels is a promising method to assess peripheral processing, including compression and high-intensity coding.

In this study, the applicability of ASSR to assess peripheral compression and to evaluate the integrity of ANF is investigated. Processing an SAM tone by a compressive system, such as the cochlea, will reduce the modulation depth of the processed signal. Since ASSR codes the envelope of the stimulus after cochlea processing, the ASSR must be also affected by the cochlear compressive nonlinearity. We hypothesize that ASSR recorded as a function of stimulation level reflects peripheral compression in NH listeners and a loss of compression at the impaired frequencies in listeners with a mild HI. For SAM tones at high intensities, different groups of ANF are required to encode the intensity fluctuations of the envelope. For shallow modulation depths and high carrier levels, especially low-SR fibers are required to encode the temporal fluctuations. Hence, ASSR recorded at higher stimulus intensities using shallow modulated SAM tones must rely mostly on the accurate temporal coding of low-SR fibers. Assuming that deafferentation is more predominant in low-SR fibers (Furman *et al.*, 2013), we hypothesize that ASSR magnitudes get reduced at higher stimulation levels and shallow modulation depths.

METHOD

ASSR were recorded using a Biosemi ActiveTwo system. The electrode placement followed the 10-10 system. The results in this study were obtained from the Cz-P10 vertical montage potential in response to right-ear stimulation and the Cz-P9 potential in response to left-ear stimulation. The acoustic stimuli were generated in MATLAB and presented to the subject though a pair of ER-2 insert earphones (Etymotic Research Inc.) mounted on an ER-10B+ low noise distortion product oto-acoustic emissions (DPOAE) microphone probe connected to a RME Fireface UCX 24-bit audio interface at a sampling rate of 48 kHz. Subjects were lying on a bed in a double-walled soundproof and electrically shielded booth.

Subjects

A total of 23 adult subjects (12 females) participated in this study. Sixteen (10 females, 26 ± 3 years old) had normal hearing at octave frequencies between 125 and 8000 Hz (threshold ≤ 15 dB HL). Seven mild HI subjects (2 females, 53 ± 14 years old) showed hearing thresholds above 20 dB HL and not higher than 45 dB HL at 4 kHz but normal thresholds at lower audiometric frequencies.

Stimuli and recordings

For the evaluation of peripheral compression, stimuli were presented at sound pressure levels (SPL) ranging from 20 to 80 dB in steps of 5 dB, using a multi-frequency stimulation paradigm. For the HI listeners, the input levels ranged from 30 to 80 dB SPL in 5 dB steps. The multi-frequency stimulus was composed by the addition of four SAM tones, each having a different carrier frequency (f_c = 498, 1000, 2005, and 4011 Hz) to excite four different regions on the BM, and modulated at a different modulation frequency (f_m = 81, 87, 93, and 98 Hz), respectively. For the evaluation of high-intensity level fluctuations, a single SAM tone with $f_c = 2005$ Hz and $f_m = 93$ Hz was used as stimulus. ASSR growth functions were recorded using four modulation depths ($m = 100$, 85, 50, and 25%) and input levels of 34, 40, 54, 60, 63, 66, 71, 74, 77, 81, and 87 dB SPL. All stimuli were generated in epochs, each lasting 1 second.

Data analysis

The recorded epochs were band-pass filtered between 60 to 400 Hz using a zero-phase fourth-order Butterworth filter and rejected if a voltage amplitude of ± 80 µV was reached. Weighted averaging was used to improve the signal-to-noise ratio (Picton *et al.*, 2003). Trials of 16 epochs were concatenated prior to analysis in the frequency domain. A fast Fourier Transform was applied to each trial, and an F-test statistic was use to determine the presence of a signal (Picton *et al.*, 2003). A significant level of 1% was used as criterion for statistical significance of the ASSR. To estimate compression, a two-slopes model similar to the one suggested by Neely *et al.* (2003) was used to estimate the slopes of the level growth function. The model was fitted exclusively to the statistically significant data points.

RESULTS

Estimates of peripheral compression

Figure 1 shows ASSR level growth functions in a representative NH subject. ASSR magnitudes (circles) were well above the background noise (crosses and grey areas), showing smooth and clear functions at all carrier frequencies (panels A-D). Significant ASSR magnitudes (solid circles) were recorded at stimulation levels as low as 20-30 dB SPL and above. ASSR level growth functions showed a compressive growth with level (slopes < 1) up to about 60 dB SPL. Above 60 dB SPL, ASSR growth functions were found to saturate. A compressive growth was found at all frequencies for all subjects, with averaged slope estimates of about 0.25 dB/dB, ranging from 0.1 to 0.5 dB/dB.

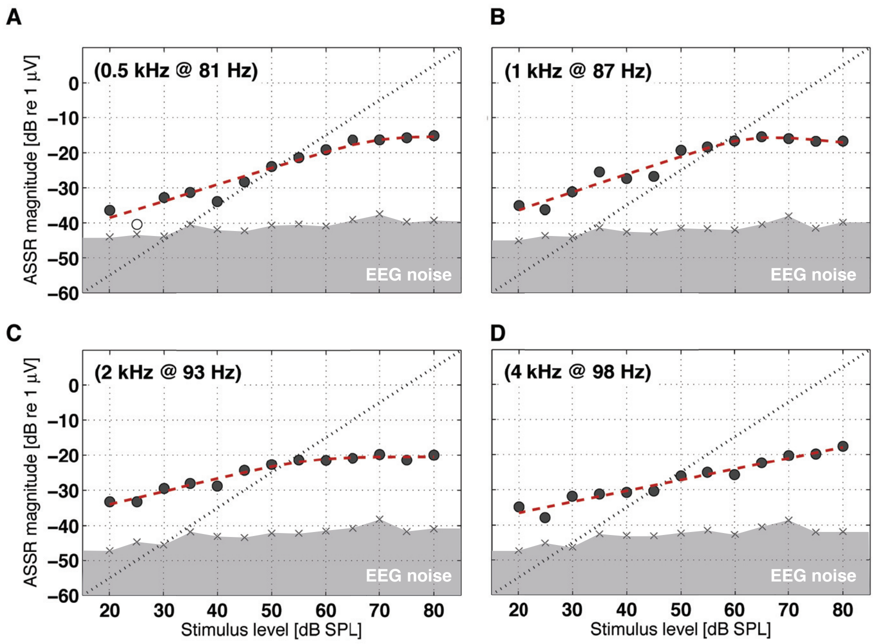


Fig. 1: ASSR growth functions in a representative NH subject. Panels A-D show frequencies at 0.5, 1, 2, and 4 kHz. Filled circles represent statistically significant ASSR magnitudes. Open circles represent non-significant ASSR. Crosses and grey areas show EEG background noise. A linear reference with slope of 1 is represented by the dashed line. The dotted line show a two-slopes fitting curve.

Figure 2 shows ASSR level growth functions in a representative HI subject. For simplicity, panels A and B show results for the 2 and 4 kHz carrier frequencies only. The results at 0.5 and 1 kHz were similar to panel A (not shown). In general, the ASSR level growth functions at the non-impaired audiometric frequencies (panel A) were similar to those in the NH subjects. Panel B shows results at the impaired frequency for this specific subject (with a 30 dB HL threshold). Open circles at the lower input levels represent statistically non-significant ASSR magnitudes.

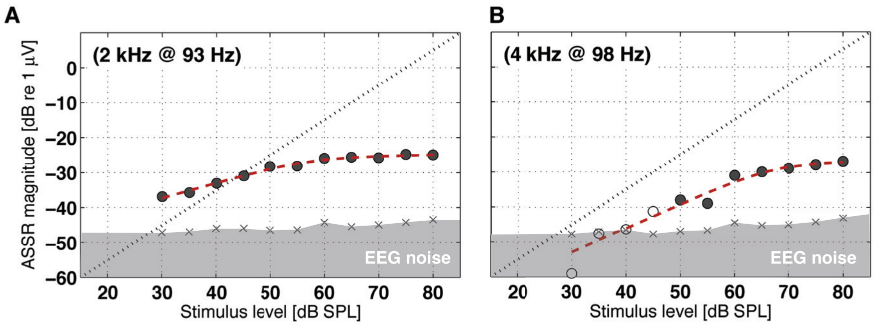


Fig. 2: Same as Fig. 1 but for a representative HI subject. Only frequencies 2 and 4 kHz (panels A and B) are shown (same as panels C and D in Fig. 1)

ASSR growth functions at different modulation depths

Figure 3 shows results from ASSR level growth functions recorded using four modulation depths for three individual NH subjects. Only the upper input level range is shown. Modulation depths ranged from fully modulated (m = 100%, circles) to shallow modulation (m = 25%, squares). Modulations at m = 85% are indicated by downwards triangles and m = 50% are shown as upwards triangles. The results from subjects NH2 and NH5 (panels A and B) were similar, with ASSR magnitudes growing monotonically with level at all modulation depths. The ASSR level growth functions in subjects NH2 and NH5 showed constant compressive slopes comparable to the slopes shown in Fig. 1. The ASSR level growth functions for 100% and 25% modulation depths were parallel, with larger magnitudes for 100% modulation results than for 25% modulation depth. The results from subject NH4 (panel C) showed similar ASSR growth functions only at the larger modulations, whereas ASSR magnitudes at 25% modulation depth (squares) were reduced.

DISCUSSION AND CONCLUSION

The results in Fig 1 showed that ASSR level functions grow compressively for stimulation levels up to 60 dB SPL. Compression could be estimated from all NH subjects and at all frequencies simultaneously using a multi-frequency paradigm. Estimates of compression were on average about 0.25 dB/dB (compression ratio of 4), which is in good agreement with behavioral estimates of cochlear compression (Plack *et al.*, 2004) and compression estimates using DPOAEs (Neely *et al.*, 2003). Above 60 dB SPL, ASSR growth functions recorded from multi-frequency stimulation saturate, probably due to the interaction between the different SAM components at the level of the cochlea and suppression mechanisms in the BM (Picton *et al.*, 2007). This may also explain why ASSR growth functions saturate less at 4 kHz (Panel D in Fig. 1), as there is not a higher frequency tone that suppresses the 4 kHz response.

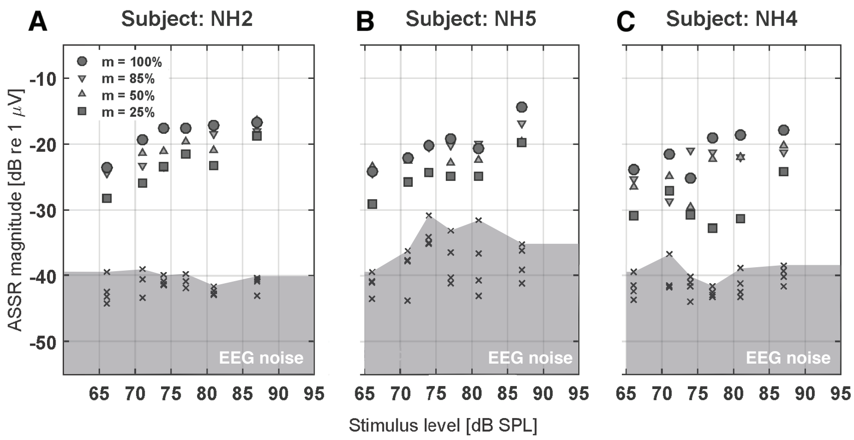


Fig. 3: ASSR growth functions for three NH subjects at four modulation depths. Circles show 100% modulation depth, downward triangles show $m = 85\%$, downward triangles show $m = 50\%$ and squares show $m = 25\%$. Crosses and grey areas show EEG background noise.

For the HI subjects, the ASSR growth functions at the non-impaired frequencies (panel A in Fig. 2) showed the same behavior as the results for the NH subjects (Fig. 1). However, at the mildly impaired frequency (panel B in Fig. 2), the ASSR magnitudes at lower input levels were statistically non-significant, which represents the loss of sensitivity or threshold elevation at this frequency. The non-significant ASSR magnitudes at low stimulus levels did not allow a proper fit of the two-slope model. The data indicate, however, a loss of compression at the impaired frequencies.

ASSR growth functions at higher supra-threshold levels and shallow modulations showed a large variability across young subjects with normal audiograms. Figure 3 showed that, in some NH subjects, ASSR growth functions at shallower modulations are reduced, in line with the initial hypothesis. At higher input levels, the rate-intensity function of high-SR fibers saturate, whereas it increases with level for the low-SR fibers (Liberman, 1978). Considering that deafferentation is more predominant in low-SR fibers (Furman *et al.*, 2013), the reduction in ASSR magnitude at these higher levels might be connected to the inability of the ANF to code the intensity fluctuations.

In addition to the use of ASSR to estimate thresholds, it is suggested here that the slope of ASSR level growth functions at low supra-threshold levels can be used to estimate peripheral compression at different frequencies simultaneously both in NH and HI listeners. It is also hypothesized that ASSR growth functions at higher stimulation levels, using shallow modulations, reflect the integrity of ANFs in special low-SR fibers, which can lead to a potential tool to evaluate individuals suffering from deafferentation.

ACKNOWLEDGMENTS

This work was funded by the Oticon Center of Excellence for Hearing and Speech Sciences (CHeSS) at the Technical University of Denmark (DTU).

REFERENCES

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.

Hind, S.E., Haines-Bazrafshan, R., Benton, C.L., Brassington, W., Towle, B., and Moore, D.R. (**2011**). "Prevalence of clinical referrals having hearing thresholds within normal limits," Int. J. Audiol., **50**, 708-716.

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.

Kumar, G., Amen, F., and Roy, D. (**2007**). "Normal hearing tests: Is a further appointment really necessary?" J. R. Soc. Med., **100**, 66.

Liberman, M.C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am., **63**, 442-455.

Lin, H.W., Furman, A.C., Kujawa, S.G., and Liberman, M.C. (**2011**). "Primary neural degeneration in the Guinea pig cochlea after reversible noise-induced threshold shift," J. Assoc. Res. Otolaryngol., **12**, 605-616.

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.

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.

Picton, T.W., John, M.S., Dimitrijevic, A., and Purcell, D. (**2003**). "Human auditory steady-state responses," Int. J. Audiol., **42**, 177-219.

Picton, T.W., van Roon, P., and John, M.S. (**2007**). "Human auditory steady-state responses during sweeps of intensity," Ear Hearing, **28**, 542-557.

Plack, C.J., Drga, V., and Lopez-Poveda, E.A. (**2004**). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am., **115**, 1684-1695.

Ruggero, M.A., Rich, N.C., Recio, A., Narayan, S.S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am., **101**, 2151-2163.

Saunders, G.H. and Haggard, M.P. (**1989**). "The clinical assessment of obscure auditory dysfunction – 1. Auditory and psychological factors," Ear Hearing, **10**, 200-208.

Winter, I.M., Robertson, D., and Yates, G.K. (**1990**). "Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibres," Hear. Res., **45**, 191-202.

Are temporary threshold shifts reflected in the auditory brainstem response?

LOU-ANN CHRISTENSEN ANDERSEN[1,2,*], TURE ANDERSEN[1],
ELLEN RABEN PEDERSEN[3], AND JESPER HVASS SCHMIDT[1,2]

[1] *Institute of Clinical Research, University of Southern Denmark, Odense, Denmark*

[2] *Department of Audiology, Odense University Hospital, Odense, Denmark*

[3] *The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, Denmark*

Background: Temporary hearing loss in connection with excessive exposure to sound is described as temporary threshold shift (TTS). The auditory cortex has neural pathways, which directly affect the medial olivocochlear system (MOCS) via the descending efferent auditory system. One of the functions of MOCS may be to protect the inner ear from noise exposure. *Objective*: To investigate the influence of a TTS measured with auditory brainstem responses (ABRs) using noise, familiar, and unfamiliar music as auditory exposure stimulus, respectively. *Method*: Normal-hearing subjects were exposed to the three different sound stimuli in randomized order on separate days. Each stimulus was 10 minutes long and the average sound pressure level was 100 dB. ABRs (4-kHz tone burst) were measured pre-exposure and also immediately after the sound exposure. *Results*: Preliminary results show a tendency towards an increase in the ABR amplitude for Jewit I and a decrease in the ABR amplitude for Jewit V for the left ear after sound exposure. Jewit I represents action potentials in the spiral ganglion neuron, and Jewit V represents action potentials further up the brainstem.

INTRODUCTION

Exposure to high sound levels may entail a temporary threshold shift (TTS), which is described as a temporary hearing loss in connection with immoderate sound exposure. If the hearing loss persists, the threshold shift is considered to be a permanent threshold shift (Quaranta *et al.*, 1998).

As a part of the auditory efferent neural pathway, the medial olivocochlear system (MOCS) originates from the medial superior olive (MSO) and project mainly onto the contralateral cochlear and forms synapses with outer hair cells (OHC) (Fig. 1). MOCS inhibits OHC motility and one of the MOCS functions may be to protect the inner ear from noise exposure (Perrot and Lionel, 2014).

*Corresponding author: loand10@student.sdu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

The auditory cortex has neural pathways, which directly affect MOCS via the descending efferent auditory system (Fig. 1) (Perrot and Lionel, 2014). Studies investigating auditory selective attention and visual attention have shown contradictive results regarding the influence on MOCS (Perrot and Lionel, 2014). Because of the negative and contradictive results further work is needed to clarify the effects of auditory attention.

It is known that different sound characters (music and noise) induce different levels of TTS when comparing noise with music (Strasser *et al.*, 2003). Maybe the character of the sound is important for MOCS activation too?

A TTS can be measured in several ways including normal audiometry and otoacoustic emissions (Kemp, 2002). The immediate change of the ABR after sound exposure has not been studied in humans. However, ABR is affected in normal hearing subjects with tinnitus (Schaette and McAlpine, 2011) and previous noise exposure within 12 months before ABR measurement (Stamper and Johnson, 2015).

This paper gives an overview over the temporary findings in a one-year research project where the temporary changes of the ABR have been studied.

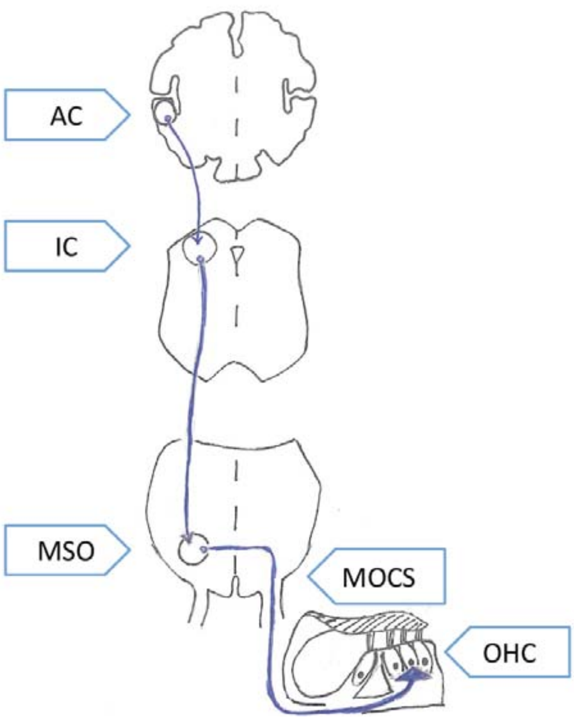


Fig. 1: A simplified representation of the descending efferent auditory system: AC (auditory cortex), IC (inferior colliculus), MSO (medial superior olive), MOCS (medial olivocochlear system), and OHC (outer hair cell).

AIM

To investigate the influence of a TTS measured with auditory brainstem responses (ABRs) using noise, familiar, and unfamiliar music as auditory exposure stimuli.

METHODS

Subjects

Thirteen normal-hearing subjects have participated in the experiment to date. They were recruited using posters at the University of Southern Denmark, at social Internet sites, and similar or related locations. The inclusion criteria were defined as normal ear canals without obstructing cerumen and hearing thresholds better than 20 dB hearing level (HL) at the frequencies 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz evaluated by a two-alternative forced-choice (2AFC) audiometry test (Schmidt *et al.*, 2014). The exclusion criteria were impaired hearing at the first 2AFC audiometry test, smoking (due to a possible influence on the nervous system), and chronic or acute disease in the middle ear.

The test subjects were between 22 and 27 years old (mean of 24 years) consisting of six males (46%) and seven females (54%). One out of the 13 test subjects was left-handed, whereas the remaining 12 were right-handed.

Exposure

The subjects were binaurally exposed to three sound stimuli: music shaped noise, known music, and unknown music. Each stimulus was 10 minutes long and the average unweighted sound pressure level was 100 dB (97 dBA), hereby sufficiently below the Danish noise regulations for work places (L_{eq} of 85 dBA for eight hours). The National Committee on Health Research Ethics have accepted the project. The known music was selected among the top 100 songs of the 500 greatest songs published by www.rollingstones.com and consisted of ten different songs. A central part of each song was played for approximately one minute. The known and unknown music was matched by rhythm. The music shaped noise was made from a white noise signal that was shaped to have same frequency composition as the known music.

While exposed, the test subjects were randomized to the task of auditory attention or non-auditory attention. For each of the two attention tasks the test subjects were exposed to the three sound stimuli in random order, i.e., each subject was exposed to six different test conditions on six different days. The different test conditions are shown in Fig. 2. The non-auditory attention task was the “Tower of Hanoi” – a mathematical, analytical, and motor cortex-demanding puzzle. The subject was presented with the task in order to avoid evoked activity of the hearing sense. The part of the study presented in this article does not investigate the effect of auditory attention and the type of sound stimuli, because a balanced randomization between attention and stimuli was not completed at this time.

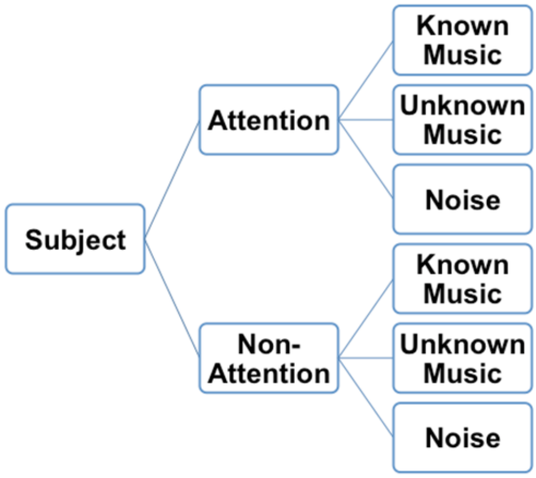


Fig. 2: An overview of the six different test conditions. The subjects were presented with the different conditions on separate days. The order in which each subject was presented with the different conditions was randomized.

Measurements

A small questionnaire and a test of musicality, called the Advanced Measures of Music Audiation (AMMA) test, were obtained before sound exposure. Pre- and post sound exposure ABRs at 4 kHz for 90 dB nHL tone bursts and 2AFC audiometry tests (pre exposure: 1, 2, 3, and 4 kHz and post exposure: 4 kHz) were conducted (Fig. 3). The pre exposure ABR test (pretest) consists of two measurements that are combined in the analysis. The post exposure ABR test consists of three measurements (test 1-3) that are used separately in the comparison to the pre exposure ABR test. After sound exposure, debriefing was used to check if the subject was paying sufficiently attention to the task and his/hers acquaintance with the music. Pre- and post sound exposure measurements of distortion product otoacoustic emissions (DPOAEs) were also a part of the overall study design but are not dealt with in this paper.

Material

A computer-controlled RM-2 processor (Tucker-Davis Technologies) was used for audiometry. ABRs were recorded with Eclipse (Interacoustics), using ER-3A insert headphones. Sennheiser HDA200 headphones were used for sound exposure.

Statistics

All data were analyzed with linear mixed models with subjects as random effects. At the time of data analysis, not all 13 subjects had completed tests for all six conditions. Their randomization in sound stimuli and attention was thus not complete. This was taken into account in the statistics.

Fig. 3: Pre and post exposure measurements in the overall study design by order. Distortion product otoacoustic emissions (DPOAEs) are not dealt with in this paper. The DPOAE, audiometry test, and auditory brainstem response (ABR) were performed before and after sound exposure. The DPOAE and ABR were measured alternately, starting with the DPOAE.

RESULTS

The ABR amplitudes (Jewit I to V) represent the action potentials from the spiral ganglion neuron and throughout the brain stem to the inferior colliculus.

With the difference in amplitude between Jewit I and Jewit V (DiffVI), calculated as V_Amplitude – I_Amplitude, as the response, a linear mixed model was created to describe the difference in DiffVI prior (pretest) and after (test 1-3) sound exposure, with the data not divided into type of sound stimuli or attention task.

Preliminary results showed a significant negative change in DiffVI for the left ear between pretest and test 1 [mean = −0.06, $p = 0.00$], and between pretest and test 2 [mean = −0.06, $p = 0.01$] (Fig. 4). No significant change of DiffVI was observed for the right ear (Fig. 4).

The AMMA score, represented as the combined tone and rhythm score (AMMA_Com), showed a significant increase of DiffVI for the left ear [mean = 0.01, $p = 0.03$] (Table 1) and the right ear [mean = 0.01, $p = 0.00$] (Table 1).

Other significant factors were age for the right ear [mean = 0.02, $p = 0.00$] (Table 1) and FMP, which is a quality number of the ABR response, for the left ear [mean = −0.00, $p = 0.01$] (Table 1).

During backwards-stepwise elimination analysis the type of stimuli variable showed no influence on the results and was thus eliminated from the final model.

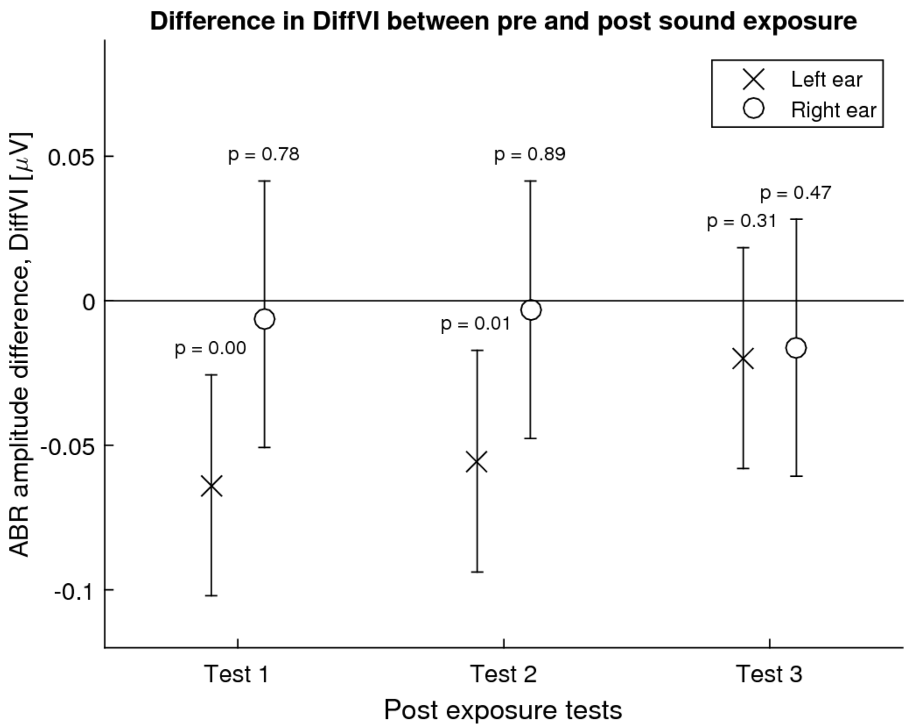


Fig. 4: The mean difference in DiffVI (V_Amplitude – I_Amplitude) with 95% confidence interval between pretest and post exposure tests 1-3, with the data not divided into type of sound stimuli or attention task. The figure shows results for both the left and the right ear.

DISCUSSION

The overall study design includes 20 subjects with complete randomization. This paper only included 13 subjects with incomplete randomization in sound stimuli and attention, therefore the data were not divided into different sound stimuli (music shaped noise, known music, and unknown music) and tasks (attention and non-attention).

Preliminary results showed a significant negative change in DiffVI for the left ear between pretest and test 1 as well as between pretest and test 2 (Fig. 4). This negative change showed a tendency towards a suppression of Jewit V and/or an excitation of Jewit I on the left ear after sound exposure.

The difference in the findings between the right and the left ear may be due to some kind of lateralization of the auditory system, which may be related to the handedness lateralization. Only one out of the 13 test subjects in this sample was left-handed. Maybe right-handed listeners have better protection of their right ear due to a more dominant descending auditory system of MOCS that mainly innervates the contralateral outer hair cells (Fig. 1). Further work is needed to investigate this consideration.

		Coef.	p	95% conf. interval
Left: DiffVI	AMMA_Com	0.01 *	0.03	0.00 ; 0.02
	Female	0.00	1.00	−0.10 ; 0.10
	Age	0.02	0.12	−0.01 ; 0.05
	Right hand	−0.06	0.38	−0.20 ; 0.08
	FMP	−0.00 *	0.01	−0.00 ; −0.00
	Constant	−0.84	0.04	−1.65 ; −0.03
Right: DiffVI	AMMA_Com	0.01 *	0.00	0.01 ; 0.01
	Female	0.02	0.41	−0.03 ; 0.06
	Age	0.02 *	0.00	0.01 ; 0.03
	Right hand	−0.05	0.14	−0.13 ; 0.02
	FMP	0.00	0.80	−0.00 ; 0.00
	Constant	−0.79	0.00	−1.12 ; −0.46

Table 1: The other covariates in the model, for the left and the right ear. AMMA_Com (tone and rhythm scores), Right hand (dominant hand), FMP (quality number of ABR response), constant (the regression constant for the linear mixed model). * Statistically significant difference at a 5% significance level.

The AMMA score had a significant positive change in DiffVI for both ears (Table 1). This shows a tendency towards the ABR can be affected by a person's level of musicality after sound exposure, which is comparable to other investigation of auditory training effects seen in musicians (Micheyl *et al.*, 1995).

The lateral asymmetry in human auditory processing is described as domain-specific lateralization (speech/music) or parameter-specific lateralization where the left and right hemispheres are specialized to process respectively rapid temporal changes and tiny changes in pitch (Tervaniemi and Hugdahl, 2003). The sound stimuli in our study contained both temporal and spectral aspects but the quantity of the two in each type of stimulus had not been matched and all sound stimuli data were unified. Maybe this influenced the results for the left and the right ear. The subjects' familiarity with the sound stimuli and their musical abilities were taken into account in the statistics, but not their training in language—all variables that may influence the pattern of lateralization (Tervaniemi and Hugdahl, 2003). The types of stimuli and number of experimental days were also considered in the study along with the variables in Table 1.

The non-attention task was chosen to avoid evoked activity of the hearing sense, by being a mathematical, analytical, and motor cortex-demanding puzzle.

CONCLUSION

A significant negative change in DiffVI (the difference in ABR amplitude between Jewit I and Jewit V) was shown between pretest and test 1 after sound exposure as well as between pretest and test 2 after sound exposure for only the left ear. This may be caused by an excitation of Jewit I and/or a suppression of Jewit V after sound exposure.

The AMMA score had a significant positive change in DiffVI for both ears. This may point at a trend towards the auditory brainstem response after sound exposure being affected by how musically trained a person is, regarding tone and rhythm.

The presented data are preliminary results from an ongoing one-year research project. Results from more test subjects will be presented at a later stage.

ACKNOWLEDGEMENTS

The authors thank the test subjects for participating, Torben Poulsen from the ISAAR 2015 Scientific Committee for valuable comments on the manuscript and Chris V. Jørgensen from DELTA for technical support. The authors also thank the Danish Association of the Hard of Hearing and Region Syddanmark for financing the project.

REFERENCES

Kemp, D. (**2002**). "Otoacoustic emissions, their origin in cochlear function, and use," Br. Med. Bull., **63**, 223-241.

Micheyl, C., Carbonnel, O., and Collet, L. (**1995**). "Medial olivocochlear wystem and loudness adaptation – Differences between musicians and non-musicians," Brain. Cogn., **29**, 127-136.

Perrot, X. and Lionel, C. (**2014**). "Function and plasticity of the medial olivocochlear system in musicians: A review," Hear. Res., **308**, 27-40.

Quaranta, A., Portalatini, P., and Henderson, D. (**1998**). "Temporary and permanent threshold shift: an overview," Scand. Audiol. Suppl., **48**, 75-86.

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

Schmidt, J., Brandt, C., Pedersen, E., Christensen-Dalsgaard J., Andersen, T., Poulsen, T., and Bælum, J. (**2014**). "A user-operated audiometry method based on the maximum likelihood principle and the two-alternative forced-choice paradigm," Int. J. Audiol., **53**, 383-391.

Stamper, G. and Johnson, T. (**2015**). "Auditory function in normal-hearing, noise-exposed human ears," Ear. Hearing, **36**,172-184.

Strasser, H., Irle, H., and Legler, R. (**2003**). "Temporary hearing threshold shifts and restitution after energy-equivalent exposures to industrial noise and classical music," Noise Health, **5**, 75-84.

Tervaniemi, M. and Hugdahl, K. (**2003**). "Lateralization of auditory-cortex functions," Brain. Res. Rev., **43**, 231-246.

Signs of noise-induced neural degeneration in humans

PERNILLE HOLTEGAARD[1,*] AND STEEN ØSTERGAARD OLSEN[2]

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

[2] *Research Laboratory, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital, Rigshospitalet, Copenhagen, Denmark*

Animal studies demonstrated that noise exposure causes a primary and selective loss of auditory-nerve fibres with low spontaneous firing rate. This neuronal impairment, if also present in humans, can be assumed to affect the processing of supra-threshold stimuli, especially in the presence of background noise, while leaving the processing of low-level stimuli unaffected. The purpose of this study was to investigate if signs of such primary neural damage from noise-exposure could also be found in noise-exposed human individuals. It was investigated: (1) if noise-exposed listeners with hearing thresholds within the "normal" range perform poorer, in terms of their speech recognition threshold in noise (SRTN), and (2) if auditory brainstem responses (ABR) reveal lower amplitude of wave I in the noise-exposed listeners. A test group of noise/music-exposed individuals and a control group were recruited. All subjects were between 18-32 years of age and had pure-tone thresholds ≤ 15 dB HL from 250-8000 Hz. Despite normal pure-tone thresholds, the noise-exposed listeners required a significantly better signal-to-noise ratio to obtain SRTN, compared to the control group. The ABR results showed significantly lower amplitude of wave I, in the left-ear, of the test group listeners. Significantly higher wave III and normal wave V were also found in the left ear of the test group listeners suggesting a compensated neural gain in the brainstem. Overall, the results from this study seem to suggest that noise exposure affects supra-threshold processing in humans before pure-tone sensitivity, raising suspicion to the hypothesis of primary neural involvement.

INTRODUCTION

For decades the outer hair cells (OHCs) have been presumed to be the primary targets of noise-exposure (Spoendlin, 1971; Lawner *et al.*, 1997), and the first auditory symptom has been assumed to be elevated pure-tone thresholds, showing a dip/noise-notch around 4 kHz. However, our current knowledge of noise-induced hearing loss is now questioned both in regards to the pathology, but also in respect to the perceptual consequences of the damage. Recent research on animal models (mice and guinea pigs) has suggested that noise exposure causing only a temporary

---

*Corresponding author: perholt@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

threshold shift (TTS) can lead to primary and extensive damage of the afferent type I nerve fibres innervating the inner hair cells (IHCs), despite the recovery of pure-tone thresholds and no evidence of OHC and IHC loss (Kujawa and Liberman, 2009; Lin *et al.*, 2011; Furman *et al.*, 2013). Also, despite normalisation of the pure-tone thresholds the wave-I component of the auditory brainstem response (ABR) in response to supra-threshold stimuli (80 dB nHL) was found to be significantly reduced in the exposed animals. This reduction is assumed to reflect that fewer peripheral afferent nerve fibres fire synchronously in response to supra-threshold sound stimuli, supporting that noise-exposure causing a TTS can cause primary damage of the peripheral synapses and nerve fibres impairing only supra-threshold processing. This synaptic and neural damage seems to be of progressive nature causing a slow degeneration of the spiral ganglion cells (Kujawa and Liberman, 2009). For the remainder of this document this noise-induced neural damage will be referred to as noise-induced neural degeneration (NIND).

In 2013, Furman *et al.* further documented this NIND to be selective of the nerve fibres with low spontaneous firing rate, i.e., low spontaneous rate fibres (LSRFs), while leaving the high-spontaneous rate fibres (HSRFs) unaffected. This finding provides a physiological explanation to why NIND does not affect pure-tone sensitivity, but primarily affects supra-threshold processing. The HSRFs, found to be largely unaffected (Furman *et al.*, 2013), are responsible for the coding of low-intensity stimuli (Liberman, 1978; Taberner and Liberman, 2005). However, the LSRFs that are suggested to be the primary targets of noise-exposure (Furman *et al.*, 2013) are responsible for the coding of mid- to high-intensity stimuli (Liberman, 1978; Taberner and Liberman, 2005). In addition to coding supra-threshold stimuli, LSRFs have been suggested to have greater resistance to the limitations of saturation that can occur in the presence of high background noise levels (Costalupes *et al.*, 1984). This suggests the LSRFs to be important for the processing of auditory stimuli in the presence of high-level background noise. Thus, assuming acoustic overexposure also causes primary NIND of the LSRFs in humans, it can be hypothesised that the first signs of a noise-induced hearing impairment is supra-threshold processing difficulties, and not elevated pure-tone thresholds.

Signs of a disorder impairing only supra-threshold processing without affecting pure-tone thresholds have been documented in humans before and it has been referred to as, e.g., the “King-Kopetzky Syndrome” (KKS; Zhao and Stephens, 1996) or more recently as “hidden hearing loss” (Schaette and McAlpine, 2011). The main symptom of KKS is difficulties with speech in noise despite normal pure-tone thresholds (Zhao and Stephens, 1996). Based on the characteristics of LSRFs this deficit can also be assumed to occur in response to NIND.

The goal of the current study was to investigate if signs of NIND could also be documented in humans with a history of acoustic exposure, and if this damage could potentially be linked to the diagnosis of KKS/hidden hearing loss. Using a combination of supra-threshold behavioral tests and electrophysiological measures we set out to test (1) if a test group with a history of acoustic exposure needs a better signal-noise-ratio (SNR) to understand speech in noise, compared to a control group, with no history of acoustic exposure despite pure-tone thresholds ≤ 15 dB HL in all subjects, and (2) if the test group has lower amplitudes of the wave-I component of their ABR compared to the control group.

METHOD

Two groups (a test and a control group) of young normal hearing listeners between 18-32 years of age and with pure-tone thresholds < 20 dB HL from 250-8000 Hz were recruited and participated in the study. The listeners were classified and divided into groups based on their present and/or past experience working in noise or music exposure. Thorough questioning regarding acoustic exposure was always completed with each listener to ensure that the control group listeners had not been exposed to any longer lasting acoustic exposures. Control group listeners with large scale usage of MP3 players or similar exposures were excluded from the study. The control group consisted of listeners with no work-related acoustic exposure. The test group listeners however, represented listeners with a history of acoustic over-exposure from their work environment. Work-related acoustic overexposure was defined as a work environment with a level of noise or music so loud that the listener felt that it would be necessary to raise his or her voice in order to conduct a conversation. Furthermore a test group with listeners categorized as having a history of acoustic exposure had to have worked in this noise or music for at least 5 hours a day, 5 days a week, for at least 6 months. The test group listeners consisted mainly of professional musicians (14 out of 16) that were recruited from the Royal Danish Navy Band. The test group consisted of 16 listeners (12 men, 4 women), with a history of acoustic exposure from their workplace. The control group consisted of 16 listeners (12 men, 4 women) with no history of acoustic exposure.

Procedure and materials

All listeners completed a test session of 2 hours. Initially a questionnaire was filled out together with the researcher. Otoscopy was performed. Pure-tone audiometry was conducted for frequencies 250-8000 Hz in a double-walled sound-proof booth, using a GN Otometrics Madsen Astera Audiometer and Sennheiser HDA 200 circumaural earphones. Speech recognition thresholds (SRTs) and word recognition score in quiet were also measured, using the “Dantale I” material, to ensure normal hearing and processing of speech in quiet.

The speech material “Dantale II” (Wagener *et al.*, 2003) was applied for testing SRTs in noise (SRTN), to investigate if the test group listeners needed a better dB SNR, compared to the control subjects, to obtain 50% speech intelligibility in noise. The speech and noise signal were presented binaurally in the sound-field environment of the sound booth with the listener seated in the center between 5 loudspeakers. The speech signal was presented from a front loudspeaker (0° azimuth), and the noise was presented from two speakers at ± 45° and two at ± 135° azimuth. Each new session started with a SNR of 0 dB SNR, i.e. a noise level of 70 dB SPL and speech at 70 dB SPL. The noise level was kept at a constant level of 70 dB SPL while the speech presentation level was adjusted according to the number of words repeated correctly in each sentence. Three training lists of 10 sentences each were always completed to familiarize the test subject with the material. The speech recognition threshold in noise was calculated from the subsequent sentences by adding the presentation levels from sentences 12-31 together, dividing by 20, and then subtracting the noise level giving the final result in dB SNR.

The measure of ABR was performed using the Interacoustics Eclipse ABR system (EP15/EP25). Disposable non-invasive electrodes were used for this purpose. 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 just below the non-inverting electrode. An impedance of maximum 3 kΩ was always ensured. Click stimuli at a level of 90 dB nHL were presented with alternating polarity at a rate of 16.1/s through ER-3A insert earphones. A time window of 0-20 ms was used and 4000 sweeps were completed for all listeners. Amplitudes of wave I, III and V were measured from peak to following trough.

RESULTS

The statistical method "Mann-Whitney *U*" was applied to investigate significant differences between the two groups. As expected the measures of speech recognition threshold in noise showed a significant difference ($p < 0.001$) between the two groups, despite normal pure-tone thresholds in all the listeners. The test group listeners needed a significantly higher speech level to recognize 50% of the speech material in noise. Figure 1 shows the SRTN data for both groups. From this figure it is seen that the test group listeners generally required a higher presentation level of the speech signal to obtain their SRTN compared to the control group.

For the ABR a significant difference of the wave amplitudes could only be documented from the left ear between the two groups. Significantly lower wave I amplitudes ($p < 0.05$) were documented from the left ear of the test subjects (M = 0.253 μV, SD = 0.107) compared to the control subjects (M = 0.326 μV, SD = 0. 092). Further analysis of the left ear ABR amplitudes showed the opposite tendency for the subsequent wave III amplitudes. The test group had significantly higher amplitude ($p < 0.05$) of wave III (M = 0.431 μV) compared to the control group (M = 0.344 μV). For wave V no significant difference was observed between the two groups. The amplitudes of wave I and III for the two groups are displayed in Fig. 2, panels A and B.

DISCUSSION

The SRTN results confirmed that the noise-exposed test group listeners needed a significantly higher speech level to recognize 50% of a speech signal in noise. This result cannot be regarded as an effect of impaired pure-tone sensitivity as normal pure-tone sensitivity (pure-tone thresholds $\leq$ 15 dB HL) was documented only minutes prior to the measure of SRTN. This finding could thus be assumed to reflect NIND affecting processing of supra-threshold stimuli in background noise.

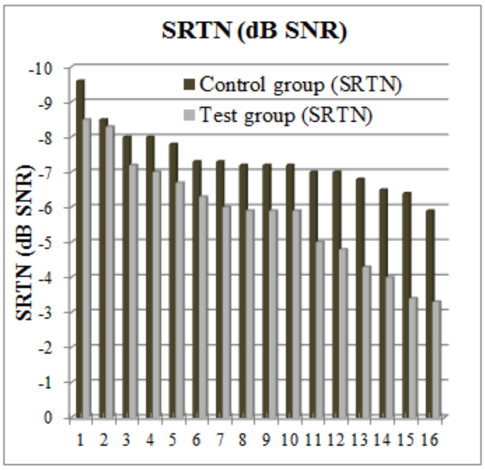


Fig. 1: The SRTN data for both groups arranged from smallest to highest value (dB SNR). The light grey pillars reflect data from the test group listeners and the dark represent the data from the control group. The height of the pillars reflect the level by which the speech signal could be reduced compared to the noise level (70 dB SPL), while still recognizing 50% of the speech.

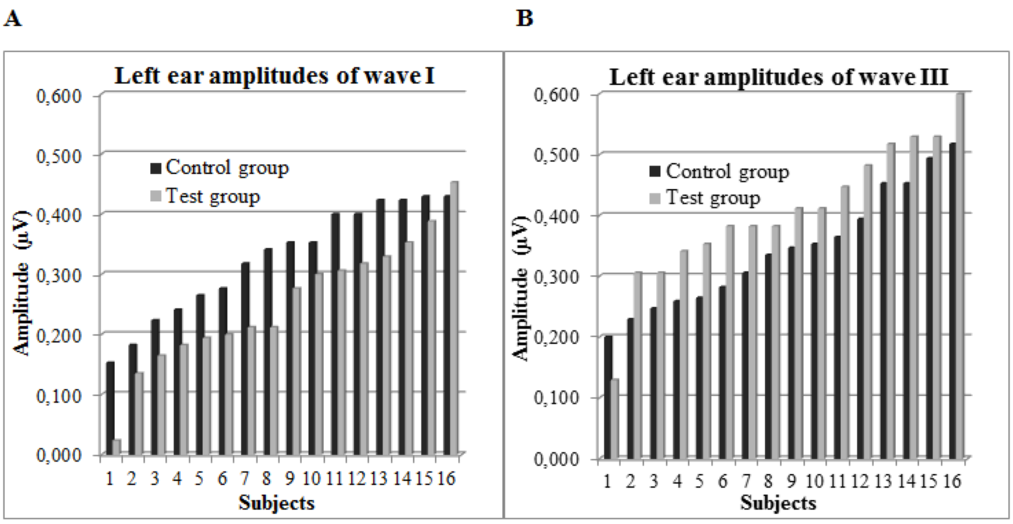


Fig. 2: Left ear ABR amplitudes of wave I and III across test and control listeners, arranged from smallest to highest value (µV). The left side (panel A) reflects the wave I amplitudes, and the right (panel B) reflects the amplitudes of wave III in response to a 90 dB nHL click. The light grey pillars reflect data from the test group and the dark pillars represent the control group.

ABR was measured to acquire objective and physiological evidence of NIND. It was hypothesised that the test group would present with lower amplitudes of the wave-I component reflecting reduced neural synchrony. The results of the ABR measurements did in fact confirm the anticipated hypothesis. Significantly lower amplitudes of wave I were found in the left ear of the test group listeners. However, no statistically significant difference was documented from the right ear. With the hypothesis only confirmed for the left ear, it can be questioned whether this asymmetric finding reflects NIND, or if it is merely a coincidence or an error. It can be argued that the asymmetry could be a potential consequence of using professional musicians in the test group. Musicians cannot be expected to be evenly exposed on each ear, thus symmetric NIND cannot be expected. Also, there have been findings suggesting the left ear to be more vulnerable to noise damage than the right ear. Binaural noise exposure has been shown to cause more severe TTS on the left ear compared to the right (Pirilä, 1991). Furthermore, tinnitus which is a common consequence of acoustic overexposure (Palmer *et al.*, 2002) has also been suggested to be more common in the left ear (Axelson and Ringdahl, 1989). Also, greater efferent activity of the medial olivocochlear bundle (MOCB) has been indicated on the right ear (Bidelman and Bhagat, 2015) and this right ear efferent activity has furthermore been suggested to be greater in musicians vs. non-musicians (Micheyl *et al.*, 1997). The MOCB has been shown to have a protective role of the ear against noise exposure (Maison *et al.*, 2013). Thus, it can be speculated that the majority of the test group listeners have greater protection against acoustic exposure from the efferent MOCB on the right ear, providing a potential explanation to why signs of NIND were only indicated from the measures of the left-ear ABR. This is highly speculative and more research is needed to explain this asymmetric finding. It cannot be affirmed with certainty that the findings of poorer SRTN and lower amplitudes of wave I in the test group is caused by NIND.

Despite the asymmetry of the findings and the lack of correlation between the scores of SRTN and the wave-I amplitudes, there are still findings raising suspicion that NIND is the potential contributor of these results. Findings of significantly enhanced wave-III amplitudes of the left ear ABR of the test group listeners can potentially support the findings of poorer SRTN and lower wave I amplitudes to be a result of NIND. Decreased synchronous sound evoked activity in the auditory nerve (reflected as a reduced wave I amplitude), as a result of loss of LSRFs has been found to lead to a compensated neural gain (hyperactivity) in the brainstem (Hickox and Liberman, 2014; Knipper *et al.*, 2013; Schaette and McAlpine, 2011). This pathological increase in the response gain is reflected in the ABR as normal or increased amplitudes of waves III and V in the presence of reduced amplitude of wave I. In the current test group with reduced amplitude of wave I, wave III was significantly enhanced in the left ear and wave V showed no significant difference between the two groups. In the presence of the significantly reduced wave-I amplitudes of the test group, the enhanced wave III and normalised wave V can be suggested to reflect compensated neural gain in the brainstem in the response to NIND. However, the results must still be analysed with caution as there are limitations of this study. Factors such as the exposure characteristics (duration, level, etc.), environmental differences and genetic factors are not accounted for or controlled. With the spiral ganglion cell (SGC) loss suggested to be slowly progressive over many years, the time elapsed since the initial exposure could play a role in the magnitude of the SGC loss. Thus, some test group listeners may suffer more progressed NIND than others while a few may have no NIND at all, despite a somewhat similar type of acoustic exposure in the test group (14 from the same workplace). Also, with no objective measure of the work-related exposure level of the test group listeners, it cannot be proven that the test group has been exposed to damaging levels and durations. However, the results do show significant deviations of supra-threshold processing in the test group listeners with a history of working in acoustic exposure compared to the control listeners with no reported history of acoustic exposure. It can thus be argued that these findings, despite not knowing the exact levels of the exposure, are a result of an exposure severe enough to affect supra-threshold processing of the left ear, suggested here to be perhaps the most vulnerable ear to noise exposure, and thus suggest NIND.

In conclusion, the test group listeners with a history of acoustic overexposure were found to need significantly higher SNRs to recognise 50% of a speech signal in noise despite normal pure-tone sensitivity. They also showed significantly reduced wave-I amplitudes of the left ear ABR. In the presence of pure-tone thresholds ≤ 15 dB HL and signs of compensatory neural gain in the brainstem (i.e., enhanced amplitude of wave III) which is found to accompany loss of LSRFs, these findings of lowered wave-I amplitudes and poorer SRTN can be suggested to reflect signs of NIND in the test group. This study cannot proclaim NIND in human listeners. However, it does show signs of impairments in agreement with the pathology of NIND in listeners with a history of acoustic overexposure. Thus these results can support the possibility that acoustic overexposure can also lead to NIND in humans. This study implicates the need of more research towards exploring this pathology and the potential auditory consequences in humans.

REFERENCES

Axelsson, A. and Ringdahl, A. (**1989**). "Tinnitus – A study of its prevalence and characteristics," Br. J. Audiol., **23**, 53-62.

Bidelman, G.M. and Bhagat, S.P. (**2015**). "Right-ear advantage drives the link between olivocochlear efferent 'antimasking' and speech-in-noise listening benefits," NeuroReport, **26**, 483-487

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.

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.

Lin, H.W., Furman, A.C., Kujawa, S.G., and Liberman, M.C. (**2011**). “Primary neural degeneration in the guinea pig cochlea after reversible noise-induced threshold shift,” J. Assoc. Res. Otolaryngol., **12**, 605-616.

Maison, S.F., Usubuchi, H., and Liberman, M.C. (**2013**). “Efferent feedback minimizes cochlear neuropathy from moderate noise exposure,” J. Neurosci., **33**, 5542-5552.

Micheyl, C., Khalfa, S., Perrot, X., and Collet, L., (**1997**). “Difference in cochlear efferent activity between musicians and non-musicians,” NeuroReport, **8**, 1047-1050.

Palmer, K.T., Griffin, M.J., Syddall, H.E., Davis, A., Pannett, B., and Coggon, D. (**2002**). “Occupational exposure to noise and the attributable burden of hearing difficulties in Great Britain,” Occup. Environ. Med., **59**, 634-639.

Pirilä, T. (**1991**). “Left-right asymmetry in the human response to experimental noise exposure: II. Pre-exposure hearing threshold and temporary threshold shift at 4 kHz frequency;” Acta Otolaryngol., **111**, 861-866.

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.

Spoendlin, H. (**1971**). “Primary structural changes in the organ of corti after acoustic overstimulation,” Acta Otolaryngol., **71**, 166-176.

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.

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.

Predictors of supra-threshold speech-in-noise intelligibility by hearing-impaired listeners

PETER T. JOHANNESEN[1,2], PATRICIA PÉREZ-GONZÁLEZ[1,2], SRIDHAR KALLURI[3], JOSÉ L. BLANCO[1], AND ENRIQUE A. LOPEZ-POVEDA[1,2,4,*]

[1] *Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, Salamanca, Spain*

[2] *Instituto de Investigación Biomédica de Salamanca, Universidad de Salamanca, Salamanca, Spain*

[3] *Starkey Hearing Research Center, Berkeley, CA, USA*

[4] *Departamento de Cirugía, Facultad de Medicina, Universidad de Salamanca, Salamanca, Spain*

The aim was to assess the relative importance of cochlear mechanical dysfunction, temporal processing deficits, and age for hearing-impaired listeners to understand supra-threshold speech in noise backgrounds. 68 hearing-aid candidates took part in the study. Intelligibility was assessed for speech-shaped noise (SSN) and reversed two-talker masker (R2TM) backgrounds. Behavioural estimates of cochlear gain loss and residual compression from a previous study were used as indicators of cochlear mechanical dysfunction. Temporal processing abilities were assessed using frequency modulation detection thresholds. Age, audiometric thresholds, and the difference between audiometric thresholds and cochlear gain loss were also included in the analyses. Stepwise multiple linear regression models of intelligibility were designed to assess the relative importance of the various factors for speech intelligibility. Results showed that (1) cochlear gain loss was unrelated to intelligibility; (2) residual cochlear compression was related to intelligibility in SSN but not in R2TM backgrounds; (3) temporal processing was strongly related to intelligibility in R2TM backgrounds and much less so in SSN backgrounds; (4) age *per se* hindered intelligibility. We conclude that all factors affect speech intelligibility but their relative importance varies across masker backgrounds.

INTRODUCTION

Hearing-impaired (HI) people vary widely in their ability to understand speech in noise backgrounds, even when their audiometric loss is compensated with frequency-specific sound amplification (e.g., Moore, 2007). The present study aimed at shedding some light on the relative importance of cochlear mechanical dysfunction, temporal processing deficits, and age as predictors of this variability.

*Corresponding author: ealopezpoveda@usal.es

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4.

Several explanations have been proposed to account for the ability of HI listeners to understand audible speech in noise backgrounds (reviewed by Lopez-Poveda, 2014). One of them is that HI listeners could suffer from outer hair cell (OHC) loss or dysfunction and this would degrade the representation of the speech spectrum in the mechanical response of the cochlea, particularly in noisy environments, for various reasons. First, OHC dysfunction reduces cochlear frequency selectivity. This can smear the cochlear representation of the acoustic spectrum, making it harder for HI listeners to separately perceive the spectral cues of speech from those of interfering sounds. Second, in the healthy cochlea, suppression might facilitate the encoding of speech in noise by enhancing the most salient frequency features of the target speech against those of the background noise. OHC dysfunction reduces suppression and this might hinder speech-in-noise intelligibility. Third, cochlear mechanical compression might facilitate the understanding of speech in fluctuating noise by amplifying the speech in the silent noise intervals, a phenomenon known as 'listening in the dips'. OHC loss or dysfunction reduces compression (i.e., linearizes cochlear responses) and thus could hinder dip listening (Gregan *et al.*, 2013).

The view that OHC dysfunction accounts for the ability of HI listeners to understand audible speech in noise is almost certainly only partially correct. First, for HI listeners, there appears to be no significant correlation between residual cochlear compression and the benefit from 'dip listening' (Gregan *et al.*, 2013), which undermines the role of compression on the intelligibility of supra-threshold speech in noise backgrounds. Second, at high intensities, cochlear tuning is comparable for healthy and impaired cochleae and yet HI listeners still perform more poorly than do normal hearing (NH) listeners in speech-in-noise intelligibility tests (reviewed in pp. 205–208 of Moore, 2007). Third, elderly listeners with normal audiometric thresholds and presumably healthy OHCs often have difficulty understanding speech in noise (CHABA, 1988), which suggests that age *per se* or mechanisms other than OHC dysfunction can limit the intelligibility of audible speech.

Another explanation for the ability of HI listeners to understand speech in noise is that HI listeners may suffer from temporal processing deficits. This view would be consistent with the reported correlation between the reduced speech-in-noise intelligibility of HI listeners and their reduced ability to use the information conveyed in the rapid temporal changes of speech sounds, known as 'temporal fine structure' (Lorenzi *et al.*, 2006; Strelcyk and Dau, 2009). It would also be consistent with evidence that temporally jittering the frequency components in speech, as might occur after auditory neuropathy (Pichora-Fuller *et al.*, 2007), or stochastic undersampling of a noisy speech waveform, as might occur after synaptopathy (Lopez-Poveda and Barrios, 2013), both decrease speech-in-noise intelligibility with negligible reductions in audibility.

The present study aimed at assessing the relative contribution of cochlear mechanical dysfunction, temporal processing deficits, and age to the performance of HI listeners understanding audible speech in noisy environments.

MATERIAL AND METHODS

Subjects

The same 68 subjects (43 males) with symmetrical sensorineural hearing losses of the study of Johannesen *et al.* (2014) participated in the present study. Speech-in-noise intelligibility was assessed in bilaterally listening conditions (see below). Indicators of cochlear mechanical status and temporal processing ability, however, were measured in one ear only. For most cases, the test ear was the ear with better audiometric thresholds in the 2-6 kHz frequency range (30 left ears, 38 right ears).

Indicators of cochlear mechanical dysfunction

OHC dysfunction linearizes cochlear mechanical responses. Johannesen *et al.* (2014) compared behaviourally inferred cochlear input/output curves for each HI listener at each one of five test frequencies (0.5, 1, 2, 4, and 6 kHz) with corresponding reference input/output curves for NH listeners. They reported three main variables from their analyses. One variable was cochlear mechanical gain loss (HL_{OHC} in dB). It was defined as the contribution of cochlear gain loss to absolute thresholds and was calculated as the difference sound level required for a pure tone at the test frequency to evoke identical mechanical responses in the cochlea of a HI and a NH listener at absolute threshold (see also Lopez-Poveda and Johannesen, 2012). A second variable was inner hair cell (IHC) loss or HL_{IHC}. It was defined as the difference (in dB) between the pure tone threshold (PTT in dB HL) and HL_{OHC}. This difference was reported after earlier studies where the audiometric loss was assumed to be the sum of a cochlear mechanical component, HL_{OHC}, and an additional component of an uncertain nature conveniently termed HL_{IHC} (Moore and Glasberg, 1997). A third variable was the basilar-membrane compression exponent (BMCE). It was defined as the slope (in dB/dB) of an inferred cochlear input/output curve over its compressive segment. See Johannesen *et al.* (2014) for further details.

PTT, HL_{OHC}, HL_{IHC}, and BMCE were taken from Johannesen *et al.* (2014) and were all considered potential predictors of speech-in-noise intelligibility. Note that the four variables had values at each of the five test frequencies.

Johannesen *et al.* (2014) reported that they could not measure input/output curves for listeners and test frequencies where the audiometric loss was too high. Here, these cases were assumed to be indicative of total cochlear gain loss and HL_{OHC} was set equal to the cochlear gain observed for NH listeners (for details, see p. 11 of Johannesen *et al.*, 2014) and BMCE was set equal to 1 dB/dB.

Frequency modulation detection thresholds

Temporal processing ability was assessed using frequency modulation detection thresholds (FMDTs). The experiment was identical to that of Strelcyk and Dau (2009). In short, an FMDT was defined as the minimum detectable excursion in frequency for a tone carrier and was estimated using a three alternative forced choice procedure. In each trial, the three intervals contained a 1500-Hz pure tone with a level of 30 dB above the detection threshold for the tone. The tones in all intervals were amplitude-modulated (AM) with a modulation depth of 6 dB and a time-varying modulation rate. In the target interval (selected at random), the tone's frequency was varied with a rate of 2 Hz and with a maximum frequency excursion. The logarithm of the maximum frequency excursion was varied in successive trials according to an adaptive one-up two-down rule to estimate the 71% point on the psychometric function (Levitt, 1971). Three FMDTs estimates were obtained and their mean was taken as the threshold.

Speech reception thresholds

Speech-in-noise intelligibility was quantified using the speech reception threshold (SRT), defined as the speech-to-noise ratio (SNR) required to understand 50% of the sentences and was measured using the hearing-in-noise test (HINT) (Nilsson *et al.*, 1994). The background noise was either a steady speech-shaped noise (SSN) or a masker that consisted of two simultaneous talkers (one male and one female) played in reverse (reversed two-talker masker, R2TM). The corresponding SRTs are referred to as SRT_{SSN} and SRT_{R2TM}, respectively.

To measure an SRT, the speech was fixed in level to a nominal value of 65 dB SPL and the masker level was varied adaptively using a one-up, one-down rule. After setting the levels of the speech and the masker, the two sounds were mixed digitally and filtered to simulate a free-field listening condition were the speech and the masker would be co-located one meter in front of the listener at eye level (Table 3 in ANSI, 1997). The resulting stimulus was linearly amplified individually for each participant according to the NAL-R rule (Byrne and Dillon, 1986) to account for the potential effect of the audiometric loss on intelligibility. The amplified stimulus was played diotically to the listeners. All other details of the procedure were as in the original HINT test (Nilsson *et al.*, 1994).

Stimuli and apparatus

For all measurements, stimuli were digitally generated or stored as digital files with a sampling rate of 44100 Hz. They were digital-to-analogue converted using an RME Fireface 400 sound card with a 24-bit resolution, and were played through Sennheiser HD580 headphones. Data were collected in a sound attenuation booth.

Statistical analyses

Pairwise Pearson correlations were first sought between each of the independent variables (PTT, HL_{OHC}, HL_{IHC}, BMCE, FMDT, and age) and each of the dependent variables (aided SRT_{SSN} and SRT_{R2TM}). Prior to the correlation analysis, variables with values at different frequencies were combined into a single value by weighting the value at each test frequency according to the frequency's importance for speech perception (ANSI, 1997) and summing the weighted values across frequencies.

Multiple linear regression (MLR) models were constructed for SRT_{SSN} and SRT_{R2TM} independently to assess the relative importance of the potential predictors for intelligibility. Sometimes several potential predictors might reflect a common underlying factor, a phenomenon known as co-linearity. To minimize the impact of co-linearity, MLR models were constructed in a stepwise fashion (i.e., by gradually adding new potential predictors to the model in each step). The final model omits co-linear variables.

RESULTS

Raw data

The mean absolute thresholds across listeners for the test ears were 37, 44, 51, 61, and 75 dB HL at 0.5, 1, 2, 4, and 6 kHz, respectively (see also Fig. 1 in Johannesen *et al.*, 2014). The standard deviations were in the range 11 to 20 dB HL across frequencies. High-frequency losses were more frequent than other types of losses.

The listeners' ages ranged from 25 to 82 years, with a mean and a standard deviation of 62 and 14 years, respectively. The 5%, 25%, 50%, 75%, and 95% percentiles of age were 38, 54, 61, 74, and 81 years, respectively.

For most listeners, SRT_{SSN} were in the range −5 to 1 dB SNR, thus in line with values reported by earlier studies for SSN maskers (Peters *et al.*, 1998; George *et al.*, 2006; Gregan *et al.*, 2013). SRT_{R2TM} values were in the range −2 to 5 dB SNR and generally higher than SRT_{SSN} values. This trend and range of values are consistent with the −4 to 2 dB range reported by Festen and Plomp (1990) for SRT_{R2TM}. The present SRT_{R2TM} values were about 3, 5, and 5 dB higher than the SRTs for interrupted or modulated-noise backgrounds reported by George *et al.* (2006), Peters *et al.* (1998), and Gregan *et al.* (2013), respectively. This shows that SRTs can be different for different types of fluctuating maskers.

FMDTs for the present participants were in the range 0.7 to 2, in units of $\log_{10}$(Hz), and thus similar to the range of values reported by Strelcyk and Dau (2009) (0.7 to 1.7, when converted to the present units).

Pairwise Pearson's correlations

Table 1 shows squared Pearson's correlation coefficients (R^2 values) for pairs of variables. HL_{OHC} and HL_{IHC} were significantly correlated with PTT but were uncorrelated with each other. This supports the idea that the people with similar audiometric losses can suffer from different degrees of mechanical cochlear gain loss (e.g., Plack *et al.*, 2004; Lopez-Poveda and Johannesen, 2012).

BMCE was positively correlated with PTT and HL_{OHC}, a result indicative that the greater the audiometric loss or the loss of cochlear gain, the more linear (less compressive) the cochlear input/output curves. The positive correlation between BMCE and PTT appears inconsistent with the study of Johannesen *et al.* (2014) that, based on the same data, reported no correlation between those two variables. Differences in the data analyses might explain this discrepancy. First, the cited studies based their conclusions on frequency-by-frequency correlation analyses whereas the present result is based on across-frequency weighted averages. Second, BMCE was set here to 1 dB/dB whenever the audiometric loss was so high that a corresponding

			PTT	HL_{OHC}	HL_{IHC}	BMCE	FMDT	SRT_{SSN}	SRT_{R2TM}
Age	years	R^2	0.01	0.02	0.00	**0.08**	0.00	**0.07**	**0.06**
		p	0.48	0.28	0.63	**0.024**	0.57	**0.032**	**0.039**
PTT	dB HL	R^2	-	**0.63**	**0.30**	**0.13**	0.03	**0.14**	**0.17**
		p	0.088	**10^{-15}**	**$1.4\cdot10^{-6}$**	**0.002**	0.13	**0.00144**	**0.00042**
HL_{OHC}	dB	R^2	-	-	0.01	**0.34**	**0.06**	**0.12**	**0.16**
		p	-	0.25	0.51	**$2.4\cdot10^{-7}$**	**0.04**	**0.0046**	**0.00077**
HL_{IHC}	dB	R^2	-	-	**-**	0.00	0.02	**0.10**	**0.08**
		p	-	-	**0.031**	0.90	0.31	**0.0102**	**0.023**
BMCE	dB/dB	R^2	-	-	-	**-**	0.00	**0.29**	**0.07**
		p	-	-	-	**0.0096**	0.81	**$3.1\cdot10^{-6}$**	**0.035**
FMDT	$\log_{10}$(Hz)	R^2	-	-	-	-	-	**0.07**	**0.28**
		p	-	-	-	-	0.26	**0.028**	**$3.4\cdot10^{-6}$**
SRT_{SSN}	dB SNR	R^2	-	-	-	-	-	-	**0.51**
		p	-	-	-	-	-	0.17	**$1.05\cdot10^{-11}$**

Table 1. Squared pairwise Pearson correlations (R^2) and significance levels (p) between all potential predictors and aided SRT_{SSN} and SRT_{R2TM}. The p-values in the diagonal indicate the probability for a non-Gaussian distribution of the corresponding variable.

input/output curve could not be measured, something that may have biased and increased the correlation slightly.

Table 1 also shows that FMDTs were not correlated with PTT, HL_{IHC}, or BMCE and were only slightly positively correlated with HL_{OHC}. Furthermore, FMDTs were not correlated with age. This suggests that FMDTs were indeed assessing auditory processing aspects unrelated (or only slightly related) to cochlear mechanical dysfunction or age, as was intended.

In addition, Table 1 shows that SRT_{SSN} and SRT_{R2TM} were significantly and positively correlated with each other. The two SRTs were measured using identical conditions and yet their R^2 (0.51) shows that only 51% of the variance in SRT_{SSN} could be explained by the SRT_{R2TM}. This suggests that different mechanisms and/or deficits mediate speech intelligibility for different masker backgrounds. If the mechanisms or deficits mediating speech intelligibility were identical for the two masker backgrounds, one would expect a higher correlation (higher R^2) between SRT_{SSN} and SRT_{R2TM} than the one found.

Potential predictors of speech-in-noise intelligibility

Table 1 shows that SRT_{SSN} and SRT_{R2TM} were significantly correlated with all of the independent variables and hence in principle they could all be contributing to the measured SRTs. The correlations (Table 1) show that PTT explained slightly more SRT_{R2TM} variance ($R^2 = 0.17$) than SRT_{SSN} ($R^2 = 0.14$) variance. This trend and values are consistent with those reported by Peters *et al.* (1998).

Table 1 suggests that PTT, HL_{OHC}, and HL_{IHC} had only a mild influence on aided SRTs, as the largest amount of variance explained by any of these three predictors on any of the two SRTs was 17% (Table 1). For both SRT_{SSN} and SRT_{R2TM}, HL_{OHC} and HL_{IHC} predicted less variance than the PTT, which suggests that specific knowledge about the proportion of the PTT that is due to cochlear mechanical gain loss (HL_{OHC}) or other uncertain factors (HL_{IHC}) does not provide more information than the PTT alone about supra-threshold speech-in-noise intelligibility deficits.

In addition, Table 1 reveals that BMCE predicted 29% of SRT_{SSN} variance but only 7% of SRT_{R2TM} variance, while FMDTs predicted 28% of the SRT_{R2TM} variance but only 7% of the SRT_{SSN} variance. This suggests that residual cochlear compression could be more important than temporal processing abilities for understanding speech in steady noise backgrounds while temporal processing abilities could be more important for understanding speech in fluctuating-masker backgrounds.

Stepwise multiple linear regression models

Stepwise MLR models for SRT_{SSN} and SRT_{R2TM} are shown in Table 2.

Priority	Predictor	Coefficient	*t*-value	*p*	Adj. accum. R^2
			SRT_{SSN}		
n/a	Intercept	-7.5	-8.0	$3.5 \cdot 10^{-11}$	-
1	BMCE	4.25	5.0	$5.6 \cdot 10^{-6}$	0.28
2	HL_{IHC}	0.097	3.3	0.0017	0.37
3	Age	0.023	2.1	0.038	0.41
4	FMDT	0.90	2.0	0.045	0.44
			SRT_{R2TM}		
n/a	Intercept	-7.1	-5.5	$7.0 \cdot 10^{-7}$	-
1	FMDT	2.24	4.8	$1.25 \cdot 10^{-5}$	0.27
2	PTT	0.061	3.5	0.008	0.38
3	Age	0.032	2.9	0.0047	0.45

Table 2. Stepwise MLR models of aided SRT_{SSN} and SRT_{R2TM}. Columns indicate the predictor's priority order and name, the regression coefficient, the *t*-value, the corresponding probability for a significant contribution *(p)*, and the adjusted accumulated proportion of total variance explained (Adj. accum. R^2), respectively. The priority order is established according to how much the corresponding predictor contributed to the predicted variance.

The top part of Table 2 shows that the most significant predictor of SRT_{SSN} was BMCE, which explained 28% of the SRT_{SSN} variance. Additional predictors were HL_{IHC}, age, and FMDT which contributed an additional 9, 4, and 3% to the predicted variance, respectively. The model predicted a total of 44% of the SRT_{SSN} variance. PTT and HL_{OHC} were not significant additional predictors.

The MLR model for aided SRT_{R2TM} was strikingly different than the model for SRT_{SSN} (compare the top and bottom parts of Table 2). The most significant predictor of SRT_{R2TM} was FMDT, which explained 27% of the SRT_{R2TM} variance. Additional predictors were PTT and age, which contributed an additional 11 and 7% to the model predicted variance, respectively. Altogether, the model accounted for 45% of the SRT_{R2TM} variance. Neither HL_{OHC}, or BMCE, or the HL_{IHC} were significant predictors of SRT_{R2TM}.

The role of audibility

Reduced audibility decreases speech-in-noise intelligibility (e.g., Peters *et al.*, 1998). Although NAL-R amplification was provided, audibility might still have been reduced and could have affected the SRTs. To discard this possibility, we calculated the speech intelligibility index (SII) (ANSI, 1997). The SII indicates the proportion of the speech spectrum that is above the absolute threshold and above the background noise (ANSI, 1997). Here, however, we calculated an SII taking into account only the absolute thresholds, the speech spectrum, and the NAL-R amplification while the background noise was disregarded (i.e., here, the SII informed of the proportion of the speech spectrum that was above absolute threshold). In all other aspects, our SII calculations conformed to ANSI (1997). The rationale behind this approach is that if the full speech spectrum were audible, then performance deficits in a masker background would be due to the presence of the masker (Peters *et al.* 1998) rather than to reduced audibility, and would thus reflect supra-threshold deficits.

For 95% of the participants, the SII values were above 0.52, a value that corresponds to an intelligibility of almost 90% for NH listeners (see, e.g., Fig. 3 in Eisenberg *et al.*, 1998). The high SII values indicate that it is unlikely that audibility affected SRT_{SSN} or SRT_{R2TM}. To further rule out the influence of reduced audibility, new MLR models of SRT_{SSN} and SRT_{R2TM} were explored including the SII as a potential predictor. The resulting models in this case were identical to those of Table 2 and the SII did not become a significant predictor in any of the final MLR models. Therefore, it is unlikely that reduced audibility have influenced the present SRTs.

DISCUSSION

The aim of the present study was to assess the relative importance of cochlear mechanical dysfunction, temporal processing deficits, and age for the ability of HI listeners understanding audible speech in noise backgrounds. The main findings were:

1) For the present sample of HI listeners, age, PTT, BMCE, and FMDTs were virtually uncorrelated with each other (Table 1) and yet they were significant predictors of aided SRT in noise backgrounds (Table 2).
2) Residual cochlear compression (BMCE) was the most important single predictor of aided SRT_{SSN}, while FMDT was the most important single predictor of aided SRT_{R2TM} (Table 2).
3) Cochlear mechanical gain loss (HL_{OHC}) was correlated with aided SRT_{SSN} and SRT_{R2TM} (Table 1) but did not increase the variance explained by the MLR

models of SRT_{SSN} or SRT_{R2TM} once the previously mentioned predictors were included in the models.

4) Age was a significant predictor of SRT_{SSN} and SRT_{R2TM}, and it was independent of FMDTs and virtually independent of BMCE (Table 1).

For the present sample, age, PTT, FMDT, and (virtually) BMCE were uncorrelated with each other. This result was incidental. Given the well-established relationship between age and PTT (reviewed by Gordon-Salant *et al.*, 2010), the absence of a correlation between those two variables was surprising. One possible explanation is that our participants were required to be hearing aid candidates (something necessary for a different aspect of the study not reported here) while having mild-to-moderate audiometric losses in the frequency range from 0.5 to 6 kHz, something necessary to infer HL_{OHC} estimates using behavioural masking methods (Johannesen *et al.*, 2014). Thus, it is possible that their hearing losses spanned a narrower range than would be observed across the same age span in a random sample. Our across-frequency weighted-averaging of audiometric thresholds (see Methods) may have contributed to wash out any correlation between age and PTT.

The absence of a correlation between age or PTTs with FMDTs was unexpected. The number of synapses between IHC and auditory nerve fibres is known to decrease gradually with increasing age, even in cochleae with normal IHC and OHC counts and thus presumably normal PTT (Makary *et al.*, 2011). Insofar as hearing impairment can be caused by noise exposure and noise exposure decreases the number of afferent synapses (Kujawa and Liberman, 2009), hearing impairment is also thought to be associated with a reduced number of synapses. A reduced synapse count (or synaptopathy) is thought to impair auditory temporal processing (Lopez-Poveda and Barrios, 2013). The absence of a correlation between age and FMDTs or between FMDTs and PTT (Table 2) suggests that either our participants did not suffer from synaptopathy (unlikely given the wide age range) or that FMDTs reflect temporal processing abilities not directly (or not solely) related to synaptopathy.

The finding that age, PTT, FMDT, and BMCE are correlated with supra-threshold speech-in-noise intelligibility (Table 1) was expected for the reasons reviewed in the Introduction. A significant though incidental aspect of the present study is, however, that for the present sample those factors were uncorrelated or poorly correlated with each other (Table 1) and yet they affected intelligibility in different proportions for different types of masker backgrounds (Table 2).

The two indicators of cochlear mechanical dysfunction (HL_{OHC} and BMCE) were correlated with speech intelligibility in the two noise backgrounds, and they were correlated with each other (Table 1). However, HL_{OHC} did not remain as a significant predictor of intelligibility in neither of the two masker backgrounds when other variables were included in the MLR models, while BMCE became the most significant predictor of intelligibility only in SSN backgrounds (Table 2). The estimates of cochlear gain loss (HL_{OHC}) and residual compression (BMCE) are indirect and based on numerous assumptions (Johannesen *et al.*, 2014). Assuming that these estimates are reasonable, the present finding suggests that cochlear mechanical gain loss and residual compression are not equivalent predictors of the impact of cochlear mechanical dysfunction on the intelligibility of speech in SSN. The finding further suggests that residual compression is more significant than cochlear gain loss, perhaps because the impact of HL_{OHC} on intelligibility may be compensated for with linear amplification but the impact of BMCE may not.

The importance of compression for understanding supra-threshold speech in SSN appears inconsistent with the findings of Summers *et al.* (2013) who reported that compression was not clearly associated with understanding loud speech (at a fixed level of 92 dB SPL) in a steady noise background. This inconsistency may be partly due to methodological differences across studies. First, Summers *et al.* (2013) assessed intelligibility using the percentage of sentences identified correctly for a fixed SNR rather than the SRT (in dB SNR). Second, Summers *et al.* (2013) reported correlations between intelligibility and estimates of compression at single frequencies while we are reporting correlations between SRTs and across-frequency weighted average of compression. Lastly, Summers *et al.* (2013) did not take into account important precautions regarding inference of compression estimates using the temporal masking curve (TMC) method. This method is based on the assumption that cochlear compression may be inferred from comparisons of the slope of TMCs unaffected by compression (linear references) with that of TMCs affected by compression. Summers *et al.* (2013) used different linear reference TMCs for different test frequencies and their linear references were TMCs for a masker frequency equal to 0.55 times the probe frequency. This almost certainly underestimates compression (e.g., Lopez-Poveda *et al.*, 2003; Lopez-Poveda and Alves-Pinto, 2008), particularly at lower frequencies and for NH listeners, something that might have contributed to 'hiding' differences in compression across listeners with different audiometric thresholds in the data of Summers *et al.* (2013).

Residual compression (BMCE) was the best single predictor of supra-threshold speech intelligibility in a SSN background while FMDT became the most significant predictor in a R2TM background (Table 2). The reason is uncertain, though it seems reasonable that temporal processing ability be more important for intelligibility in fluctuating than in steady masker backgrounds.

CONCLUSIONS

1) Cochlear gain loss is unrelated to understanding audible speech in noise.
2) Residual cochlear compression is related to speech understanding in speech-shaped steady noise but not in reversed two-talker masker backgrounds.
3) Auditory temporal processing ability is strongly related to speech understanding in fluctuating masker backgrounds but has relatively minor importance in a steady noise background.
4) Age hinders the intelligibility of supra-threshold speech in any of the two masker backgrounds tested here, regardless of absolute thresholds, cochlear mechanical dysfunction, or temporal processing deficits.

ACKNOWLEDGEMENTS

We thank William S. Woods, Olaf Strelcyk, and Brent Edwards for useful suggestions. Work supported by Starkey Ltd. (USA), by Junta de Castilla y León and MINECO (ref. BFU2012-39544-C02).

REFERENCES

ANSI (**1997**). *S3.5 Methods for Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).

Byrne D. and Dillon H. (**1986**). "The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid," Ear Hearing, **7**, 257-265.

CHABA (**1988**). "Speech understanding and aging. Working Group on Speech Understanding and Aging. Committee on Hearing, Bioacoustics, and Biomechanics, Commission on Behavioral and Social Sciences and Education, National Research Council," J. Acoust. Soc. Am., **83**, 859-895.

Eisenberg L.S., Dirks D.D., Takayanagi S., and Martinez A.S. (**1998**). "Subjective judgements of clarity and intelligibility for filtered stimuli with equivalent speech intelligibility index predictions," J. Speech Lang. Hear. Res., **41**, 327-339.

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.

George E.L., Festen J.M., and Houtgast T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am., **120**, 2295-2311.

Gordon-Salant, S., Frisina R.D., Popper A.N., and Fay R.R. (**2010**). *The Aging Auditory System* (Springer, New York).

Gregan M.J., Nelson P.B., and Oxenham A.J. (**2013**). "Behavioral measures of cochlear compression and temporal resolution as predictors of speech masking release in hearing-impaired listeners," J. Acoust. Soc. Am., **134**, 2895-2912.

Johannesen P.T., Perez-Gonzalez P., and Lopez-Poveda E.A. (**2014**). "Across-frequency behavioral estimates of the contribution of inner and outer hair cell dysfunction to individualized audiometric loss," Front. Neurosci., **8**, 214.

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.

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

Lopez-Poveda, E.A., Plack, C.J., and Meddis, R. (**2003**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am., **113**, 951-960.

Lopez-Poveda, E.A. and Alves-Pinto A. (**2008**). "A variant temporal-masking-curve method for inferring peripheral auditory compression," J. Acoust. Soc. Am., **123**, 1544-1554.

Lopez-Poveda E.A. and Johannesen, P.T. (**2012**). "Behavioral estimates of the contribution of inner and outer hair cell dysfunction to individualized audiometric loss," J. Assoc. Res. Otolaryngol., **13**, 485-504.
Lopez-Poveda, E.A. and Barrios, P. (**2013**). "Perception of stochastically undersampled sound waveforms: a model of auditory deafferentation," Front. Neurosci., **7**, 124.
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**, 348.
Lorenzi C., Gilbert G., Carn H., Garnier S., and Moore B.C.J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. USA, **103**, 18866-18869.
Makary C.A., Shin, J., Kujawa, S.G., Liberman, M.C., and Merchant, S.N. (**2011**). "Age-related primary cochlear neuronal degeneration in human temporal bones," J. Assoc. Res. Otolaryngol., **12**, 711-717.
Moore, B.C.J. and Glasberg, B.R. (**1997**). "A model of loudness perception applied to cochlear hearing loss," Auditory Neuroscience, **3**, 289-311.
Moore, B.C.J. (**2007**). *Cochlear Hearing Loss* (John Wiley & Sons, Ltd., Chichester).
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.
Peters, R.W., Moore, B.C.J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am., **103**, 577-587.
Pichora-Fuller, M.K., Schneider, B.A., MacDonald, E., Pass, H.E., and Brown, S. (**2007**). "Temporal jitter disrupts speech intelligibility: a simulation of auditory aging," Hear. Res., **223**, 114-121.
Plack, C.J., Drga, V., and Lopez-Poveda, E.A. (**2004**). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am., **115**, 1684-1695.
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.
Summers, V., Makashay, M.J., Theodoroff, S.M., and Leek, M.R. (**2013**). "Suprathreshold auditory processing and speech perception in noise: Hearing-impaired and normal-hearing listeners," J. Am. Acad. Audiol., **24**, 274-292.

Neural modelling to relate individual differences in physiological and perceptual responses with sensorineural hearing loss

MICHAEL G. HEINZ[1,2,*]

[1] *Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN, USA*

[2] *Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA*

A great challenge in diagnosing and treating hearing impairment comes from the fact that people with similar degrees of hearing loss often have different speech-recognition abilities. Many studies of the perceptual consequences of peripheral damage have focused on outer-hair-cell (OHC) effects; however, anatomical and physiological studies suggest that many common forms of sensorineural hearing loss (SNHL) arise from mixed OHC and inner-hair-cell (IHC) dysfunction. Thus, individual differences in perceptual consequences of hearing impairment may be better explained by a more detailed understanding of differential effects of OHC/IHC dysfunction on neural coding of perceptually relevant sounds. Whereas it is difficult experimentally to estimate or control the degree of OHC/IHC dysfunction in individual subjects, computational neural models provide great potential for predicting systematically the complicated physiological effects of combined OHC/IHC dysfunction. Here, important physiological effects in auditory-nerve (AN) responses following different types of SNHL and the ability of current models to capture these effects are reviewed. In addition, a new approach is presented for computing spike-train metrics of speech-in-noise envelope coding to predict how differential physiological effects may contribute to individual differences in speech intelligibility.

INTRODUCTION

In the last 35 years, our knowledge of the physiological aspects of sensorineural hearing loss (SNHL) has expanded tremendously; however, despite these advances, very little physiological knowledge of SNHL goes into the design or fitting of hearing aids today. Although it is often difficult to relate experimentally measured physiological and perceptual findings, computational modelling provides great promise for quantitatively relating physiological and perceptual effects of SNHL in translational applications (Heinz, 2010). In fact, long before the accuracy of sensory models allowed such potential to be realized, a general theoretical framework (Fig. 1) was described for using mathematical models in the development of sensory

*Corresponding author: mheinz@purdue.edu

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

prostheses (Biondi, 1978; Biondi and Schmid, 1972). Fortunately, in the last 20 years advances in the complexity and accuracy of computational models of normal and impaired auditory systems have dramatically improved their potential for use in the quantitative design and fitting of hearing aids (e.g., Bondy *et al.*, 2004).

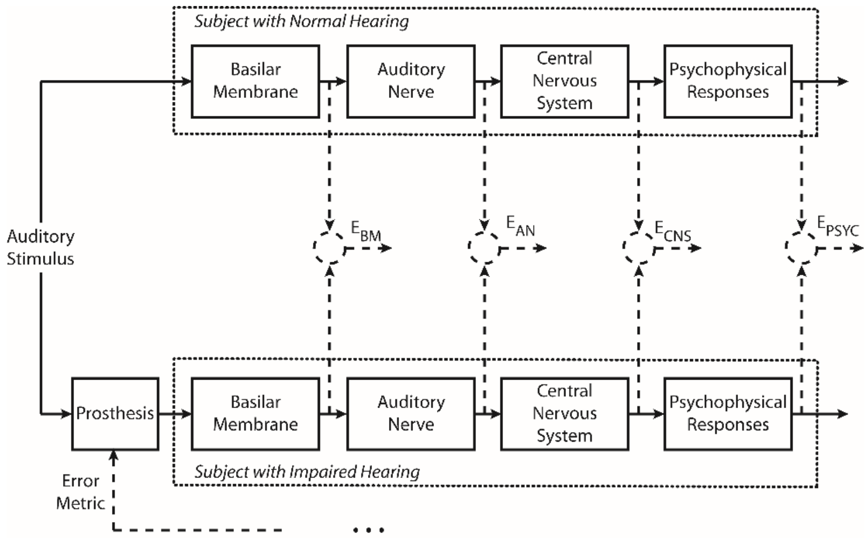


Fig. 1: Modelling framework for the design of auditory prostheses. The goal of the prosthesis is to restore normal auditory responses in a subject with impaired hearing. Computational models allow this goal to be optimally pursued by adjusting the prosthesis to minimize an error metric that quantifies the difference between normal and aided-impaired model responses. Error metrics can be derived based on models of responses at different levels of the auditory system, ranging from basilar-membrane to psychophysical responses. Modified and extended from Biondi (1978).

Current audiological diagnoses classify all types of SNHL into a single category, despite clear individual differences within this one category (e.g., different speech recognition among patients with similar audiograms). It has long been believed (and perhaps still is in some places) that mild-moderate SNHL is primarily outer-hair-cell (OHC) based (with degraded frequency selectivity responsible for difficulty understanding speech), and that inner-hair-cell (IHC) effects only play a role in cases where threshold shifts are greater than 60 dB (e.g., Edwards, 2004; Moore, 1995). In fact, much insight into perceptual effects of SNHL has been derived from considering the effects of OHC dysfunction on basilar-membrane responses (Moore, 1995; Oxenham and Bacon, 2003). However, anatomical and physiological evidence suggests that many common forms of SNHL are likely to involve mixed OHC/IHC dysfunction, and that IHC dysfunction can significantly affect perceptually relevant response properties in the auditory nerve (AN) related to intensity and speech coding. Thus, applications of computational models to account for sources of individual physiological differences in SNHL require modelling at the AN (rather than basilar membrane).

PHYSIOLOGICAL EFFECTS OF SNHL COMMONLY DERIVE FROM A MIXTURE OF IHC AND OHC DYSFUNCTION

Liberman and Dodds (1984) established a strong correlation between AN tuning-curve shapes following cochlear damage and the status of the underlying hair cells. Hair-cell stereocilia condition provided a much stronger correlation with threshold shift than did hair-cell survival. Unlike the well-defined “tip” and “tail” region of normal AN tuning curves, significant OHC stereocilia loss produced reduced tip sensitivity and broadened tuning, whereas damage to IHC stereocilia produced threshold elevations at all frequencies with little effect on frequency selectivity. In contrast to the longstanding belief that OHC dysfunction is the primary correlate of mild-moderate SNHL, anatomical evidence from these noise-induced hearing loss studies showed major overlap in the cochlear regions with OHC and IHC stereocilia damage, and in fact often showed broader regions of IHC stereocilia damage (see Figs. 4, 5, 7, 8, and 9 in Liberman and Dodds, 1984).

In addition to noise-induced hearing loss, age-related hearing loss (or presbycusis) is also likely to include a mixture of IHC/OHC dysfunction. Schmiedt *et al.* (2002) demonstrated that young gerbils with furosemide-induced endocochlear-potential (EP) reductions showed physiological audiograms that matched those of aged gerbils (with similar reductions in EP). Furthermore, these audiograms showed the typical sloping high-frequency hearing loss characteristic of age-related hearing loss in humans. Reductions in EP have been shown to produce physiological AN responses (e.g., broadened tuning, reduced spontaneous and driven firing rates) consistent with mixed OHC/IHC dysfunction (Sewell, 1984). A mixed hair-cell loss fits with the view that the EP provides the battery that drives transduction in both types of hair cells. In summary, the available anatomical and physiological evidence suggests that many common forms of SNHL (e.g., noise and age) involve mixed OHC/IHC dysfunction.

MODEL REQUIREMENTS TO RELATE INDIVIDUAL DIFFERENCES IN PHYSIOLOGICAL AND PERCEPTUAL EFFECTS OF SNHL

The combined (and sometimes confounding) effects of OHC and IHC dysfunction in the same cochlear frequency region are likely to be quite complicated. Computational neural models provide great potential for predicting systematically the complicated physiological effects of combined OHC/IHC dysfunction. Based on anatomical and physiological knowledge of peripheral effects of SNHL, general requirements are now clear for computational modelling approaches to relate individual differences in physiological and perceptual responses with SNHL: 1) inclusion of both OHC and IHC dysfunction, since each is likely to occur in common forms of SNHL (e.g., age and noise); 2) ability to predict responses to arbitrary complex signals, since deficits often occur in complex listening situations (e.g., speech in noise); 3) accurate representation of temporal responses (both rapid and slow), since both timescales are likely to be perceptually relevant in many tasks for which listeners with SNHL have particular difficulty (e.g., source segregation, speech intelligibility in real-world background noises); 4) ability to evaluate cochlear synaptopathy, since a reduction in the number of IHC synapses occurs with age and moderate noise exposure, even in cases without permanent threshold shift (Kujawa and Liberman, 2015); and 5) ability to relate spike-time responses to perceptually relevant metrics, since responses at the level of the AN are required to capture the known physiological SNHL effects.

Computational models now exist that incorporate the salient response properties (both rate and timing) that are important for modelling individual differences in OHC/IHC dysfunction (reviewed by Heinz, 2010), and numerous modelling frameworks exist for relating physiological and perceptual responses (e.g., Elhilali *et al.*, 2003; Heinz *et al.*, 2001; Hines and Harte, 2012). This chapter will focus on a well-established AN model (Bruce *et al.*, 2003; Carney, 1993; Zilany *et al.*, 2009; 2014) to review how various physiological effects of OHC/IHC dysfunction can be accounted for in computational models. This model accounts for a wide range of response properties measured from both normal and hearing-impaired animals for a wide range of stimuli (e.g., tones, noise, and speech). The model takes as input an arbitrary acoustic waveform and produces an output of spike times for a single AN fibre with a specified characteristic frequency (CF) (see Fig. 2 in Zilany *et al.*, 2009). The model allows independent control of OHC and IHC function through two parameters, C_{OHC} and C_{IHC}, ranging from 1 (normal) to 0 (fully dysfunctional).

Modelling OHC dysfunction

Damage to OHCs has been shown to result in numerous correlated effects: increased thresholds (reduced cochlear gain at CF), broadened tuning, reduced cochlear compression, reduced two-tone suppression, and reduced level dependence in phase responses. Each of these properties is believed to be associated with a single mechanism (sometimes called the cochlear amplifier), for which OHCs play a major role. Thus, a key insight into modelling the effects of OHC dysfunction is to include a single signal-processing mechanism that accounts for all of these effects together (Carney, 1993; Kates, 1991; Patuzzi, 1996), and for which the effects of OHC dysfunction can be included in a single-parameter fashion (Bruce *et al.*, 2003; for review, see Heinz, 2010). By modelling OHC dysfunction as a single parameter that controls the maximum cochlear gain at low sound levels, partial OHC damage reduces each of these nonlinear properties by an amount that is proportional to the fractional reduction in cochlear gain (see Fig. 4 in Bruce *et al.*, 2003). While insight can be gained from simpler models that isolate some of these effects, such models are limited in their generality for complex stimuli, which are likely to be critical for SNHL model applications such as hearing-aid design.

Modelling IHC dysfunction

IHC damage has often been thought of primarily in terms of complete IHC loss (i.e., dead regions); however, it is clear that dysfunction of remaining IHCs can have significant effects on perceptually relevant neural responses. Although moderate IHC dysfunction does not significantly affect tuning, there are other consistent effects on

AN response properties following IHC damage, including elevated thresholds, reduction in spontaneous and driven rates, and reduction in the slopes of rate-level functions (Liberman and Dodds, 1984; Liberman and Kiang, 1984; Wang *et al.*, 1997). It is typically believed that AN rate-level functions are steeper following SNHL, consistent with the effects of loudness recruitment; however, this has been shown not to be a consistent effect (Harrison, 1981; Heinz and Young, 2004; see Fig. 2A). IHC dysfunction was hypothesized to be responsible for reducing the rate of response growth based on one of two mechanisms. The first is that damage to IHC stereocilia may reduce the number of transduction channels that can open, resulting in a reduced maximum transduction current and a reduced response-growth slope. Second, if OHC function remains normal, IHC damage may elevate thresholds enough so that the AN-fibre rate-level function is made shallower by remaining BM compression. Thus, IHC damage can confound the effects of OHC damage on basic response properties, such as response growth with level (Fig. 2B).

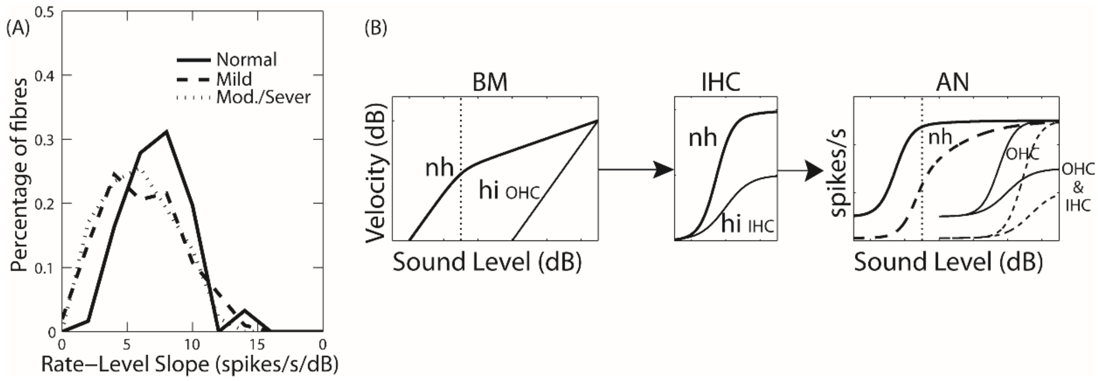


Fig. 2: (A) CF-tone AN rate-level functions were shallower than normal following noise-induced hearing loss. (B) Damage to the IHC transducer can produce shallower AN response growth despite steeper BM responses due to OHC damage. Modified from Heinz and Young (2004), Heinz *et al.* (2005).

IHC dysfunction was modelled by Bruce *et al.* (2003) with a shallower input/output transduction function, under the control of C_{IHC}. The IHC module with SNHL accounted for the wide range of rate-level functions observed experimentally (Heinz and Young, 2004; see Fig. 6 in Zilany and Bruce, 2006), and contributed to reduced synchrony capture in vowel responses (Bruce *et al.*, 2003; Miller *et al.*, 1997).

Further insight into the effects of IHC dysfunction comes from studies in chinchillas, where the platinum-based chemotherapy drug carboplatin specifically affects IHCs while leaving OHCs intact. While dose-dependent IHC loss is observed along the cochlea following carboplatin, structural damage has also been observed (e.g., to stereocilia) in remaining IHCs (Wake *et al.*, 1994), with reduced spontaneous and driven discharge rates despite normal tuning (Wang *et al.*, 1997). Although remaining AN fibres show normal temporal coding to modulated tones (quantified by average vector strength), fewer spikes in response to sound alters the response statistics and degrades predicted detection/discriminability when both response mean and variance are considered (i.e., in terms of a *d'* metric; Axe and Heinz, 2015). Thus, reduced spike rates from IHC dysfunction may be perceptually relevant due to reduced information from less reliable responses.

Challenges in modelling mixed hair-cell dysfunction

While this signal-processing filter-bank model approach allows for implementation of an IHC transduction function that accounts for the observed range of rate-level shapes (Zilany and Bruce, 2006), it does not currently account for the reduction in spontaneous rate observed with carboplatin toxicity, noise induced hearing loss, and metabolic age-related hearing loss. Because this reduced spike count may be perceptually relevant, this is likely to be an important component to include in future models of IHC dysfunction. Also, the current independent control of OHC/IHC dysfunction does not capture directly the effects of metabolic hearing losses associated with presbycusis (Schmiedt *et al.*, 2002), where OHC and IHC function are both dependent on the same EP "battery". More biophysically based models with EP control of both OHC and IHC function have been shown to capture the main effects of metabolic presbycusis in a more physiologically constrained approach (Saremi and Stenfelt, 2013), but this is not currently possible with the phenomenological signal-processing model approach of Zilany *et al.* (2009).

COMPUTING PERCEPTUAL METRICS FROM SPIKE-TIME RESPONSES

The influence of inherent fluctuations on speech intelligibility: SNR_{ENV}

Recent psychophysically based modelling has demonstrated that the signal-to-noise ratio (SNR_{ENV}) at the output of a modulation filter bank provides a robust measure of speech intelligibility (Jørgensen and Dau, 2011). The effect of the noise (N) on speech (S) coding is assumed to: 1) reduce envelope power of S+N by filling in the dips of clean speech, and 2) introduce a noise floor due to intrinsic fluctuations in the noise itself. Changes in the SNR_{ENV} metric with acoustic processing/distortion can be related to a change in speech reception threshold (SRT). An ideal-observer framework is used to convert SNR_{ENV} to percent correct. The central hypothesis of this modelling framework is that the predicted change in intelligibility arises because the processing (or in this case SNHL) changes the input (acoustic) SNR needed to obtain the SNR_{ENV} corresponding to a given percent correct. SNR_{ENV} predicted speech intelligibility across a wider range of degraded conditions than many long-standing speech-intelligibility models (e.g., STI). Key insight into the effect of spectral subtraction on speech intelligibility was garnered by consideration of the modulation-domain SNR, which factors in the inherent fluctuations within the noise. Although spectral subtraction increased the envelope power in the noisy-speech (leading STI-based metrics to predict improvements), it also increased the envelope power in the noise-alone response to a greater degree such that SNR_{ENV} decreased, consistent with the observed performance degradation.

Extending the envelope power spectrum model (SNR_{ENV}) to neural responses

While the promise of the SNR_{ENV} metric has been demonstrated for normal-hearing listeners (Jørgensen and Dau, 2011), it has yet to be thoroughly extended to hearing-impaired listeners because of limitations in our physiological knowledge of how SNHL affects the envelope coding of speech in noise relative to noise alone. Here, envelope coding to non-periodic stimuli (e.g., speech in noise) was quantified from model neural spike trains using shuffled-correlogram analyses, which were analysed in the modulation frequency domain to compute modulation-band based estimates of signal and noise envelope coding (e.g., a neural SNR_{ENV} metric).

Neural spike-train responses were obtained from the most recent version of the AN model (Zilany *et al.*, 2014), with responses from medium-spontaneous-rate fibres considered here. Figure 3 shows the single sentence considered in this initial study, along with predicted AN-fibre discharge rate waveforms for a CF=1 kHz fibre to clean speech, noisy speech, and noise alone. The noise used was broadband Gaussian noise, spectrally matched to the sentence. Strong speech modulations are seen in the clean-speech response, with only the largest and slowest modulations apparent in the noisy-speech response. Inherent fluctuations in the noise-alone response are seen, which are an important factor in the envelope power spectrum model analyses (Jørgensen and Dau, 2011). Shuffled correlogram analyses were used to quantify envelope coding in each condition (Louage *et al.*, 2004; Swaminathan and Heinz, 2011). By averaging correlograms from positive- and negative-polarity versions of each stimulus, the *sumcor* (Fig. 3B top) quantifies the temporal envelope coding in terms of an autocorrelation function, whereas the Fourier transform of the sumcor estimates the envelope power-spectral density (Fig. 3B bottom). As in the envelope power spectrum model analyses, the SNR_{ENV} metric was computed for each fibre CF and modulation filter band by computing the ratio of the response envelope power for speech (estimated as the envelope power to noisy speech (S+N) minus the envelope power to noise alone) divided by the envelope power to noise alone (see Eqs. 2 and 4 in Jørgensen and Dau, 2011). Here, envelope power was computed within seven modulation-frequency bands (Fig. 3C) by integrating the envelope power spectral density within different modulation-frequency ranges (a low-pass range at and below 1 Hz, and six octave-spaced bands centred at 2 to 64 Hz with a bandwidth equal to the centre frequency, i.e., Q=1). Although not implemented directly as modulation filters, these seven modulation bands correspond closely to the seven original modulation bands (Jørgensen and Dau, 2011). A total SNR_{ENV} (see Fig. 4) was computed by combining the individual SNR_{ENV} values from each modulation band and each acoustic filter (as in Eq. 6 of Jørgensen and Dau, 2011). In this initial study, four AN-fibre CFs were used (0.5, 1, 2, and 4 kHz) to compute SNR_{ENV} as a function of acoustic SNR (Fig. 4) for several versions of the AN model (normal hearing, 30-dB OHC loss, and mixed 15-dB/15-dB IHC/OHC loss). For SNHL model versions, a speech level of 80 dB SPL was used so that all comparisons were at equal sensation level (SL).

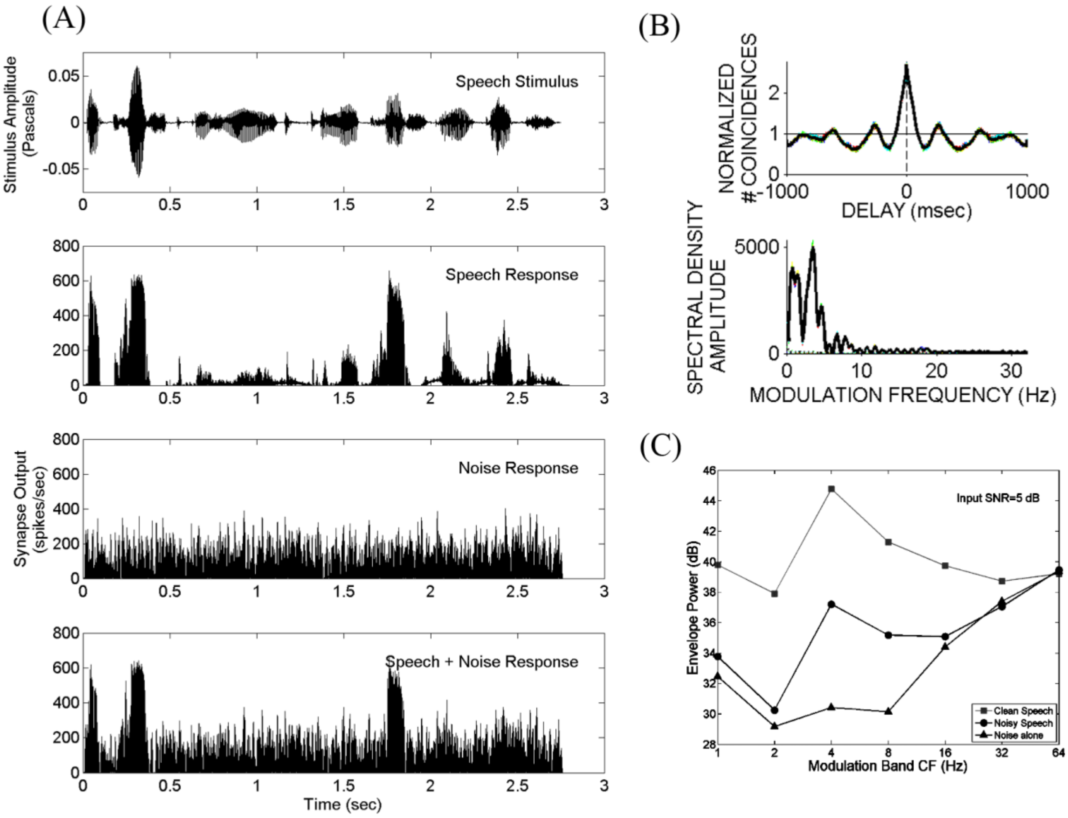


Fig. 3: Extending the envelope power spectrum model analysis of *SNR_ENV* (Jørgensen and Dau, 2011) to neural spike-train responses. (A) One speech sentence with overall sound level of 50 dB SPL (best modulation level for this fibre; top row) was presented to a medium-spontaneous-rate model AN fibre with CF = 1 kHz (2nd row). Noise-alone (3rd row) and noisy-speech responses (bottom row) are shown for a 5-dB acoustic SNR condition. (B) The shuffled-correlogram *sumcor* (top) was used to quantify temporal envelope coding in each response, with the *envelope power spectral density* estimated as the Fourier transform of the sumcor (bottom). (C) Envelope power as a function of modulation-band centre frequency computed from the envelope spectral density for clean speech, noisy speech, and noise alone.

Preliminary predictions of the effect of individual differences on SNR_{ENV}

Overall, many aspects of the SNR_{ENV} predictions computed here from neural spike trains showed close similarities to the psychoacoustical model predictions motivating this work (Jørgensen and Dau, 2011). 1) Envelope power excitation patterns (e.g., Figs. 3C and 5) showed the same relative position across conditions, with the highest envelope power for clean speech, the lowest envelope power for noise alone, and noisy speech in between. 2) The peak in speech envelope power was observed in the

4-Hz modulation band. 3) Negligible differences were predicted between noisy-speech and noise-alone envelope power (i.e., zero SNR_{ENV}) for the 16-64 Hz modulation bands. 4) The total SNR_{ENV} varied from about 2 dB to 14 dB as acoustic (input) SNR varied from −9 to 9 dB, with these neural values being above the neural noise floor in all conditions.

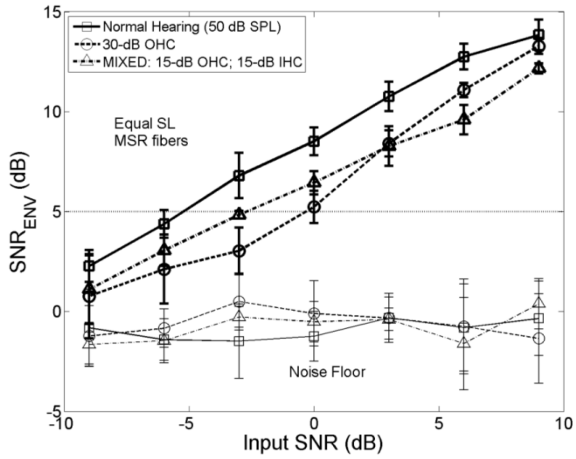


Fig. 4: Individual differences in speech intelligibility are predicted with varying degrees of OHC/IHC dysfunction. Neural-based predictions of total SNR_{ENV} are shown as a function of input (acoustic) SNR for three versions of the AN model that varied in OHC/IHC dysfunction. All comparisons were made at equal SL, using medium-spontaneous-rate (MSR) fibres. A neural noise floor is shown based on randomized spike times.

Predictions of total SNR_{ENV} as a function of acoustic SNR (Fig. 4) varied across the three AN-model versions with different degrees of OHC/IHC dysfunction. Both SNHL versions predicted reduced SNR_{ENV} relative to normal hearing, with a ~5-dB acoustic (input) SNR loss (estimated at SNR_{ENV} = 5 dB) for 30-dB OHC dysfunction and ~2.5-dB SNR loss for 30-dB mixed OHC/IHC dysfunction. A cross-over between OHC and mixed predictions was observed at 3-dB input SNR, with OHC dysfunction predictions nearly matching normal hearing at the highest SNR. The cross-over is due to a greater SNR loss (rightward shift) in the OHC function for low SNRs (e.g., more noise through broad filters) and less SNR loss for cleaner speech.

This spike-train approach allows the exploration of individual differences in the modulation domain for noisy-speech encoding. For these 30-dB hearing losses and equal-SL comparisons, there was (Fig. 5): 1) less difference across model versions for clean speech, 2) reduced envelope power for noisy speech, 3) reduced (but less so than noisy speech) envelope power for noise alone (intrinsic fluctuations), 4) resulting overall reductions in SNR_{ENV} (from 2nd and 3rd points), and 5) differences predicted when IHC dysfunction (i.e., shallower transduction) was included.

SUMMARY AND IMPLICATIONS

Modelling at the AN level is required to include the physiological factors known to influence neural coding of complex sounds: 1) OHC dysfunction, 2) IHC dysfunction, 3) IHC loss (dead regions), 4) EP reduction in presbycusis, 5) cochlear synaptopathy. Preliminary spike-train analyses show strong similarities to the speech envelope power spectrum model of Jørgensen and Dau (2011), which has shown the importance of SNR_{ENV} for predicting speech intelligibility across a wide range of processing conditions. While these preliminary neural predictions are shown here primarily to demonstrate the feasibility of neural SNR_{ENV} computations from spike-train responses, the cross-over in Fig. 4 suggests that individual differences may occur based on differential degrees of OHC/IHC dysfunction in listeners currently diagnosed into the single category of SNHL. These neural computations will be applied in future animal studies to quantify the effects of various types of SNHL on coding of speech and inherent noise modulations, which may provide valuable insight for understanding individual differences in speech-in-noise intelligibility.

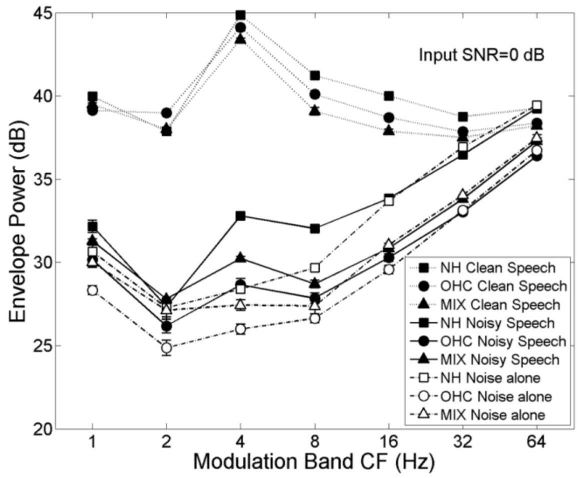


Fig. 5: Envelope power as a function of modulation-band centre frequency for clean-speech, noisy-speech, and noise-alone responses from AN-model versions that varied in OHC/IHC dysfunction (as in Fig. 4).

ACKNOWLEDGMENTS

Varsha Rallapalli assisted with the neural SNR_{ENV} computations. Søren Jørgensen, Torsten Dau, and Christoph Scheidiger are thanked for discussions on the speech envelope power spectrum model. This work was supported by an NIH grant (R01-DC009838) and an International Project Grant from Action on Hearing Loss.

REFERENCES

Axe, D.R. and Heinz, M.G. (**2015**). "The effects of carboplatin-induced ototoxicity on temporal coding in the auditory nerve," Assoc. Res. Otolaryngol. Abstr., **38**, 76-77.

Biondi, E. and Schmid, R. (**1972**). "Mathematical models and prostheses for sense organs," in *Theory and Applications of Variable Structure Systems*. Eds. R.R. Mohler and A. Ruberti (Academic Press, London), pp. 183-211.

Biondi, E. (**1978**). "Auditory processing of speech and its implications with respect to prosthetic rehabilitation. The bioengineering viewpoint," Audiology, **17**, 43-50.

Bondy, J., Becker, S., Bruce, I.C., Trainor, L., and Haykin, S. (**2004**). "A novel signal-processing strategy for hearing-aid design: Neurocompensation," Signal Process., **84**, 1239-1253.

Bruce, I.C., Sachs, M.B., and Young, E.D. (**2003**). "An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses," J. Acoust. Soc. Am., **113**, 369-388.

Carney, L.H. (**1993**). "A model for the responses of low-frequency auditory-nerve fibers in cat," J. Acoust. Soc. Am., **93**, 401-417.

Edwards, B. (**2004**). "Hearing aids and hearing impairment," in *Speech Processing in the Auditory System*. Ed. S. Greenberg, W.A. Ainsworth, A.N. Popper, and R.R. Fay (Springer, New York), pp. 339-421.

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.

Harrison, R.V. (**1981**). "Rate-versus-intensity functions and related AP responses in normal and pathological guinea pig and human cochleas," J. Acoust. Soc. Am., **70**, 1036-1044.

Heinz, M.G., Colburn, H.S., and Carney, L.H. (**2001**). "Evaluating auditory performance limits: I One-parameter discrimination using a computational model for the auditory nerve," Neural Comput., **13**, 2273-2316.

Heinz, M.G. and Young, E.D. (**2004**). "Response growth with sound level in auditory-nerve fibers after noise-induced hearing loss," J. Neurophysiol., **91**, 784-795.

Heinz, M.G., Issa, J.B., and Young, E.D. (**2005**). "Auditory-nerve rate responses are inconsistent with common hypotheses for the neural correlates of loudness recruitment," J. Assoc. Res. Otolaryngol., **6**, 91-105.

Heinz, M.G. (**2010**). "Computational Modeling of Sensorineural Hearing Loss," in *Computational Models of the Auditory System*, Eds. R. Meddis, E.A. Lopez-Poveda, A.N. Popper, and R.R. Fay (Springer, New York), pp. 177-202.

Hines, A. and Harte, N. (**2012**). "Speech intelligibility prediction using a Neurogram Similarity Index Measure," Speech Commun., **54**, 306-320.

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.

Kates, J.M. (**1991**). "A time-domain digital cochlear model," IEEE Trans. Signal Process., **39**, 2573-2592.

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.
Liberman, M.C. and Dodds, L.W. (**1984**). "Single-neuron labeling and chronic cochlear pathology III Stereocilia damage and alterations of threshold tuning curves," Hear. Res., **16**, 55-74.
Liberman, M.C. and Kiang, N.Y.S. (**1984**). "Single-neuron labeling and chronic cochlear pathology IV Stereocilia damage and alterations in rate- and phase-level functions," Hear. Res., **16**, 75-90.
Louage, D.H., van der Heijden, M., and Joris, P.X. (**2004**). "Temporal properties of responses to broadband noise in the auditory nerve," J. Neurophysiol., **91**, 2051-2065.
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.
Moore, B.C.J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford University Press, New York).
Oxenham, A.J. and Bacon, S.P. (**2003**). "Cochlear compression: Perceptual measures and implications for normal and impaired hearing," Ear Hearing, **24**, 352-366.
Patuzzi, R. (**1996**). "Cochlear micromechanics and macromechanics," in *The Cochlea*. Eds. P. Dallos, A.N. Popper, and R.R. Fay (Springer, New York), pp. 186-257.
Saremi, A. and Stenfelt, S. (**2013**). "Effect of metabolic presbyacusis on cochlear responses: A simulation approach using a physiologically-based model," J. Acoust. Soc. Am., **134**, 2833-2851.
Schmiedt, R.A., Lang, H., Okamura, H.O., and Schulte, B.A. (**2002**). "Effects of furosemide applied chronically to the round window: A model of metabolic presbyacusis," J. Neurosci., **22**, 9643-9650.
Sewell, W.F. (**1984**). "The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats," Hear. Res., **14**, 305-314.
Swaminathan, J. and Heinz, M.G. (**2011**). "Predicted effects of sensorineural hearing loss on across-fiber envelope coding in the auditory nerve," J. Acoust. Soc. Am., **129**, 4001-4013.
Wake, M., Takeno, S., Ibrahim, D., and Harrison, R. (**1994**). "Selective inner hair cell ototoxicity induced by carboplatin," Laryngoscope, **104**, 488-493.
Wang, J., Powers, N., Hofstetter, P., Trautwein, P., Ding, D., and Salvi, R. (**1997**). "Effects of selective inner hair cell loss on auditory nerve fiber threshold, tuning and spontaneous and driven discharge rate," Hear. Res., **107**, 67-82.
Zilany, M.S.A. and Bruce, I.C. (**2006**). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," J. Acoust. Soc. Am., **120**, 1446-1466.
Zilany, M.S.A., Bruce, I.C., Nelson, P.C., and Carney, L.H. (**2009**). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics," J. Acoust. Soc. Am., **126**, 2390-2412.
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.

Individual speech recognition in noise, the audiogram and more: Using automatic speech recognition (ASR) as a modelling tool

BIRGER KOLLMEIER*, MARC RENÉ SCHÄDLER, ANNA WARZYBOK, BERND T. MEYER, AND THOMAS BRAND

Medizinische Physik and Cluster of Excellence Hearing4all, Universität Oldenburg, Oldenburg, Germany

To characterize the individual patient's hearing impairment, a framework for auditory discrimination experiments (FADE, Schädler *et al.*, 2015) was extended here using different degrees of individualization. FADE has been shown to predict the outcome of both speech recognition tests and psychoacoustic experiments based on simulations using an automatic speech recognition (ASR) system which requires only few assumptions. It builds on the closed-set matrix sentence recognition test which is advantageous for testing individual speech recognition in a way comparable across languages. Individual predictions of speech recognition thresholds in stationary and in fluctuating noise were derived using the audiogram and an estimate of the internal detector noise ("level uncertainty"). Either "typical" audiogram shapes with or without a "typical" level uncertainty or the individual data were used for individual predictions. As a result, the individualisation of the level uncertainty was found to be more important than the exact shape of the individual audiogram to accurately model the outcome of the German matrix test in stationary or fluctuating noise for listeners with hearing impairment.

INTRODUCTION

Recent progress in computational modelling of the normal and impaired auditory system nurtures the hope that a better understanding is achieved of how hearing impairment affects speech communication in daily life. This will help to construct and assess more effective hearing devices. A first approach to provide a model framework which might be developed into an "objective yard stick" in rehabilitative audiology is considered here: The prediction of speech recognition thresholds (SRTs) in noise for an individual based on known audiological data (such as, e.g., the audiogram or measures of supra-threshold processing deficits). By comparing the predictions with the individual empirical SRTs, any special problems of the patient in understanding speech in noise other than explainable from his/her audiogram (such as, e.g., due to auditory neuropathy or more central or cognitive components of hearing impairment) may become obvious.

---

*Corresponding author: birger.kollmeier@uni-oldenburg.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

Traditional modelling approaches for speech recognition are either based on predefined features (like an energy increase in a certain auditory band) or on instrumental measures that are calibrated using a set of reference thresholds (like the Articulation Index or Speech Intelligibility Index-based methods, see ANSI 1997; Meyer and Brand, 2013). More sophisticated approaches are based on psychoacoustical processing models (e.g., Holube and Kollmeier, 1996, Dau *et al.*, 1997; Jürgens and Brand, 2009), but require an "optimal detector" that possesses perfect prior knowledge about the to-be-recognized signals. The strong assumption of an optimal detector provides the model with an unfair advantage over human listeners that perform the same task, and may even weaken the need of an optimum auditory-system-inspired processing front end to achieve human-like performance, which, in turn, could be crucial to accurately model human sound perception.

An alternative way of predicting both sentence recognition thresholds and psychoacoustic performance using automatic speech recognition (ASR) without requiring a predefined reference or an optimal detector was recently proposed by Schädler *et al.* (2015; 2016). They predicted the outcome of the German matrix sentence recognition test (Kollmeier *et al.*, 2015) for different types of stationary background noise using Mel-frequency cepstral coefficients (MFCCs) as a front-end and whole-word Gaussian mixture/hidden Markov models (HMMs) as a back-end. By training and testing the ASR system with noisy matrix sentences on a broad range of signal-to-noise ratios (SNRs) they were able to predict SRTs for listeners with normal hearing with a remarkably high precision, outperforming SII-based predictions. In a second study, they extended the so-called simulation framework for auditory discrimination experiments (FADE) to successfully simulate basic psychoacoustical experiments as well as more complex Matrix sentence recognition tasks with a range of feature sets (front-ends). Schädler *et al.* (2015) concluded that the proposed FADE framework is able to predict empirical data from the literature with a single set of parameters, less assumptions compared to traditional modelling approaches, and without the need of an empirical reference condition.

The aim of the current study is to extend the FADE approach to model the effect of hearing impairment on speech recognition thresholds obtained with the German Matrix test in stationary and fluctuating noise. Therefore, different degrees of individualization for the model predictions were employed and compared with the empirical results for 99 normal-hearing and hearing-impaired listeners (198 ears).

METHODS

FADE approach

The simulation framework for auditory discrimination experiments (FADE) from Schädler *et al.* (2016) was used to simulate the outcome of the German Matrix test in a stationary and a fluctuating noise condition (see Schädler *et al.*, 2015, for details). The speech material consists of 120 recorded semantically unpredictable sentences with a fixed syntax (name-verb-number-adjective-object, like "Peter sees eight wet chairs".) For each word class, ten alternatives exist. The adaptively determined SRT denotes the SNR that corresponds to 50%-words-correct performance. To obtain SRTs with FADE, an automatic speech recognizer (ASR) was trained and tested with noisy sentences on a broad range of SNRs (−24dB to +6dB), and the lowest SNR which resulted in 50%-words-correct recognition performance was interpolated and used as the predicted SRT. The ASR system used modified MFCCs as a front-end. On the back-end side, HMMs were used to model speech with whole-word models based on a "parametrically hearing-impaired" acoustical representation provided by the front-end. Hearing impairment was modelled in the front-end and implemented in the log Mel-spectrogram (logMS) from which the MFCC features were derived. A frequency-dependent attenuation was used to model an attenuation-loss (A) by clipping the amplitude values in each channel to the corresponding (interpolated) threshold from the audiogram. To model a supra-threshold distortion loss (D), a level uncertainty was implemented in the logMS by adding a Gaussian white noise with a standard deviation of u_L.

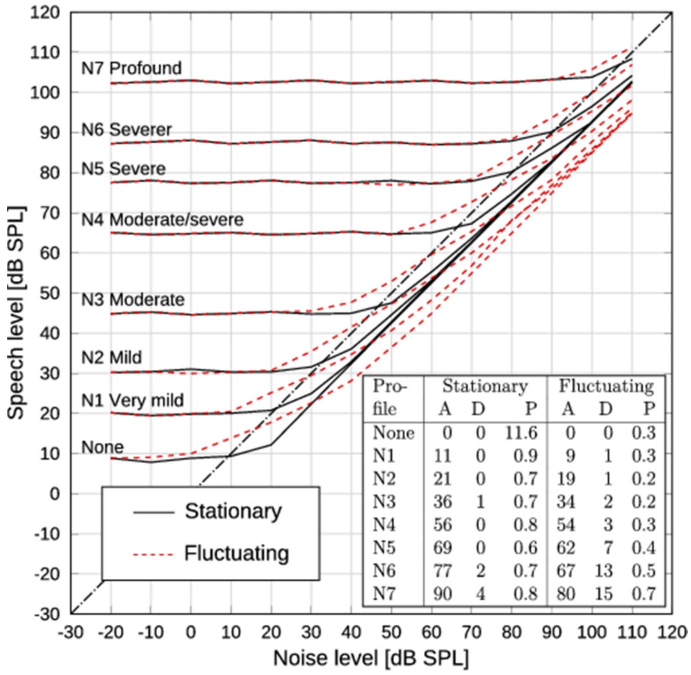
Audiological Data

Results from Brand and Kollmeier (2002) were used for comparing the predictions with empirical data. The data included measurements from 99 listeners (198 separately measured ears) ranging in age from 23 to 82 years (mean and standard deviation: 61.4 ± 13.2 years) and covering a broad range of hearing loss with the PTA varying from 0 to 80 dB HL (mean: 40.5 ± 16.1 dB HL). SRTs were obtained with the German matrix test in stationary ICRA1 and fluctuating ICRA5-250 noise. The ICRA5-250 noise is a speech-like modulated noise which simulates the long-term frequency spectrum and modulation properties of a single male speaker with silent intervals limited to 250 ms (Wagener *et al.*, 2006). The same noise condition was used in a study of Meyer and Brand (2013) with 113 listeners (of whom the 99 listeners considered here are a subgroup). They considered three extensions of the Speech Intelligibility Index (SII) for predicting SRT in stationary and fluctuating noise: A) original SII, B) considering frequency-independent level fluctuation of the noise, C) considering frequency-dependent level fluctuations of the noise, and D) considering frequency- dependent fluctuations of the speech and the noise.

RESULTS AND DISCUSSION

Audiogram-based predictions without suprathreshold distortions

Figure 1 shows the simulated SRTs for the 7 typical audiograms for flat and moderately sloping hearing loss defined by Bisgaard *et al.* (2010) as a function of the level of the stationary, test specific noise (solid lines). The simulations for the remaining 3 typical audiograms are not shown here to preserve the separability across curves. In general, the curves follow the pattern proposed by Plomp (1978) who separated an "Attenuation" component (A) from a "Distortion" component (D) of the hearing loss to derive the SRT as a function of noise level (NL). A power-law additivity parameter P was also introduced here to better reflect the fluctuating noise condition:



Pro-file	Stationary A	Stationary D	Stationary P	Fluctuating A	Fluctuating D	Fluctuating P
None	0	0	11.6	0	0	0.3
N1	11	0	0.9	9	1	0.3
N2	21	0	0.7	19	1	0.2
N3	36	1	0.7	34	2	0.2
N4	56	0	0.8	54	3	0.3
N5	69	0	0.6	62	7	0.4
N6	77	2	0.7	67	13	0.5
N7	90	4	0.8	80	15	0.7

Fig. 1: Speech recognition thresholds (SRTs) for the German matrix sentence test in the test-specific, stationary noise condition as a function of the noise level from simulations with FADE (solid lines). The curves correspond to different grades of hearing impairment based on the seven standard audiograms for flat and moderately sloping hearing loss from Bisgaard *et al.* (2010). The dashed lines show the same results for the fluctuating ICRA5-250 noise. The embedded table reports the attenuation (A) and distortion (D) components (in dB) and the power coefficient P of the best-fitting Plomp curves.

$$SRT_{Plomp}=10\log_{10}\left(10^{\frac{(A+D)*P}{10}}+10^{\frac{(NL+D)*P}{10}}\right)/P \qquad \text{(Eq. 1)}$$

For a given hearing loss, the SRT in quiet is dominated by A+D (horizontal part of the curves). With increasing noise level NL, a transition region (controlled by P) occurs until a constant SNR at SRT is achieved across a wide range of noise levels which reflects the D-value. The A-, D- and P- values fitted to the simulated curves using the Plomp (1978) formula for the different typical audiograms are given in the insert table in Fig. 1. Note that most of the variation across the typical audiograms are captured by the variation in the "Attenuation" component, whereas only the more severe hearing losses require an additional "Distortion" component which also reflects some deviation of the audiogram shape from the standard speech spectrum.

To test the non-individualized SRT predictions based on the audiogram alone (i.e., without suprathreshold processing impairment), Fig. 2A displays the predictions from the "typical" audiograms in Fig. 1 for the individual SRT in stationary ICRA1-noise. The SRT predictions obtained by interpolating across the 10 prototype audiograms

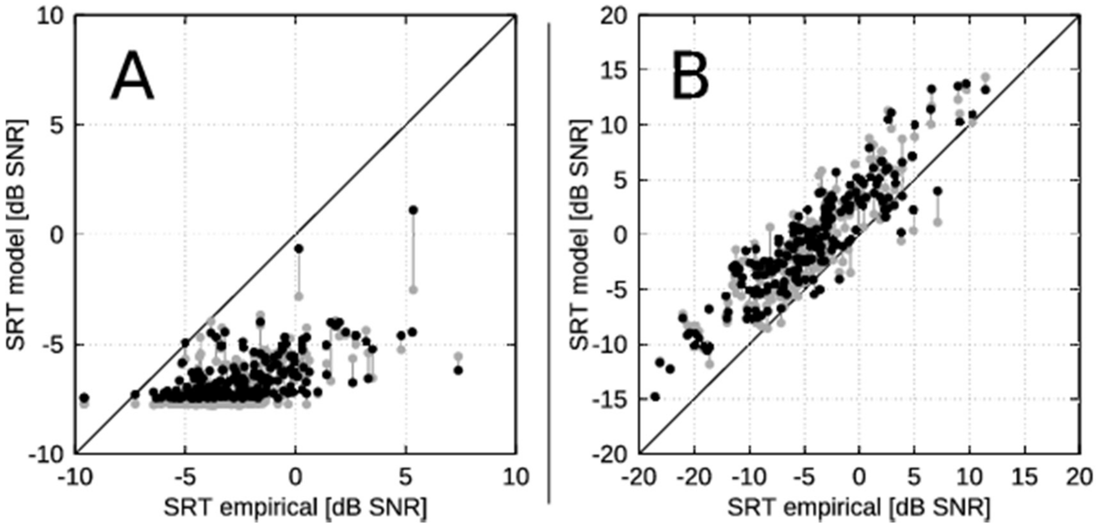


Fig. 2: Modelled speech recognition thresholds (SRTs) for 198 ears from 99 subjects plotted against the empirical data (x-axis). Panel A: Stationary noise, non-individualized predicted SRT (black dots) obtained from the respective best-fitting typical audiogram compared to the individually simulated SRT (grey dots) using the FADE approach with the individual audiogram data. Panel B: Fluctuating noise, predicted SRT from typical audiogram (black dots) vs. individually simulated (grey dots) taking into account the individual level uncertainty u_L estimated from the stationary noise condition.

(black dots) are plotted against the empirical values (given on the x-axis). For comparison, the individualized FADE simulations are given as grey symbols using the individual audiogram. The connection lines between the predicted values (that require only a very small computational load) and the simulated values (that are computationally expensive) indicate already a high coincidence in SRT prediction between both methods. However, neither method is able to model the empirical SRT in stationary noise in a satisfactory way since the large spread in the empirical data (ranging from −9 dB to + 7dB in SNR) is not reflected in the predictions based on the audiogram alone.

Modelling suprathreshold distortion as level uncertainty

Figure 3 displays the simulated SRT using the FADE approach for a normal audiogram with a set of fixed "level uncertainty parameter" u_L-values in order to model an increasing amount of supra-threshold distortions. Note that the curves exhibit a parallel shift to higher SRT values with increasing parameter u_L which is very similar to the effect of the D-parameter of the Plomp model. However, an increase by 10 dB in the level uncertainty parameter u_L does not translate directly into an equally-spaced increase of the D-parameter fitted to the curves in Fig. 3 (see inlaid table in Fig. 3): At low and high u_L-values the largest resulting difference in D for a 10-dB step in u_L is observed, whereas in the midrange the simulations exhibit a higher robustness against an increase in level uncertainty.

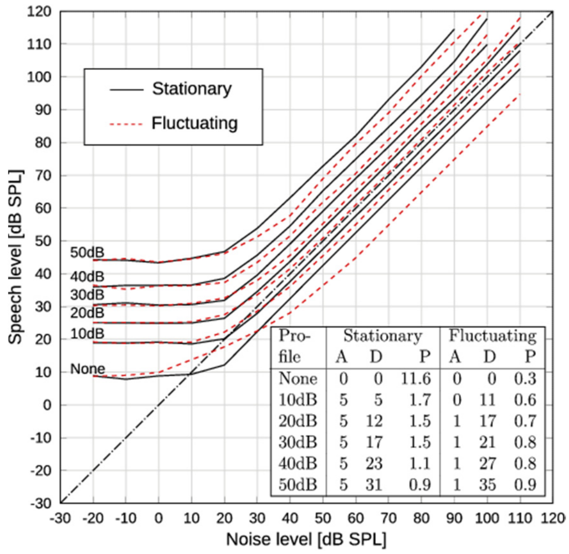


Fig. 3: Speech recognition thresholds (SRTs) in the test-specific, stationary noise condition as a function of the noise level from simulations with FADE for different values of level uncertainty u_L. The dashed lines show the corresponding results for the fluctuating ICRA5-250 noise. The embedded table reports the attenuation (A) and distortion (D) components (in dB) and the power coefficient P of the best-fitting Plomp curves according to Eq. 1.

Combining individualization of audiogram and suprathreshold distortion correction

To assess the effect of including a distortion correction based on estimates of the level uncertainty parameter u_L into the modelling of the SRT data, Table 1 shows the correlation coefficients (Pearson's R^2) between modelled SRTs for stationary and fluctuating noise and the empirical data. *Predictions* indicate an interpolation method based on computations for the 10 typical audiograms only, while *simulations* refer to computations performed for each individual audiogram. The individual supra-threshold distortion effect was not individually computed with the whole FADE approach, but rather estimated in two ways:

- For the "typical" estimate of the level uncertainty parameter u_L, a group of at least 5 and up to 32 listeners, characterized by the same "typical" audiogram, was considered. Their deviation between prediction and empirical SRT was averaged either for the stationary or for the fluctuating noise. This deviation in SNR was converted into a u_L value using the relation shown in Fig. 3, thus leading to the "typical stationary noise-based" or "typical fluctuating noise-based" individualization of u_L. The predicted or simulated SRT was obtained as before, but corrected by an appropriate SRT shift read out from the respective curve in Fig. 3.
- The "individual" estimate of u_L was determined from the individual deviation between modelled and empirical result in the stationary noise condition and then used to correct for suprathreshold distortions in fluctuating noise and vice versa. Note that estimating the "typical" u_L values from the stationary or fluctuating

noise condition provides approximately the same prediction accuracy in both cases (i.e., both for predicting the stationary and the fluctuating noise data) as indicated by the very similar R^2. This suggests that the individual distortion effect is estimated to be very similar for both types of noises – which is a desired property for a universally applicable parameter characterizing the impaired individuals performance. Using the "typical" audiogram and distortion correction already outperforms the SII prediction accuracy for the fluctuating noise case. In the stationary noise case, the individual distortion correction is required to outperform the SII predictions.

		Stationary noise			Fluctuating noise		
Model	Distortion correction	R^2	B	RMS	R^2	B	RMS
FADE prediction	none	0.31	-4.1	4.6	0.48	-4.6	6.1
	typical stat.-based	0.44	0.0	1.9	0.57	3.7	5.4
	typical fluc.-based	0.42	-1.9	2.7	0.56	0.0	3.8
	individual stat.-based	-	-	-	0.78	3.4	4.3
	individual fluc.-based	0.63	-1.6	2.3	-	-	-
FADE simulation	none	0.36	-4.3	4.7	0.57	-4.5	5.9
	typical stat.-based	0.49	-0.1	1.8	0.63	3.8	5.2
	typical fluc.-based	0.46	-2.0	2.7	0.63	0.1	3.5
	individual stat.-based	-	-	-	0.83	3.8	4.5
	individual fluc.-based	0.70	-1.9	2.4	-	-	-
SII version A		0.55	-	-	0.24	-	-
SII version B		0.59	-	-	0.42	-	-
SII version C		0.51	-	-	0.42	-	-
SII version D		0.35	-	-	0.52	-	-

Table 1: Statistical analysis of the predicted/simulated speech recognition thresholds (SRTs). Pearson's correlation coefficients (R^2) are reported along with the root-mean-square (RMS) prediction error and the bias (B) for predicted/simulated SRTs with different distortion correction methods and SII-based predictions from Meyer and Brand (2013).

Overall, the highest prediction and simulation accuracy is achieved if not typical parameter sets, but individualized audiogram and u_L values are employed: Fig. 2B shows the individually modelled SRT in fluctuating noise using the individually obtained u_L estimates from the stationary noise condition either predicted from the typical audiogram data (black dots) or individually simulated (grey dots). The graph demonstrates the high prediction accuracy observed for the individualized suprathreshold distortion parameter u_L even if not an individualized, but typical audiogram is used.

CONCLUSIONS

The ASR-based, reference-free FADE approach can be used as a theoretical counterpart of the empirical Plomp (1978) model to quantitatively assess the effect of hearing impairment on SRTs in stationary and fluctuating noise.

Suprathreshold processing deficiencies can be modelled by the level uncertainty parameter u_L which should be individually determined for a high prediction accuracy.

The prediction accuracy achieved (expressed by Pearson's R^2) is much higher than the prediction accuracy achieved with modified and optimized SII-based measures (e.g., data presented by Mayer and Brand, 2013).

Hence, the FADE approach is not only more versatile and makes much less assumptions than the SII, but also yields much higher prediction accuracy.

ACKNOWLEDGMENTS

Supported by DFG (SFB/TRR 31 & EXC Hearing4all).

REFERENCES

ANSI (**1997**). *S3.5 Methods for Calculation of the Speech Intelligibility Index.* Standards Secretariat, Acoustical Society of America.

Bisgaard, N., Vlaming, M.S., and Dahlquist, M. (**2010**) "Standard audiograms for the IEC 60118-15 measurement procedure," Trends Amplif., **14**, 113-120.

Brand,T. and Kollmeier, B. (**2002)**. "Vorhersage der Sprachverständlichkeit in Ruhe und im Störgeräusch aufgrund des Reintonaudiogramms," Proc. 5. Jahrestagung der Deutschen Gesellschaft für Audiologie, Zürich.

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.

Holube, I. and Kollmeier, B. (**1996**). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," J. Acoust. Soc. Am., **100**, 1703-1716.

Jürgens, T. and Brand, T. (**2009**). "Microscopic prediction of speech recognition for listeners with normal hearing in noise using an auditory model," J. Acoust. Soc. Am., **126**, 2635-2648.

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M., Uslar, V.N., Brand, T., and Wagener, K.C. (**2015**). "The multilingual matrix test: principles, applications and comparison across languages - a review," Int. J. Audiol. Suppl., **54**, 3-16.

Meyer, R.M. and Brand, T. (**2013**). "Comparison of different short-term Speech Intelligibility Index procedures in fluctuating noise for listeners with normal and impaired hearing," Acta Acust. United Ac., **99**, 442-446.

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

Schädler, M., Warzybok, A., Hochmuth, S., and Kollmeier, B. (**2015**). "Matrix sentence intelligibility prediction using an automatic speech recognition system," Int. J. Audiol. Suppl., **54**, 100-107.

Schädler, M.R., Warzybok, A., Ewert, S.F., and Kollmeier, B. (**2016**). "A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception," Submitted.

Wagener, K.C., Brand, T., Kollmeier, B. (**2006**). "The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners," Int. J. Audiol., **45**, 26-33.

Modeling individual loudness perception in cochlear implant recipients with normal contralateral hearing

JOSEF CHALUPPER*

Advanced Bionics GmbH, European Research Center, Hannover, Germany

Use of acoustic and electric models may make the fitting of bimodal patients more efficient. The electric loudness model (McKay *et al.*, 2003) was extended to account for simultaneous and high-rate stimulation. Both acoustic and electric loudness models require clinical audiometric data for individualization. While the availability of an individual's thresholds is essential to achieve accurate model predictions, average values of electric field spread can be used for calculating group data. The use of individual spatial spread functions may further improve model predictions, allowing individual predictions and hence automating bimodal loudness balancing.

INTRODUCTION

Due to widening candidacy criteria for provision of cochlear implants (CIs), more and more CI recipients have aidable contralateral hearing and, thus, wear a hearing aid (HA) in addition to their CI. This bimodal configuration is used worldwide by about 30% of all CI recipients (Scherf and Arnold, 2014) and has been shown to improve speech understanding in noise, localization, sound quality and music perception (Ching *et al.*, 2007). In order to achieve the maximum bimodal benefit for an individual patient, balancing of loudness across ears is regarded to be important (Francart and McDermott, 2013; Dorman *et al.*, 2014). Manual adjustment of HA and CI, however, is very tedious and time-consuming as there are many parameters to be optimized, such as channel gains, compression ratio and knee-points, M- and T-Levels, input dynamic range and sensitivity. Typically, two different fitting modules are used by the clinician. As a consequence, in clinical practice, loudness is often not balanced and, thus, the patient may not obtain the maximum bimodal benefit.

A possible way to speed up bimodal loudness balancing is to use loudness models. Ideally, such models would be individualized for a given patient. This would then help predict both individual acoustic and electric loudness, automatically finding the HA and CI parameters that lead to balanced loudness for a large variety of relevant stimuli. Research on acoustic loudness models started in the late 1950s (Zwicker, 1958). Hence, today acoustic loudness is probably the best understood hearing sensation and broadly validated loudness models for normal-hearing and hearing-impaired listeners are available (e.g., Moore and Glasberg 1997; Chalupper and Fastl, 2002). In contrast, work on electric loudness models started rather recently (McKay *et al.*, 2003) and, thus, a large number of effects in electric loudness perception still

*Corresponding author: josef.chalupper@advancedbionics.com

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

needs further investigation. The general structure of electric and acoustic loudness models, however, seems to be the same (Fig. 1). First, an acoustic free-field stimulus is converted into the respective spatial excitation pattern inside the cochlea. Next excitation is transformed into specific loudness in Sone per critical band. Finally, specific loudness is summed across all critical bands to calculate overall loudness in Sone. As categorical loudness scaling is often used in behavioural measurements of loudness, Sone have to be converted into categorical units (CU).

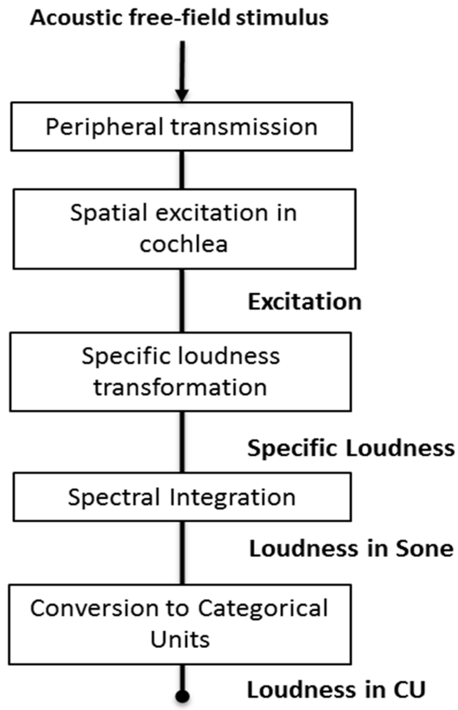


Fig. 1: General structure of loudness models.

The "practical" electric loudness model of McKay *et al.* (2003) is based on the simplifying assumption that each electrode contributes independently to overall loudness and thus, any explicit modelling of electric field interactions is not required. As a consequence, this model does not comprise a stage to calculate spread of excitation across electrodes, but directly converts current amplitudes of a CI's pulse pattern into specific loudness. This approach is valid for channel rates between 200 pulses per second (pps) and 1000 pps and overall pulse rates between 500 pps and 4000 pps, but is not valid for simultaneous or analog electric stimulation. At present, however, some advanced coding strategies use simultaneous stimulation for current steering and pulse rates of more than 1000 pps (e.g., Advanced Bionics HiRes Fidelity 120, Nogueira *et al.*, 2009). Chalupper et al. (2015) used the "practical" model to predict loudness summation of CI recipients using Advanced Bionics' HiRes Fidelity 120 and concluded that the model overestimates the loudness summation effect and that spread of excitation needs to be accounted for. Additionally, lack of behavioural M- and T-levels and use of fluctuating noises complicate the model calculations.

The purpose of the present study was to investigate whether a modified electric loudness model could be used for balancing electric and acoustic loudness. CI recipients with normal contralateral hearing were studied as special case of bimodal users to avoid modelling the signal processing and coupling acoustics of hearing aids.

ELECTRIC AND ACOUSTIC LOUDNESS MODEL

Structure

For the calculation of acoustic loudness the dynamic loudness model (DLM, Chalupper and Fastl, 2002) was used. As only stationary stimuli and unaided acoustic hearing were used in this study, dynamic blocks of the DLM (forward masking, temporal integration) and HA signal processing were not included in the calculation (see Fig. 2). The "practical" electric loudness model was extended by explicitly modelling electric field spread according to Hamacher (2004). A simulation of the signal processing blocks of HiRes Fidelity 120 was used to calculate electrodograms for acoustic free-field stimuli. An individualized specific loudness transform was employed to calculate specific loudness patterns across critical band rate from acoustic excitation patterns and spatial electric field, respectively. To convert Sone into CU, a cubic fit function using four parameters as suggested by Heeren *et al.* (2013) was applied.

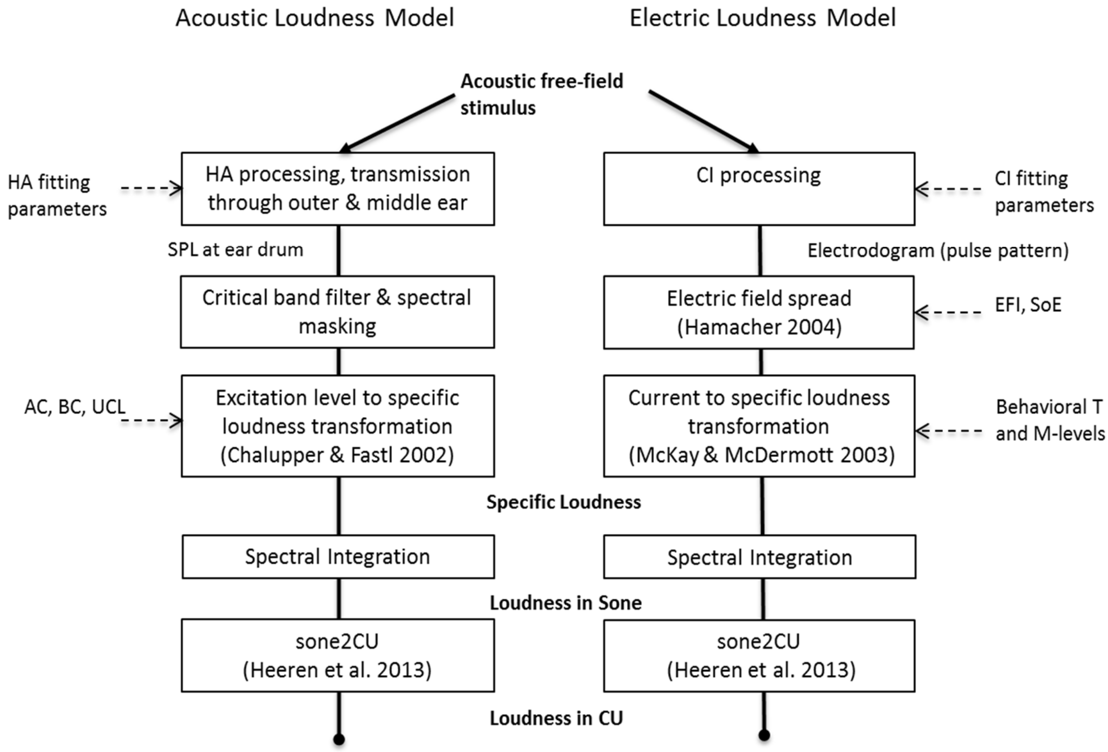


Fig. 2: Block diagram of electric and acoustic loudness models.

Electric field spread

The electric field spread was calculated using a double-sided one-dimensional exponentially decreasing function with a spread constant lambda. It is generally assumed that the spread constant varies substantially across CI patients, type of electrode arrays and electrodes within an array. Based on considerations in Fredelake and Hohmann (2012), lambda values (exponential spatial decay constants) of 1 mm and 10 mm were included in loudness model calculations. Figure 3 shows the resulting electric fields for simultaneous stimulation of two adjacent electrodes.

Individualization

To individualize the acoustic loudness model, unaided air conduction thresholds (AC) were used to adapt the parameters of the specific loudness transformation. For individualization of the electric loudness model, the individual maps of the CI recipients were employed in the simulation of the CI signal processing. Additionally, the parameters of loudness growth function given by McKay *et al.* (2003) were individually fitted to subjects' behavioural T-level.

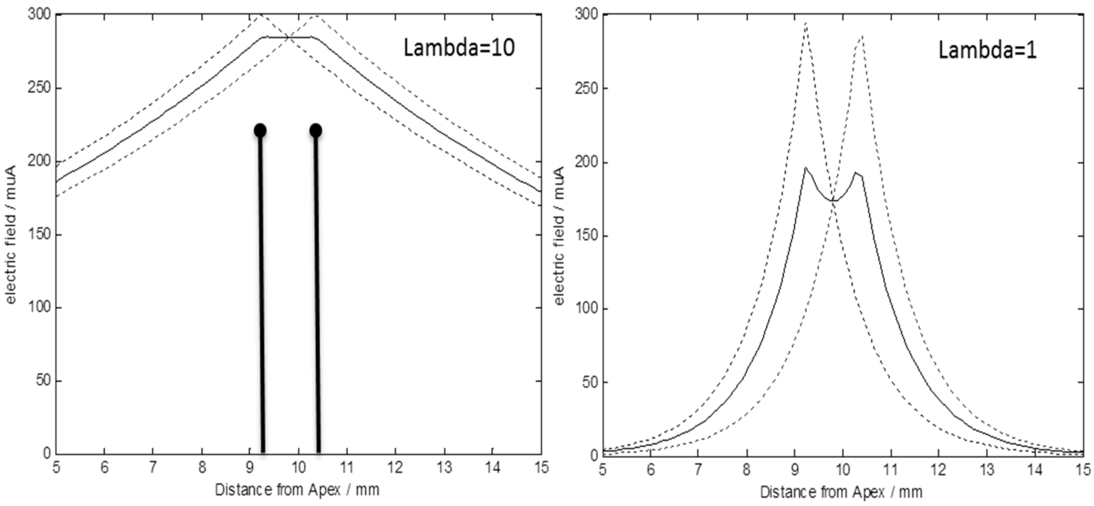


Fig. 3: Spatial spread of electric field for simultaneous stimulation. Left: spread function with lambda = 10. Right: spread function with lambda = 1.

EVALUATION

Data from a study with single-sided deaf CI-recipients conducted by Büchner et al. (2013) were used to evaluate the acoustic and electric loudness models.

Methods

Five CI recipients with contralateral thresholds of better than 30 dB HL below 4 kHz participated in this study. All stimuli were presented via direct audio input (DAI) and headphone to CI and normal-hearing ear, respectively. The fitting of the CI was adjusted to achieve balanced loudness perception: T-levels were set to behaviourally measured T-levels. M-levels were adjusted until subjects indicated the same interaural loudness for narrow band noises presented at 80 dB SPL. Input Dynamic Range (IDR) was modified to balance loudness for speech shaped noise presented at 50 and 80 dB SPL. To verify the result of this fitting approach, loudness scaling was administered for both the electric and acoustic ears separately.

Results

The modified electric loudness model is able to predict loudness of narrowband stimuli with a similar accuracy as acoustic loudness models. Typical cases for acoustic modelling and electric modelling are shown in Fig. 4 and Fig. 5, respectively. The selection of lambda did not affect the accuracy of the electric model predictions.

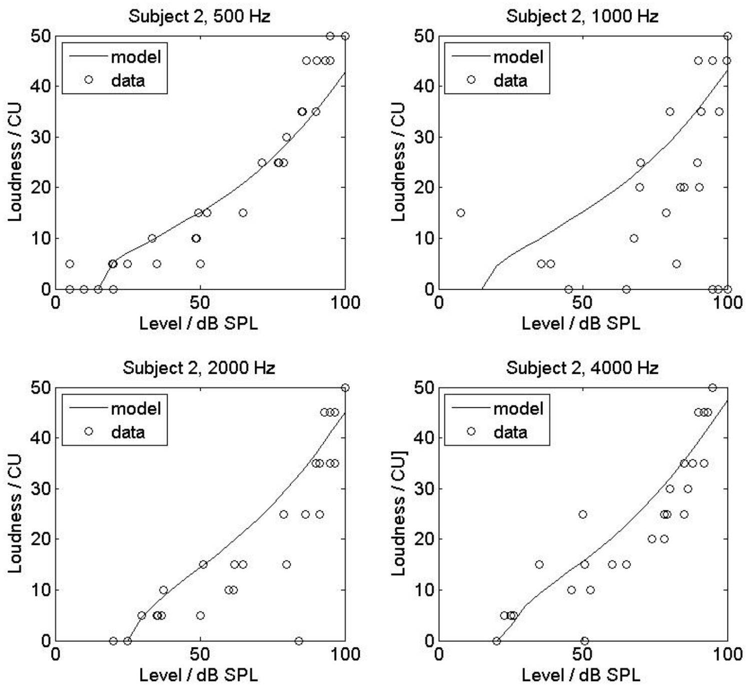


Fig. 4: Individual acoustic model predictions and behavioural data for narrowband stimuli.

In order to evaluate the models for broad-band sounds, acoustic and electric loudness growth curves for speech were calculated using the same parameters as for modelling the loudness of narrowband stimuli. Recall that during fitting, loudness of speech noise was balanced for 50 dB SPL and 80 dB SPL. Assuming that both models are valid for narrowband stimuli, the level difference at the same calculated loudness for speech can be used to evaluate model predictions for broadband sounds. Figures 6 and 7 indicate that selection of lambda can make a substantial difference for some subjects. Thus, further individualization of the electric model by using individual and electrode-specific spread functions has the potential to improve the model predictions for the individual listener. Individual spread functions can be derived from impedance measurements between electrodes ("electric field imaging") or individual electrically evoked compound action potential (ECAP) data.

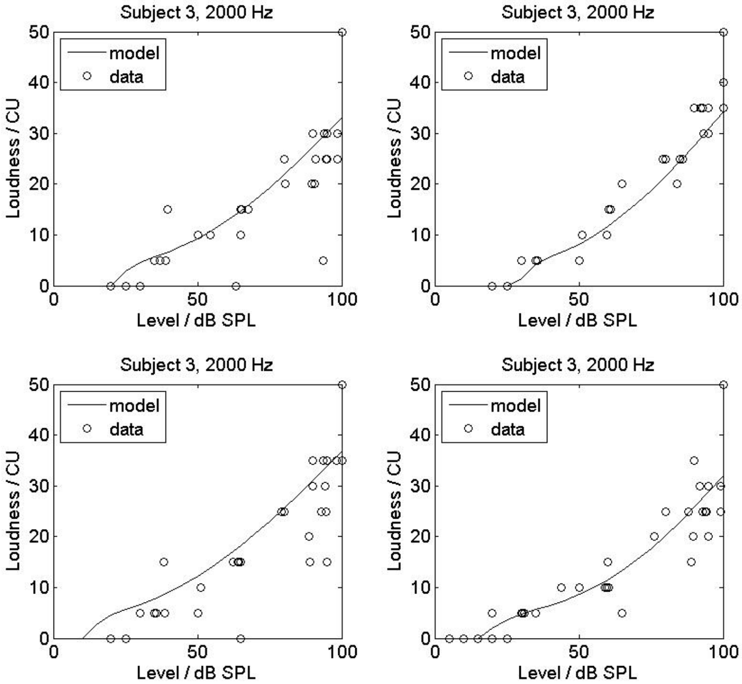


Fig. 5: Individual electric model predictions and behavioural data for narrowband stimuli.

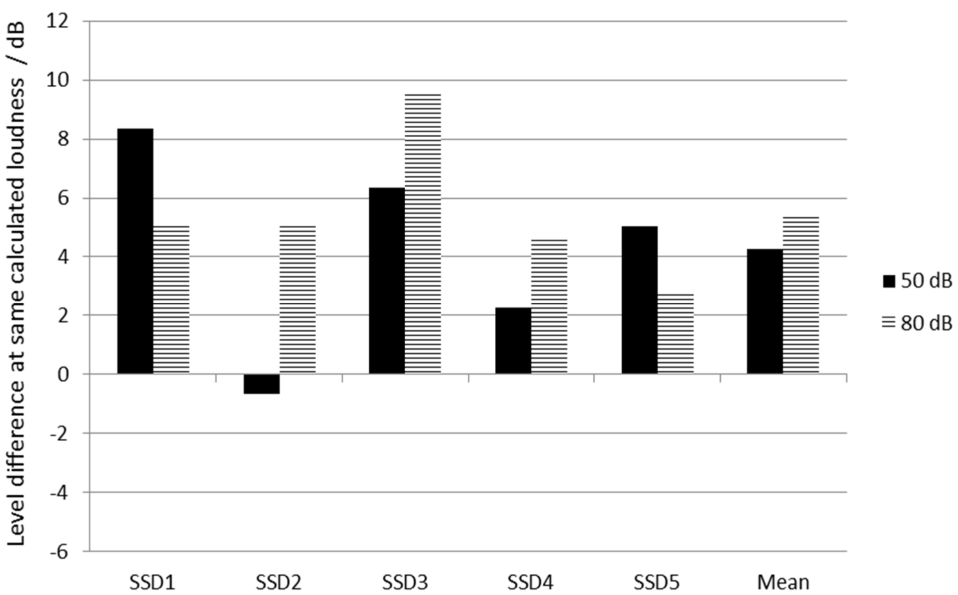


Fig. 6: Individual model predictions and behavioural data for loudness summation with lambda = 10.

On average, with a spatial spread of 10 mm, the loudness summation effect for broad-band stimuli is underestimated, while a spread of 1 mm results in an average prediction error of less than 2.5 dB. While this should be sufficient for modelling group differences, this approach presumably is not accurate enough to automate loudness balancing, as there are deviations for individuals by more than 5 dB. Moreover, the prediction of loudness of time-varying stimuli needs to be modelled and verified (Francart *et al.*, 2014).

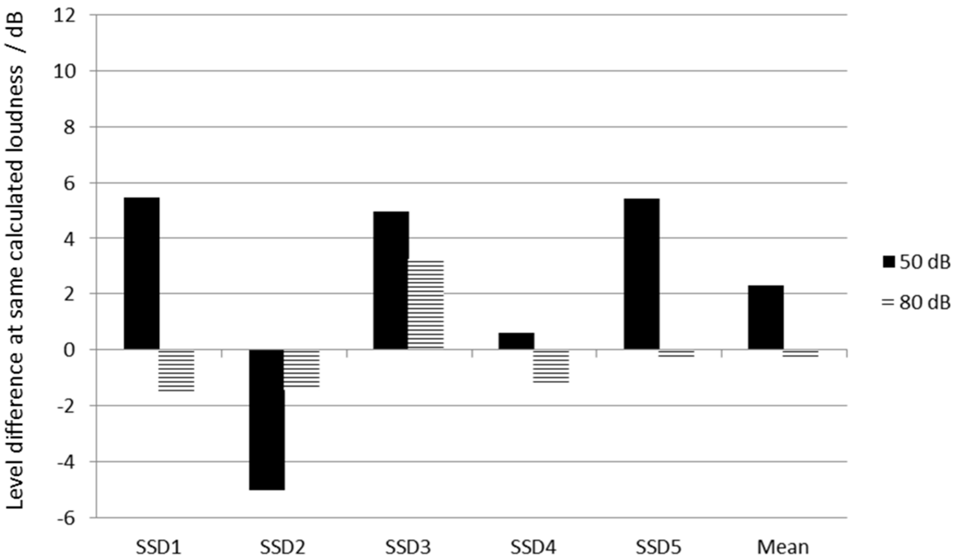


Fig. 7: Individual model predictions and behavioural data for loudness summation with lambda = 1.

CONCLUSIONS

In order to predict the loudness of CI coding strategies employing current steering and high pulse rates, electric loudness models must incorporate a stage to simulate the spatial electric field within the cochlea. Using standard audiometric data (electric and acoustic thresholds) allows the prediction of loudness for stationary stimuli on a group level. For the application of loudness models to automatically adjust fitting parameters of CI and HA to achieve a balanced interaural loudness, however, further individualization appears to be required. Electric field imaging, or electrically evoked compound action potentials, could be used to individualize spatial spread functions and, thus, improve the accuracy of individual loudness predictions.

REFERENCES

Büchner, A., Geissler, G., Lenarz, T., Lesinski-Schiedat, A., Chalupper, J., Brendel, M., and Fredelake, S. (**2014**). "Lautheitswahrnehmung bei einseitig tauben Cochlea-Implantat-Trägern," 17. Jahrestagung der Deutschen Gesellschaft für Audiologie (DGA), Oldenburg.

Chalupper, J. and Fastl, H., (**2002**). "Dynamic Loudness Model (DLM) for normal and hearing-impaired listeners", Acta Acust. United Ac., **88**, 378-386.

Chalupper, J., Schulz, J., Elke Devocht, E., George, E., and Fredelake, S. (**2015**). "Modeling the effects of pulse width and acoustic bandwidth on electric loudness perception," 18. Jahrestagung der Deutschen Gesellschaft für Audiologie (DGA), Bochum.

Ching, T.Y., 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.

Dorman, M.F., Loizou, P., Wang, S., Zhang, T., Spahr, A., Loiselle, L., and Cook, S. (**2014**). "Bimodal cochlear implants: The role of acoustic signal level in determining speech perception benefit," Audiol. Neurotol., **19**, 234-238.

Francart, T. and McDermott, H.J. (**2013**). "Psychophysics, fitting, and signal processing for combined hearing aid and cochlear implant stimulation. Review," Ear. Hearing, **34**, 685-700

Francart, T., Innes-Brown, H., McDermott, H.J., and McKay, C.M. (**2014**). "Loudness of time-varying stimuli with electric stimulation," J. Acoust. Soc. Am., **35**, 3513-3519.

Fredelake, S. and Hohmann, V. (**2012**). "Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation," Hear. Res., **287**, 76-90.

Hamacher, V. (**2004**). *Signalverarbeitungsmodelle des elektrisch stimulierten Gehörs*. Ph.D. thesis, RWTH Aachen. Wissenschaftsverlag Mainz in Aachen, 1. Edition.

Heeren, W., Hohmann, V., Appell, J.E., and Verhey, J.L. (**2013**). "Relation between loudness in categorical units and loudness in phons and sones," J. Acoust. Soc. Am., **133**, EL314-319.

McKay, C.M., Henshall, K.R., Farrell, R.J., and McDermott, H.J. (**2003**). "A practical method of predicting the loudness of complex electrical stimuli," J. Acoust. Soc. Am., **113**, 2054-2063.

Moore, B.C.J. and Glasberg, B. (**1997**). "A model of loudness perception applied to cochlear hearing loss," Audit. Neurosci., **3**, 289-311.

Nogeira, W., Litvak, L., Edler, B., Ostermann, J., and Büchner, A. (**2009**). "Signal processing strategies for cochlear implants using current steering," EURASIP J. Adv. Sig. Proc., **2009**, 1-20.

Scherf, F. and Arnold, L. (**2014**). "Exploring the clinical approach to the bimodal fitting of hearing aids and cochlear implants: Results of an international survey," Acta Otolaryngol., **134**, 1151-1157.

Zwicker, E. (**1958**). "Über psychologische und methodische Grundlagen der Lautheit", Acustica, **8**, 237-258.

Modelling the effect of individual hearing impairment on sound localisation in sagittal planes

ROBERT BAUMGARTNER*, PIOTR MAJDAK, AND BERNHARD LABACK

Acoustics Research Institute, Austrian Academy of Sciences, Austria

Normal-hearing (NH) listeners use monaural spectral cues to localize sound sources in sagittal planes, including up-down and front-back directions. The salience of monaural spectral cues is determined by the spectral resolution and the dynamic range of the auditory system. Both factors are commonly degraded in impaired auditory systems. In order to simulate the effects of outer hair cell (OHC) dysfunction and loss of auditory nerve (AN) fibres on localisation performance, we incorporated a well-established model of the auditory periphery [Zilany *et al.*, 2014, J. Acoust. Soc. Am. 135] into a recent model of sound localisation in sagittal planes [Baumgartner *et al.*, 2014, J. Acoust. Soc. Am. 136]. The model was evaluated for NH listeners and then applied on conditions simulating various degrees of OHC dysfunction. The predicted localisation performance is hardly affected by a moderate OHC dysfunction but drastically degrades in case of a severe OHC dysfunction. When further applied on conditions simulating loss of AN fibres with specific spontaneous rates (SRs), predicted localisation performance degrades if only high-SR fibres are preserved.

INTRODUCTION

Monaural spectral cues enable sound localisation where binaural cues are ambiguous. This ambiguity occurs approximately in planes orthogonal to the interaural axis, the sagittal planes, and thus concerns localisation of both up-down and front-back directions. The mapping of direction-dependent spectral cues to perceived spatial direction is considered as being implemented in the auditory system as a comparison process between the incoming sound spectrum and learned spectral templates.

The extraction of spectral localisation cues relies on a proper functioning of the auditory periphery. Sensorineural hearing loss is known to degrade localisation performance especially within the sagittal planes (Otte *et al.*, 2013; Rakerd *et al.*, 1998). Degradation of localisation performance with high-frequency attenuation is relatively well understood (Baumgartner *et al.*, 2014; Best *et al.*, 2005), but is not sufficient to accurately explain individual localisation performance of hearing-impaired listeners (Noble *et al.*, 1994). Additional deficits that potentially affect localisation performance are a dysfunction of olivocochlear efferents for modulation of outer hair cell (OHC) activity causing broadened auditory filters (May *et al.*, 2004) as well as a noise-induced loss of auditory nerve (AN) fibres with low to moderate spontaneous

*Corresponding author: robert.baumgartner@oeaw.ac.at

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

firing rates (SRs) resulting in a reduced dynamic range of AN responses (Furman *et al.*, 2013). These deficits are, however, hard to assess and quantify for real listeners.

In order to better understand the effect of peripheral deficits on localisation performance in the sagittal planes, this study aimed at simulating their consequences by means of an auditory model. To this end, we integrated a model of the auditory periphery (Zilany *et al.*, 2009; 2014) into a model of sound localisation in sagittal planes (Baumgartner *et al.*, 2014). First, we evaluated the modified model for normal-hearing listeners, and then, investigated effects of OHC dysfunctions and loss of specific SR fibres at various sound pressure levels (SPLs).

METHODS

We use the interaural-polar coordinate system to describe auditory localization. The lateral angle in the horizontal plane ranges from −90° at the left hand side to 90° at the right hand side. The polar angle in the sagittal plane ranges from −90° to 270° with 0° corresponding to the front, 90° to the top, and 180° to the back of the listener.

Sagittal-plane localisation model

The model from Baumgartner *et al.* (2014) follows a template-based comparison procedure. First, the incoming sound and the template head-related transfer functions (HRTFs) are processed by an approximation of the auditory periphery in order to obtain internal excitation patterns. Then, across-frequency differentiation with 1-equivalent-rectangular-bandwidth spacing and restriction to positive gradients yields gradient profiles. This gradient extraction is a functional approximation of the rising spectral edge sensitivity observed in the dorsal cochlear nucleus of cats (Reiss and Young, 2005). The incoming sound is, then, compared to the template HRTFs on the basis of the corresponding gradient profiles. Finally, the model yields polar-angle response probabilities that can be used to calculate expectancy values of performance measures. More details about the model stages are described in Baumgartner *et al.* (2014).

Integration of auditory-periphery model

In Baumgartner *et al.* (2014), we used a linear Gammatone filterbank and were able to predict several effects of HRTF modifications and spectral variations of the sound source on localisation performance. A more realistic model of the auditory periphery is required for modelling level dependencies and the effects of individual hearing impairment. Thus, in the present study, we replaced the Gammatone filterbank by the humanized auditory-periphery model from Zilany *et al.* (2009; 2014), in the following called the Zilany model. In order to obtain internal excitation patterns, we temporally integrated the instantaneous firing rates from the synapse output and then averaged the outputs across fibre types according to their physiologically assumed frequency of 61% high-, 23% medium-, and 16% low-SR fibres (Liberman, 1978). Within the Zilany model, the frequency range from 0.7-18 kHz was represented by 100 auditory-nerve fibres, the internal sampling rate was 100 kHz, and the approximate implementation of the power-law functions was used.

Listeners and stimuli

We simulated 14 female and nine male listeners aged between 19 and 46 years. Simulated stimuli were Gaussian white noise bursts with a duration of about 170 ms. Informal tests have shown that the excitation patterns differed only marginally for longer bursts. Targets were simulated in the midsagittal plane and polar angles between −30° and 270°.

Simulated conditions of hearing impairments

We simulated three degrees of OHC dysfunction and three conditions of SR-specific activity of AN fibres at SPLs of 50 dB and 80 dB in each case (see Table 1). Assuming perfect adaptation to the hearing impairment, model templates were processed according to the same OHC dysfunction and fibre types.

Factor	Levels	Meaning
C_{OHC}	$C_{OHC} = 1.0$	Intact OHC functionality
	$C_{OHC} = 0.5$	Moderate OHC dysfunction
	$C_{OHC} = 0.1$	Severe OHC dysfunction
FT	LMH	Activity of low-, medium-, and high-SR fibres
	MH	Activity of medium-, and high-SR fibres
	H	Activity of only high-SR fibres

Table 1: Simulated conditions of impaired auditory peripheries.

Data analysis

Localisation performance was quantified by the following measures. The quadrant error rate (QER) is the percentage of polar-angle errors larger than 90°. The local polar error (LPE) is the root mean square (RMS) of polar-angle errors smaller than 90°. The polar gain is the slope of a linear regression line fitted separately to responses in the front and the rear hemispheres. Quasi-veridical responses are defined as responses deviating less than 45° from the regression line, and the variability is the RMS of deviations between quasi-veridical responses and the regression line.

In order to quantify the predictive power of the model, the following two metrics were used: e denotes the RMS difference between actual and predicted listener-specific performance measures, and r denotes Pearson's correlation coefficient between actual and predicted listener-specific performance measures.

For statical analysis of main effects, we performed 3-way repeated-measures analyses of variance (ANOVAs) with Greenhouse-Geisser correction for departure from sphericity. For post-hoc analysis, we used Tukey's honest significance difference test. All effects are reported as significant at $p < .05$.

RESULTS AND DISCUSSION

Model evaluation

First, we simulated three experimental conditions with moderate and approximately constant SPL from Baumgartner *et al.* (2014) and compared the predictive power of the new model with that from the Gammatone-based model. In order to reduce computational effort, simulations were only conducted for the median plane. Table 2 shows the results for a baseline condition testing localisation of Gaussian white noise bursts with reference HRTFs, a condition testing spectral warping of HRTFs in combination with band limitation (Majdak *et al.*, 2013), and a condition testing limited spectral resolution of HRTFs (Goupell *et al.*, 2010). Both model variants performed very similar, though the predictive power of the Gammatone-based model was slightly better in most cases.

Model of auditory periphery	Baseline				Warping				Resolution			
	QER (%)		LPE (deg)		QER (%)		LPE (deg)		QER (%)		LPE (deg)	
	e	*r*	*e*	*r*	*e*	*r*	*e*	*r*	*e*	*r*	*e*	*r*
Zilany	2.39	0.93	2.19	0.75	8.73	0.84	6.22	0.78	6.81	0.72	5.44	0.71
Gammatone	2.60	0.95	2.64	0.83	8.18	0.83	5.45	0.78	8.24	0.69	4.67	0.76

Table 2: Model performance for different approximations of the auditory periphery.

However, the Gammatone model cannot represent any SPL dependencies. In order to test the predictive power of the new model with respect to changes in SPL, we simulated the experiments from Sabin et al. (2005). Sabin et al. (2005) showed their results only as functions of the sensation level (SL), that is, the level relative to the detection threshold for a frontal target. For transferring our SPL-specific predictions to SLs, we generally assumed the SL to be 10 dB less than the SPL. Figure 1 shows the model predictions and the actual results replotted from Sabin et al. (2005). Predicted performance is more similar to the actual performance for the front than the rear. Especially in the front, the model predictions show a generally larger variability than the actual listeners from Sabin et al. (2005). The comparison of the quasi-veridical response rate for rear targets with an SL of less than 15 dB should be treated with caution because Sabin et al. (2005) considered a default regression line with a slope of one and intercept of 180° if less than five were audible and not confused between front and rear. In our simulations, we did not explicitly consider audibility and simulated many more trials. Hence, there were always enough responses to estimate a regression line.

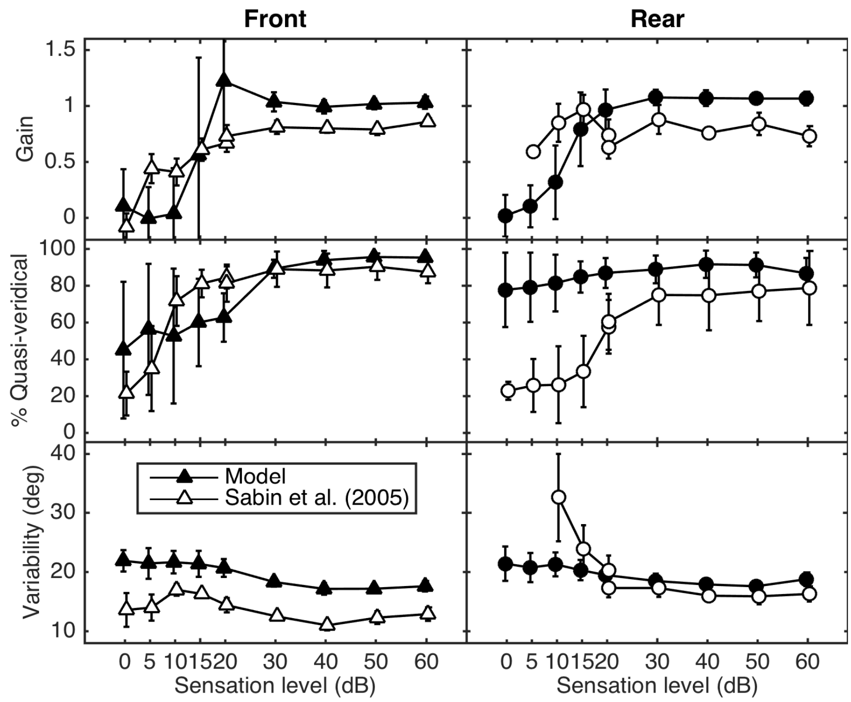


Fig. 1: Results of modelling the level dependence. Error bars represent standard deviations. The data are slightly shifted along the abscissa for better visibility. Results of two experimental blocks from Sabin *et al.* (2005) are shown at SL of 20 dB.

Effects of OHC dysfunction and SR-specific loss of AN fibres

Predicted localisation performance for impaired auditory peripheries is shown in Fig. 2. There are significant main effects of the OHC dysfunction for QER [$F(1.40,30.81) = 405.57$] and LPE [$F(1.54,33.87) = 321.07$]. Paired comparisons revealed that QER and LPE are significantly larger for the severe dysfunction than for the moderate dysfunction and intact functionality. The moderate dysfunction and the intact functionality are not significantly different for QER but different for LPE. The main effects of fibre type activity (FT) are also significant for QER [$F(1.92,42.19) = 840.74$], and LPE [$F(1.25,27.51) = 691.76$]. In paired comparisons, all FT conditions are significantly different to each other. MH performed best, LMH slightly worse, and H worst. There is no significant main effect of the SPL for QER [$F(1,22) = 0.78$], but LPE increases with SPL [$F(1,22) = 13.46$]. There are significant interactions between OHC dysfunction and SPL both for QER [$F(1.63,35.60) = 535.17$] and LPE [$F(1.57,34.59) = 170.7$], as well as between OHC dysfunction and FT also both for QER [$F(2.80,61.55) = 249.52$] and LPE [$F(2.63,57.94) = 188.53$]. The QER and LPE degradation caused by severe OHC dysfunction is less severe for louder sounds and if only high-SR fibres are activated.

The ability of the AN fibre population to accurately represent spectral cues depends on the dynamic range they are able to capture. In order to obtain estimates of the

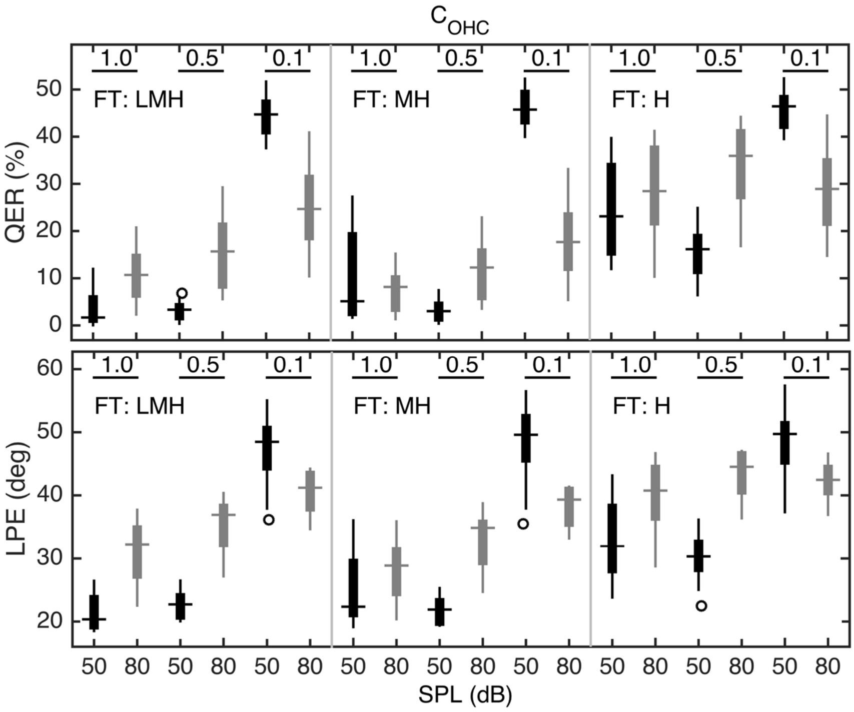


Fig. 2: Effects of OHC dysfunctions and SR-specific loss of AN fibres. Horizontal line: mean. Thick bar: interquartile range (IQR). Thin bar: data range within 1.5 IQR. Open circle: outlier.

dynamic ranges, we simulated AN responses to broadband noise at various SPLs in steps of 10 dB, averaged the predicted firing rates across the relevant frequency range between 700 Hz and 18 kHz, and finally differentiated these average rates across the SPL steps. The simulated rate differences are shown in Fig. 3 for the different fibre type combinations and degrees of OHC dysfunction. The range of rate differences being larger than zero indicates the dynamic range of the AN fibre population and the maximum rate difference indicates best sensitivity. Dynamic range bounds and sensitivity maxima are predicted to increase with OHC dysfunction. Within certain degrees of OHC dysfunction, predicted rate-difference curves are slightly shallower and peak at larger SPLs, the more fibre types are preserved. The consequently slightly larger dynamic range in the LMH and MH conditions compared to the H condition for the intact OHC functionality appears important to accurately represent the templates at 80 dB SPL with a level variability of at least 40 dB across frequencies and directions and thus explains that the model predicts degraded performance if only high-SR fibres are preserved. Very similar dynamic ranges but mostly higher sensitivities might be the reason for the slightly better performance in the MH condition compared to the LMH condition.

Localisation performance was shown to be relatively robust with respect to moderate OHC dysfunction, but, especially for silent sounds, a severe OHC dysfunction drastically degraded the performance. Figure 4 shows the positive spectral gradient profiles as functions of the polar angle for intact OHC functionality as well as moderate

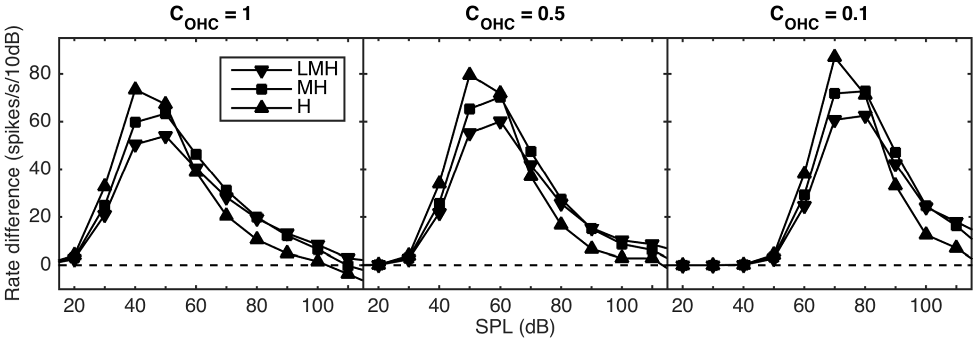


Fig. 3: Rate difference vs. level functions predicted for different combinations of fibre types and OHC dysfunctions. Frequency range: 0.7-18 kHz.

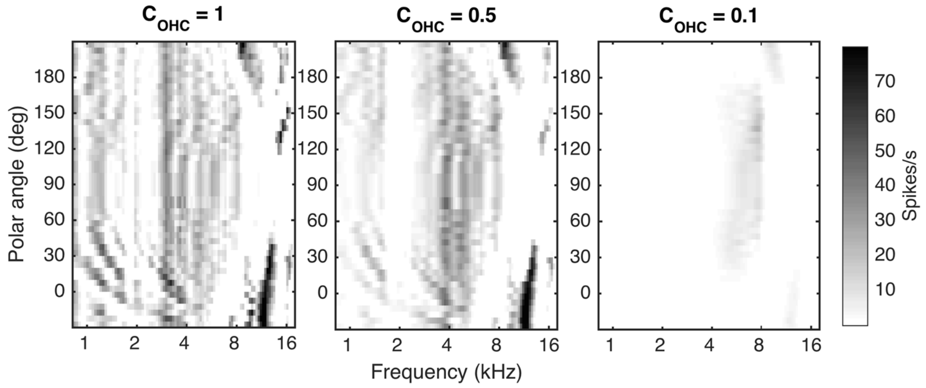


Fig. 4: Effect of OHC dysfunction on positive spectral gradients. Simulation for 50 dB SPL and all fibre types preserved (FT: LMH). Note that spectral gradients are almost absent for the most severe OHC dysfunction ($C_{OHC} = 0.1$).

and severe OHC dysfunction from left to right. While the profiles for the intact and moderate cases are very similar and reveal direction-specific patterns, there are only very few and small gradients with little directionality in the severe case. In this severe condition, auditory filters are too broad to resolve most of the spectral information. Note that our investigations focussed on listening conditions without any background noise. The presence of background noise might increase the importance of proper OHC functionality (May *et al.*, 2004).

CONCLUSIONS

In order to study the effect of peripheral hearing disorders on sound localisation in quiet in sagittal planes, we integrated the auditory periphery model from Zilany *et al.* (2009; 2014) into the sagittal-plane sound localisation model from Baumgartner *et al.* (2014). The initial model evaluation for normal-hearing listeners showed that replacing the Gammatone filterbank with the nonlinear Zilany model preserves the predictive power of the localisation model and enables level-dependent simulations. The model predicts poor localisation performance if only high-SR fibres are preserved in the auditory periphery and low- and medium-SR fibres are lost because then the dynamic range of AN responses is too limited to represent HRTF-filtered sounds at various SPLs. Model predictions suggest that OHC dysfunctions are critical only if the dysfunction is quite severe. Moderate OHC dysfunctions seem to provide auditory filters sharp enough to capture spectral cues.

ACKNOWLEDGMENTS

Thanks to Christian Kaseß for assistance with the statistical analyses.

REFERENCES

Baumgartner, R., Majdak, P., and Laback, B. (**2014**). "Modeling sound-source localization in sagittal planes for human listeners," J. Acoust. Soc. Am., **136**, 791-802.

Best, V., Carlile, S., Jin, C., and van Schaik, A. (**2005**). "The role of high frequencies in speech localization," J. Acoust. Soc. Am., **118**, 353-363.

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.

Goupell, M.J., Majdak, P., and Laback, B. (**2010**). "Median-plane sound localization as a function of the number of spectral channels using a channel vocoder," J. Acoust. Soc. Am., **127**, 990-1001.

Liberman, M.C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am., **63**, 442-455.

Majdak, P., Walder, T., and Laback, B. (**2013**). "Effect of long-term training on sound localization performance with spectrally warped and band-limited head-related transfer functions," J. Acoust. Soc. Am., **134**, 2148-2159.

May, B.J., Budelis, J., and Niparko, J.K. (**2004**). "Behavioral studies of the olivocochlear efferent system: Learning to listen in noise," Arch. Otolaryngol. Neck Surg., **130**, 660-664.

Noble, W., Byrne, D., and Lepage, B. (**1994**). "Effects on sound localization of configuration and type of hearing impairment," J. Acoust. Soc. Am., **95**, 992-1005.

Otte, R.J., Agterberg, M.J.H., Wanrooij, M.M. V., Snik, A.F.M., and Opstal, A.J.V. (**2013**). "Age-related hearing loss and ear morphology affect vertical but not horizontal sound-localization performance," J. Assoc. Res. Otolaryngol., **14**, 261-273.

Rakerd, B., Vander Velde, T.J., and Hartmann, W.M. (**1998**). "Sound localization in the median sagittal plane by listeners with presbyacusis," J. Am. Acad. Audiol., **9**, 466-479.

Reiss, L.A.J. and Young, E.D. (**2005**). "Spectral edge sensitivity in neural circuits of the dorsal cochlear nucleus," J. Neurosci., **25**, 3680-3691.

Sabin, A.T., Macpherson, E.A., and Middlebrooks, J.C. (**2005**). "Human sound localization at near-threshold levels," Hear. Res., **199**, 124-134.

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

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.

Auditory-model based assessment of the effects of hearing loss and hearing-aid compression on spectral and temporal resolution

BORYS KOWALEWSKI[1,*], EWEN MACDONALD[1], OLAF STRELCYK[2], AND TORSTEN DAU[1]

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

[2] *Sonova U.S. Corporate Services, Warrenville, IL, USA*

Most state-of-the-art hearing aids apply multi-channel dynamic-range compression (DRC). Such designs have the potential to emulate, at least to some degree, the processing that takes place in the healthy auditory system. One way to assess hearing-aid performance is to measure speech intelligibility. However, due to the complexity of speech and its robustness to spectral and temporal alterations, the effects of DRC on speech perception have been mixed and controversial. The goal of the present study was to obtain a clearer understanding of the interplay between hearing loss and DRC by means of auditory modeling. Inspired by the work of Edwards (2002), we studied the effects of DRC on a set of relatively basic outcome measures, such as forward masking functions (Glasberg and Moore, 1987) and spectral masking patterns (Moore *et al.*, 1998), obtained at several masker levels and frequencies. Outcomes were simulated using the auditory processing model of Jepsen *et al.* (2008) with the front end modified to include effects of hearing impairment and DRC. The results were compared to experimental data from normal-hearing and hearing-impaired listeners.

INTRODUCTION

Many studies have investigated whether amplification with multi-channel compression can be beneficial for speech intelligibility compared to linear amplification. While some studies have reported that multi-channel compression provides an advantage (Moore *et al.*, 1999; Souza and Bishop, 1999; Souza and Turner, 1999), others have reported no benefit or even a detrimental effect relative to linear gain (Franck *et al.*, 1999; Stone *et al.*, 1999). Thus, the benefit of multi-channel compression for improving speech intelligibility remains unclear.

Edwards (2002) suggested using a set of relatively simple outcome measures, based on narrowband signals, for the evaluation of hearing-aid signal processing. This also allows the use of computational models of the auditory system, which have recently been proven to be able to account for the effects of sensorineural hearing loss on

*Corresponding author: bokowal@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

auditory signal detection and discrimination (Jepsen and Dau, 2011; Panda *et al.*, 2014).

The purpose of the current study was to follow this approach in a systematic manner. Two psychoacoustic experiments were chosen, one to evaluate temporal resolution and the other to evaluate spectral resolution. Both normal-hearing (NH) and hearing-impaired (HI) subjects were tested to evaluate how hearing loss affects these outcome measures. In addition to behavioral experiments, the results were also simulated in a computational model of the auditory system that can account for detection and masking data from NH (Jepsen *et al.*, 2008) and HI (Jepsen and Dau, 2011) listeners. The modeling framework allowed the evaluation of multi-channel dynamic range compression without tedious retesting. To simulate an aided-impaired auditory system, a preprocessing stage was added to the model.

METHODS

Subjects

Three NH subjects, aged 24-29 and three HI subjects, aged 69-74, were tested. The hearing loss of all HI subjects was mild to moderate and sensorineural in nature.

Stimuli and procedure

The decay of forward masking (Glasberg *et al.*, 1987) and spectral masking patterns (Moore *et al.*, 1998) were measured in both subject groups in the unaided condition.

In the first experiment, the masker was a band of noise, centered either at 1 or 4 kHz with a bandwidth of 500 and 2000 Hz, respectively. Its sound pressure level (SPL) was fixed at 75 dB and its duration was 220 ms including a 10-ms rise and a 5-ms fall raised-cosine ramp. The probe was a short, 20-ms long pure tone gated for its entire duration (no steady state) at the masker central frequency. Different time intervals between the offset of the masker and the onset of the probe (zero voltage points) were tested. Positive offset values therefore relate to a purely forward masking condition, whereas the negative values mean that there was either a full or partial temporal overlap of the probe and the masker (simultaneous masking).

In the second experiment, the masker was an 80-Hz-wide band of noise centered either at 1 or 4 kHz. Its level was fixed at 75 dB SPL and its duration was 220 ms including 10-ms rise and fall raised-cosine ramps. The probes were pure tones at various frequencies, gated simultaneously with the masker.

In both experiments, the level of the probe at each frequency was varied adaptively using a 3AFC 1-up-2-down paradigm, until the masked threshold was reached. Absolute thresholds (in the absence of any masker) were also measured.

Additionally, for the HI subjects, temporal masking curves (TMCs; Nelson *et al.*, 2001) were measured in order to obtain basilar membrane input/output functions (BMIO). The probe was a 16-ms long pure tone gated for its entire duration (no steady state portion) at the frequency of interest (1 or 4 kHz) with a fixed level of 8 dB sensation level (SL), based on a prior measurement. For each probe, the masker was a 220 ms-long (gated with 8-ms rise and fall raised-cosine ramps) pure tone at the probe frequency (on-frequency condition) or one octave below (off-frequency condition). The masker-signal temporal gap and the masker level were varied adaptively in two dimensions using the Grid method (Fereczkowski *et al.*, 2015). Plotting the measured off-frequency versus the on-frequency TMC, a behavioural estimate of the BMIO at the probe frequency was obtained.

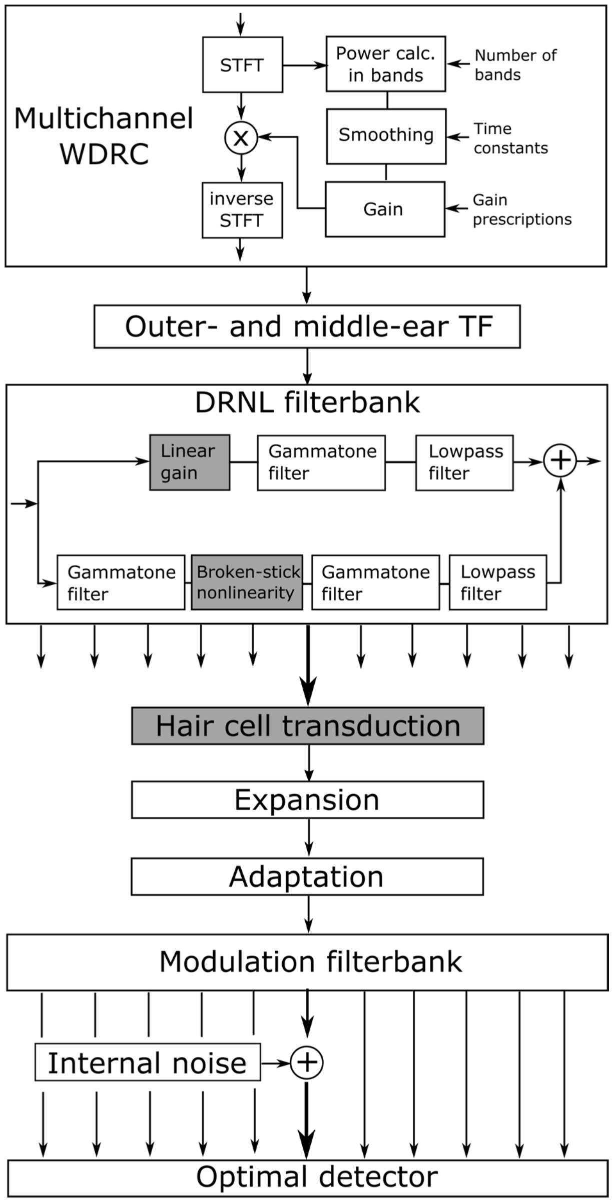


Fig. 1: Structure of the CASP model (Jepsen and Dau, 2011) with the preprocessing stage simulating a hearing-aid multi-channel wide dynamic range compression system.

Model of the auditory system

Simulations of the first two experiments were performed in the unaided condition for the NH and the HI listeners and in the aided condition for the HI only. The CASP model (Jepsen *et al.*, 2008) was used. The model structure is shown in Fig. 1. To simulate the effects of individual hearing loss, changes were made in the dual-resonance nonlinear (DRNL) filterbank (Lopez-Poveda and Meddis, 2001) and the inner hair cell (IHC) stage, based on the TMC data (Jepsen and Dau, 2011). First, the parameters of the DRNL broken-stick nonlinearity and the linear path gain at 1 and 4 kHz were adjusted to best fit the measured BMIO. Then a linear interpolation of the parameters across frequencies was performed. The IHC loss was estimated as the difference between the total loss (from the audiogram) and the OHC loss (inferred from the fitted DRNL input-output function).

Hearing-aid simulator

A multi-channel DRC hearing-aid simulator was developed for this study (Fig. 1). In the simulator, the input signal is broken down into time-frequency units using the short time Fourier transform (STFT). Then, for each time slice, the fast Fourier transform (FFT) bins are assigned to bands (channels). The spacing was quasi-logarithmic, based on the equivalent-rectangular-bandwidth scale. In this study, the number of channels was set to 19. In each frequency band, the power was calculated. Then the power was smoothed by a one-pole filter using the desired time constants. Attack and release times were set to 1 and 10 ms, respectively (RC time constants). Based on the smoothed power estimate, a gain matrix was calculated. The amount of gain in each channel depends on the prescription. Here, the NAL-NL1 targets were used as a starting point. The band-specific values were projected back onto the original FFT bins. The resulting gain matrix was then multiplied by the STFT representation of the input signal and STFT synthesis was performed to obtain the output time signal.

RESULTS

Figure 2 shows the masked thresholds for the decay of forward masking. Figure 3 shows the masked thresholds for the simultaneous spectral masking patterns.

In both figures, the left column shows results for the masker centered at 1 kHz while the right column indicates results for the 4-kHz case. The average data and model simulations for the NH listeners are shown in the top panel. The consecutive three panels show individual unaided data with both unaided and aided model simulations.

The absolute thresholds (in absence of the masker) are not shown in the figures. Due to hearing loss, for all of the HI subjects, these thresholds are elevated. Introducing hearing-aid amplification leads to a significant decrease in these thresholds.

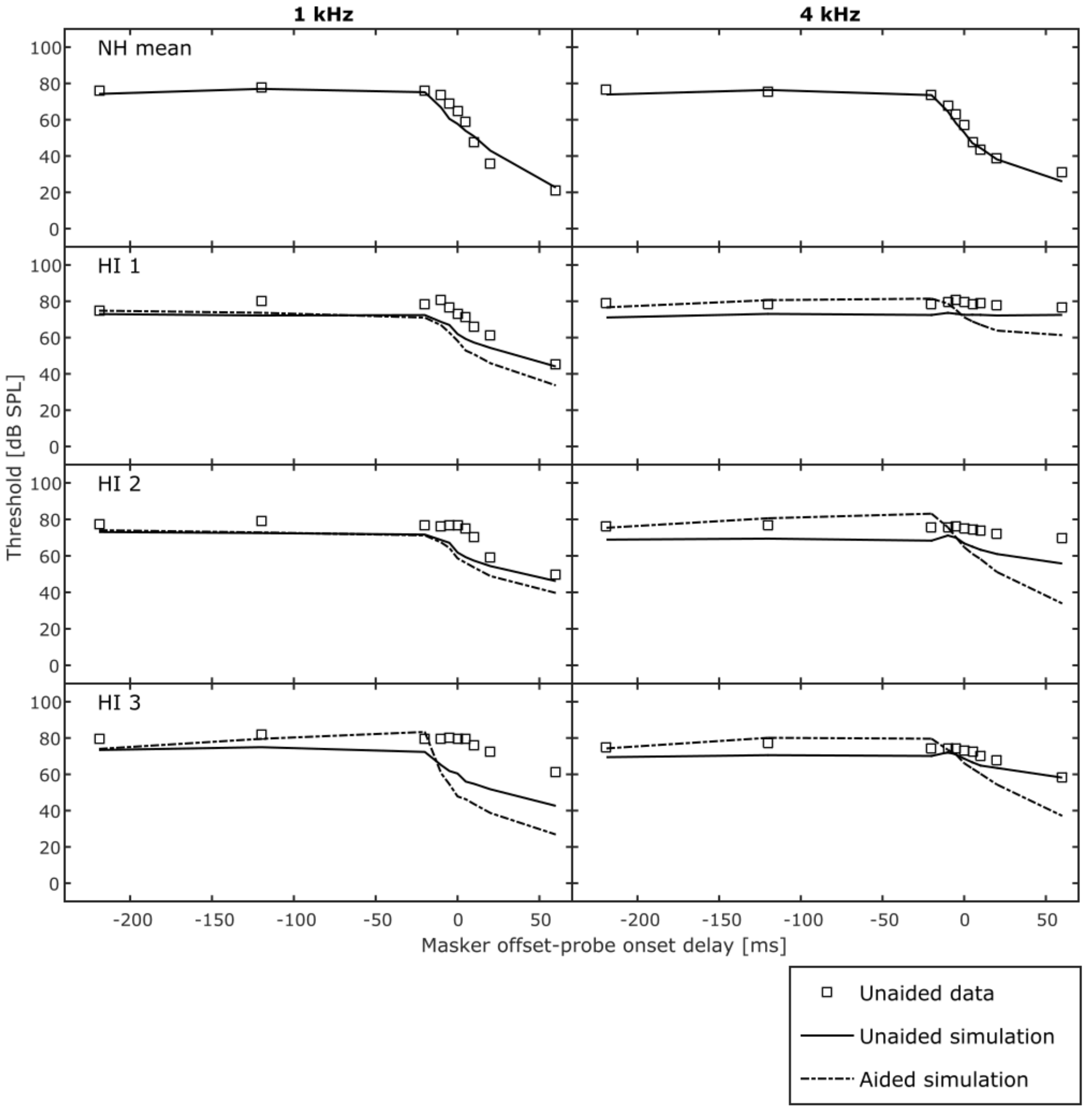


Fig. 2: Masked thresholds in the decay of forward masking experiments. Left and right panels show results for probe frequency of 1 and 4 kHz respectively. The top panel shows average unaided data and model simulations for the NH subjects. The consecutive panels show the unaided data with model simulations and aided simulations individually for each of the HI subjects.

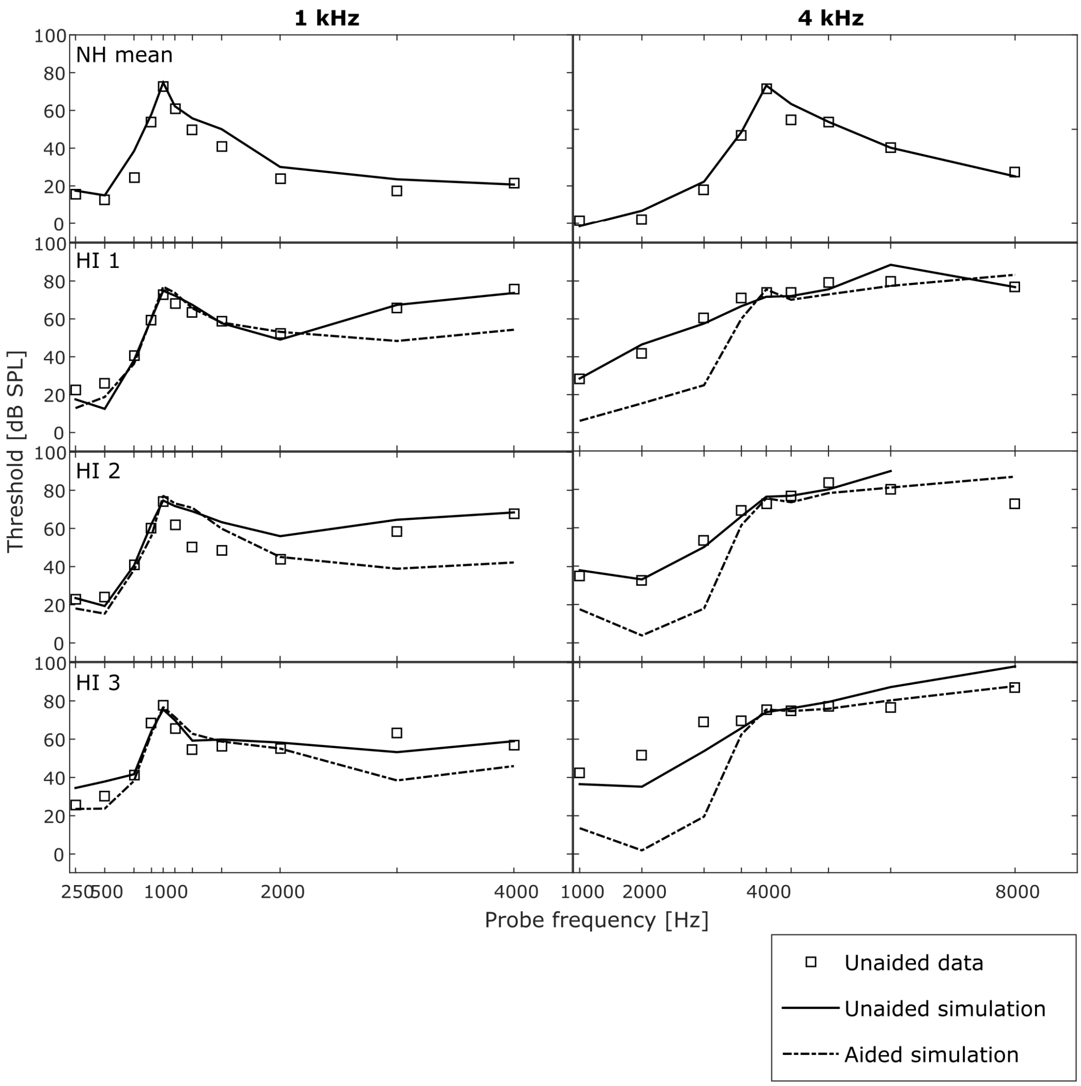


Fig. 3: Masked thresholds in the simultaneous spectral masking experiments. Left and right panels show results for probe frequency of 1 and 4 kHz respectively. The top panel shows average unaided data and model simulations for the normal hearing listeners. The consecutive panels show the unaided data with model simulations and aided simulations individually for each of the hearing-impaired listeners.

DISCUSSION

Decay of forward masking

Relative to the NH data, the HI listeners show elevated absolute thresholds (for detection of the short probe in silence) and a slower rate of recovery from forward masking. There is no elevation of the masked threshold in the full-overlap region (−219 to −20 ms). At 1 kHz, the audibility of the probe does not seem to be the only factor limiting the decay. However, at 4 kHz for HI1 and HI2, the curves are approximately at the level of the absolute threshold (not shown) regardless of the masker-probe separation. Simulated and measured decay curves have a similar dynamic range (except for HI3 at 1 kHz) but the model generally performs better, which is also the case for several subjects in Jepsen and Dau (2011). Introducing compression results in faster rates of recovery.

Spectral masking patterns

At 1 kHz, the lower skirt of the masking pattern does not differ significantly from the normal shape. However, the upper skirt is elevated. Only a part of this can be attributed to increased absolute thresholds. The rest indicates an increase in the upward spread of masking, related to broadening of auditory filters. In most cases, the model predicts the unaided masked threshold data. The thresholds are overestimated for probes above the masker frequency for subject HI2 at 1 kHz and subject HI3 at 4 kHz and underestimated below the masker frequency for HI3 at 4 kHz. Introducing multi-channel compression with the 1-kHz-centered masker leads to an improved performance only for the probe frequencies above 2 kHz. With the 4-kHz masker, the most significant improvement (decrease) in masked thresholds occurs for probe frequencies below the masker central frequency. There is also a slight reduction in the masking effect at probe frequencies above 4 kHz. In all cases, the absolute thresholds are significantly lower. Therefore, if the amount of masking (the difference between the masked and absolute thresholds) was plotted, the multi-channel compression would appear to restore the masking patterns to a more normal shape. However, this restoration would be due to the increased audibility and could be achieved using only linear gain.

CONCLUSIONS

Sensorineural hearing loss results in a decreased audibility of pure tones, slower rate of decay of forward masking and flattened spectral masking patterns, consistent with earlier studies. Multi-channel compression appears capable of restoring, to some extent the performance of HI subjects in the above-mentioned tasks back to normal. More research is needed to disentangle the effects of linear (increased audibility) and level-dependent gain.

ACKNOWLEDGMENTS

This research is supported by Phonak and the Technical University of Denmark. The contribution was also supported by the ISAAR Scholarship. The authors would like to thank Michal Fereczkowski and Henrik Gert Hassager for help in data collection and analysis.

REFERENCES

Edwards, B. (**2002**). “Signal processing, hearing aid design, and the psychoacoustic Turing test,” Proc. International Conference on Acoustics, Speech, and Signal Processing, 2002, Orlando, Florida.

Fereczkowski, M., Kowalewski, B., Dau, T., and MacDonald, E.N. (**2015**). “Time-efficient multidimensional threshold tracking method,” J. Acoust. Soc. Am., **137**, 2228-2228.

Franck B.A., van Kreveld-Bos C.S., Dreschler, W.A., and Verschuure, H. (**1999**). “Evaluation of spectral enhancement in hearing aids, combined with phonemic compression,” J. Acoust. Soc. Am., **106**, 1452-1464.

Glasberg, B.R., Moore, B.C.J., and Bacon, S. (**1987**). “Gap detection and masking in hearing-impaired and normal-hearing subjects,” J. Acoust. Soc. Am., **81**, 1546-1556.

Jepsen, M.L., Ewert, S., 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.

Lopez-Poveda, E.A. and Meddis, R. (**2001**). “A human nonlinear cochlear filterbank,” J. Acoust. Soc. Am., **110**, 3107-3118.

Moore, B.C.J., Alcántara, J. I., and Dau. T. (**1998**). “Masking patterns for sinusoidal and narrow-band noise maskers,” J. Acoust. Soc. Am., **104**, 1023-1038.

Moore, B.C.J., Peters, R.W., and Stone, M.A. (**1999**). “Benefits of linear amplify-cation and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips,” J. Acoust. Soc. Am., **105**, 400-411.

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.

Panda, M.R., Lecluyse, W., Tan, C.M., Jürgens, T., and Meddis, R. (**2014**). “Hearing dummies: Individualized computer models of hearing impairment,” Int. J. Audiol., **53**, 699-709.

Souza, P.E. and Bishop, R.D. (**1999**). “Improving speech audibility with wide dynamic range compression in listeners with severe sensorineural loss,” Ear Hearing, **20**, 461-470.

Souza, P.E. and Turner, C.W. (**1999**). “Quantifying the contribution of audibility to recognition of compression-amplified speech,” Ear Hearing, **20**, 12-20.

Stone, M.A., Moore, B.C.J., Alcántara, J.I., and Glasberg, B.R. (**1999**). “Comparison of different forms of compression using wearable digital hearing aids,” J. Acoust. Soc. Am., **106**, 3603-3619.

Simulating hearing loss with a transmission-line model for the optimization of hearing aids

PETER VAN HENGEL*

INCAS³, Assen, The Netherlands

Modern hearing aids provide many parameters that can be adjusted to optimize the hearing experience of the individual user. Optimization of these parameters can be based on a comparison of an internal representation of sound processed by the hearing aid and the impaired hearing system with the representation in a non-impaired ear. Models that can represent the most common types of hearing loss and can be adjusted to fit individual hearing loss can play a crucial role in such optimization procedures. Simulations are presented that show the potential of a transmission line model in such a procedure. The model is extended to remap cochleogram energy based on estimations of the local instantaneous frequency. This 'remapping' of the cochleogram gives an advantage in tone-in-noise detection that may be related to neural deafferentation.

INTRODUCTION

Modern hearing aids often contain multiple listening programs for different situations. Each of these programs contains a multitude of parameters that can be adjusted. Finding optimal values for all these parameters often requires more time than is available in a clinical setting of, e.g., an audiological center. To facilitate the optimization process most manufacturers offer first fit settings. Based on the feedback from the user, parameters are adjusted to arrive at an individualized setting. Several contributions in these proceedings address the difficulties in arriving at optimal, or even satisfactory, settings in this manner (e.g., Edwards, 2015).

As described by Biondi (1978), computational models can play an important role in the optimization of parameters in a first fit procedure by simulating the effects of hearing impairment. As shown in Fig. 1, the use of such a computational model allows the comparison of the internal representation of a sound in a normal hearing ear (top path) with the internal representation of the same sound processed first by a hearing aid (prosthesis) and then by a hearing impaired ear.

Optimization of the parameter set (P) of the hearing aid now can be formulated as

$$P_{opt} = \arg\min \varepsilon_2 \qquad \text{(Eq. 1)}$$

where ε_2 is some difference measure. The parameter set P_{opt} that minimizes this distance gives the optimal setting of the hearing aid.

*Corresponding author: petervanhengel@incas3.eu

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

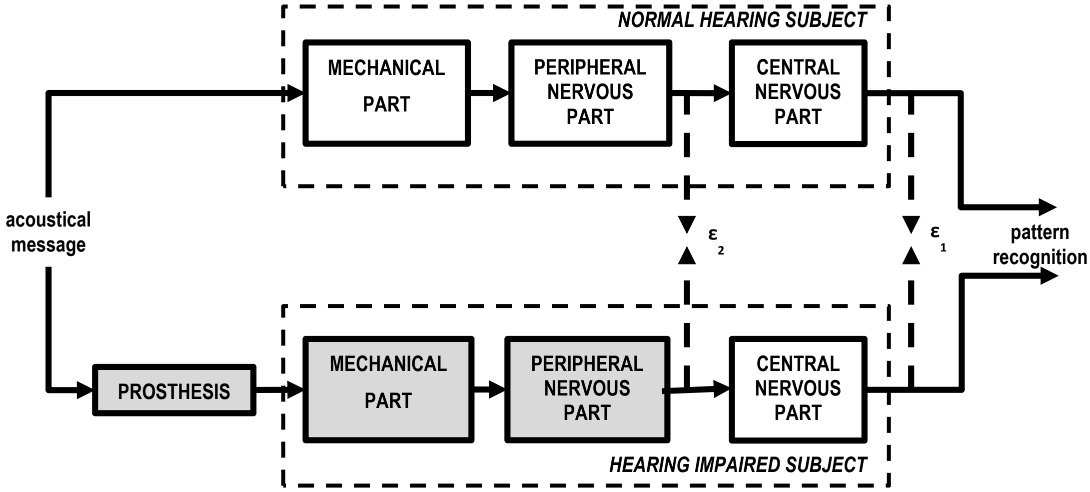


Fig. 1: Block diagram for an optimization procedure for first settings of hearing aids. Adapted from fig 2 of Biondi (1978). The top branch shows the processing of sound in a normal hearing model, the bottom branch in a hearing impaired model with hearing aid. Shaded parts indicate differences with the normal hearing model.

Numerous nonlinear cochlea models have been developed for normal hearing (reviewed, e.g., by Lopez-Poveda, 2005) that could be used. Many of these models offer the possibility to generate internal representations and include options for modelling cochlear hearing loss. Most are implemented as strictly one-way processing. This implies that they do not offer the possibility to simulate otoacoustic emissions (OAEs). Since OAEs are expected to provide essential objective information about active processing in the cochlea the present study uses a transmission-line cochlea model. This model was originally described by Duifhuis *et al.* (1985) and has been used to simulate a variety of physiological and psychoacoustic data (e.g., Epp *et al.*, 2010).

METHODS

Model parameter settings

The parameter settings were taken from Epp *et al.* (2010), including the Greenwood place-frequency map, the Zweig-impedance function (Zweig, 1991) with 1% roughness in the stiffness term and a double Boltzmann nonlinearity in the damping term *d* as well as the delayed feedback stiffness term *s'* (see Epp *et al.*, 2010, for details).

Types of hearing loss simulated

Initially, two types of hearing loss were considered:

- a loss of mechanical to neural conversion, referred to as type I loss;
- a loss of cochlear amplification, referred to as type II loss.

Type I loss was implemented by multiplication of the simulated cochlear partition excitation with an attenuation factor:

$$\vartheta(x) = 1 - D\,e^{-\left(\frac{x-0.5}{0.1}\right)^2} \qquad \text{(Eq. 2)}$$

where x is the relative (longitudinal) distance along the cochlea.

Equation 2 describes an attenuation or loss factor ϑ varying as a function of the (scaled) cochlear length x. In this case, the loss is centred at half of the cochlear length and has a width corresponding to 10% of the cochlear length. Only the depth D is varied in the simulations.

Type II loss was implemented by regarding the nonlinear damping d and feedback stiffness s' (see Epp *et al.*, 2010, for details) as a combination of a passive linear part and an active nonlinear part.

$$d = (1 - \vartheta(x))\,d_{NL} + \vartheta(x)\,d_L \qquad \text{(Eq. 3)}$$

$$s' = (1 - \vartheta(x))\,s'_{NL} \qquad \text{(Eq. 4)}$$

with $\vartheta(x)$ as in Eq. 2.

An additional stage was added to the output of the model. Following the work of Violanda *et al.* (2009), the phase information of the cochlear partition movement was used to extract the local instantaneous frequency at each oscillator in each time frame. Instead of the phase extraction method described in their work, zero crossings of both velocity and displacement were used to estimate local instantaneous frequencies. No correction for group delay was made. Excitation values were remapped to oscillators with resonance frequencies that were closest to the local instantaneous frequencies that were found. A third type of hearing loss – type III loss – could now be simulated, consisting of a loss of this 'remapping' stage.

Output measures

Three types of output were generated from the model to simulate data that can be obtained from hearing impaired subjects in a clinical setting:

- pure tone audiograms;
- distortion-product OAE (DPOAE) levels;
- tone-in-noise detection thresholds.

The pure tone audiograms were computed by comparing the excitation – in the case of loss of mechanical to neural conversion after attenuation – for a 50-ms sinusoid with a 20-ms rising window, to a fixed threshold 3 dB above the excitation for a normal-hearing model without roughness. An iterative procedure was used to find the level of the sinusoid required to match the threshold within 1 dB. Thresholds were computed for frequencies from 500 Hz to 8 kHz using 50 points per octave.

Previous work on simulating hearing loss using the transmission line cochlea model indicated that DPOAE levels were affected by simulated hearing loss (Mauermann *et al.*, 1999). Therefore, DPOAE levels were computed with a simulated probe in the ear canal for the $2f_1$-f_2 component in the range from 500 Hz to 8 kHz using 50 points per octave. The primary levels were at 20 dB and f_2/f_1 was fixed at 1.2.

Tone-in-noise detection thresholds were computed by adding the 50-ms sinusoids (as used for the pure tone audiograms) to a 50 ms snippet of the threshold equalizing noise (TEN, Moore *et al.*, 2000) at a fixed level of 60 dB. Detection thresholds were set at 4 standard deviations above the average noise level, where mean and standard deviations were computed for each oscillator over 50 random 50-ms snippets of the noise. An iterative procedure was used to find signal to noise levels required to match the threshold within 0.1 dB.

RESULTS

Figure 2 shows the pure tone audiograms for simulated loss of type I using values for *D* of 0, 20%, and 100%. The fact that thresholds are found for all frequencies with *D*=100% is due to off-frequency listening. No DPOAEs or tone in noise detection thresholds were calculated for this type of loss, since neither will be affected by an attenuation of the excitation profiles.

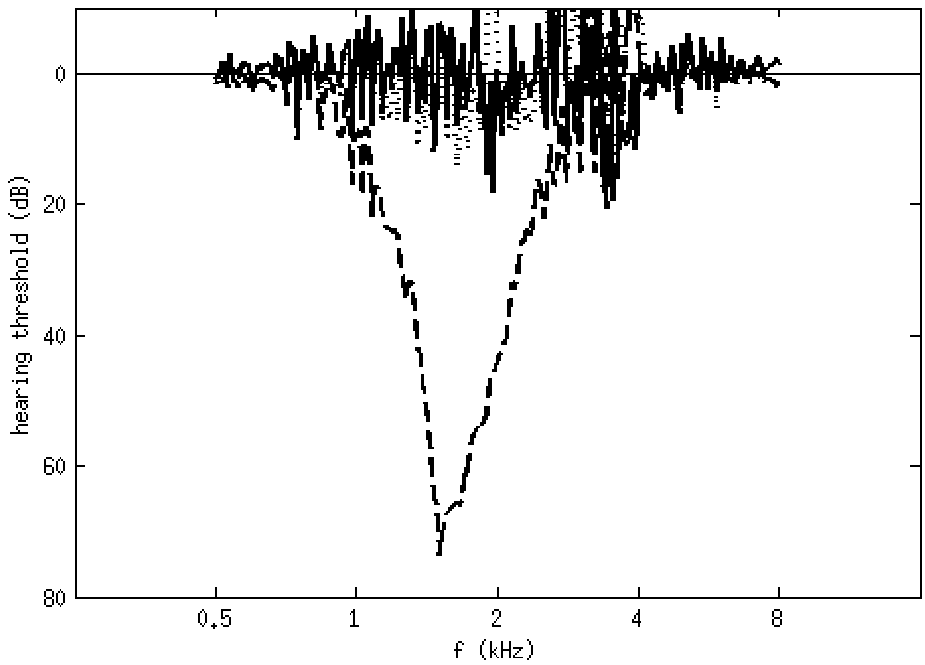


Fig. 2: Pure tone thresholds for type I loss. *D*=0 (solid line), *D*=20% (dotted line), and *D*=100% (dashed line).

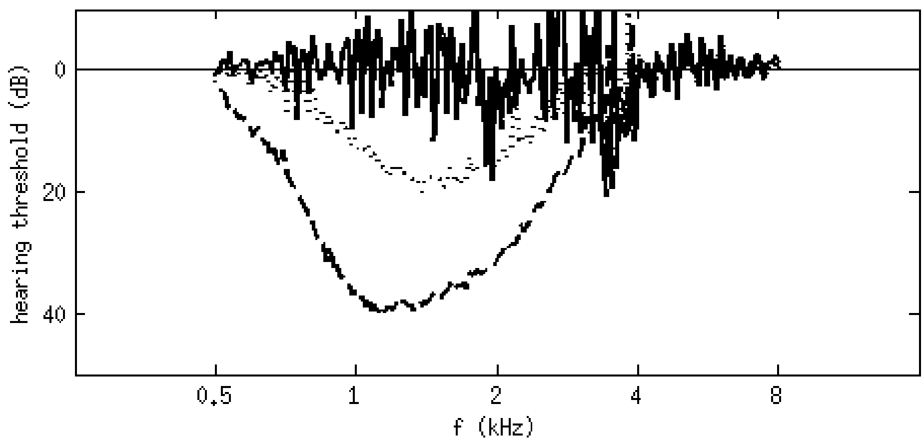


Fig. 3: Pure tone thresholds for type II loss. *D*=0 (solid line), *D*=20% (dotted line), and *D*=100% (dashed line).

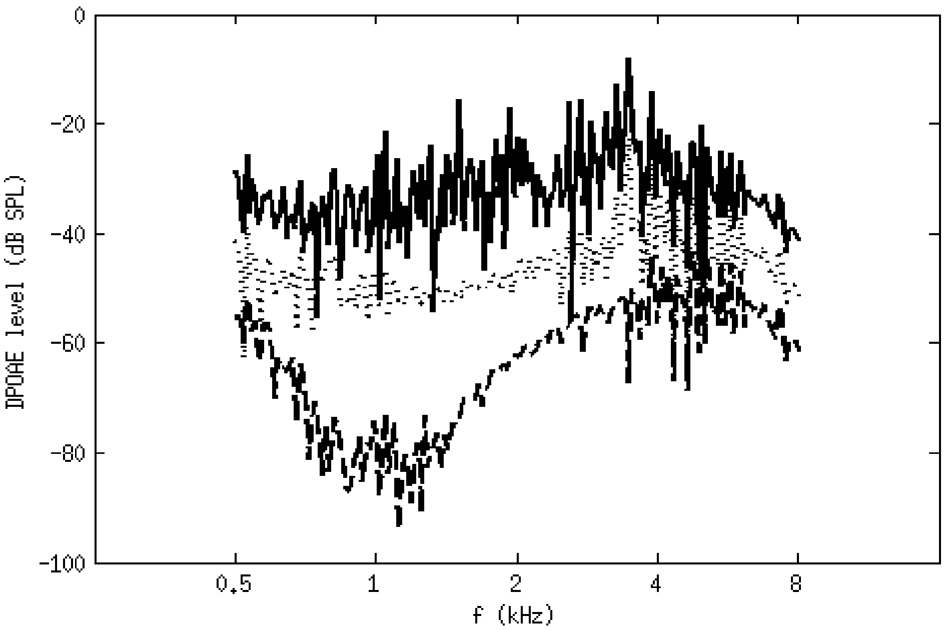


Fig. 4: $2f_1$-f_2 DPOAE levels for type II loss. *D*=0 (solid line), *D*=20% (dotted line), and *D*=100% (dashed line). Lines for *D*=20% and *D*=100% are offset by −10 dB and −20 dB respectively for reasons of clarity.

The curve for no loss (0%) shows threshold fine structure, as is expected from a model with Zweig impedance and roughness (see Epp *et al.*, 2010). At 20% loss the fine structure changes slightly but there is no overall level effect. At 100% loss, the threshold overall level and fine structure are affected, as expected.

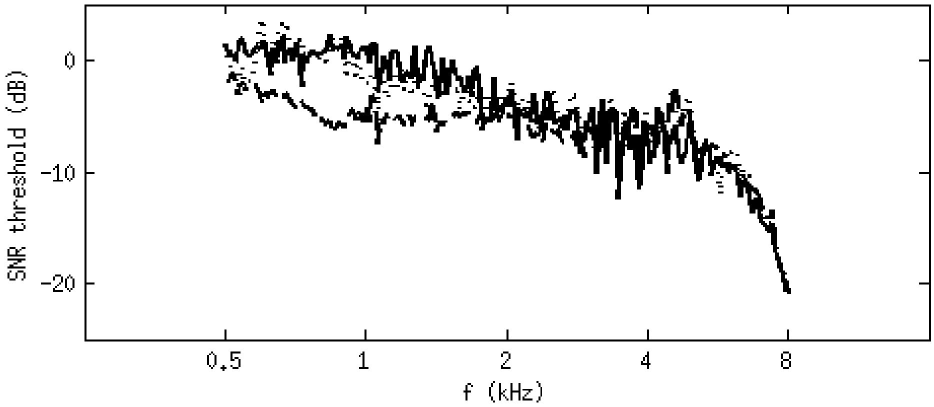


Fig. 5: Tone-in-noise detection thresholds for type II loss. *D*=0 (solid line), *D*=20% (dotted line), and *D*=100% (dashed line).

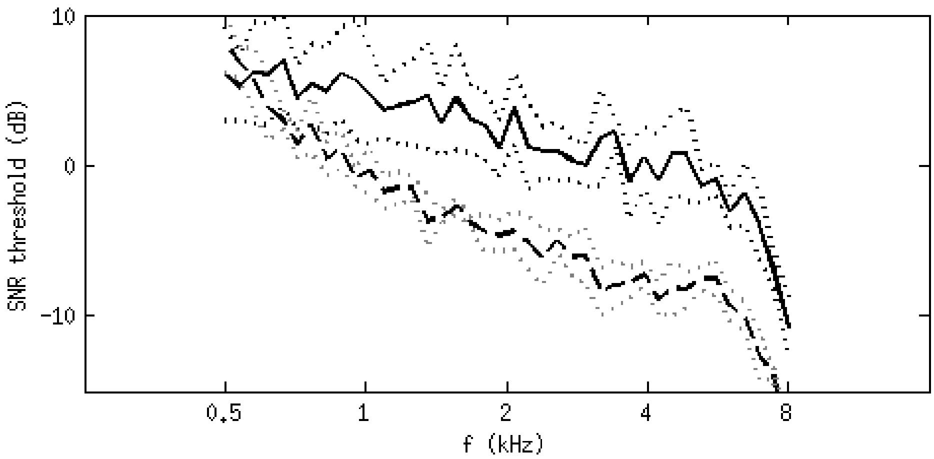


Fig. 6: Tone-in-noise detection thresholds based on excitation profiles without remapping (solid line) and with frequency remapping (dashed line). 20 repetitions were computed for each of the curves showing the mean. Dashed lines indicate standard deviations relative to the mean.

Figures 3, 4, and 5 show the pure tone audiogram, DPOAE levels, and tone-in-noise de-tection thresholds, respectively for simulated loss of type II, using the same values for *D*. As in Fig. 2, the line for no loss (0%) shows fine structure in all three figures.

Figure 3 shows that 20% and 100% loss in terms of pure tone detection results in a reduction in the fine structure and an overall level effect. The maximum loss is approximately 40 dB, which corresponds to the amplification caused by the Zweig impedance function.

Figure 4 shows that the effect of the simulated type II loss on DPOAEs is a loss of fine structure and an overall decrease in level, centred at the $2f_1$-f_2 location and the f_2 location, respectively, as was also found in Mauermann *et al.* (1999).

Figure 5 shows only a limited effect of simulated type II loss on tone-in-noise detection thresholds. Only for 100% loss there is a loss of fine structure and a decrease in thresholds. This contradicts the view that loss of OHC amplification causes difficulties with detection of signals in noise.

Finally, Fig. 6 shows the tone-in-noise detection thresholds based on excitation profiles as computed without and with frequency remapping based on local instantaneous frequencies. In Fig. 5, the same noise snippet was used for all calculations. Because there is a dependence of the choice of noise snippet on the resulting threshold, in this case, 20 repetitions were computed for each frequency and Fig. 6 gives means and standard deviations based on these 20 repetitions. The thresholds after remapping are clearly lower, indicating that the remapping process focusses the tone energy in a single frequency channel, whereas the noise energy remains distributed, making the tone easier to detect.

DISCUSSION

It can be seen from Fig. 2 that small amounts of type I loss will be difficult to detect. Only the curve for the complete loss (100%) shows values that differ significantly from the normal hearing curve. The pure tone audiograms in Fig. 3, and especially the DPOAE levels in Fig. 4 clearly show the effect of a type II loss. Already at 20% loss, a substantial reduction in the fine structure can be observed, as well as a level effect. Since DPOAE levels are relatively easy to measure in a clinical setting, the parameters for type II losses can be easily fitted to individual ears. The tone-in-noise detection thresholds in Fig. 5 hardly show any effect of type II loss, which was one of the reasons to implement type III loss, as described in the Methods section. Only results for the situation without and with remapping were compared.

Figure 6 showed that the remapping of energy provided an improvement in the detection thresholds of about 6 dB for frequencies above 1 kHz. The observation that the improvement was less at the lower frequencies may be due to the relatively short duration of the probe tones. The fact that type III loss did not affect pure tone thresholds or DPOAEs, together with results indicating that deafferentation is associated with a loss of temporal information at the first stages of neural processing, suggests that remapping may occur in the auditory system and a loss of remapping may be the cause of a ‘hidden hearing loss’.

CONCLUSIONS

The transmission line model can be used to simulate three types of hearing loss, associated with loss of mechanical to neural transduction, loss of amplification, and loss of neural coding temporal accuracy. Linking these types of losses to specific damage of hair cells or synapses is difficult since model parameters do not directly relate to structures in the cochlea. In particular, the relation between the Zweig-impedance function's negative damping and delayed feedback force as well as the electromotility of outer hair cells in the 3D geometry of the organ of Corti are far from clear yet.

However, for the purpose of optimizing hearing aid settings as described by Biondi (1978), a direct link to specific damage in the cochlea may not be required. If the parameters describing the hearing loss can be accurately estimated, the model properly represents the impaired hearing system. This may provide a sufficient base for the optimization procedure in Eq. 1.

REFERENCES

Biondi, E. (**1978**). "Auditory processing of speech and its implications with respect to prosthetic rehabilitation. The bioengineering viewpoint," Audiology, **17**, 43-50.

Duifhuis, H., Hoogstraten, H.W., van Netten, S.M., Diependaal, R.J., and Bialek, W. (**1985**). "Modelling the cochlear partition with coupled Van der Pol oscillators," in *Peripheral Auditory Mechanisms.* Eds. J.B. Allen, J.L. Hall, A.E. Hubbard, S.T. Neely, and A. Tubis (Springer, New York), pp. 290-297.

Edwards, B., (**2015**). "Individualizing hearing aid fitting through novel diagnostics and self-fitting tools," in Proceedings of ISAAR 2015: *Individual hearing loss – Characterization, modelling, compensation strategies*. 5th International Symposium on Auditory and Audiological Research, Nyborg, Denmark. Eds. S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen, The Danavox Jubilee Foundation.

Epp, B., Verhey, J.L., and Mauermann, M. (**2010**). "Modeling cochlear dynamics: Interrelation between cochlea mechanics and psychoacoustics," J. Acoust. Soc. Am., **128**, 1870-1883.

Lopez-Poveda, E.A. (**2005**). "Spectral processing by the peripheral auditory system: facts and models." Int. Rev. Neurobiol., **70**, 7-48.

Mauermann, M., Uppenkamp, S., van Hengel, P.W.J., and Kollmeier, B. (**1999**). "Evidence for the distortion product frequency place as a source of DPOAE fine structure in humans II. Fine structure for different shapes of cochlear hearing loss," J. Acoust. Soc. Am., **106**, 3484-3491.

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

Violanda, R.R., van de Vooren, H., van Elburg, R.A.J., and Andringa, T.C. (**2009**). "Signal component estimation in background noise," Proc. NAG/DAGA 2009, **347**, 1588-1591.

Zweig, G. (**1991**). "Finding the impedance of the organ of Corti," J. Acoust. Soc. Am., **89**, 1229-1254.

Statistical representation of sound textures in the impaired auditory system

RICHARD MCWALTER* AND TORSTEN DAU

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

Many challenges exist when it comes to understanding and compensating for hearing impairment. Traditional methods, such as pure tone audiometry and speech intelligibility tests, offer insight into the deficiencies of a hearing-impaired listener, but can only partially reveal the mechanisms that underlie the hearing loss. An alternative approach is to investigate the statistical representation of sounds for hearing-impaired listeners along the auditory pathway. Using models of the auditory periphery and sound synthesis, we aimed to probe hearing impaired perception for sound textures – temporally homogenous sounds such as rain, birds, or fire. It has been suggested that sound texture perception is mediated by time-averaged statistics measured from early auditory representations (McDermott *et al.*, 2013). Changes to early auditory processing, such as broader "peripheral" filters or reduced compression, alter the statistical representation of sound textures. We show that these changes in the statistical representation are reflected in perception, where listeners can discriminate between synthetic textures generated from normal and impaired models of the auditory periphery. Further, a simple compensation strategy was investigated to recover the perceptual qualities of a synthetic sound texture generated from an impaired model.

INTRODUCTION

The healthy auditory system is capable of processing many sounds with varying spectral and temporal features. These sounds range from the simplest artificial stimuli, such as a tone, to the most complex auditory scene, composed of such elements as the "cocktail party", music, or environmental sounds. A sensorineural hearing-impaired system, on the other hand, demonstrates weakness in processing almost all sounds as compared to the normal, healthy ear. The simple artificial tones are no longer audible for particular levels and frequencies. The auditory scene becomes overwhelming as the attention-driven source separation is no longer able to track the target sound. These changes are mostly attributed to the degradation of early auditory processing, such as broadening of "peripheral" filters and loss of compression, which in turn modifies the representation of sounds at higher stages of the auditory system.

Although environmental sounds have been used in speech-in-noise experiments, their processing and perception remains relatively unstudied in the impaired auditory system. Investigating the perception of environmental sounds in the impaired auditory

*Corresponding author: rmcw@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

system could prove beneficial for understanding the difficulties such listeners have in complex listening environments. One possible avenue is to explore the representation of sound textures – temporally homogeneous sounds such as rain, birds chirping or fire – that are composed of the superposition of many similar acoustic events. It has been shown that the perceptual qualities of sound textures can be captured using a standard model of the auditory system and a set of *texture* statistics (McDermott and Simoncelli, 2011).

In this study, we investigated the auditory systems' sensitivity to synthetic sound textures generated with various impaired models of the auditory periphery. Using normal-hearing listeners we probed the response to two major factors in sensorineural hearing loss; broader peripheral filters and loss of compression. In addition, we quantified the effects of the impaired synthetic textures by parametrically varying the synthesis system statistics. Lastly, we developed a compensation strategy to optimize the texture statistics in an attempt to regain the perceptual qualities of sounds generated from impaired models towards that of an original texture.

SOUND TEXTURE ANALYSIS AND SYNTHESIS

The generation of sound textures can be accomplished by *shaping* Gaussian noise with original sound texture statistics measured from a standard model of the auditory system (McDermott and Simoncelli, 2011). The model accounts for fundamental spectral and temporal processing by using a set of cascaded filter banks. The *texture* statistics are measured on the envelope of a filtered original sound texture, which capture the time-averaged envelope distributions as well as the covariance between pairs of neighboring filterbank channels. A companion synthesis component accepts the statistics and modifies a Gaussian noise signal, such that the statistics of the original sound texture are imposed on the synthetic sound. The synthesis process facilitates the exploration of the model structure and the statistical parameters to investigate the change in texture representation and their consequences on perception.

The auditory model is composed of three main components; peripheral frequency filtering, compression and envelope extraction, and modulation filtering as shown in Fig. 1: Analysis System. The peripheral filtering is accomplished by means of a gammatone filterbank, where the normal-hearing system uses equivalent rectangular bandwidth (ERB) spaced filters (Glasberg and Moore, 1990). A power-law compression is applied to the output of each peripheral filter signal followed by computing the absolute value of the discrete time analytic signal, resulting in the subband envelope (Harte *et al.*, 2005). The final stage is a modulation filterbank, which is composed of octave-spaced bandpass filters (Dau *et al.*, 1997).

Statistics that capture many perceptually significant features of sound textures have been identified by McDermott and Simoncelli (2011). These include marginal moments and pair-wise correlations, measured on the envelope signals of the peripheral filters and modulation filters. The envelope signals are down-sampled to 400 Hz at the output of the peripheral filter, as shown in Fig. 1: Synthesis System. The statistics

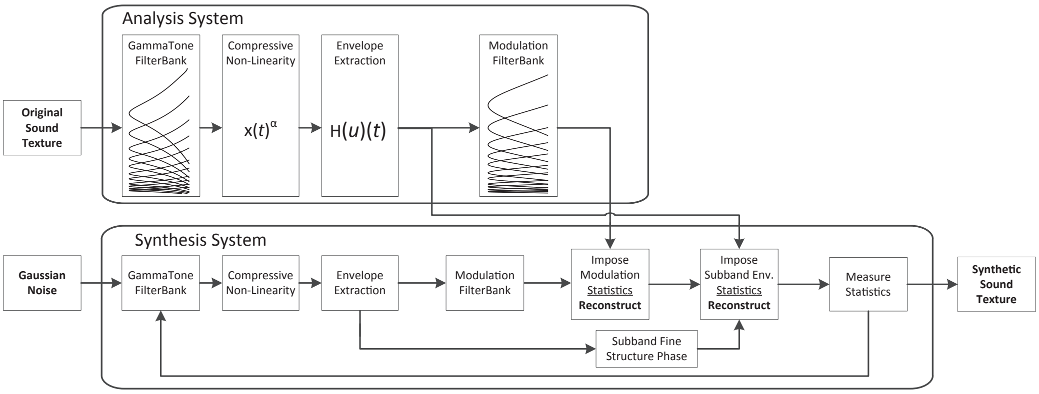


Fig. 1: Implementation of the texture synthesis system (McDermott and Simoncelli, 2011). The system is comprised of an auditory inspired analysis component, which measures marginal moments and pair-wise correlations. The statistics are passed to the synthesis component, which imposes the *texture* statistics on a noise input.

can be grouped into two main categories; the subband envelope statistics and the modulation statistics. The subband envelope statistics include marginal moments (mean, coefficient of variance, skewness, and kurtosis) and pair-wise correlations measured across the eight neighboring subbands. The modulation statistics include the modulation power measured at the output of each modulation filter, as well as pair-wise correlations measured for a specific modulation filter center frequency across the neighboring peripheral subbands.

The synthesis of sound textures is accomplished by imposing the statistics measured from the auditory model (Analysis System) to a Gaussian noise input. The synthesis system operates in two domains; the subband envelope and modulation domain. The synthesis system begins by deconstructing the noise signal to the modulation domain and applying both the modulation power statistics and modulation correlation statistics. The modulation filtered signals are then reconstructed to the subband envelope form, where the marginal moments and pair-wise correlation statistics are imposed. The subband envelope signals are then recombined with the subband fine structure phase signal and reconstructed to the time-domain signal.

Synthetic textures were generated to functionally account for changes to the auditory system caused by sensorineural hearing loss. The limited frequency selectivity is modeled by broadening the peripheral gammatone filters and the loss of compression is modeled as an increase in the power-law compression (Moore, 2007; Rosengard *et al.*, 2005). Figures 2A and 2B show the filter bandwidth and compression ratio used to generate the synthetic textures. The cross-over level for neighboring filters was preserved in all models, which resulted in fewer peripheral filters being used for the impaired hearing model. In turn, this reduced number of peripheral filters reduces the number of parameters measured for each textures. A comparison of the peripheral filterbank structure is shown in Fig. 2C.

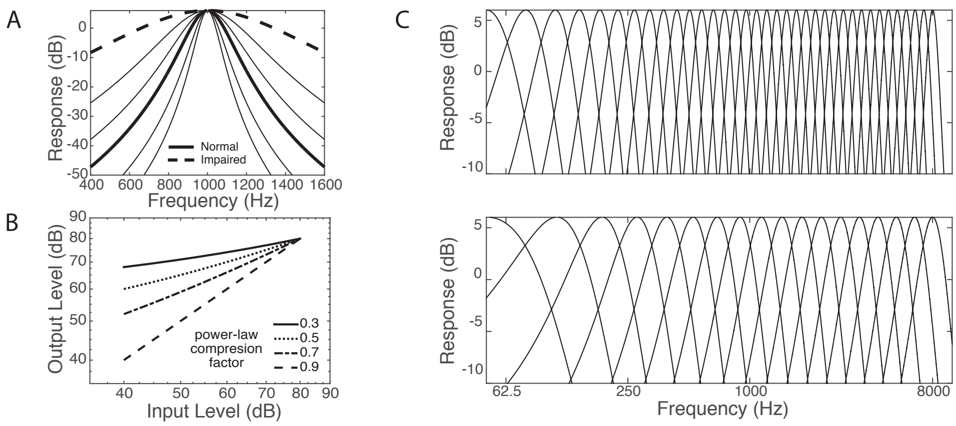


Fig. 2: Comparison of normal and impaired model configurations. (A) Simulated peripheral filter bandwidth for normal and impaired (4×) listeners. (B) Power-law compression ratio input-output level between normal ($\alpha = 0.3$) and impaired ($\alpha = 0.9$). (C) Filterbank model of frequency selectivity for normal (upper) and impaired (lower) hearing.

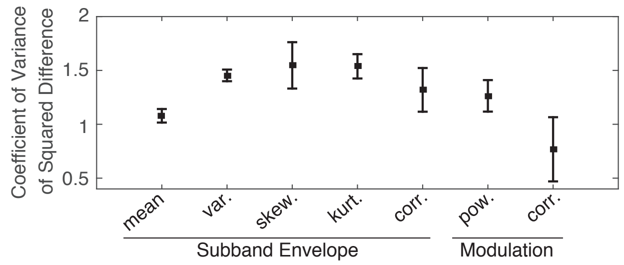


Fig. 3: Normalized coefficient of variance comparing the normal and impaired synthetic texture statistics.

Textures synthesized with impaired models of the auditory periphery alter the representation of the sound textures, as shown in Fig. 3. In order to characterize this change, we generated 45 different textures with a normal and an impaired model with four times broader filters. The textures, including birds chirping, babble, river flowing, and jackhammer, were selected to span the space of statistics, and therefore also covered a broad range of perception. The synthetic sounds were then analyzed using a reference normal auditory model. To make the normal and impaired synthetic textures more comparable, parameters were transformed such that they varied linearly. The coefficient of variance was computed on the individual statistics. As can be seen in Fig. 3, the variation is not consistent for all textures suggesting that some statistical groups are more affected by changes in the early auditory processing than others.

Although it is valuable to compare the averaged variation in texture statistics between normal and impaired auditory models, it is perhaps more intuitive to examine the individual statistics for a given texture. Figure 4 shows this comparison for the sound texture *birds chirping*. The marginal statistics vary (Fig. 4A), particularly for the high frequency channels and higher-order marginal moments. However, for this texture, the time-averaged frequency spectrum is well preserved, as shown by the similarity between the normal and impaired mean statistics. The correlation statistics (Fig. figure:5B) vary as well, showing a noticeable increase in the co-variance of neighboring peripheral channels. This was expected for the hearing-impaired filters, as there is considerably more overlap between neighboring filters (see Fig. 2C). Lastly, the modulation power (Fig. 4C) reveals a difference between the two synthetic textures, particularly in the frequency region around 1.5 kHz for slow modulations.

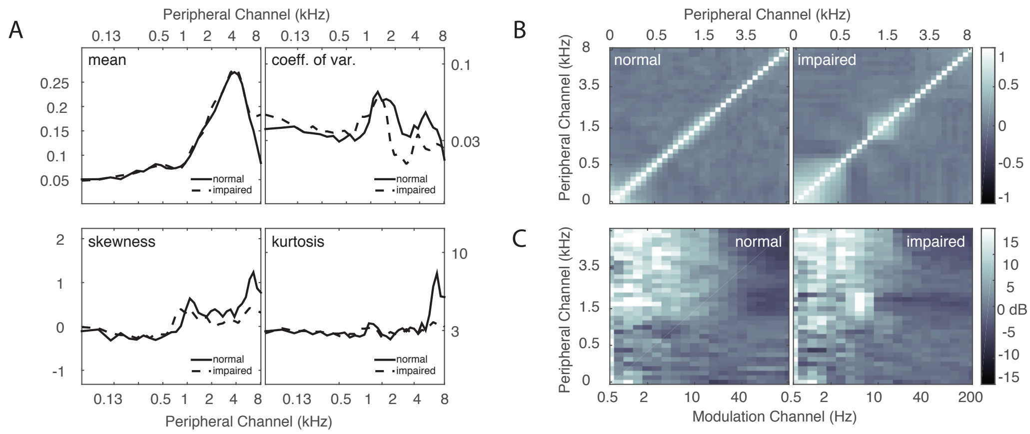


Fig. 4: Comparison of normal and imparied texture statistics for *birds chipring*. (A) Marginal moments (mean, coeff. of variance, skewness, kurtosis), (B) pair-wise correlations for subband envelope, (C) modulation power. Note the modulation pair-wise correlation statistics are not shown.

EXPERIMENTS

In order to investigate the significance of frequency selectivity and compression in sound texture perception, we asked listeners to discriminate between synthetic textures generated with normal and modified auditory models. The listeners were presented with three intervals, each 2 seconds in duration, and required to find the *odd* or modified interval, where two intervals were generated with a normal hearing model and the *odd* interval was generated with a modified hearing model. The stimuli were presented via open-ear headphones at a sound pressure level (SPL) of 65 dB. The modified texture could either be the first interval or the last interval. The two intervals generated from a normal hearing model were from the same texture family, but different sound instances, ensuring that listeners could not use unique acoustic features in their judgments.

Figure 5A shows the results for textures generated with broader as well as narrower peripheral filters, where the textures generated from ERB-spaced filters are the reference. Fifteen self-reported normal-hearing listeners participated in the experiment. The results show an increase in discrimination performance as the model deviated

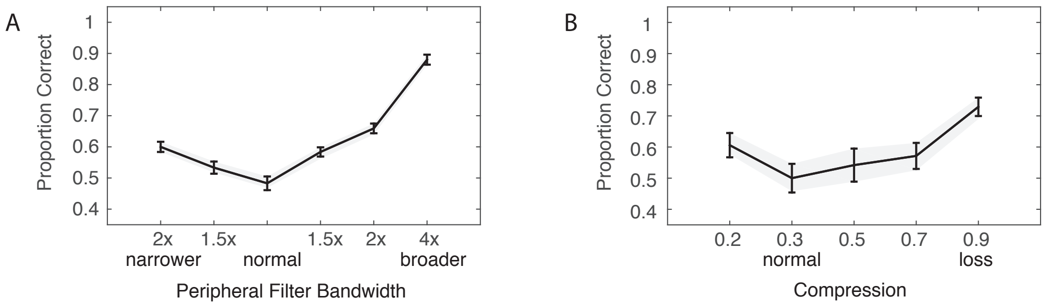


Fig. 5: Discrimination results for synthetic textures generated with impaired models of the auditory periphery shown as a proportion correct for (A) broader/narrower peripheral filter and (B) loss/change in compression. Error bars show standard error.

from the reference. This is particularly the case when the synthetic textures were generated with broader filters. However, it can also be seen that performance increases with narrower filters, suggesting that the higher number of filters may capture some additional frequency cues. Figure 5B shows the results for textures generated with reduced compression. Eight self-reported normal-hearing listeners participated in the experiment. The results show an increase in discrimination performance as the auditory model parameters deviated from normal hearing. The listeners reported audible artifacts in some of the intervals, and indeed, the change in compression seemed to offer cues when listening to modified compression settings. In addition, the synthesis process applies the compression during the analysis and removes the compression during the synthesis process, essential by reversing the effects of the compression. Therefore, the synthesis process seems to negate the possibility of exploring the perceptual consequences of compression with texture synthesis.

To better quantify the contribution of the texture statistics to the perception of normal and impaired synthetic textures, we designed a preference task experiment with stimuli that impaired particular statistical groups; marginal moments, pair-wise correlations, or modulation power. The listeners' were presented an original sound texture which was compared to two synthetic sounds generated from a normal and parametrically impaired auditory model. The three intervals were each 4 seconds in duration. The presentation of the synthetic intervals was randomized. The stimuli were presented via headphones at a level of 65 dB SPL.

The results from the parametrically impaired auditory model with 4x broader filters are shown in Fig. 6A. Twelve self-reported normal hearing listeners participated. The figure shows the pair-wise correlation parameter group was the most sensitive to impairment, as 72% of synthetic textures generated from a normal-hearing model were preferred over a pair-wise correlation-impaired model. The impaired marginal moments parameter group also showed an effect on the perception followed by the modulation power. It should be highlighted that a common modulation selective

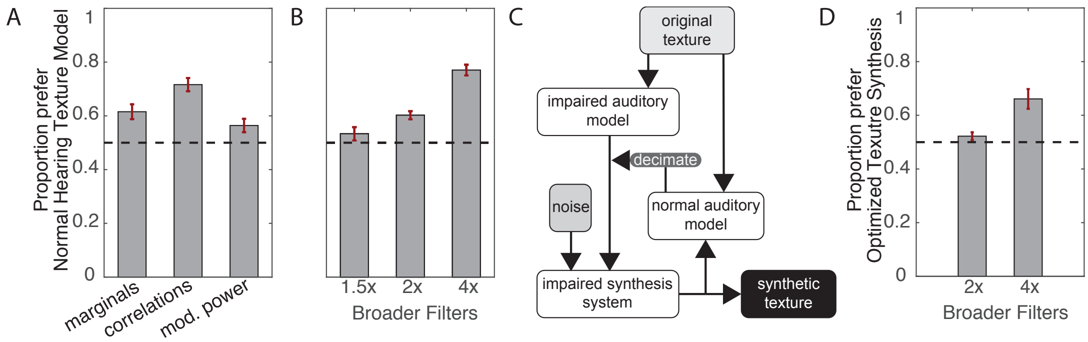


Fig. 6: Results of listeners who prefer textures synthesized with normal hearing model for (A) impairing individual parameter groups and (B) varying severity of the model impairment. (C) Compensation stategy for impaired sound texture synthesis. (D) Optimized impaired texture statistics results for $2\times$ and $4\times$ broader peripheral filters. Error bars show standard error.

filterbank structure was used for all synthetic textures. These results highlight the impact of the individual impaired parameter groups on hearing impairment.

As a control, we also asked listeners to perform a preference task with the wholly impaired auditory system with 3 configurations of peripheral filter broadening – $1.5\times$, $2\times$, and $4\times$ – shown in Fig. 6B. The results are consists with the results shown in Fig. 5A as well as the parametrically varied impaired auditory model results. The results show that the perceptual quality declined as the auditory model deviated from that of a normal system.

COMPENSATION STRATEGY

Given that the representation and perception of synthetic sound textures change with the impairment of the peripheral auditory model, the question is whether it possible to modify the statistical representation to regain the perceptual quality towards the original texture. The results from experiments 1 and 2 revealed that a broadening of peripheral filters is salient for synthetic sound textures and most affected by the changes in the representation of pair-wise correlation statistics. A possible optimization strategy for an impaired auditory system could be a decimated version of the normal hearing statistics. However, the textures synthesized with an impaired model and decimated normal hearing statistics yielded poor synthetic versions, and often the synthesis failed. A different structure was implemented that used parallel normal and impaired model analysis systems, which is shown in Fig. 6C. The coupled analysis adjusts the impaired statistics such that the synthetic output is optimized to yield a synthetic texture similar to the original texture as measured by a normal auditory model. This can be thought of as *nudging* the impaired model representation to output a texture with similar perceptual qualities to the original texture.

Listeners performed a preference task to reveal the significance of the impaired auditory model optimization system. In each trial, listeners were presented with an original texture recording followed by two randomly presented synthetic textures; one synthesized with an impaired auditory model and another synthesized with the impaired auditory model optimization system. The stimuli were presented via headphones at a level of 65 dB SPL and each interval was 4 seconds in duration. The results from the impaired auditory model texture optimization system in Fig. 6D show a modest improvement in subjective performance for the 4$\times$ broader peripheral filter case. In the case of the 2$\times$ broader filters, no improvements were found. Although the performance of the optimization system did not yield comparable results to the original, there is modest benefit and the method does warrant further investigation.

SUMMARY

Sound textures offer a novel avenue for investigating the changes in representation due to hearing impairments, as well as the perceptual consequences of those changes. The differences in sound textures synthesized with auditory models that deviated from the normal hearing system were identifiable by normal-hearing listeners. The model impairments introduced changes to the statistical representation of sound textures, which related to perception to varying degrees. The results showed that pair-wise correlation statistics offer a primary auditory cue that affects the quality of the texture synthesis. Understanding how such *noise* signals are represented in the normal and impaired auditory system may offer some insight into the processing involved in "cocktail party" scenarios, where the auditory system separates a target signal from the noise.

REFERENCES

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.

Glasberg, B.R. and Moore, B.C.J. **(1990)**. "Derivation of auditory filter shapes from notched-noise data," Hear. Res., **47**, 103-138.

Harte, J.M., Elliott, S.J., and Rice, H.J. **(2005)**. "A comparison of various nonlinear models of cochlear compression," J. Acoust. Soc. Am., **117**, 3777-3786.

McDermott, J.H. and Simoncelli, E.P. (**2011**). "Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis," Neuron, **71**, 926-940.

McDermott, J.H., Schemitsch, M., and Simoncelli, E.P. **(2013)**. "Summary statistics in auditory perception," Nat. Neurosci., **16**, 493-498.

Moore, B.C.J. **(2007)**. *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues.* John Wiley and Sons.

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.

Auditory model responses to harmonic and inharmonic complex tones: Effects of the cochlear amplifier

VÁCLAV VENCOVSKÝ*

Musical Acoustics Research Center, Academy of Performing Arts in Prague, Prague, Czech Republic

Hopkins and Moore [J. Acoust. Soc. Am. 122, 1055-1068 (2007)] measured the ability of hearing-impaired (HI) listeners to discriminate harmonic (H) from inharmonic (I) – all harmonics shifted upwards by the same amount in Hz – complexes. The complexes were composed of many bandpass filtered harmonics (shaped stimuli) or five equal-amplitude harmonics (non-shaped stimuli). HI listeners performed worse with the shaped stimuli than with the non-shaped stimuli. Since shaping of the complexes should minimize envelope and spectral cues, listeners should discriminate H from I stimuli mainly using temporal fine structure (TFS) cues even when the harmonics are not resolved. This ability seems to be worsened in HI listeners. This study employed an auditory model with a physical cochlear model to show how the cochlear amplifier affects responses to H and I stimuli. For the shaped stimuli, the TFS of the simulated neural signals for H and I stimuli differed, represented by low cross-correlation coefficients computed from the shuffled cross-polarity correlograms. However, for the passive auditory model (simulating HI), the inter-spike intervals smaller than half of the stimulus period were similar. This could explain the poor performance for HI listeners. For the non-shaped stimuli, differences in the inter-spike intervals were observed even for the passive model, which could contribute to the improved performance.

INTRODUCTION

Hopkins and Moore (2007) investigated the ability of hearing-impaired (HI) listeners to discriminate harmonic (H) complex tones from inharmonic (I) complex tones, i.e. the complexes created by shifting all the spectral components in the H complexes upwards by the same amount in Hz. They used two types of stimuli: “shaped” and “non-shaped”. The shaped stimuli were composed of many harmonics filtered by a bandpass filter. The bandpass filter should eliminate changes in the excitation patterns (spectral cues), which could be used to discriminate H from I complexes. Thus, if the stimuli contain unresolved spectral components (high relative to the fundamental frequency F0), listeners should discriminate between H and I complexes by using temporal fine structure (TFS) cues, i.e., the intervals between peaks of the TFS close to envelope maximums. On the other hand, the non-shaped stimuli were composed

*Corresponding author: vaclav.vencovsky@gmail.com

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

of five equal-amplitude harmonics. The listeners could thus use spectral cues to discriminate between H and I complexes. Hopkins and Moore (2007) showed that the performance of HI listeners in the H-I discrimination tasks was much poorer for the shaped than for the non-shaped stimuli. They therefore interpreted the results such that the HI listeners could not use TFS cues to discriminate the stimuli. However, some researchers have questioned that TFS information may code fundamental frequency (e.g., Oxenham *et al.*, 2009).

The aim of this study was to analyze the H and I complexes by an auditory model. The auditory model was composed of known models of the outer/middle ear processing, a physical cochlear model and algorithms simulating the physiology of inner hair cells and auditory-nerve synapses. The physical cochlear model simulated the active function of outer hair cells by a feedback force. This cochlear amplifier was removed to simulate hearing impairment.

METHODS

Stimuli

The stimuli were the same as those used in Hopkins and Moore (2007). The shaped stimuli were composed of the fundamental and higher harmonics up to 20 kHz, each starting in sine phase. The complexes were filtered by a bandpass filter centered at nominal harmonic number $N = 11$. The bandwidth of the bandpass filter was set such that the components between $N-2$ and $N+2$ had an amplitude of 1 and the amplitude of the remaining components decreased at a rate of 30 dB/octave. The non-shaped stimuli contained only five harmonics centered at $N = 11$, each starting in sine phase. The fundamental frequency, F0, of the stimuli was 100 Hz. The I complexes (shaped and non-shaped) were created by shifting all the spectral components in the H complexes upwards by $\Delta f = 35$ Hz, which should allow normal-hearing (NH) listeners to discriminate between the H and I complexes (Hopkins and Moore, 2007). The overall level of the shaped and non-shaped stimuli was 65 dB SPL.

Auditory model

The auditory model was composed of known algorithms simulating different parts of the peripheral ear: an outer- and middle-ear model from the Matlab Auditory Periphery (MAP), a physical cochlear model designed by Nobili *et al.* (2003) and algorithms simulating the function of inner hair cells and auditory-nerve synapse (Meddis, 2006). The model input was an acoustic waveform at the entrance of the outer ear; the model outputs were neural discharges (spikes) fired into the auditory nerve.

The individual stages of the auditory model were used with the parameters described in the above given references. The physical cochlear model had 300 channels with characteristic frequencies (CFs) distributed between 20 Hz and 17 kHz. The model simulated the active function of outer hair cells by a feedback force. This

level (dB SPL)	characteristic frequency (kHz) 0.125	0.25	0.5	1	2	4
	active model (NH), ERB (Hz)					
20	43	62	89	141	225	390
40	43	62	90	148	245	521
60	43	70	122	201	337	818
80	54	98	168	307	528	1107
	passive model (HI), ERB (Hz)					
	90	170	311	529	982	1640
ERB_{GM}(Hz)	38	52	79	133	241	456

Table 1: Equivalent rectangular bandwidth (ERB) of simulated cochlear filters (physical cochlear model). ERB_{GM} are psychoacoustical data given in Glasberg and Moore (1990).

cochlear amplifier was removed to simulate hearing impairment. Since the model with the parameters given in Nobili *et al.* (2003) had wider cochlear filters than was measured psychophysically (e.g., Glasberg and Moore, 1990), its damping parameter was adjusted here. Table 1 shows the measured equivalent rectangular bandwidth (ERB) of the simulated cochlear filters: active model (with the cochlear amplifier); and passive model (without the cochlear amplifier). ERB_{GM} are the psychoacoustically measured ERBs given in Glasberg and Moore (1990).

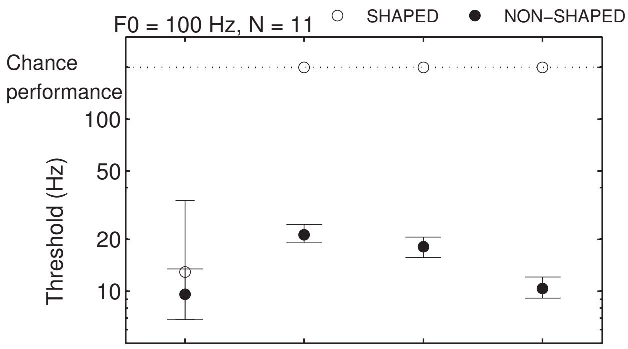


Fig. 1: Thresholds (Δf) in H-I discrimination tasks. The data were reproduced from Hopkins and Moore (2007). Open circles are the thresholds for the shaped stimuli, filled circles for the non-shaped stimuli, and the error bars are standard deviations of the mean. Abscissa denotes the listeners: "mean NH" are the mean values across the NH listeners; HI3, HI6, and HI7 are data from the HI listeners. The notation is the same as in Hopkins and Moore (2007).

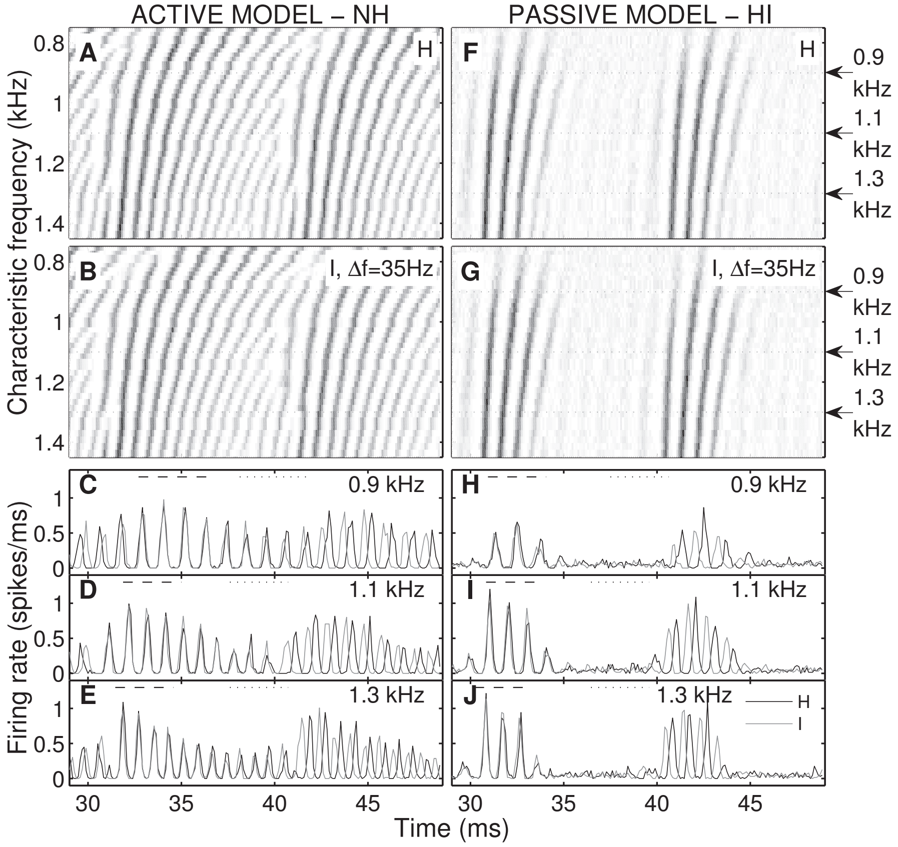


Fig. 2: Auditory model responses to the shaped stimuli. Panels in the left column (A to E) show the data for the active auditory model (NH); panels in the right column (H to J) for the passive auditory model (HI). (A, F) The responses to the H complexes. (B, G) The responses to the I complexes ($\Delta f = 35$ Hz). (C, D, E) The responses of the active model (in the channels with CF of 0.9, 1.1 and 1.3 kHz, respectively). (H, I, J) The responses of the passive model (in the same channels as in C, D, E).

RESULTS AND DISCUSSION

Fig. 1 shows psychoacoustical thresholds (Δf) from the H-I discrimination tasks, measured by Hopkins and Moore (2007). Open circles show the thresholds for the shaped stimuli, filled circles for the non-shaped stimuli, error bars show standard deviation of the mean. The abscissa of the graphs shows the listeners: the mean values calculated across the NH listeners are denoted as “mean NH”; HI3, HI6, and HI7 denote the data from the HI listeners. The notation is the same as used in Hopkins and Moore (2007). For the shaped stimuli, the performance of the HI listeners was at a chance level, which may indicate their inability to use TFS cues in these tasks.

Figures 2 and 3 show the auditory model responses – post stimulus time histograms (PSTHs), binwidth = 10/(sample frequency), 600 repetitions – to the shaped and non-

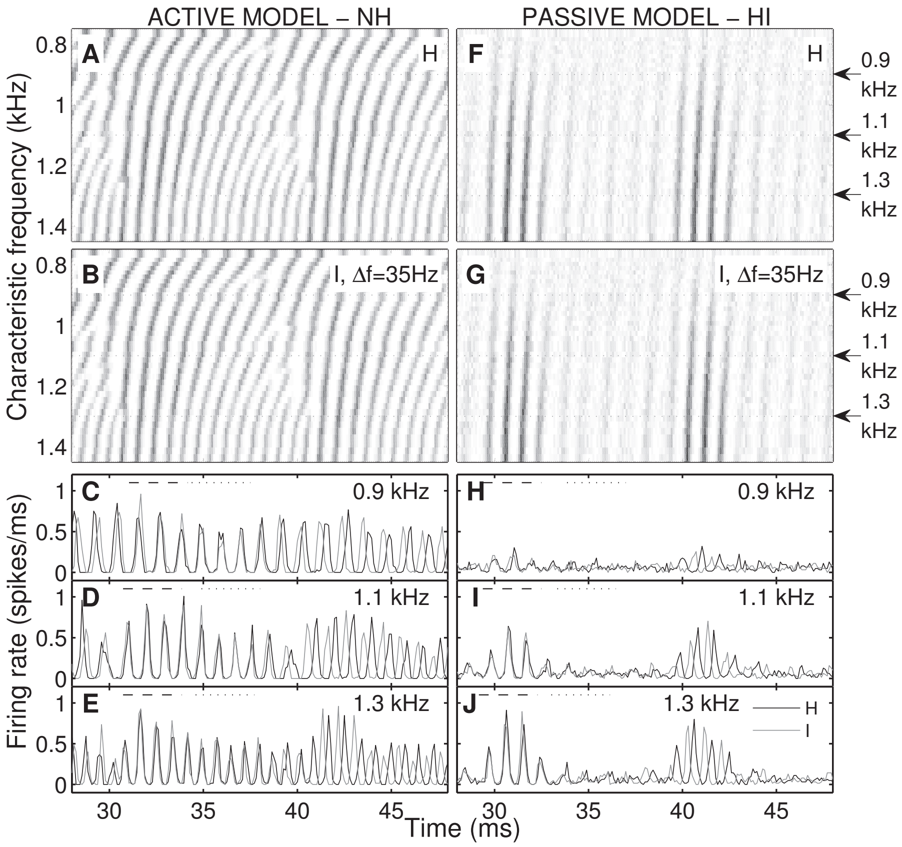


Fig. 3: Auditory model responses to the non-shaped stimuli. Description of the panels is similar to Fig. 2.

shaped stimuli, respectively. In each figure, the panels in the left column show the responses of the active model (simulating NH); the panels in the right column show the responses of the passive model (simulating HI). Panels A and F show the PSTHs (in the several adjacent channels) for the H complexes; panels B and G show the PSTHs for the I complexes ($\Delta f = 35\,\text{Hz}$). Panels C, D, E (for the active model) and H, I, J (for the passive model) show the PSTHs in three discrete model channels with CF given in each panel; the black solid lines show the PSTHs for the active model, and the gray solid lines for the passive model.

The responses to the H and I complexes differ – the corresponding PSTHs do not exactly overlap (see Figs. 2 and 3). This difference leads to low (close to zero) across-stimulus neural cross-correlation coefficient calculated according to the method described in Kale *et al.* (2014), which indicates perceptible changes in the stimuli. However, the TFS cues suggested by Hopkins and Moore (2007) are also visible in the responses of the passive model. These results would indicate that HI listeners also have TFS information in the activity patterns.

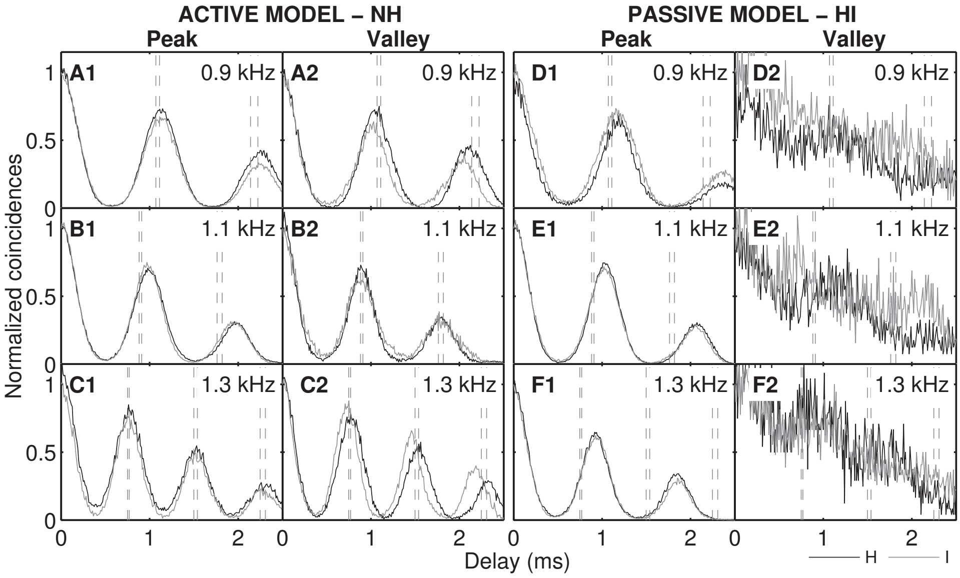


Fig. 4: SACs of the responses to the shaped stimuli.

Removing the cochlear amplifier affected the responses to the shaped and non-shaped stimuli in quite similar way: the valleys of the PSTHs became very pronounced with much smaller amplitude in comparison to the envelope maximum. The responses seem to lack TFS information in the valleys. To find out whether the intervals between successive peaks in the PSTHs are involved, shuffled auto-correlations (SACs) across the spike trains were calculated by tallying inter-spike intervals in the responses (Joris, 2003).

Figures 4 and 5 show the SACs for the shaped and non-shaped stimuli, respectively. The columns denoted "Peak" show the SACs calculated from the portions of the auditory model responses (shown in Figs 2 and 3) marked by the horizontal dashed lines; those denoted "Valley" show the SACs calculated from the responses marked by the horizontal dotted lines. The vertical dashed lines in each panel of Fig. 4 and Fig. 5 show the delays corresponding to k/CF and $k/(\mathrm{CF}+\Delta f)$, where k is 1, 2, or 3.

The positions of the first peak in the SACs (at delay about 1/CF) calculated from the active model responses to the shaped H and I complexes seem to differ only in the valleys. Since only the information from the peaks is available in the responses of the passive auditory model, this could explain why the HI listeners performed poorly in the H-I discrimination tasks. In contrast to this, the SACs at "Peak" portions of the responses which were calculated for the non-shaped stimuli are not overlapping even for the passive auditory model. Therefore the HI listeners may use these TFS cues in addition to spectral cues to discriminate between the H and I non-shaped stimuli. This could contribute to their better performance (see Fig. 1). However, all these results only show that TFS of the responses may differ, but do not relate the TFS information with the perceived pitch of the complexes.

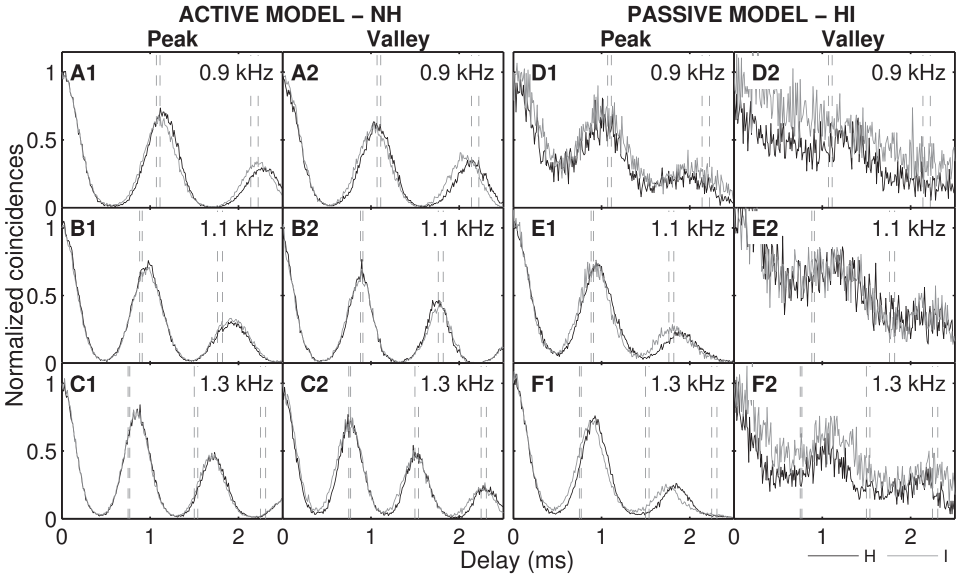


Fig. 5: SACs of the responses to the non-shaped stimuli.

The above shown results cannot rule out the possibility that listeners may use combination tones to discriminate between H and I complexes (Oxenham *et al.*, 2009) – hearing loss may eliminate the combination tones. Another possibility which cannot be ruled out is that the spectral components of the complexes were resolved in the NH listeners and unresolved – because of the wider cochlear filters – in the HI listeners (Santurette *et al.*, 2012). On the other hand, the above shown results would also explain the better performance of the HI listeners with the non-shaped stimuli. Since only the cochlear amplifier was removed from the model, all the simulation only holds for HI listeners with hearing deficits solely based on a loss of outer hair cells.

SUMMARY

Both, the shaped and the non-shaped H and I complexes were analyzed using the auditory model – the active auditory model and the passive auditory model (without a cochlear amplifier). The results can be summarized as follows:

1. The responses to the H and I complexes showed that the stimuli differed in the intervals between peaks of TFS (the intervals long as about the period of the complexes). As already suggested in Hopkins and Moore (2007), these TFS cues could be used to discriminate the H and I complexes with unresolved spectral components. This study showed that these TFS cues are available also in the responses to the non-shaped stimuli and that the cues may also be available for the HI listeners. However, this would not explain the poor performance of the HI listeners – with deficits based solely on a loss of outer hair cells – for the shaped stimuli.

2. The SACs calculated from the short portions of the responses showed the largest differences (shifts in the position of the first peak in the SACs) in the valleys of the responses to the shaped stimuli. Since there is no TFS information in the valleys of the passive auditory model responses, this could explain the poor performance of the HI listeners if their deficits were solely based on a loss of outer hair cells. However, for the non-shaped stimuli TFS information seem to be available in the envelope maximums of the passive auditory model responses. This may help the HI listeners to discriminate the H from I complexes.

ACKNOWLEDGMENTS

I thank Aleš Vetešník for sending parameters of the cochlear model and the ISAAR committee for the scholarship allowing me to attend the ISAAR symposium. This research was supported by the Ministry of Education, Youth and Sports of the Czech Republic in the Long Term Conceptual Development of Research Institutes grant of the Academy of Performing Arts in Prague: The "Sound quality".

REFERENCES

Glasberg, B.R. and Moore, B.C.J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res., **47**, 103-138.

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.

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

Kale, S., Micheyl, C., and Heinz, M.G. (**2014**). "Implications of Within-Fiber Temporal Coding for Perceptual Studies of F0 Discrimination and Discrimination of Harmonic and Inharmonic Tone Complexes," J. Assoc. Res. Otolaryngol., **15**, 465-482.

Matlab Auditory Periphery (MAP). University of Essex, Hearing Research Laboratory: `http://www.essex.ac.uk/psychology/department/hearinglab/modelling.html`.

Meddis, R. (**2006**). "Auditory-nerve first-spike latency and auditory absolute threshold: A computer model," J. Acoust. Soc. Am., **119**, 406-417.

Nobili, R., Vetešník, A., Turicchia, L., and Mammano, F. (**2003**). "Otoacoustic emissions from residual oscillations of the cochlear basilar membrane in a human ear model," J. Assoc. Res. Otolaryngol., **4**, 478-494.

Oxenham, A., Micheyl, C., and Keebler, M.V. (**2009**). "Can temporal fine structure represent the fundamental frequency of unresolved harmonics," J. Acoust. Soc. Am., **125**, 2189-2199.

Santurette, S., Dau, T., and Oxenham, A.J. (**2012**). "On the possibility of a place code for the low pitch of high-frequency complex tones," J. Acoust. Soc. Am., **132**, 3883-3895.

Loss of speech perception in noise – causes and compensation

JORGE MEJIA[1,2], HARVEY DILLON[1,2,*], RICHARD VAN HOESEL[1,4], ELIZABETH BEACH[1,2], HELEN GLYDE[1,2], INGRID YEEND[1,2], TIM BEECHEY[1,2], MARGO MCLELLAND[1,2], ANNA O'BRIEN[1,4], JÖRG BUCHHOLZ[1,2,3], MRIDULA SHARMA[1,3], JOAQUIN VALDERRAMA[1,2], AND WARWICK WILLIAMS[1,2]

[1] *The HEARing Cooperative Research Centre, Australia*

[2] *National Acoustic Laboratories, Macquarie University, Sydney, Australia*

[3] *Department of Audiology, Macquarie University, Sydney, Australia*

[4] *Department of Audiology and Speech Pathology, The University of Melbourne, Melbourne, Australia*

Any damage within the cochlea, whether affecting hearing thresholds or high threshold nerve fibres, that affects the resolving power of the cochlear, necessitates a higher input signal-to-noise ratio to achieve normal speech understanding in noise. Other than wireless remote microphone systems, super-directional beamformers are the most effective way to achieve this. To optimise their performance, they should have beam widths that are neither too narrow nor too broad, attenuate off-beam signals in a way that preserves spatial awareness of the environment, and adapt to changing competing signals fast enough to suppress them but not so fast as to distort the target signal. This paper reports on the advantages and limitations of super-directional beamformers as measured in six different experiments.

INTRODUCTION

It is now well established in animal studies that high levels of noise, even for a few hours, can damage the auditory system in ways that are not evident in the audiogram. In particular, high threshold, low spontaneous rate, afferent fibres originating at inner hair cells are destroyed, starting with destruction of the synapse within days of the noise exposure (Furman *et al.*, 2013; Kujawa and Liberman 2009). There is some uncertainty about how this finding translates to humans, and if so, what the consequences for humans are. The first part of this paper shows the context in which this question is being comprehensively investigated. A likely consequence is that some people with little or no elevation in hearing thresholds require a better signal-to-noise ratio (SNR) to communicate than do their peers with the same hearing thresholds (Plack *et al.*, 2014).

The second part of the paper very briefly summarises one reason, which turns out to be simple audibility, why people with elevated hearing thresholds also require a better SNR than people with normal hearing. Although the reason may be simple, the solution is not, as there is a limit to how much amplification a person with hearing loss will tolerate. We are not yet at the stage of being able to analyse precisely why, other than inadequate audibility, damage to the hearing system creates difficulties in recognising and

*Corresponding author: harvey.dillon@nal.gov.au

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

understanding speech for an individual, especially in noisy situations. We are even further from being able to build inverse processes (in the unlikely event that is even possible) into hearing devices to restore normal functioning. It seems extremely likely, however, that anyone with a hearing problem, whatever its underlying origin, will benefit from devices that provide them with a better SNR than they would have access to without any device. The final, and major, section of the paper therefore provides an overview of a series of experiments designed to evaluate a novel method of improving SNR. The method is based on a binaural beamformer that provides a greater degree of directivity than hearing aids working in isolation on each side of the head can provide.

NOISE EXPOSURE, SPEECH RECOGNITION, COGNITION, AND COCHLEAR FUNCTIONING

Our primary interest is in understanding the relationship between noise exposure, cochlear functioning, and the consequences of the latter for speech recognition. Our hypothesis is that any relationship between noise exposure and speech recognition will be completely mediated by the effect that the noise exposure has had on cochlear functioning, and of course its downstream effects on auditory nerve fibres and higher centres (Bramhall *et al.*, 2015; Schaette and McAlpine, 2011). Speech recognition, however, is very likely to be affected by cognitive abilities (Helfer and Jesse, 2015). It is also possible that it is affected by musical training or experience, either by improving auditory brainstem functioning (Skoe and Kraus, 2013; Slater *et al.*, 2015) or by improving cognitive abilities, so consequently we need to measure these as well.

Figure 1 shows the relationships that we are hypothesising may exist between the quantities measured. Lifetime noise exposure, estimated from a questionnaire (Beach *et al.*, 2013) is assumed to damage high-threshold nerve fibres, outer hair cell (OHC) functioning, and possibly low-threshold nerve fibres. The latter two forms of damage (along with any reduction in stria vascularis effectiveness that affects their functioning) are presumed to determine hearing thresholds. OHC damage should be observable in the levels of otoacoustic emissions both transient (TEOAE) and distortion product (DPOAE). High threshold fibre damage should be observable behaviourally in the detection of tones in threshold equalizing noise (TEN test; Moore *et al.*, 2012), in reduced sensitivity to temporal fine structure (TFS; Moore and Sek, 2009), and in elevated thresholds for detection of amplitude modulation (AM). In the latter two tests, lower level background noise is used to limit the ability of low and medium threshold fibres to contribute to the task. Damage to high threshold fibres should also be observable as a reduced growth of envelope following response as modulation depth increases (Bharadwaj *et al.*, 2014), reduced amplitude of wave I in a click ABR (Schaette and McAlpine, 2014; Stamper and Johnson, 2014), and a decreased magnitude and coherence of a speech ABR (Anderson *et al.*, 2013). Stimuli for these electrophysiological tests will also be masked to maximise sensitivity to high threshold fibre activity.

The three cochlear variables are hypothesised to affect speech recognition, whether measured behaviourally with the Listening in Spatialised Noise Sentences test (LiSN-S; Cameron and Dillon, 2007) or via self report with the Speech Spatial Qualities test (SSQ12; Noble *et al.*, 2013). Each of the measures of speech recognition may be affected by verbal memory, attention (the Test of Every Day Attention;

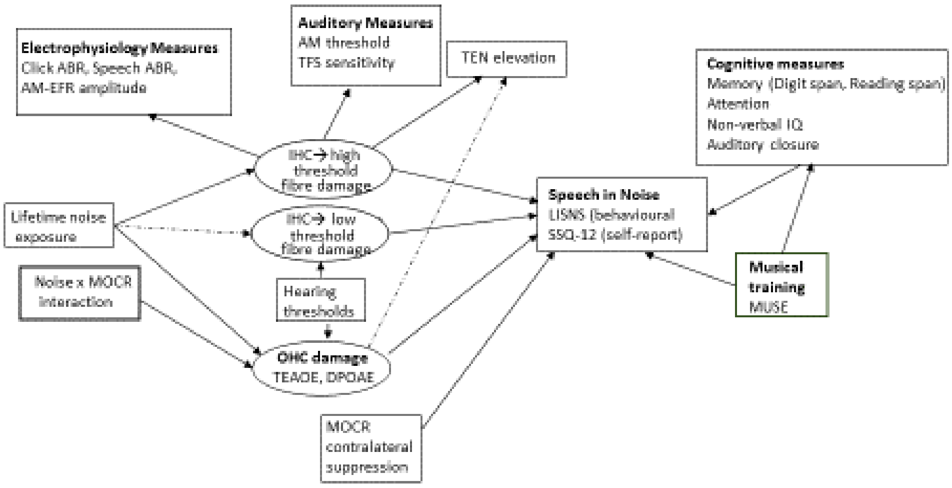


Fig. 1: Hypothesised relationships between factors affecting the perception of speech in noise. The diagram shows latent variables (that cannot directly be observed) as ellipses, and indicators of those variables or other measurable quantities as rectangles.

Robertson *et al.*, 1996), non-verbal intelligence and auditory closure ability. Verbal working memory can be assessed with digit span forward and reverse, and the Reading Span Test (Daneman and Carpenter, 1980). Musical experience can be estimated with the Music Use (MUSE) questionnaire (Chin and Rickard, 2012), which provides indices relating to both level of musical training and experience in playing music. Finally, the medial olivo-cochlear response (MOCR) is assumed to assist in recognising speech in noise, and to help protect OHCs against noise damage, hence the box showing an interaction between MOCR strength and noise exposure.

At the time of writing, behavioural data has been measured on 78 adults aged 30 to 55 years with hearing thresholds in the normal or 'near to normal' range. Their noise exposure varied greatly, up to an estimated 62,000 Pa^{2}hrs. Musical training and experience likewise varied over a wide range. Electrophysiological data have so far been obtained on only 12 participants. Analysis of the results will be reported on in later publications, when further behavioural and electrophysiological data have been collected.

HEARING IMPAIRMENT AND SPEECH PERCEPTION IN SPATIALLY SEPARATED COMPETITION

Investigation into the impact of hearing impairment on speech recognition in spatially separated distractors further demonstrates the need for improved SNRs. Glyde et al (2013a) tested 80 people, aged 7-89 years with hearing levels ranging from normal to moderately-severe, on the LiSN-S and found increasing hearing impairment correlated with worsening speech reception thresholds in noise (SRTn) (see Fig. 2). This relationship existed despite the use of NAL-RP amplification. The deficit was strongest in the test conditions in which the target speech was spatially separated from the distractors due to decreasing spatial release from masking (SRM) with increasing hearing loss (partial $r^2 = 0.66$).

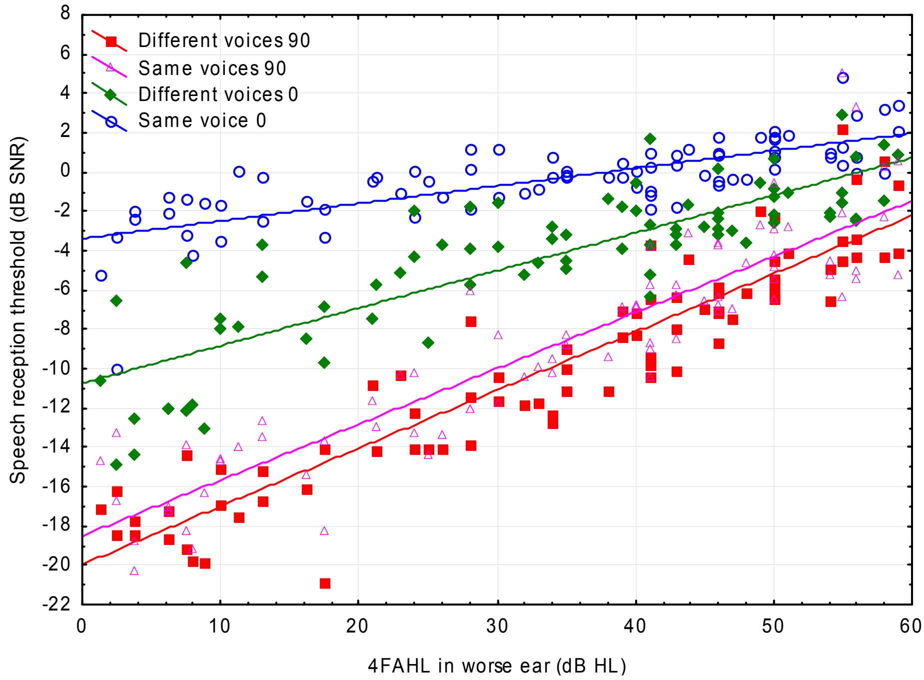


Fig. 2: Variation of SRTn measured with the LiSN-S test as four-frequency average hearing thresholds in the worse ear vary from normal to 60 dB HL. The four conditions of the LiSN-S comprise the talker being the same or different voice as the distractors, combined with the distractors being at the same (0°) location as the talker or at different (±90°) locations. Reprinted with permission from Glyde et al (2013a).

The underlying cause of this apparent loss of ability to use spatial cues was investigated in a subsequent series of experiments. Firstly, by creating versions of the LiSN-S test stimuli which contained only interaural level differences (ILDs) or interaural time differences (ITDs), and comparing normal-hearing adults' performance on these versions to performance with both cues available, it was ascertained that ILDs alone provided as great SRM as the two cues together (Glyde *et al.*, 2013b). This result suggested that ILD interpretation or transmission was the most likely barrier to achieving SRM. Given ILD's dominance in the high frequencies, limited audibility of the SNR benefits arising from ILDs could explain the results shown in Fig. 2.

This hypothesis was examined in Glyde *et al.* (submitted) where frequency-specific filtering was applied to the stimuli so that sensation levels were matched between a sample of normal-hearing and hearing-impaired adults. Speech reception thresholds were compared at three amplification levels (NAL-RP, NAL-RP+25%, NAL-RP+50%). Increased amplification significantly improved SRM ($p< 0.001$). Therefore if better audibility could be provided to hearing-impaired individuals, better performance in spatially separated competition is expected. However, high-frequency gain considerably in excess of that provided by NAL-RP would be needed to enable close to normal SRM, and this much high-frequency gain is generally not acceptable to hearing aid wearers, and is often not possible because of feedback oscillation.

IMPROVING SNR THROUGH BINAURAL BEAMFORMING

The most promising method for improving SNR is to use directivity to provide greater amplification for sounds coming from a target direction than to sounds coming from other directions. Directional microphones mounted within a hearing aid are very limited in the extent to which they can do this, as the sounds arriving at two closely spaced ports differ very little in either time of arrival or level. Signals arriving at one side of the head, however, differ greatly in both time of arrival and level from signals arriving at the other side, for sources other than those directly in front or behind the listener. Inputs to the beamformer applied to each ear come from the output of conventional directional microphones. Essentially, at any moment in time, the beamformer gives maximum amplification to those frequency components of the signal that have the same level and phase (i.e., time of arrival) at the two sides of the head, and progressively less amplification to those components that differ in either amplitude or phase. Many variations are possible while still conforming to this general principle. Variations include:

- The azimuth variation from straight ahead beyond which sounds are attenuated (and hence the target beam width);
- The degree to which off-beam signals are attenuated;
- The rate at which the characteristics of the beamformer are allowed to adapt, and the frequency resolution with which the characteristics are determined;
- The extent to which the original time and level differences at each ear are retained in the outputs fed to each ear (and hence the extent to which spatial awareness is retained);
- The relative reliance placed on inter-aural time differences versus inter-aural level differences in determining the weights given to each component; and
- The way in which each of the above considerations is varied with frequency.

The results reported in this paper were obtained with beamformers that were progressively improved by fine tuning these variations over several years to optimise the combination of SNR enhancement, lack of perceptible distortion, and retention of spatial information. The choices that affect each of these also affect the other two desired characteristics, so optimising the trade-off is necessary. An audio-visual example of the performance that is possible with such beamformers can be accessed at www.hearingcrc.org/xc/xc4-applications-of-binaural-signal-processing/.

Describing beamformer performance

Unlike a conventional, static directional microphone, the CRC beamformers are adaptive, so do not have a single polar pattern or directivity index that captures their performance. Figure 3 shows how dramatically the polar diagram can change when other signals are present in addition to the target signal, the sensitivity to which the polar diagram represents.

Beam width

Just how super-directional should a beamformer be? The narrower the beam-width, the greater the SNR enhancement that is possible, especially when the dominant competing sound(s) come from the frontal hemi-field. However, the narrower the beam-width, the greater the chance that listeners will misalign their heads, thus decreasing sensitivity to the target, or that targets will be distorted if, due to the effects

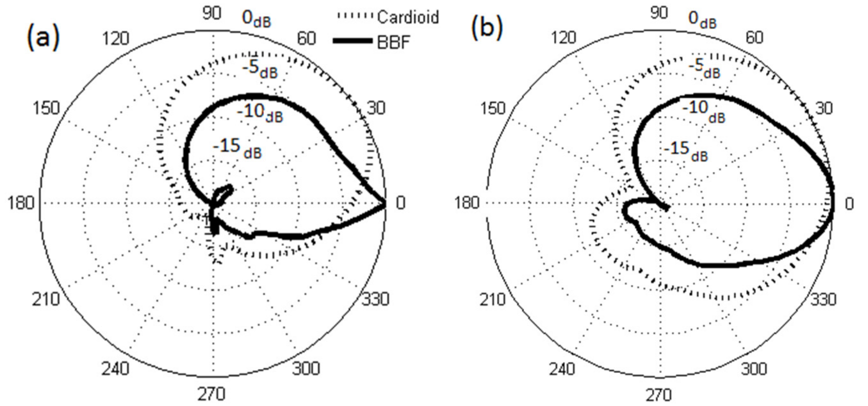


Fig. 3: Polar diagrams measured for a beamformer (a) when a single target source is varied in azimuth, and (b) when in addition to the target signal, speech babble comes from eight loudspeakers spaced every 5 degrees around the listener in the horizontal plane. In both diagrams, the dotted line shows the pattern for a conventional directional microphone as a comparison.

of other signals, some target components are assessed as coming from within the beam aperture and others as coming from outside the beam aperture.

Figure 4 shows, for three different beam-widths, how the beamformer improved intelligibility relative to two independent cardioid directional microphones for seven listeners with mild to moderate hearing loss. Each listener was tested at the SNR for which he or she obtained 50% of items correct when listening to the cardioid microphones. The narrowest beamformer gave the worst performance for those listeners most able to communicate at very poor SNRs. The remaining results in this paper were therefore obtained using beamformers that did *not* have extremely narrow beams.

Retention of spatial cues

Spatial cues are important to listeners for many reasons: awareness of one's surroundings, localization of desired target sounds, and spatial separation of desired targets from unwanted competition. To investigate their effect on intelligibility, spatial cues were intentionally removed from both the beamformer and the reference condition (independent cardioid directional microphones). In both cases, the signals normally applied separately to each ear were mixed and the mixture applied to both ears – that is, diotic presentation.

Figure 5 shows the results. When spatial cues were present for both processing types, beamformer performance was 17 percentage points higher than cardioid performance (significant with p=0.003). Removal of spatial cues from the cardioid microphone outputs decreased performance by 39 percentage points. By contrast removal of the spatial cues from the beamformer decreased its performance by only 8 percentage points. The interpretation of this in unclear. One possibility is that the beamformer processing was already taking advantage of spatial cues, much as listeners do when listening to separate left and right ear signals, so that a loss of spatial cues is of less consequence for intelligibility. A second possibility is that the beamformer was not adequately retaining spatial cues in the

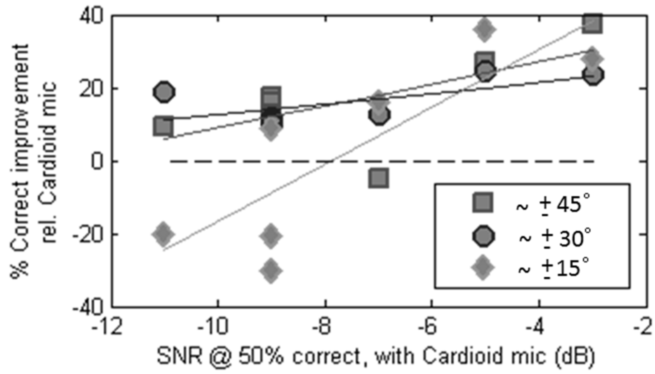


Fig. 4: Intelligibility improvement (percentage points) relative to independent cardioid microphones for beamformers with different beam widths. The target was presented to the front, and independent speech babble was presented from each of the remaining 45-degree intervals around the listener.

dichotic condition, so there were fewer spatial cues to remove. Our interpretation, based on careful but informal listening trials was that both of these possibilities were occurring, and we strengthened the retention of spatial cues within the beamformer for subsequent experiments.

Dynamic listening situations

Dynamic listening situations, where the direction of arrival of the target sound changes rapidly, such as in a group discussion, are potentially challenging for beamformers. To investigate this, we compared performance for a single frontal talker to performance with two talkers engaged in natural conversation. In the latter case, one talker was presented from the front and the second was randomly presented from either −45° or +45°. Listeners were encouraged to turn towards each talker throughout the conversation. Strong competing talkers were included at −45° and +135°, or in a second configuration at −90° and +90°. In both cases, weaker background (uncorrelated) cafeteria noises were placed at all other multiples of 45° around the circle. In this experiment, rather than measure speech intelligibility, we measured the acceptable noise level (ANL) for cardioid microphone and for beamformer processing. Listeners first adjusted the gain for the target sound to give a comfortable level, and then the competing sounds to the loudest level they would be willing to tolerate for sustained listening. The ANL was the SNR at this just-tolerable noise level.

Figure 6 shows the improvement in ANL offered by the beamformer over the independent cardioid microphones for 4 listeners with normal hearing and 11 listeners with mild to moderate sensorineural hearing loss. For reasons that will become apparent later in this paper, we think that the beamformer offers the greatest advantages to those with the greatest hearing loss. For the two-talker condition, however, there appears to be less advantage for the listeners with a mild loss than for those with either normal hearing or moderate loss. We therefore fitted a quadratic curve to the data. For the single frontal talker condition, the advantage of the beamformer varies less markedly, but again a quadratic curve was fitted. The advantage is, nonetheless, about 2 dB, almost independent of hearing loss over this range of hearing losses. This 2-dB improvement in ANL enabled by the beamformer is smaller than we have obtained in other single talker experiments, a difference we ascribe to the strong competition being only 45° away from the target talker in this experiment, rather than being equally distributed across azimuths. A possible interpretation of the quadratic variation of benefit in the two talker conversation is that:

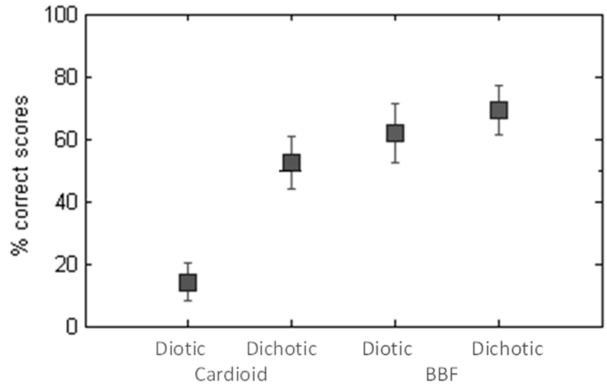


Fig. 5: Percent correct intelligibility, with 95% confidence intervals for independent cardioid directional microphones and the binaural beamformer under dichotic and diotic conditions. Target was presented to the front, and independent two-talker noise was presented from each of the remaining 45 degree intervals around the listener.

- those with normal hearing had sufficiently good hearing to quickly and accurately track the talker location and hence orient their head optimally, even if the salience of localization cues was reduced by the beamformer;
- those with mild loss had their ability to track the target talker negatively impacted by the beamformer; which the improved SNR offered by the beamformer only just made up for; and
- those with moderate loss had reduced ability to track the target talker even with the cardioid microphones, and so were less affected by the reduced spatial cues in the beamformer.

Real-life noises, reverberation and distances

The performance of all directional microphones is adversely affected by increasing reverberation times and distance from the source, as directivity cannot be useful if effectively all sounds come from all directions. Evaluating beamformers in real-world conditions is therefore important to get a proper view of their capabilities. To achieve this in a controlled manner, recordings of the background sounds picked up by dual omnidirectional microphones inside behind-the-ear (BTE) hearing aid cases worn on each side of the first author's head were made in 30 real-world locations. At each location, the impulse response from an imaginary talker's position to each of the four microphones was also recorded. These impulse responses were later convolved with anechoic speech to provide the target talker signal that would have been received in each of these situations had a talker been present at the appropriate distance directly in front of the listener. Target to background signal to noise ratios were set appropriate to the actual SPL of the background noise based on the data in Pearsons *et al.* (1977). The combined target and background noise signals were then processed to provide stereo signals corresponding to:

- omnidirectional microphones;
- cardioid directional microphones;
- binaural beamformer, with retention of some spatial information;
- an "ideal" beamformer, formed by using the cardioid directional microphones and simply then increasing the SNR by 5 dB.

Listeners (12 with normal hearing and 24 with hearing loss) were asked to rate, using a slider scaled from 0 (very poor) to 1 (perfect), each listening situation for listening effort, naturalness, noisiness, smoothness, distortion and overall acceptability.

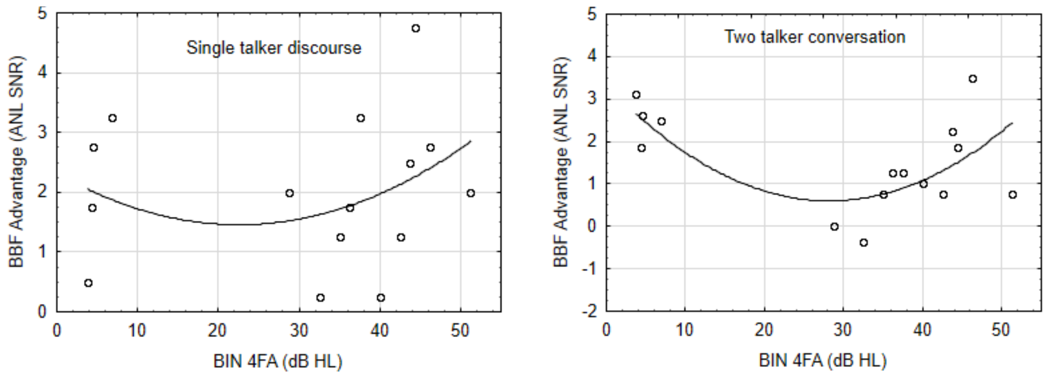


Fig. 6: Improvement in ANL for the beamformer relative to cardioid microphones in the one-talker and two-talker situations.

Figure 7 shows the sound quality rating differences relative to the cardioid. On all measures beamformer processing was preferred to cardioid processing. The extent of the preferences varied from 0.15 scale points (distortion) to 0.55 scale points (listening effort). In each case, preference for the beamformer was similar to that for the ideal beamformer, indicating that the beamformer *sounded* like it was giving about a 5-dB improvement in SNR. One of the benefits that was evident for the beamformer was a marked reduction in wind noise in those outdoor situations where wind noise was present. Although the omni microphone was rated below the cardioid microphone for 5 out of the 6 qualities, the difference is always small. This reflects the very limited advantage that a standard directional microphone can provide in reverberant listening situations. The higher directivity obtainable with a beamformer substantially increases the range of situations in which directivity is beneficial.

Application to cochlear implants

It seems likely that the net benefit offered by beamformers reflects the advantage achieved by increasing the SNR, offset by the disadvantage caused by any loss of spatial information and any distortions created by the beamformer that are perceived by the listeners. Because of the limited auditory ability of listeners with severe loss, including those listening through cochlear implants, these disadvantages should be smaller, thus creating a larger net benefit for these listeners. Performance of the beamformer was evaluated for 10 users of bilateral cochlear implants, under conditions of sparse competition (competing talkers at 60°, 90°, and 270°) and diffuse competition (competing talkers at 45°-intervals from 45° to 315°).

Figure 8 shows the SRTn values achieved with the beamformer. Depending on performance with the omni mic, the improvement in SRTn was on average 8.8 dB SNR for the sparse talker condition, and varied from 4 to 8 dB for the diffuse competing talker condition. Although it was not possible to use a cardioid reference condition in this experiment, the benefit relative to omni microphones considerably exceeds the benefit in SRTn typically offered by cardioid microphones relative to omni.

SRTn benefit at positive SNRs

The results so far, especially in combination with the subjective impression of beamforming, contain a paradox. The subjective impression is of a very marked

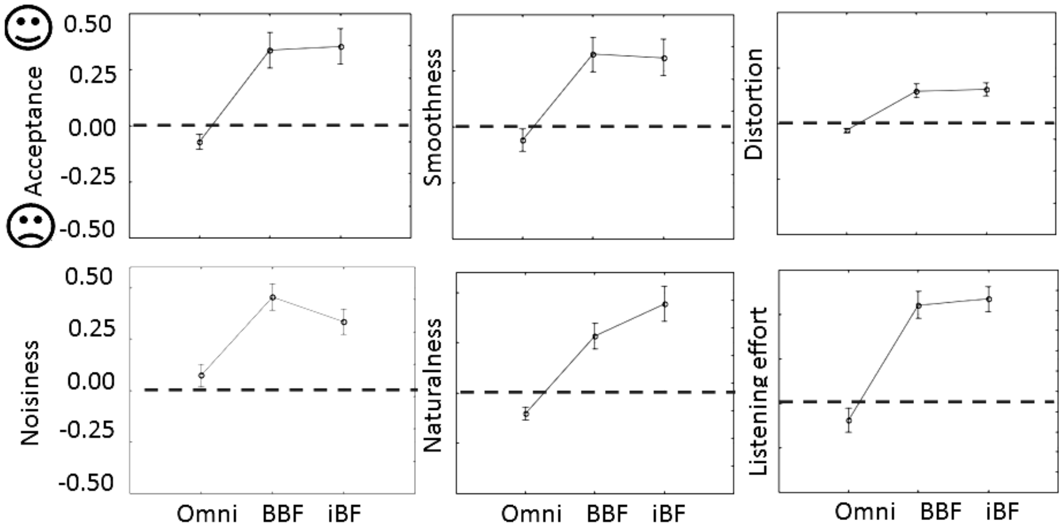


Fig. 7: Sound quality ratings relative to cardioid for the omni, beamformer (BBF) and ideal beamformer (IBF) microphone systems, averaged across the 30 listening situations.

improvement relative to cardioid microphones, which is consistent with the SRTn improvement for cochlear implantees (Fig. 8) and the quality ratings for hearing aid wearers (Fig. 7). The intelligibility improvements re cardioid at SRTn for hearing aid wearers are, however, "only" around 20 percentage points (Figs. 4 and 5), equivalent to about a 2-dB improvement in SRTn to cardioid microphones at 0 dB SNR, decreasing to 3 dB at −15 dB SNR. Is the smaller benefit measured in SRTn because SRTn typically occurs at very negative SNRs, or is it because the improved SNRn is offset by some distortions introduced by the beamformer, such as a reduction in the salience of spatial cues?

To investigate this, we performed an additional experiment in which we made the test material difficult by using casually articulated nonsense CVC syllables, and in which we targeted the 50% point on the psychometric function which was 20% lower than the scores obtained in quiet. This was evaluated in a diffuse background formed from competing talkers at 45° intervals from 45° to 315°.

Figure 9 shows the SRTn and acceptance scores for beamformer relative to cardioid. As shown in the figure, although we created the speech test with the aim of hearing impaired subjects obtaining SRTn at SNRs at or above 0 dB (typical of real-life conversational levels), SRTn with the cardioid microphone nonetheless ranged from -10 to +2 dB across the 26 participants with mild to moderate hearing loss. Beamformer SRTn benefit relative to cardioid was, on average, 1.8 dB SNR. Participants also rated acceptability of the amplified sound on a 1 to 10 scale, when measured at a SNR of 0 dB. On average, the beamformer produced a score 1.5 scale points higher than the cardioid microphone.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the Australian Department of Health, the Cooperative Research Centre scheme, the National Health and Medical Research Council, and the Hearing Industry Research Consortium.

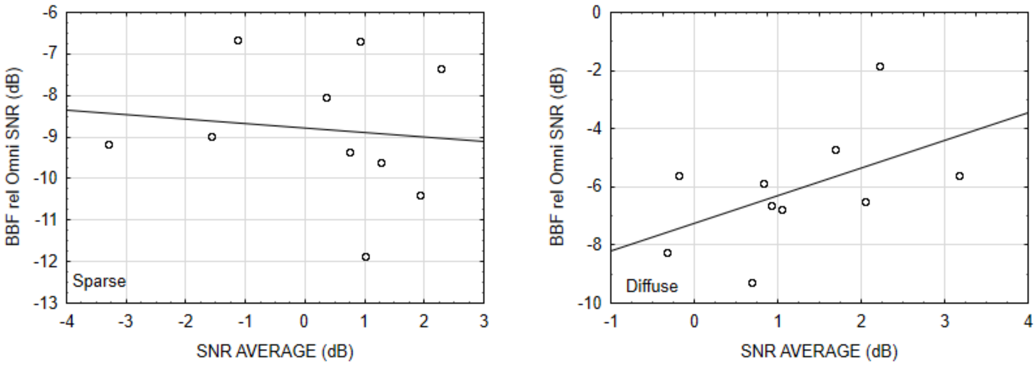


Fig. 8: SRTn in noise (individual data points and regression lines) for the beam-former SRTn minus omnidirectional microphone SRTn versus performance averaged across the SRTn values for BBF and omnidirectional microphone.

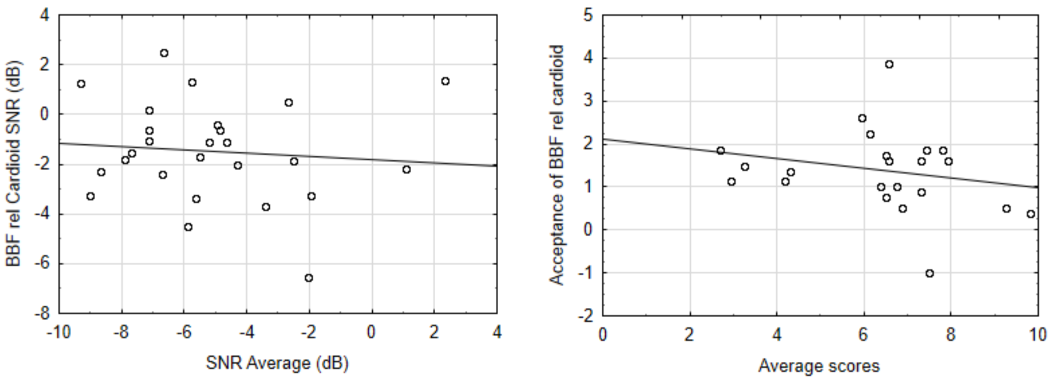


Fig. 9: SRTn performance of BBF relative to cardioid for each listener relative to the scores averaged across BBF and cardioid. (a) shows difference in SRTn and (b) shows difference in acceptance ratings. The solid lines show the corresponding regressions.

REFERENCES

Anderson, S., Parbery-Clark, A., White-Schwoch, T., and Kraus, N. (**2013**). "Auditory brainstem response to complex sounds predicts self-reported speech-in-noise performance," J. Speech Lang. Hear. Res., **56**, 31-43.

Beach E., Gilliver, M., and Williams, W. (**2013**). "Sound Check Australia: A citizen science approach to noise and hearing conservation research," National Hearing Conservation Association Conference, St Petersburg, Florida.

Bharadwaj, H.M., Verhulst, S., Shaheen, L., Liberman, M.C., and Shinn-Cunningham, B.G. (**2014**). "Cochlear neuropathy and the coding of supra-threshold sound," Front. Sys. Neurosci., **8**, 26.

Bramhall, N., Ong, B., Ko, J., and Parker, M. (**2015**). "Speech perception ability in noise is correlated with auditory brainstem response wave I amplitude." J. Am. Acad. Audiol., **26**, 509-517.

Cameron, S. and Dillon, H. (**2007**). "Development of the Listening in Spatialized Noise – Sentences Test (LISN-S)." Ear. Hearing, **28**, 196-211.
Chin, T. and Rickard, N. (**2012**). "The music USE (MUSE) questionnaire: An instrument to measure engagement in music." Music Percept., **29**, 429-446.
Daneman, M. and Carpenter, P. (**1980**). "Individual differences in working memory and reading." J. Verb. Learn. Verb. Be., **19**, 450-466.
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.
Glyde, H., Cameron, S., Dillon, H., Hickson, L., and Seeto, M. (**2013a**). "The effect of hearing impairment and aging on spatial processing ability." Ear. Hearing, **34**, 15-28.
Glyde, H., Buchholz, J.M., Dillon, H., Cameron, S., and Hickson, L. (**2013b**). "The importance of interaural time differences and level differences on spatial release from masking." J. Acoust. Soc. Am. **134**, EL147-152.
Glyde, H., Buchholz, J.M., Nielsen, L., Best, V., Dillon, H., Cameron, S., and Hickson, L. (**submitted**). "Effect of audibility on spatial release from speech-on-speech masking."
Helfer, K.S. and Jesse, A. (**2015**). "Lexical influences on competing speech perception in younger, middle-aged, and older adults." J. Acoust. Soc. Am., **138**, 363-376.
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.
Moore, B.C.J. and Sęk, A. (**2009**). "Sensitivity of the human auditory system to temporal fine structure at high frequencies." J. Acoust. Soc. Am., **125**, 3186-3193.
Moore, B.C.J., Creeke, S., Glasberg, B.R., Stone, M.A., and Sęk, A. (**2012**). "A version of the TEN Test for use with ER-3A insert earphones." Ear. Hearing, **33**, 554-557.
Noble, W., Jensen, N.S., Naylor, G., Bhullar, N., and Akeroyd, M.A. (**2013**). "A short form of the Speech, Spatial and Qualities of Hearing scale suitable for clinical use: The SSQ12," Int. J. Audiol., **52**, 409-412.
Pearsons, K.S., Bennett, R.L., and Fidell, S. (**1977**). "Speech levels in various noise environments." Washington, D.C., U.S. Environmental Protection Agency.
Plack, C.J., Barker, D., and Prendergast, G. (**2014**). "Perceptual consequences of "hidden" hearing loss," Trends Hear., **18**, 1-11.
Robertson, I.H., Ward, T., Ridgeway, V., and Nimmo-Smith, I. (**1996**). "The structure of normal human attention: The Test of Everyday Attention." J. Int. Neuropsychol. Soc., **2**, 525-534.
Schaette, R. and McAlpine, D. (**2011**). "Tinnitus with a normal audiogram: Physiological evidence for hidden hearing loss and computational model." J. Neurosci., **31**, 13452-13457.
Skoe, E. and Kraus, N. (**2013**). "Musical training heightens auditory brainstem function during sensitive periods in development." Front. Psychol., **4**, 622.
Slater, J., Skoe, E., Strait, D.L., O'Connel, S., Thompson, E., and Kraus, N. (**2015**). "Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program." Behav. Brain Res., **291**, 244-252.
Stamper, G.C. and Johnson, T.A. (**2014**). "Auditory function in normal-hearing, noise-exposed human ears," Ear Hearing, **36**, 172-184.

Cognitive compensation of speech perception in hearing loss: How and to what degree can it be achieved?

DENIZ BAŞKENT[1,2,*], PRANESH BHARGAVA[1,2], JEFTA SAIJA[1,2], JEANNE CLARKE[1,2], MICHEL R. BENARD[1,3], CARINA PALS[1,2], ANASTASIOS SARAMPALIS[2,4], ANITA WAGNER[1,2], AND ETIENNE GAUDRAIN[1,2,5]

[1] *University of Groningen, University Medical Centre Groningen, Department of Otorhinolaryngology/Head and Neck Surgery, Groningen, The Netherlands*

[2] *University of Groningen, Graduate School of Medical Sciences, Research School of Behavioural and Cognitive Neurosciences, Groningen, The Netherlands*

[3] *Pento Speech and Hearing Center Zwolle, Zwolle, The Netherlands*

[4] *University of Groningen, Department of Psychology, Groningen, The Netherlands*

[5] *Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics, CNRS, Université de Lyon, Lyon, France*

In daily life, speech is often degraded due to environmental factors, but its perception can be enhanced using cognitive mechanisms. Such compensation not only relies on increased cognitive processing (listening effort), but also makes use of context, linguistic knowledge and constraints. In hearing impairment, the speech signal is additionally and intrinsically degraded due to loss of audibility and/or suprathreshold deficiencies. In cochlear implants, the signal transmitted is spectro-temporally degraded. Hence, it has not been clear if hearing-impaired individuals and hearing-device users can as successfully use the cognitive compensation mechanisms, due to the interactive effects of these degradations with aging and hearing device front-end processing. The speech intelligibility tests are not capable of characterizing the cognitive compensation mechanisms. In our research, reviewed here, we have employed new approaches (phonemic restoration, dual-task paradigm, eye tracking, verbal response times) to answer this research question. Our results have shown that there is a fine balance between the speech degradations and their top-down compensation. This can be broken in advanced degrees of hearing impairment or due to inadequate device settings. With degraded speech, sentential context can still be used. Yet, this may come at the cost of delayed processing, likely drawing on more cognitive resources then timely integration of semantic information by normal-hearing listeners. Aging does not always have to have a negative effect; long-term linguistic and lexical knowledge may be successfully employed to achieve compensation. These findings indicate that new measures of cognitive processes need to be developed and used in clinics and device development, to comprehensively capture speech comprehension abilities and to improve diagnostic and rehabilitation procedures and tools.

*Corresponding author: d.baskent@umcg.nl

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

INTRODUCTION

Understanding speech under ideal conditions presents little ambiguity. As a result, lexical activation is automatic, requiring minimal cognitive processing for the decoding of the message (e.g., Marslen-Wilson and Welsh, 1978). In real life, listening conditions are hardly ideal. The speech signal is usually distorted by poor room acoustics, masked by background sounds, and heavily reduced in acoustic speech cues. Resolving the increased ambiguity due to these factors calls for cognitive mechanisms to be engaged (e.g., attention, use of grammatical and syntactical constraints, context). This disambiguation must be accomplished in a rapid pace so that the conversation may continue. As a result, top-down mechanisms play an important role in compensating for factors complicating daily life speech communication (Mattys *et al.*, 2012), especially for hearing-impaired (HI) individuals. Similar to the external or articulation-related factors listed above, hearing impairment is another factor that can negatively affect speech intelligibility. This may be the direct result of missing speech cues due to reduced audibility, or as the consequence of distortions due to supra-threshold factors related to hearing impairment. Hearing devices can also change the speech signals, for example, due to front-end processing, or due to the limitations of the speech transmission to the auditory nerve, such as the case for cochlear implants (CIs). A further compromise may occur due to age-related changes in cognitive processes (Salthouse, 1996).

Cognitive processes of speech perception have been of special interest to our group. The speech intelligibility test commonly used for speech audiometry in the clinic provides only a partial picture of an individual's speech communication skills. This score only provides one number for speech perception, tested under ideal conditions of one (clearly articulated) word or sentence presented at a time, without revealing any of the underlying processes of the comprehension. In our research, we have employed new approaches to explore if the HI individuals can still benefit from top-down compensation mechanisms, or if the cognitive processes of speech comprehension would differ for them. If latter, this difference could be one of the factors contributing to difficulties HI listeners experience in perceiving speech in noise. However, because such differences are not yet fully studied and only poorly understood, no adequate solutions can yet be offered.

TOP-DOWN RESTORATION OF INTERRUPTED SPEECH

In perception, pieces of information that belong to a common object are segregated (from others), and grouped together (Wagemans *et al.*, 2012), making perception easier and more efficient. This tendency for forming a perceptual object from perceived pieces can also enhance perception of degraded speech. As early as in the 1950s, Miller and Licklider (1950) observed that interrupted speech remained highly intelligible for a wide range of interruptions (from very slow interruptions of 0.1 Hz to as high as 10 kHz), despite a large amount of missing speech information. This is partially due to the acoustic redundancy in speech signals, where speech cues are coded in multiple ways (Best *et al.*, 1981; Lippmann, 1996), and the linguistic redundancy, which comes from rich sentential context (Gillette and Wit, 1998).

Hence, the brain can overcome missing speech information with top-down restoration. The restoration can be so strong that, under specific circumstances, listeners may not even be aware of the missing part of a speech signal. Warren (1970), for the first time, demonstrated this with speech with a silent gap that was filled with a coughing sound. While such non-speech filler does not contribute to speech information, it nonetheless serves to create a continuity illusion, due to the strong grouping tendency of the human perceptual system to form an object.

Adding a filler (usually a broadband noise) in the gaps of interrupted speech can also lead to an increase in intelligibility (Fig. 1). In this case, the filler noise hides the spurious cues from the silent gaps that can be wrongfully attributed to an incorrect word. It also increases the ambiguity, perhaps also increasing reliance on context cues. The resulting intelligibility improvement provides a measure of phonemic restoration benefit, which we have frequently used in our research to quantify the top-down compensation with hearing impairment.

Fig. 1: Speech stimuli used in phonemic restoration experiments. In the top panel, the speech is interrupted with silent intervals. In the bottom panel, the silent intervals are filled with filler noise, triggering phonemic restoration.

Top-down restoration and hearing impairment

In one of the earliest studies we have conducted, we have measured phonemic restoration effect with normal-hearing (NH), mildly HI, and moderately HI individuals. Our results (Fig. 2, left panel) showed that while mildly HI individuals could benefit from phonemic restoration, moderately HI individuals could not (Başkent, 2010; Başkent *et al.*, 2010). This observation implies that in mild HI (and with adequate amplification) top-down mechanisms can still be effectively used. However, as the degree of hearing impairment increases, and perhaps also as a result of suprathreshold factors coming into play (as it can happen in moderate to severe hearing loss), these mechanisms seem to lose their effect.

Top-down restoration and aging

Because many HI individuals tend to be older, we have also studied age effects on phonemic restoration (Saija *et al.*, 2014). Previous research had shown a negative effect of age on perception of interrupted speech with silent intervals (Bergman *et al.*, 1976), mostly attributed to the age-related decline in temporal processing (Gordon-Salant and Fitzgibbons, 1993). However, it was not clear if older listeners could still effectively use the top-down restoration mechanisms. Our expectations were twofold. If the age-related decline in cognitive factors such as processing speed or working memory is an important factor, age would work against the restoration ability. If the cognitive and linguistic skills, such as long-term world and linguistic knowledge, as well as good use of context (Pichora-Fuller, 2008; Salthouse, 2004), are important factors, age should not negatively affect restoration ability. Our results showed that phonemic restoration benefit was just as strong as with younger group (Fig. 2, right panel), supporting the latter. Benard *et al.* (2014) later confirmed that linguistic skills indeed seem to play an important role on perception of interrupted speech in general. If these findings can be corroborated with further studies, this is good news for older and HI population, as linguistic knowledge and skills can be improved with proper training.

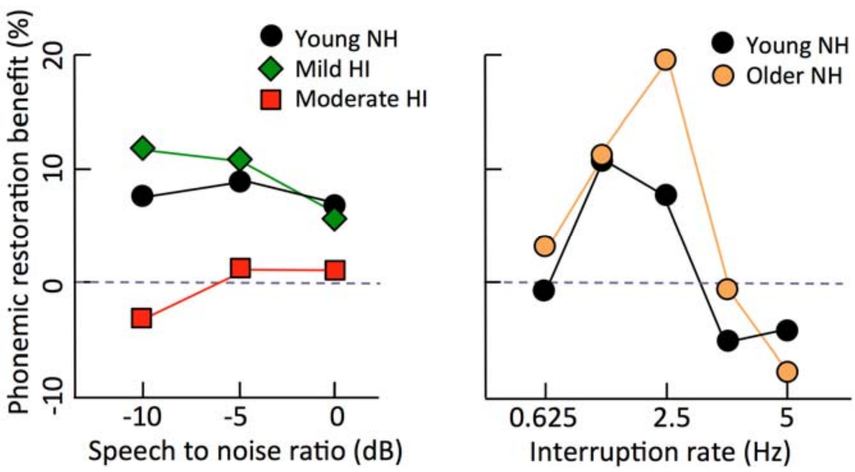


Fig. 2: Phonemic restoration benefit, shown for the effect of hearing impairment, as a function of the filler noise level (left panel; adapted from Başkent *et al.*, 2010), and shown for the effect of aging, as a function of interruption rate (right panel; adapted from Saija *et al.*, 2014).

Top-down restoration and hearing devices

In CIs, the speech signal is directly delivered to the auditory nerve via electric stimulation. This signal, mainly limited by the electrode-nerve interface, retains gross spectral information and temporal envelope, while all spectro-temporal fine structure is lost. The re-learning of the degraded speech requires substantial adaptation following the surgery (Lazard *et al.*, 2014). While many CI users reach acceptable speech intelligibility in quiet, this is not universal, with large variation across individuals (Blamey *et al.*, 2013). Further, perception of speech in complex environments with interfering background sounds remains a challenge (Friesen *et al.*, 2001; Stickney *et al.*, 2004).

As CI users have to cope with the degraded speech on a daily basis, top-down restoration mechanisms would especially be important for them. However, it is not clear if they could manage to benefit from top-down restoration given the impoverished CI speech. Earlier studies had shown that CI users have difficulty with perception of interrupted speech (Bhargava *et al.*, 2015; Chatterjee *et al.*, 2010; Nelson and Jin, 2004), and data from acoustic CI simulations implied no restoration benefit (Başkent, 2012). Data from actual CI users, however, presented a more complicated picture (Bhargava *et al.*, 2014). On average, CI users did not show phonemic restoration benefit in conditions where such benefit was observed in NH, as was expected from simulations. However, individual data showed that CI users with highest speech intelligibility scores also showed restoration benefit (Fig. 3, left panel). The causality in these data is not clear, i.e., are these good users because they use their top-down mechanisms better in general or is there a third factor that makes them good user overall? Yet, the data hint at the large variation in the use of top-down mechanisms within hearing-device users, and the importance of investigating the individual differences in such data.

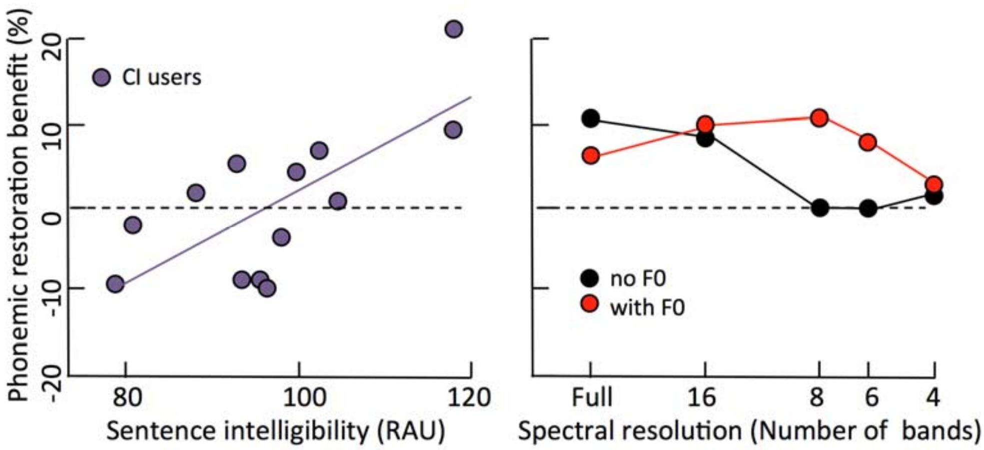


Fig. 3: Phonemic restoration benefit shown for individual CI users, as a function of baseline sentence identification score (left panel; adapted from Bhargava *et al.*, 2014), for NH listeners tested with an acoustic CI simulation, as a function of number of spectral bands (right panel; adapted from Clarke *et al.*, 2015).

The voice pitch, namely F0, is a very important cue for perceptual organization in general, and for grouping speech segments. However, this cue is only weakly delivered in CIs (Moore and Carlyon, 2005), perhaps contributing to reduced ability to separate speech from background sounds. As an exploration into the effects of device features on restoration benefit, we have used TANDEM-STRAIGHT (Kawahara and Morise, 2011) to produce noise-excited speech, a new approach to acoustic CI simulations, where we could simultaneously vary the spectral resolution and the presence/absence of F0 (Clarke *et al.*, 2015). Our results with NH listeners showed a highly interactive picture (Fig. 3, right panel). When spectral resolution was high (16 bands), where there was restoration benefit, or low (4 bands), where there was no benefit, absence or presence of F0 did not seem to matter. However, in the mid ranges of spectral resolution (6 and 8 bands), where the actual CI users functionally perform most similarly (e.g., Friesen *et al.*, 2001; Bhargava *et al.*, 2015), absence/presence of F0 seems to play a significant role in benefiting from restoration. Hence, the simulation results are in line with the observations from actual CI users, indicating that the device features can affect how a CI user can benefit from top-down restoration.

LISTENING EFFORT

Perception of degraded speech requires allocating more cognitive resources, especially that of working memory (Baddeley and Hitch, 1974), i.e., an increase in listening effort. This is a useful mechanism for maintaining a high-level intelligibility. However, it can also come at the cost of affecting other cognitive processes, such as remembering what is said (Rabbitt, 1968), as cognitive resources are limited (Kahneman, 1973).

Clinical diagnostic tools in audiological practice currently only include speech audiometry, which reveals an intelligibility score. While this score shows the capacity of the HI individual or hearing-device user for recognizing speech, it does not reveal the underlying cognitive processes. Some patients complain that they suffer from listening fatigue, likely a result of extended duration of increased listening effort (Hornsby, 2013; McGarrigle *et al.*, 2014). However, no clinical tool currently exists to quantify listening effort in clinical settings, other than attempts made in research (Mackersie and Cones, 2011; Rudner *et al.*, 2011; Sarampalis *et al.*, 2009; Zekveld *et al.*, 2010), despite a long history of general use of response times in sensory perception and speech recognition in general (Hecker *et al.*, 1966; Koga and Morant, 1923).

Recently, we have conducted a number of studies to show that simple audiometric speech scores may fail to capture the cognitive processes and listening effort needed for understanding speech via a CI. In an earlier study (Pals *et al.*, 2013), we have used a dual-task paradigm, where the participants had to simultaneously conduct a secondary visual task while also conducting the primary task of speech intelligibility. Based on the idea of limited cognitive resources and an interaction of the two tasks, this way one can measure the changes in the effort required for differing speech intelligibility conditions in the response times of the second task. We have used an acoustic CI simulation to change the quality and intelligibility of speech, by changing the number of spectral channels. As the number of channels increased, intelligibility, measured by accuracy, increased, and listening effort, measured by response time to the secondary task, decreased (Fig. 4, left and right panels, respectively). However, while intelligibility plateaued at 6 channels, listening effort continued to improve to 8 channels. Hence, while a clinical speech audiometry would indicate the same speech performance for both 6- and 8-channel settings, only the listening effort measure would indicate the additional benefit.

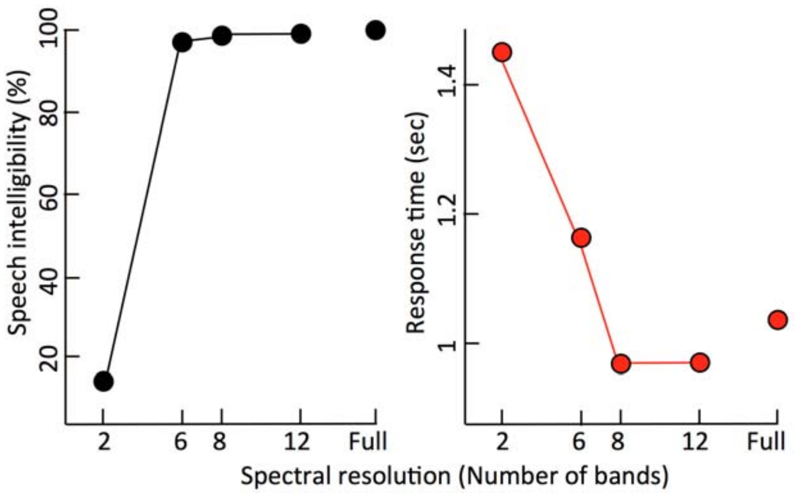


Fig. 4: Speech intelligibility from primary task (left) and response time from secondary task (right), shown as a function of the number of spectral channels of the acoustic CI simulation. Adapted from Pals *et al.* (2013).

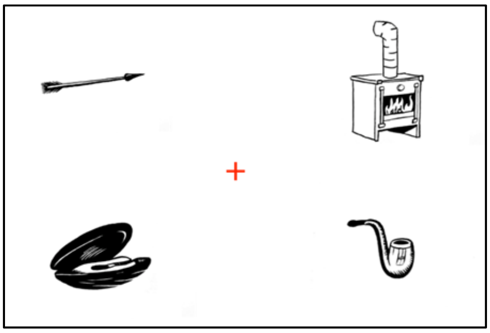


Fig. 5: Visual world paradigm screenshot used by Wagner *et al.* (2015) in measuring gaze fixations as a quantification of time course of speech comprehension.

CONTEXT EFFECT AND COGNITIVE PROCESSES

Recently, we have used an eye tracker for an online measure of lexical decision making (Wagner *et al.*, 2015). Specifically, we have measured gaze fixation, to quantify the time course of speech perception, and pupil dilation, to measure listening effort. Here, again using CI simulations, we have asked the questions if sentential context can help resolving ambiguity in word identification, despite the degradations of CI speech, and if yes, would the time course be the same. The gaze fixations were measured using visual world paradigm (Dahan and Gaskell, 2007), where the target word of a sentence ("pijp [pipe]") would be presented on the screen (Fig. 5), along with a word similar in sound ("pijl [arrow]"; phonological competitor), a word similar in meaning ("kachel [stove]"; semantic distractor), and an unrelated distractor (“mossel [mussel]”). When there is no context, the main confusion would come from the phonological competitor. When there is context, the confusion would come from the semantic distractor.

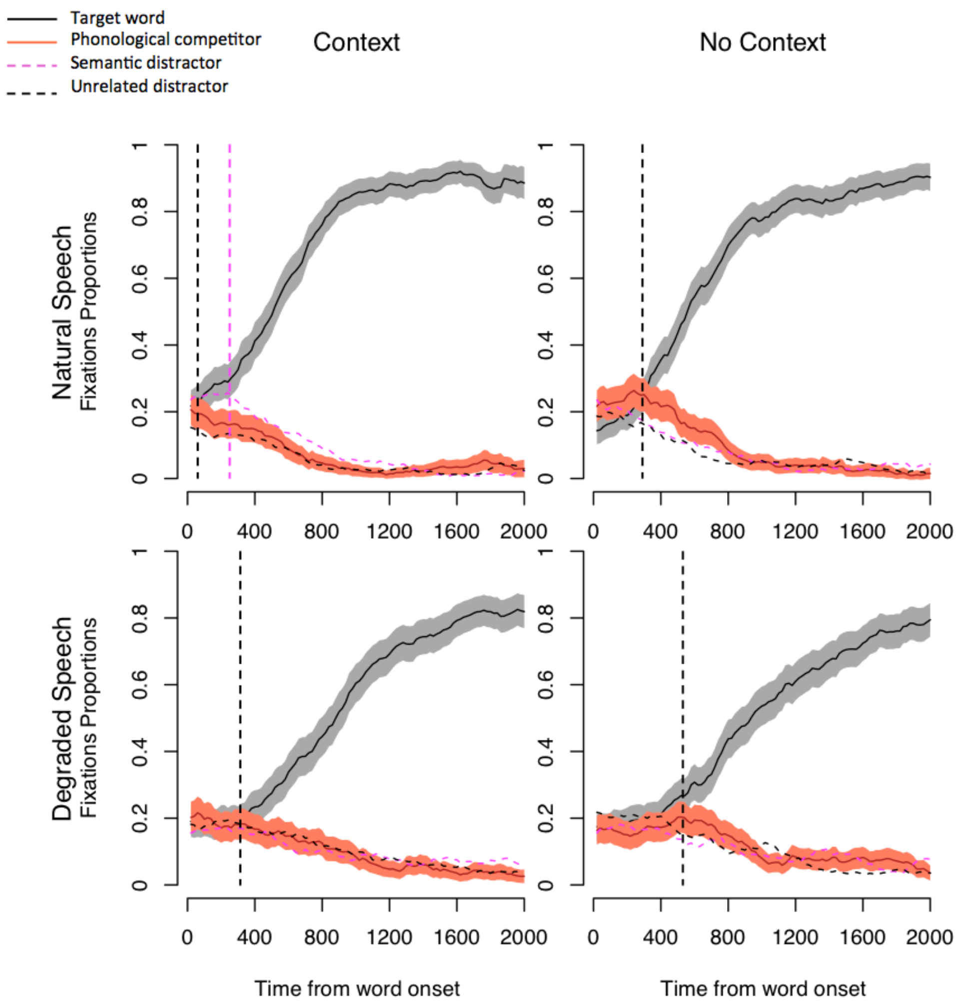


Fig. 6: Gaze fixations, as measured by a visual world paradigm, are shown for high- and no-context sentences (left and right, respectively), and for natural and degraded speech (top and bottom, respectively). Adapted from Wagner *et al.* (2015).

Fig. 6 shows the data from gaze fixations, with high- and no-context sentences (left and right panels, respectively), and without and with acoustic CI simulation (top and bottom panels, respectively). The most important data is the disambiguation point (marked with darker colour vertical dashed lines), where the target fixation (shown in grey in upper part of each panel) splits from the rest of the fixations. In natural speech, the disambiguation occurs much faster with context than with no context (comparison of left to right panels on top). With degraded speech, a similar effect is observed, but the disambiguation point comes at a significantly later time (lower panels). This observation implies that context is still helpful in dissolving the ambiguity despite the degradation. However, the caveat is that the semantic distractor is not showing an effect in degraded speech (indicated by lighter colour dashed line overlapping darker colour continuous line and darker colour dashed line in the 3 panels other than the top-left one) while it does in natural speech (top-left panel, also indicated by the vertical lighter colour dashed line). This implies that the semantic integration is not efficient, and considerably delayed, which likely would cause problems in real-life fast conversations. In short, while in NH listeners the use of semantic integration leads to a relief of resources needed for lexical access (or word finding), this source of relief is not functioning when processing degraded speech. As a result, the degraded speech cues at the early stages of speech processing seem to affect the later stages, possibly (and negatively) affecting higher-level functions. For example, the delayed processing will likely draw more on memory resources relative to NH listeners.

Currently, we are systematically investigating simpler measures that can be used in clinics, for example, simple measures of verbal response times (Pals et al., 2015). While the dual-task paradigm is proven a robust measure of listening effort, it is relatively difficult to set up. The two tasks have to interact just the right way. If one is too easy or too difficult, no effect will be observed. Further, a dual task can be too taxing for an older HI person. Similarly, eye tracking and pupillometry are robust methods for quantifying cognitive mechanisms of speech perception and listening effort. While these require expensive hardware, for populations where behavioural measures may be difficult to apply (such as in very young children), eye tracker still remains as a good potential option.

CONCLUSIONS

Overall, there seems to be a fine balance between the amount of bottom-up speech degradations and the effectiveness of the top-down compensation mechanisms. Our studies have shown that this balance can be broken in hearing impairment and/or use of hearing devices, making this population extra vulnerable in real-life noisy listening environments. There also is a strong effect of age, an important factor due to many HI individuals tending to be older, however, this effect is not easily predictable. While in some situations, such as perception of interrupted speech, age has a negative effect, in some others, such as phonemic restoration, there is no such effect. The latter is a very positive finding, as we have attributed the lack of age effect to vocabulary and linguistic knowledge that seem to be retained in advanced age, and these are entities that can potentially be improved with proper training. Hence, our results also indicate potential training tools for improving perception of degraded speech in HI individuals (e.g., Benard and Başkent, 2014).

Such complex and interactive effects of cognitive factors in speech perception with hearing loss cannot be readily captured with the existing traditional speech tests used in the audiological practice. Measures for online speech processes and for cognitive factors may reveal more to speech comprehension and communication, especially in real-life conditions, than intelligibility scores alone. New methods (such as proposed by Pals *et al.*, 2015; Wagner *et al.*, 2015; Winn *et al.*, 2015; Zekveld *et al.*, 2010) need to be incorporated into these practices, as well as into research and development of new hearing devices. With such methods, device features may be optimized and customized better for individuals, by taking into account more complex mechanisms of speech perception. Similarly, manufacturers may be able to better assess new device features. There is a possibility that some features are currently under-assessed, due to lack of such measures, and are perhaps discarded when they do not show a clear benefit in speech intelligibility. And lastly, new rehabilitation and training programs can be developed that take into account the cognitive processes of speech.

REFERENCES

Baddeley, A.D. and Hitch, G.J. (**1974**). "Working memory," in *The Psychology of Learning and Motivation*, Vol. 8, pp. 47-89.

Başkent, D. (**2010**). "Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores," J. Acoust. Soc. Am., **128**, EL169-EL174.

Başkent, D., Eiler, C.L., and Edwards, B. (**2010**). "Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss," Hear. Res., **260**, 54-62.

Başkent, D. (**2012**). "Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electric–acoustic stimulation," J. Assoc. Res. Otolaryngol., **13**, 683-692.

Benard, M.R. and Başkent, D. (**2014**). "Perceptual learning of temporally interrupted and spectrally degraded speech," J. Acoust. Soc. Am., **136**, 1344-1351.

Benard, M.R., Mensink, J.S., and Başkent, D. (**2014**). "Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities," J. Acoust. Soc. Am., **135**, EL88-EL94.

Bergman, M., Blumenfeld, V.G., Cascardo, D., Dash, B., Levitt, H., and Margulies, M.K. (**1976**). "Age-related decrement in hearing for speech: Sampling and longitudinal studies," J. Gerontol., **31**, 533-538.

Best, C.T., Morrongiello, B., and Robson, R. (**1981**). "Perceptual equivalence of acoustic cues in speech and nonspeech perception," Percept. Psychophys., **29**, 191-211.

Bhargava, P., Gaudrain, E., and Başkent, D. (**2014**). "Top–down restoration of speech in cochlear-implant users," Hear. Res., **309**, 113-123.

Bhargava, P., Gaudrain, E., and Başkent, D. (**2015**). "The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners," J. Assoc. Res. Otolaryngol. (in press).

Blamey, P., Artieres, F., Başkent, D., et al. (**2013**). "Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients," Audiol. Neurotol., **18**, 36-47.

Chatterjee, M., Peredo, F., Nelson, D., and Başkent, D. (**2010**). "Recognition of interrupted sentences under conditions of spectral degradation," J. Acoust. Soc. Am., **127**, EL37-EL41.

Clarke, J., Başkent, D., and Gaudrain, E. (**2015**). "Pitch and spectral resolution: a systematic comparison of bottom-up cues for top-down repair of degraded speech," J. Acoust. Soc. Am. (under revision).

Dahan, D. and Gaskell, G.M. (**2007**). "The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition," J. Mem. Lang., **57**, 483-501.

Friesen, L.M., Shannon, R.V., Başkent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am., **110**, 1150-1163.

Gillette, M. and Wit, E.-J.C. (**1998**). "What is linguistic redundancy: Technical report," (http://www. maths. unsw. edu. au/statistics/preprints/1998/s98-25).

Gordon-Salant, S. and Fitzgibbons, P. J. (**1993**). "Temporal factors and speech recognition performance in young and elderly listeners," J. Sp. Hear. Res., **36**, 1276-1285.

Hecker, M.H., Stevens, K.N., and Williams, C.E. (**1966**). "Measurements of reaction time in intelligibility tests," J. Acoust. Soc. Am., **39**, 1188-1189.

Hornsby, B.W. (**2013)**. "The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands," Ear Hearing, **34**, 523-534.

Kahneman, D. (**1973**). *Attention and Effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kawahara, H. and Morise, M. (**2011**). "Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework," SADHANA Acad. P. Eng. S., **36**, 713-722.

Koga, Y. and Morant, G. (**1923**). "On the degree of association between reaction times in the case of different senses," Biometrika, **15**, 346-372.

Lazard, D.S., Innes-Brown, H., and Barone, P. (**2014**). "Adaptation of the communicative brain to post-lingual deafness. Evidence from functional imaging," Hear. Res., **307**, 136-143.

Lippman, R.P. (**1996**). "Accurate consonant perception without mid-frequency speech energy," IEEE T. Speech Audi. P., **4**, 66-69.

Mackersie, C.L. and Cones, H. (**2011**). "Subjective and psychophysiological indices of listening effort in a competing-talker task," J. Am. Acad. Audiol., **22**, 113-122.

Marslen-Wilson, W.D. and Welsh, A. (**1978**). "Processing interactions and lexical access during word recognition in continuous speech," Cog. Psychol., **10**, 29-63.

Mattys, S.L., Davis, M.H., Bradlow, A.R., and Scott, S.K. (**2012**). "Speech recognition in adverse conditions: A review," Lang. Cog. Proc., **27**, 953-978.

McGarrigle, R., Munro, K.J., Dawes, P., Stewart, A.J., Moore, D.R., Barry, J.G., and Amitay, S. (**2014**). "Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'," Int. J. Audiol., **53**, 433-440.

Miller, G.A. and Licklider, J.C.R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am., **22**, 167-173.

Moore B.C.J. and Carlyon R.P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in Plack C.J., Oxenham A.J., Fay R.R., Popper A.N. (eds) *Pitch: Neural Coding and Perception.* Springer, New-York, NY, pp. 234-277.

Nelson, P.B. and Jin, S.H. (**2004**). "Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am., **115**, 2286-2294.
Pals, C., Sarampalis, A., and Başkent, D. (**2013**). "Listening effort with cochlear implant simulations," J. Speech Lang. Hear. Res., **56**, 1075-1084
Pals, C., Sarampalis, A., Van Rijn, H., and Başkent, D. (**2015**). "Validation of a simple response-time measure of listening effort," J. Acoust. Soc. Am. (in press).
Pichora-Fuller, K.M. (**2008**). "Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing," Int. J. Audiol., **47**, 72-82.
Rabbitt, P.M. (**1968**). "Channel-capacity, intelligibility and immediate memory," Q. J. Exp. Psychol., **20**, 241-248.
Rudner, M., Ng, H.N., Rönnberg, N., Mishra, S., Rönnberg, J., Lunner, T., and Stenfelt, S. (**2011**). "Cognitive spare capacity as a measure of listening effort," J. Hear. Sci., **1**, 47-49.
Saija, J.D., Akyürek, E.G., Andringa, T.C., and Başkent, D. (**2014**). "Perceptual restoration of degraded speech is preserved with advancing age," J. Assoc. Res. Otolaryngol., **15**, 139-148.
Salthouse, T.A. (**1996**). "The processing-speed theory of adult age differences in cognition," Psychol. Rev., **103**, 403-428.
Salthouse, T.A. (**2004**). "What and when of cognitive aging," Curr. Dir. Psychol. Sci., **13**, 140-144.
Sarampalis, A., Kalluri, S., Edwards, B., and Hafter, E. (**2009**). "Objective measures of listening effort: Effects of background noise and noise reduction," J. Speech Lang. Hear. Res., **52**, 1230-1240.
Stickney, G.S., Zeng, F.G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am., **116**, 1081-1091.
Wagemans, J., Elder, J.H., Kubovy, M., Palmer, S.E., Peterson, M.A., Singh, M., and von der Heydt, R. (**2012**). "A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization," Psych. Bull., **138**, 1218-1252.
Wagner, A., Pals, C., de Blecourt, C., Sarampalis, A., and Başkent, D. (**2015**). "Does signal degradation affect top-down processing of speech?" Proc. International Symposium on Hearing, Groningen, The Netherlands.
Warren, R.M. (**1970**). "Perceptual restoration of missing speech sounds," Science, **167**, 392-393.
Winn, M.B., Edwards, J.R., and Litovsky, R.Y. (**2015**). "The impact of auditory spectral resolution on listening effort revealed by pupil dilation," Ear Hearing, **36**, e153-e165.
Zekveld, A.A., Kramer, S.E., and Festen, J.M. (**2010**). "Pupil response as an indication of effortful listening: The influence of sentence intelligibility," Ear Hearing, **31**, 480-490.

Preference for compression speed in hearing aids for speech and music and its relationship to sensitivity to temporal fine structure

BRIAN C.J. MOORE[1,*] AND ALEKSANDER P. SĘK[2]

[1] *Department of Experimental Psychology, University of Cambridge, Cambridge, England*

[2] *Institute of Acoustics, Adam Mickiewicz University, Poznań, Poland*

Multi-channel amplitude compression is widely used in hearing aids. The preferred compression speed varies across individuals. Moore (2008) suggested that reduced sensitivity to temporal fine structure (TFS) may be associated with preference for slow compression. This idea was tested using a simulated hearing aid. We also assessed whether preferences for compression speed differ for speech and music. Eighteen hearing-impaired subjects were tested, and the stimulated hearing aid was fitted individually using the CAM2 method. On each trial a given segment of speech or music was presented twice, once processed with fast compression and once with slow compression, in random order. The subject indicated which segment was preferred and by how much. On average, slow compression was preferred over fast compression, more so for music, but there were distinct individual differences, which were highly correlated for speech and music. Sensitivity to TFS was assessed using the difference limen for frequency at 2 kHz and by two measures of sensitivity to interaural phase at low frequencies. The results for the DLFs, but not the measures of sensitivity to interaural phase, provided some support for the suggestion that preference for compression speed is affected by sensitivity to TFS.

INTRODUCTION

People with cochlear hearing loss usually experience loudness recruitment, and the associated reduced dynamic range (Fowler, 1936; Moore, 2007). Most modern hearing aids incorporate some form of amplitude compression or automatic gain control (AGC) to deal with this. In principle, AGC can make low-level sounds audible while preventing high-level sounds from becoming uncomfortably loud. However, controversy continues about the "best" way to implement AGC, and in particular whether it should be fast acting or slow acting (Gatehouse *et al.*, 2006a; 2006b). In this study we assessed the preferences of 18 hearing-impaired subjects for fast relative to slow compression, using a simulated hearing aid. The study was

*Corresponding author: bcjm@cam.ac.uk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

intended to answer two questions: (1) Are preferences for slow versus fast compression consistent for speech and music stimuli? For example, if an individual prefers slow compression for speech, will they also prefer slow compression for music? (2) Are preferences for compression speed related to sensitivity to temporal fine structure (TFS), as hypothesized by Moore (2008)?

Moore (2008) suggested that individual differences in “best” compression speed might be related to sensitivity to the temporal fine structure (TFS) of the waveforms evoked by sounds on the basilar membrane. Hearing-impaired subjects perform more poorly than normal-hearing subjects on tasks that are thought to rely on sensitivity to TFS, for example discrimination of harmonic and frequency-shifted tones (Hopkins and Moore, 2007; 2010b; Moore, 2014), interaural phase discrimination (Lacher-Fougère and Demany, 2005; Hopkins and Moore, 2011), and detection of low-rate frequency modulation (Moore and Skrodzka, 2002; Strelcyk and Dau, 2009). Hopkins *et al.* (2008) and Hopkins and Moore (2010b) reported high variability in the ability of hearing-impaired subjects to use TFS information, some being completely insensitive to TFS information and others having a similar ability to use TFS as people with normal hearing. Moore (2008) suggested that hearing aid users with good TFS sensitivity may benefit more from fast than from slow compression, as TFS information may be important for listening in the dips of a fluctuating background (Moore and Glasberg, 1987), and fast compression increases the audibility of signals in the dips (Moore *et al.*, 1999). However, people with poor TFS sensitivity may rely mainly on temporal envelope information in different frequency channels, and for them it may be important to avoid the temporal envelope distortion that can be introduced by fast compression (Stone and Moore, 1992; 2004; Stone *et al.*, 2009).

The present study used hearing-impaired subjects to assess whether relative preferences for fast versus slow compression were related to sensitivity to TFS. A previous study did not support that hypothesis, but that study used simulated hearing loss and simulated loss of sensitivity to TFS (Hopkins *et al.*, 2012). Since hearing aids are often used for listening to music as well as for listening to speech (Leek *et al.*, 2008; Kochkin, 2010; Madsen and Moore, 2014), we used both speech stimuli and music stimuli. The results were intended to determine whether individual preferences for compression speeds were consistent across speech and music stimuli. All subjects were assessed for their sensitivity to TFS, using three tasks.

METHOD

Subjects

Eighteen subjects (11 male) with moderate-to-severe sensorineural hearing loss were paid to participate. Their ages ranged from 56 to 87 years. Sixteen were current users of multi-channel compression hearing aids and two did not use hearing aids. Audiometric thresholds were measured for all audiometric frequencies from 0.25 to 10 kHz. Only the better ear of each subject was tested using the paired-comparison procedure. The hearing loss in the test ear ranged from 8 to 60 dB at 500 Hz, 6 to 64 dB at 1000 Hz, 26 to 70 dB at 2 kHz, 48 to 74 dB at 4000 Hz, and 54 dB to >100 dB at 8000 Hz.

Simulated hearing aid

The simulated hearing aid was the same as described by Moore *et al*. (2010a) and Moore and Sek (2013). Briefly, the aid included a digital filter for overall shaping of the frequency response prior to splitting the signal into five channels, with independent compression in each channel. The insertion gains for a 65-dB speech-shaped noise and the compression ratios (CRs) for the five channels were set according to the CAM2 prescription method (Moore *et al.*, 2010b), modified slightly as described in Moore and Sek (2013). The compression thresholds were set to 49, 41, 40, 34, and 28 dB SPL for channels 1-5, respectively.

To simulate fast compression, the attack/release times (ANSI, 2003) were set to 10/100 ms for all channels. To simulate slow compression, the attack/release times were set to 50/3000 ms for all channels. The CR was limited to 3 when fast compression was used, since there is evidence that with fast compression high CRs can lead to reduced speech intelligibility (Verschuure *et al.*, 1996). The CR was allowed to have any value up to 10 when slow compression was used.

Stimuli

The speech stimuli were digitally recorded segments of running speech (connected discourse) obtained from one male and one female talker of British English. One 4.8-s segment of speech was selected for each talker. The music signals were: a 7.3-s segment of a jazz trio (piano, bass, and drums); a 5.6-s segment of an orchestra (including brass instruments and cymbals) performing Bizet's Carmen; a 3.5-s segment of a xylophone playing the "Sabre Dance" by Khachaturian (anechoic recording); and an 8.4-s segment of a counter-tenor accompanied by guitar and recorder. For all signals, the diffuse-field equivalent level at the input to the simulated hearing aid was 50, 65, or 80 dB SPL.

Paired-comparison procedure

The procedure was similar to that described by Moore and Sek (2013). On each trial the same segment of sound was presented twice in succession, once processed with fast compression and once with slow. The possible orders were used equally often and the order was randomized across trials. Within a given pair of sounds, the only difference between the sounds was in the compression speed; the input level was always the same. The subject was asked to indicate which of the two was preferred and by how much, using a slider on the screen. The continuum was labelled "1 much better", "1 moderately better", "1 slightly better", "equal", "2 slightly better", "2 moderately better", and "2 much better".

For a given trial, if fast compression (FAST) was preferred the slider position was coded as a negative number and if slow compression (SLOW) was preferred the slider position was coded as a positive number. The overall score for each compression speed and stimulus type (e.g., classical music) was obtained by averaging all of the sub-scores obtained for that speed and stimulus type. A score of –3 would indicate a very strong and perfectly consistent preference for FAST whereas a score of +3 would indicate a very strong and perfectly consistent preference for SLOW. A score of 0 would indicate no preference.

Measurement of sensitivity to TFS

To estimate sensitivity to TFS at medium frequencies, we measured the difference limen for frequency, DLF, using a method similar to that described by Moore and Ernst (2012). It is widely believed that the DLF is based on a temporal rather than a place mechanism for low and medium frequencies (Moore, 2014). A two-interval, two-alternative forced-choice task was used. One interval contained four successive 2-kHz tones. The other interval contained four successive tones whose frequency alternated between 2 kHz and 2 kHz + Δf. The subject had to choose the interval in which they heard a fluctuation in pitch. The value of Δf was varied adaptively to determine the DLF corresponding to 70.7% correct.

To estimate sensitivity to TFS at low frequencies, we used the TFS-LF test (Hopkins and Moore, 2010a; Sek and Moore, 2012), which estimates the threshold for discriminating an interaural phase (IP) of 0° from an IP of $\Delta\varphi$. For this test, the tones had a frequency of 500 Hz and the starting value of $\Delta\varphi$ was 180°. In addition, we used a new test, in which the IP difference was fixed at 180° and the frequency of the test tone was adaptively varied to determine the highest frequency at which the task could be performed (Füllgrabe *et al.*, 2015). The starting frequency was 500 Hz. The time pattern of the stimuli was the same as for the TFS-LF test. All subjects could perform the task when the frequency was made sufficiently low. We refer to the modified task as the TFS-AF task, where AF stands for adaptive frequency.

For all three tests, each tone lasted 400 ms, including 20-ms raised-cosine ramps. The silent gap between the tones within an interval was 100 ms. The gap between intervals was 400 ms. The stimuli were presented at 30 dB sensation level (SL).

RESULTS

Compression speed preferences for speech

The preference scores were averaged across the three levels. The average preference scores for the male talker and the female talker were highly correlated ($r = 0.93$, $p < 0.001$). This indicates that the subjects were consistent in their ratings across talkers. In what follows, only the mean ratings across talkers are considered. Fig. 1 shows individual and mean preferences for speech. On average, SLOW was preferred over FAST, but only by 0.46 scale units. There were distinct individual differences. Eight subjects showed a preference for SLOW of 0.5 scale units or more, while four subjects showed a preference for FAST of 0.5 scale units or more.

Compression speed preferences for music

The preference scores were averaged across the three levels. The scores were reasonably consistent across music types except the solo percussion instrument, for which the scores were not significantly correlated with scores for the other music types. Hence, we consider only the mean scores across the three other music types. Fig. 2 shows individual and mean preferences for music. On average, SLOW was preferred over FAST, by 0.57 scale units. Seven subjects showed a preference for SLOW of 0.6 scale units or more, seven showed no clear preference (ratings within the range –0.014 to +0.18), and no subject showed a clear preference for FAST.

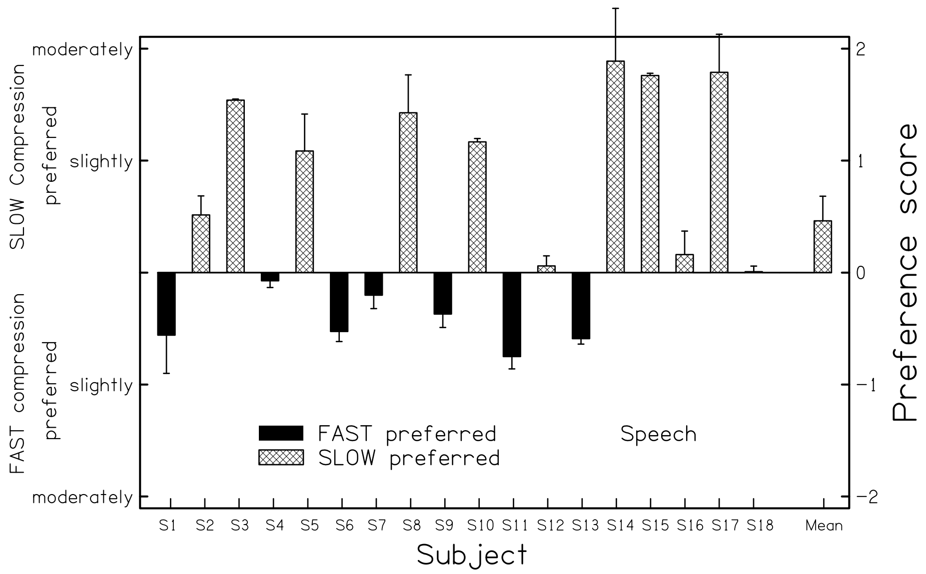


Fig. 1: Mean preference scores for speech for each subject. Error bars show ±1 SD. The bar at the right shows the mean.

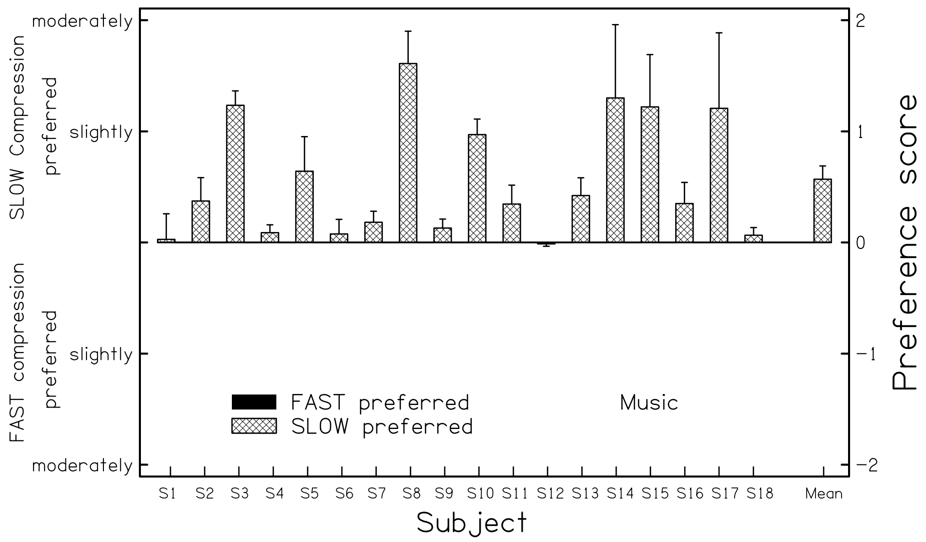


Fig. 2: As Fig. 1, but for music (percussion excluded).

Similarity of preferences for speech and music

Although the preference for SLOW relative to FAST was slightly greater for the music than for the speech stimuli, the pattern of preferences across subjects was highly correlated for the two stimulus types ($r = 0.89$, $p < 0.01$), as can be seen by comparing Figs. 1 and 2.

Relationship of preferences to sensitivity to TFS

Since we were testing the hypothesis that the relative preference for slow compression would increase with decreasing sensitivity to TFS, one-tailed tests were used to assess the significance of correlations. The DLFs for the test ears were weakly correlated with preference scores for music: $r = 0.4$, $p < 0.05$. Large DLFs, indicating poor sensitivity to TFS, were associated with greater preference for SLOW. However, the correlation of DLFs with preferences for speech failed to reach significance: $r = 0.31$, $p > 0.05$.

Six subjects were not able to complete the TFS-LF task, because the adaptive procedure called for a value of $\Delta\varphi$ greater than 180°. For the 12 subjects who were able to complete both the TFS-LF and the TFS-AF tasks, there was a strong negative correlation between the two ($r = -0.93$, $p < 0.01$), indicating good consistency across the two tests; good interaural phase sensitivity was associated with a low threshold in degrees on the TFS-LF test and a high threshold in hertz on the TFS-AF test. Scores on the TFS-AF task, which could be completed by all subjects, were not significantly correlated with compression-speed preferences for either speech or music (both $r = -0.1$, $p > 0.05$).

DISCUSSION AND CONCLUSIONS

Consistent with the research reviewed in the introduction, there were distinct individual differences in preferences for SLOW relative to FAST. On average, the relative preference for SLOW was slightly greater for music than for speech, but the pattern of preferences across subjects was similar for speech and music. The use of slow compression seems to be a “safe” option for music listening, since several subjects showed relatively clear preferences for SLOW, while none showed a clear preference for FAST. However, for speech four subjects showed a clear preference for FAST.

The preferences were not related to the measures of sensitivity to interaural phase at low frequencies. A possible reason is that some of the subjects had near-normal hearing at low frequencies, and for them little compression was applied at low frequencies. There was a weak correlation between the DLFs at 2 kHz and preferences for music but not preferences for speech. Thus, while sensitivity to TFS may have a weak influence on preferences for compression speed, other factors, such as cognitive ability (Gatehouse *et al.*, 2006a; 2006b; Lunner and Sundewall-Thoren, 2007), appear to have a more important influence.

REFERENCES

ANSI (**2003**). *ANSI S3.22-2003, Specification of hearing aid characteristics* (American National Standards Institute, New York).

Fowler, E.P. (**1936**). "A method for the early detection of otosclerosis," Arch. Otolaryngol., **24**, 731-741.

Füllgrabe, C., Harland, A.J., Sęk, A.P., and Moore, B.C.J. (**2015**). "Development of a new method for determining binaural sensitivity to temporal fine structure," Int. J. Audiol. (in preparation).

Gatehouse, S., Naylor, G., and Elberling, C. (**2006a**). "Linear and nonlinear hearing aid fittings – 1. Patterns of benefit," Int. J. Audiol., **45**, 130-152.

Gatehouse, S., Naylor, G., and Elberling, C. (**2006b**). "Linear and nonlinear hearing aid fittings – 2. Patterns of candidature," Int. J. Audiol., **45**, 153-171.

Hopkins, K. and Moore, B.C.J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am., **122**, 1055-1068.

Hopkins, K., Moore, B.C.J., and Stone, M.A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am., **123**, 1140-1153.

Hopkins, K. and Moore, B.C.J. (**2010a**). "Development of a fast method for measuring sensitivity to temporal fine structure information at low frequencies," Int. J. Audiol., **49**, 940-946.

Hopkins, K. and Moore, B.C.J. (**2010b**). "The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am., **127**, 1595-1608.

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.

Hopkins, K., King, A., and Moore, B.C.J. (**2012**). "The effect of compression speed on intelligibility: simulated hearing-aid processing with and without original temporal fine structure information," J. Acoust. Soc. Am., **132**, 1592-1601.

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

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.

Leek, M.R., Molis, M.R., Kubli, L.R., and Tufts, J.B. (**2008**). "Enjoyment of music by elderly hearing-impaired listeners," J. Am. Acad. Audiol., **19**, 519-526.

Lunner, T. and Sundewall-Thoren, E. (**2007**). "Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid," J. Am. Acad. Audiol., **18**, 604-617.

Madsen, S.M.K. and Moore, B.C.J. (**2014**). "Music and hearing aids," Trends Hear., **18**, 1-29.

Moore, B.C.J. and Glasberg, B.R. (**1987**). "Factors affecting thresholds for sinusoidal signals in narrow-band maskers with fluctuating envelopes," J. Acoust. Soc. Am., **82**, 69-79.

Moore, B.C.J., Peters, R.W., and Stone, M.A. (**1999**). "Benefits of linear amplification and multi-channel compression for speech comprehension in backgrounds with spectral and temporal dips," J. Acoust. Soc. Am., **105**, 400-411.

Moore, B.C.J. and Skrodzka, E. (**2002**). "Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation," J. Acoust. Soc. Am., **111**, 327-335.

Moore, B.C.J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues, 2nd Ed.* (Wiley, Chichester), pp. 332.

Moore, B.C.J. (**2008**). "The choice of compression speed in hearing aids: Theoretical and practical considerations, and the role of individual differences," Trends Amplif., **12**, 103-112.

Moore, B.C.J., Füllgrabe, C., and Stone, M.A. (**2010a**). "Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task," J. Acoust. Soc. Am., **128**, 360-371.

Moore, B.C.J., Glasberg, B.R., and Stone, M.A. (**2010b**). "Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF," Int. J. Audiol., **49**, 216-227.

Moore, B.C.J. and Ernst, S.M. (**2012**). "Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code," J. Acoust. Soc. Am., **132**, 1542-1547.

Moore, B.C.J., and Sęk, A. (**2013**). "Comparison of the CAM2 and NAL-NL2 hearing-aid fitting methods," Ear Hearing, **34**, 83-95.

Moore, B.C.J. (**2014**). *Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss* (World Scientific, Singapore), pp. 1-182.

Sęk, A. and Moore, B.C.J. (**2012**). "Implementation of two tests for measuring sensitivity to temporal fine structure," Int. J. Audiol., **51**, 58-63.

Stone, M.A. and Moore, B.C.J. (**1992**). "Syllabic compression: Effective compression ratios for signals modulated at different rates," Br. J. Audiol., **26**, 351-361.

Stone, M.A. and Moore, B.C.J. (**2004**). "Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task," J. Acoust. Soc. Am., **116**, 2311-2323.

Stone, M.A., Moore, B.C.J., Füllgrabe, C., and Hinton, A.C. (**2009**). "Multi-channel fast-acting dynamic-range compression hinders performance by young, normal-hearing listeners in a two-talker separation task," J. Audio Eng. Soc., **57**, 532-546.

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.

Verschuure, J., Maas, A.J.J., Stikvoort, E., de Jong, R.M., Goedegebure, A., and Dreschler, W.A. (**1996**). "Compression and its effect on the speech signal," Ear Hearing, **17**, 162-175.

Individual factors in speech recognition with binaural multi-microphone noise reduction: Measurement and prediction

TOBIAS NEHER[1,3,*], JACOB ADERHOLD[1,3], DANIEL MARQUARDT[2,3], AND THOMAS BRAND[1,3]

[1] *Medizinische Physik, Oldenburg University, Oldenburg, Germany*

[2] *Signal Processing Group, Oldenburg University, Oldenburg, Germany*

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

Multi-microphone noise reduction algorithms give typically rise to large signal-to-noise ratio improvements, but they can also severely distort binaural information and thus compromise spatial hearing abilities. To address this problem Klasen *et al*. (2007) proposed an extension of the binaural multi-channel Wiener filter (MWF), which suppresses only part of the noise and, in this way, preserves some binaural information (MWF-N). The current study had three aims: (1) to assess aided speech recognition with MWF(-N) for a group of elderly hearing-impaired listeners, (2) to explore the impact of individual factors on their performance, and (3) to test if outcome can be predicted using a binaural speech intelligibility model. Sixteen hearing aid users took part in the study. Speech recognition was assessed using headphone simulations of a spatially complex speech-in-noise scenario. Individual factors were assessed using audiometric, psychoacoustic (binaural), and cognitive measures. Analyses showed clear benefits from both MWF and MWF-N and also suggested sensory and binaural influences on speech recognition. Model predictions were reasonably accurate for MWF but not MWF-N, suggesting a need for some model refinement concerning supra-threshold processing.

INTRODUCTION

Recently, hearing aids have become available that can wirelessly exchange audio signals across the user's head. This has opened up possibilities for 'binaural' signal processing, such as multi-microphone noise reduction, which can lead to large signal-to-noise ratio (SNR) improvements but also to distortions of binaural information (e.g., Doclo *et al*., 2010). Because binaural information plays an important role for speech understanding in complex listening situations (e.g., Bronkhorst, 2015) and because hearing-aid users can differ substantially in terms of their residual binaural hearing abilities (e.g., Neher *et al*., 2011; 2012), it is of interest to relate individual factors to benefit, or lack thereof, from this type of processing.

---

*Corresponding author: tobias.neher@uni-oldenburg.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

The purpose of the current study was to address this issue for one type of multi-microphone noise reduction: binaural multi-channel Wiener filtering (MWF). MWF perfectly preserves the binaural cues of the target signal, but undesirably changes the binaural cues of the noise to those of the target (e.g., Doclo *et al.*, 2006). To address this problem, Klasen *et al.* (2007) proposed an extension of MWF, which suppresses only part of the noise and, in this way, retains some binaural information (MWF-N). For a group of young normal-hearing participants, van den Bogaert *et al.* (2008) found that MWF-N improved localisation while speech recognition was unaffected.

In the current study, we aimed to extend this research by pursuing the following three aims:

1. To assess aided speech recognition with MWF(-N) for elderly hearing aid users
2. To explore the influence of individual factors on their performance
3. To investigate if outcome can be predicted using a state-of-the-art binaural speech intelligibility model

METHODS

Speech stimuli

Our speech stimuli were based on recordings from the Oldenburg sentence test (Wagener *et al.*, 1999). To simulate a realistic complex listening situation, we convolved these recordings with pairs of head-related impulse responses, which were measured in a reverberant cafeteria using a head-and-torso simulator equipped with two behind-the-ear hearing aid shells (Kayser *et al.*, 2009). Specifically, we used the measurements made with the front and rear microphones of each hearing aid shell and a frontal source at a distance of 1 m from, and at the same height as, the head-and-torso simulator. For the interfering signal, we used a (spatially complex) recording made in the same cafeteria with the same setup during a busy lunch hour. During the measurements, we presented this signal at a nominal sound pressure level of 65 dB and mixed it with the target sentences, the level of which we adjusted to produce a given SNR.

MWF(-N) processing

The MWF(-N) processing we tested mimicked that of van den Bogaert *et al.* (2008). There were two main algorithmic parameters: μ and η. μ determines the strength of spectral post-filtering and thus trades off noise reduction against speech distortion. It was set to 1 here to result in standard MWF. η is a scaling factor between 0 and 1 that determines how much of the unprocessed input signal is mixed back into the noise-reduced output signal. For $\eta = 0$, nothing of the input is mixed back into the output, resulting in standard MWF with full noise suppression but no binaural cue preservation. For $\eta = 1$, the input is mixed completely into the output, resulting in full binaural cue preservation but no noise suppression. In the current study, we tested the three η-settings also tested by van den Bogaert *et al.* (2008): 0, 0.2, and 1. In the following, we will refer to these as the MWF, MWF-N, and reference conditions. Furthermore, as in the study of van den Bogaert *et al.* we used a perfect voice activity detector (i.e., we assumed access to the clean speech signal).

To quantify the physical effects of MWF(-N) we estimated the resultant speech-weighted SNR improvement (ΔAI-SNR) as a function of the input SNR. As expected, ΔAI-SNR increased with higher input SNRs (see Fig. 1). Furthermore, ΔAI-SNR was up to 0.8 dB larger for MWF than for MWF-N. Figure 1 also shows the SNRs during the speech recognition measurements (see below). Across participants, ΔAI-SNR amounted to 2.2 dB (σ = 0.5 dB) for MWF-N and to 2.7 dB (σ = 0.7 dB) for MWF.

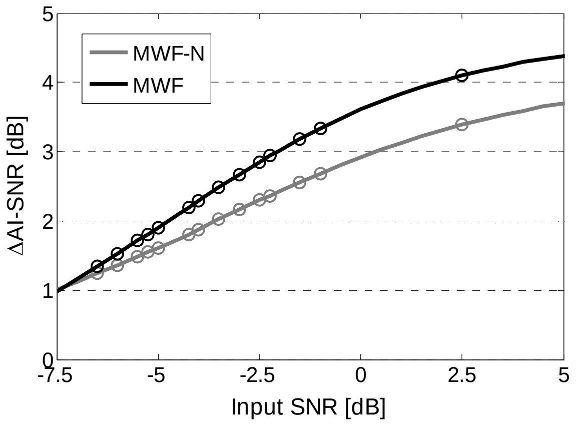


Fig. 1: ΔAI-SNR for MWF (black) and MWF-N (grey) as a function of input SNR. Circles denote individual test SNRs.

In addition, we estimated the interaural coherence (IAC) of our speech stimuli for the three processing conditions with the help of the auditory model of Dietz *et al.* (2011). The IAC can be interpreted as a measure of binaural complexity. As expected, binaural complexity decreased with MWF-N and especially MWF, i.e., the stimuli became increasingly interaurally correlated (see Fig. 2).

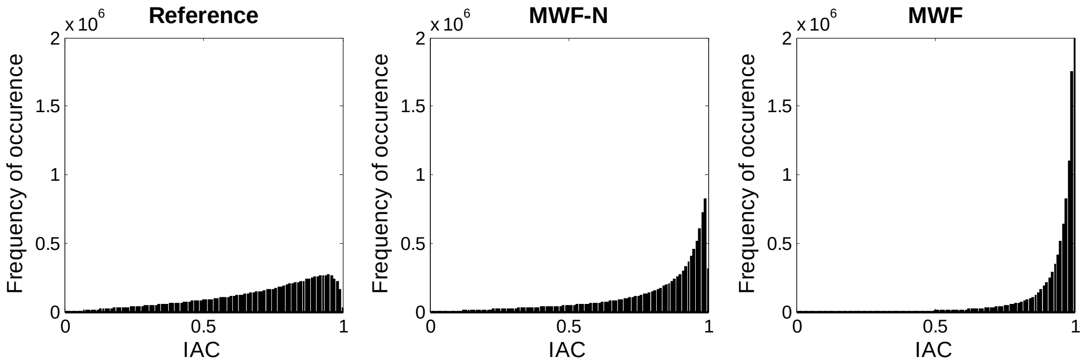


Fig. 2: Histograms of the estimated IAC for an example speech stimulus with an input SNR of –4 dB for the reference, MWF-N, and MWF conditions.

Participants and individual factors

Sixteen experienced hearing-aid users with symmetrical, gently sloping sensori-neural hearing impairments participated in the experiment. Their mean age was 74 yr (range: 56-86 yr). Their mean pure-tone average hearing loss from 500 Hz to 4 kHz (PTA) was 46 dB HL (range: 38-53 dB HL), while from 125 Hz to 750 Hz (PTA_{LF}) it was 30 dB HL (range: 17-41 dB HL).

To characterise our participants' binaural hearing abilities we performed binaural masking level difference (BMLD) measurements (test and retest) at 500 Hz with a broadband noise masker. In addition, we performed interaural phase difference frequency range (IPD_{FR}) measurements (test and retest). These measurements mark the highest frequency for which a participant is still able to detect interaural phase changes of 180° in a sinusoidal stimulus (e.g., Neher *et al*., 2011). Furthermore, we administered the reading span test (RST; Carroll *et al*., 2015) to our participants to also determine their working memory capacity.

Speech recognition measurements

Because we were interested in *aided* speech recognition performance we spectrally shaped the speech stimuli in accordance with the NAL-RP prescription rule (Byrne *et al*., 1991). We started our measurements with three training runs and then determined the individual speech reception threshold (SRT_{ind}) for the reference condition. In all subsequent measurements, we then kept the SNR fixed at the SRT_{ind}. In this manner, we obtained speech recognition rates (in percent correct) for our three processing conditions.

Binaural speech intelligibility model

For the prediction of the participants' speech scores we used the binaural speech intelligibility model (BSIM) of Beutelmann *et al*. (2010). BSIM combines a multi-channel equalization cancellation stage according to Durlach (1963) with the Speech Intelligibility Index (SII; ANSI, 1997). In the current study, we individualised BSIM based on the hearing thresholds of each participant and carried out the predictions based on the amplified speech stimuli. Furthermore, because we measured speech recognition rates at a fixed SNR for each processing condition (rather than one SRT per processing condition) we restricted the predictions to the computation of SII (rather than SRT) values and related these to the speech scores of our participants.

RESULTS

Individual factors

Analysis of the test-retest data showed that the BMLD and IPD_{FR} measurements were reliable (both $r > 0.7$, $p < 0.01$). Figure 3 provides an overview of the BMLD, IPD_{FR}, and RST data. Averaged across participants, the BMLD was 11.2 dB (range: 4-20 dB), while the IPD_{FR} was 770 Hz (range: 342-1196 Hz). In terms of RST performance, the participants were on average able to recall 37.4% of all target words (range: 28-52%). Altogether, the BMLD and IPD_{FR} data were in good agreement with the literature, while the RST data exhibited less spread toward the 'poor' end (cf. Neher *et al*., 2011; Santurette and Dau, 2012).

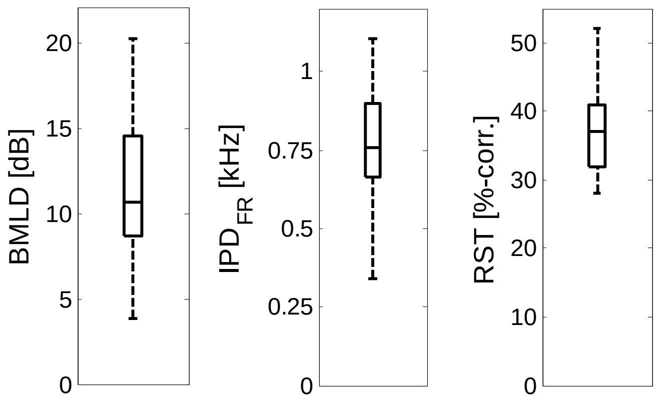


Fig. 3: Boxplots of the BMLD, IPD_{FR}, and RST data.

Speech recognition

Analysis of the SRT_{ind} data revealed a mean threshold of −3.7 dB SNR and a range of almost 10 dB (see Fig. 1). Figure 4 shows the speech scores for the three processing conditions. In the reference condition, participants could recognise 52.7% of the target speech. In the MWF-N and MWF conditions, they were able to recognise 80.5% and 78.4%, respectively. An analysis of variance with post hoc comparisons confirmed highly significant differences between the reference condition and MWF(-N), while MWF-N and MWF did not differ from each other.

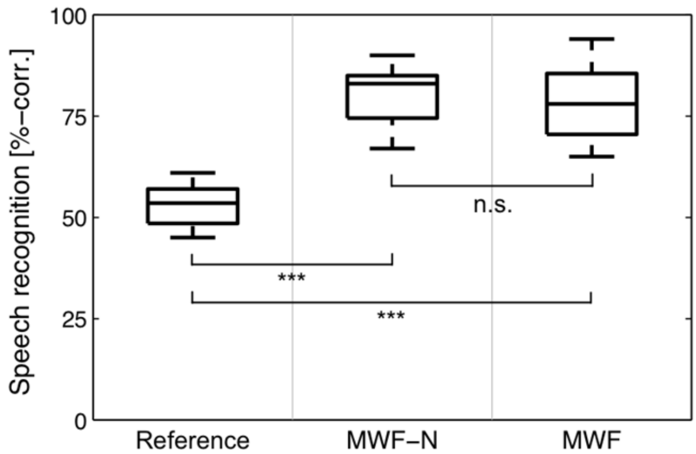


Fig. 4: Boxplots of the speech scores for the three processing conditions. *** $p < 0.001$, n.s. = non-significant.

Relations among speech outcomes and individual factors

To assess potential relations between SRT_{ind} and the individual factors we calculated a series of Pearson's *r* correlation coefficients. We observed correlations with age, PTA_{LF}, and BMLD (see Table 1). A regression model based on these three factors could account for 62.1% (adjusted R^2 = 53%) of the variance in the SRT_{ind} data ($p_{model} < 0.01$, $p_{age} > 0.05$, $p_{BMLD} < 0.05$, $p_{PTA\_LF} < 0.05$).

To assess potential relations between speech recognition (SR) with MWF(-N) and the individual factors, we calculated a series of partial correlation coefficients with ΔAI-SNR as control variable. In this manner, we controlled for the SNR-dependent effects of MWF(-N) related to speech audibility (see Fig. 1). As can be seen in Table 1, there was only a correlation between $SR_{MWF\text{-}N}$ and RST.

	Age	**PTA_{LF}**	**BMLD**	**IPD_{FR}**	**RST**
SRT_{ind}	0.53*	0.64**	–0.61*	–0.38	–0.30
$SR_{MWF\text{-}N}$	–0.22	0.51	0.35	–0.23	0.62*
SR_{MWF}	0.53	0.51	–0.15	–0.18	–0.04

Table 1: Correlation coefficients for the speech scores from the reference (SRT_{ind}), MWF-N ($SR_{MWF\text{-}N}$), and MWF (SR_{MWF}) condition and the individual factors, with ΔAI-SNR partialled out in the case of $SR_{MWF(\text{-}N)}$. * $p < 0.05$, ** $p < 0.01$.

Outcome prediction

Figure 5 summarises the results of the outcome prediction. In each case, the abscissa shows ΔSII values [MWF(-N) − reference condition], while the ordinate shows corresponding Δ speech scores.

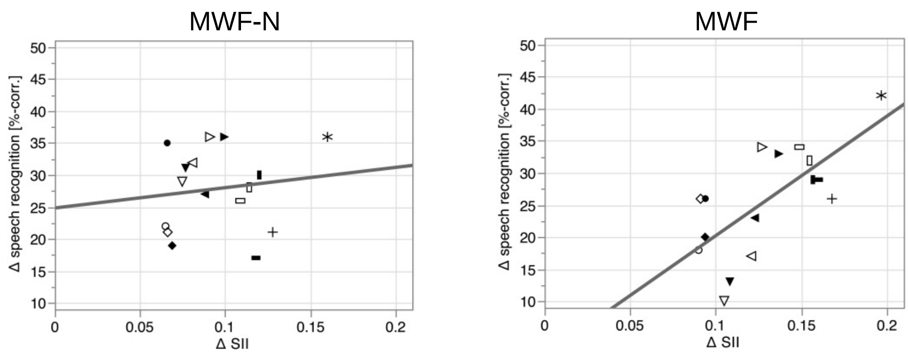


Fig. 5: Scatter plots of ΔSII values against Δ speech scores for MWF-N (left) and MWF (right). Symbols denote individual participants.

As can be seen, the accuracy was reasonably high for MWF ($r = 0.70$, $p < 0.01$) but not for MWF-N ($r = 0.14$). Partialling out ΔAI-SNR removed the correlation between the predicted and measured (relative) outcome for MWF ($r = 0.11$). Performing these predictions on short time segments of the speech stimuli and averaging across results (the "short-time BSIM"; cf. Beutelmann *et al*., 2010) did not improve the accuracy.

Altogether, these results suggest that, while BSIM is largely able to account for performance with MWF where the main effect is improved speech audibility, it fails to do so for MWF-N which due to its greater binaural complexity (see above) presumably invokes additional supra-threshold factors.

SUMMARY

With respect to the three aims outlined above, the results of the current study can be summarised as follows:

1. MWF(-N) led to significant improvements (on the order of 25%) in speech recognition performance. The benefit from MWF-N was comparable to that from MWF, despite the addition of background noise.
2. PTA_{LF} and BMLD were related to aided speech recognition in the reference condition, independent of the effects of age. For speech recognition with MWF-N, a relation with RST was found. For MWF, none of the individual factors tested here was predictive.
3. Outcome predictions were accurate for MWF, suggesting that BSIM could account for the main effect of improved speech audibility. In the case of MWF-N, outcome prediction was poor, suggesting that BSIM failed to account for certain supra-threshold effects.

Given that our study was limited to 16 participants who were tested at markedly different SNRs and that we used a perfect voice activity detector, the above findings must be regarded as preliminary. Future studies will investigate these issues in more detail, with particular emphasis on the role that individual factors play for aided outcome prediction.

ACKNOWLEDGEMENTS

This research was funded by the DFG Cluster of Excellence EXC 1077/1 "Hearing4all". We thank Christopher Hauth, Birger Kollmeier, and Simon Doclo for their support.

REFERENCES

ANSI (**1997**). "Methods for calculation of the speech intelligibility index (S3.5-1997)," New York, NY: American National Standards Institute.

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.

Bronkhorst, A.W. (**2015**). "The cocktail-party problem revisited: Early processing and selection of multi-talker speech," Atten. Percept. Psycho., **77**, 1465-1487.

Byrne, D., Parkinson, A., and Newall, P. (**1991**). "Modified hearing aid selection procedures for severe/profound hearing losses," in *The Vanderbilt Hearing Aid Report II*. Eds. G.A. Studebaker, F.H. Bess, and L.B. Beck (York Press, Parkton, NC), pp. 295-300.

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.

Dietz, M., Ewert, S.D., and Hohmann, V. (**2011**). "Auditory model based direction estimation of concurrent speakers from binaural signals," Speech Comm., **53**, 592-605.

Doclo, S., Klasen, T.J., Wouters, J., Haykin, S., and Moonen, M. (**2006**). "Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions," Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC), Paris, France, Sept. 12-14.

Doclo, S., Gannot, S., Moonen, M., and Spriet, A. (**2010**). "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*. Eds. S. Haykin and K.J.R. Liu (Wiley-IEEE Press, Hoboken, NJ), pp. 269-302.

Durlach, N.I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am., **35**, 1206-1218.

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., **298605**, DOI: 10.1155/2009/298605.

Klasen, T.J., van den Bogaert, T., Moonen, M., and Wouters, J. (**2007**). "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," IEEE Trans. Signal Process., **55**, 1579-1585.

Neher, T., Laugesen, S., 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.

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," J. Acoust. Soc. Am., **131**, 2561-2564.

Santurette, S. and Dau, T. (**2012**). "Relating binaural pitch perception to the individual listener's auditory profile," J. Acoust. Soc. Am., **131**, 2968-2986.

van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. (**2008**). "The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids," J. Acoust. Soc. Am., **124**, 484-497.

Wagener, K., Brand, T., and Kollmeier, B. (**1999**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache. I-III: Design, Optimierung und Evaluation des Oldenburger Satztests", Zeitschrift für Audiologie, **38**, 4-15, 44-56, 86-95.

Can individualised acoustical transforms in hearing aids improve perceived sound quality?

SØREN LAUGESEN*, NIELS SØGAARD JENSEN, FILIP MARCHMAN RØNNE, AND JULIE HEFTING PEDERSEN

Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark

This paper presents an experiment which aimed to clarify whether benefits in terms of perceived sound quality can be obtained from fitting hearing aids according to individualised acoustical transforms instead of average transforms. Eighteen normal-hearing test subjects participated, and hearing-aid sound processing with various degrees of individualisation was simulated and applied to five different sound samples. Stimuli were presented over insert phones and evaluated in an A/B test paradigm. Data were analysed with the Bradley-Terry-Luce model. The key result is that hearing aids individualised according to a real-ear insertion gain (REIG) target were preferred over hearing aids individualised according to a real-ear aided response (REAR) target.

INTRODUCTION

When listening with open ears, the sounds that arrive at the eardrum are coloured by the presence of the body, the head, and the detailed structure of the pinna and the ear canal. This colouration is unique to the individual ear. When a hearing aid (HA) is fitted to a person's ear, this colouration is changed, and these changes are taken into account in the hearing aid's amplification in terms of so-called acoustical transforms (ATs). The ATs are typically described in terms of three components: the microphone location effect (MLE), the open ear gain (OEG), and the real ear to coupler difference (RECD). In spite of the aforementioned individual variation, most hearing aids are fitted using average acoustical transforms. By using standardised measures, the individual variation in all three components of the ATs is disregarded. This variation can be quite large, especially at high frequencies. For example, Saunders and Morgan (2003) showed that for the RECD alone, deviations of more than 10 dB are very common at high frequencies. The combined effect of the variation in all three components of the ATs may thus be a substantial difference between the prescribed gain of a HA and the fitting's target.

The individual ATs may be taken into account in the HA fitting by means of real-ear measurements (REMs), but the use of REMs is not widespread (Dillon and Keidser, 2003; Mueller and Picou, 2010). It should be noted that there are (at least) two schools of thought regarding individualisation of HA fittings using REMs. For example, the NAL family of prescriptions (Dillon, 2012) is defined with a real-ear

*Corresponding author: slau@eriksholm.com

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

insertion gain (REIG) target. Thus, an REM-based NAL fitting set to a nominal 0-dB insertion gain will seek to recreate the exact same sound pressure level (SPL) in front of the eardrum as would be found in free-field listening with a sound source directly in front of the listener. In contrast, e.g., the DSL prescription (Seewald *et al.*, 2005) is defined with a real-ear aided response (REAR) target. This means that the individual OEG component of the total individual AT will be deliberately ignored and replaced by the standard OEG assumed by DSL. The REAR-target strategy is especially relevant when fitting small children or people with surgically modified ears, that is, when it can be argued that an extreme OEG exists for other reasons than audition (Dillon, 2012).

There are several studies showing that the use of REMs improves on a HA fitting's match to target, (e.g., Aazh *et al.*, 2012; Nelson, 2013). However, as highlighted by Humes (2012) and Mueller (2014), there has been very little research into whether or not using REM leads to any self-perceived benefits for the HA user. In fact, the present authors have identified only one public article (Abrams *et al.*, 2012) that demonstrates a self-perceived end-user benefit of using individually measured ATs in HA fitting. In that study, the APHAB questionnaire and overall preference were used to evaluate two HA fittings tested in the field: one based on standardised ATs and one based on individualised ATs (with REAR targets).

A couple of studies have examined the sound-quality disruptions perceived by listeners due to generic modifications to the frequency response of a reproduction system (van Buuren *et al.*, 1996; Moore and Tan, 2003). Their results indicate that the expected magnitude of differences between individualised and standardised ATs should be perceptible by both normal-hearing and hearing-impaired listeners.

Overall approach

The goal of the experiment reported in this paper was to investigate whether or not it would be possible to measure a sound-quality benefit from using individualised ATs, in the sense that listening through a HA programmed according to individualised ATs would be consistently preferred over listening through a HA programmed according to standardised ATs.

Being a first step, the most advantageous conditions for finding such a benefit were sought for. This involved using test subjects not requiring amplification (implying that the question of selecting gain rule for hearing-loss compensation could be neglected), using laboratory-grade equipment to measure the total individual AT, and ignoring the direction-dependence of the MLE-component of the AT by considering only sound presentation corresponding to the frontal direction. Finally, the experimental stimuli were delivered from Matlab to the test subjects' ears through insert phones.

METHOD AND MATERIAL

Test subjects

N = 18 test subjects (12 female, 6 male; age-range 22-55 years, mean age 37) were recruited. All test subjects had hearing threshold levels (HTLs) at 25 dB HL or better at all audiometric frequencies up to and including 8 kHz, except one test subject who had HTLs of 30 and 35 dB HL at 6 kHz and two test subjects who had 8-kHz HTLs of 35 dB in one ear.

Sound quality experiment

The procedure used for the sound quality assessment was based on an A/B paired-comparison approach (Bramsløw, 2010), where all pairs of the five processing conditions were compared. For each comparison (trial), the test subject had to listen to a stimulus of about 15 seconds duration, which was played back in a continuous loop. The test subject could switch between settings A and B as much and as frequently as he or she desired, using a touch screen. The test subject's task was to determine the preferred setting. When the preferred setting had been indicated, the test subject could start the next trial by pressing the 'Next' button. The user interface also included a 'Pause' button, allowing the test subject to take a break (at any time). This option was chosen by some (but not all) subjects during the main 150-trial test. There was no 'Don't know' option. Thus, the test subjects were instructed to make an arbitrary choice in cases where they had no preference.

Real-ear measurements

The laboratory-grade REM set-up was built around the Brüel&Kjær PULSE audio analyzer system, which was set to carry out two-channel FFT spectrum averaging (6400 spectral lines, 20-kHz bandwidth). The measurements were performed in an anechoic room with the test subject seated in an adjustable chair and sound delivered from a Genelec 8030A loudspeaker. Sound was recorded in the test subjects' ears through probe microphones. The probe microphones were taken from a modified Interacoustics Affinity system, which also served as power supply and conditioning amplifier for the probe microphones. From the Affinity system the microphone signals were routed through a Brüel&Kjær 5935 Dual Microphone Supply to the PULSE system. The measurements comprised loudspeaker free-field response, individual probe-microphone free-field calibration, and individual measurements of open-ear responses as well as aided responses, as described below.

In addition, a standard Interacoustics Affinity system was used for REMs in the clinic, which were used to obtain the $HA0_{REAR1}$ setting, see below.

Processing conditions

Five different processing conditions were created, representing different degrees of AT individualisation in a hypothetical HA prescribed to deliver linear amplification with 0-dB insertion gain at all frequencies. The conditions are described in Table 1.

$HA0_{REIG}$	Mimicking a 0-dB insertion gain HA fitted with individual ATs according to a REIG target.
$HA0_{avg}$	This condition was meant to mimic a 0-dB insertion gain HA fitted according to average ATs. However, due to a programming mistake the results from this condition are disregarded.
$HA0_{REAR1}$	Mimicking a 0-dB insertion gain HA fitted with individual ATs according to a REAR target. Based on the automatic AutoFit function in the Genie HA-fitting software together with the Affinity REM system.
$HA0_{REAR2}$	Mimicking a 0-dB insertion gain HA fitted with individual ATs according to a REAR target. Derived directly from the standard OEG used in the Genie fitting software.
$HA0_{REIGlowres}$	Similar to $HA0_{REIG}$, except that the individual ATs were realised with a frequency detail similar to what is available in the Genie fitting software.

Table 1: Labels and description of processing conditions.

Stimuli

The chosen sound samples were recordings of ‘Classical’, ‘Rock’, and ‘Jazz’ music, ‘Speech’ in quiet, and a dialogue in a ‘Canteen’ background. The samples were cut to allow seamless looping and they were scaled to produce reasonable playback levels ranging from 70 to 78 dB SPL (predicted free-field levels). To produce the individual stimuli, the sound samples were convolved with the processing-condition filters described above. In addition, the stimuli were shaped to compensate for the individually measured response of the Etymotic Research ER-2 insert phones used in the experiment. Then, the magnitude-smoothing approach suggested by Schärer and Lindau (2009) was applied using a ¼-octave band filter. Finally, the five condition-specific stimuli for each test person and each sound sample were scaled to have the same predicted A-weighted free-field level, in order to remove any loudness differences which are known to dominate sound-quality evaluations if present.

Test design and protocol

A test design based on counterbalancing of condition pairs and sound samples and with three repetitions was used, which amounts to a total of 10×5×3 = 150 trials. Prior to the actual test trials, 20 practice trials were performed. In each trial, the assignment of the two conditions to the A and B buttons was random.

The experiment comprised one visit for each test subject, starting with the clinical HA fitting session, then the PULSE-based REM session, and finally the sound-quality session.

RESULTS

The A/B preference data were analysed with the Bradley-Terry-Luce (BTL) model (Bradley and Terry, 1952; Luce, 1959), basically following the approach described by Wickelmaier and Schmid (2004). In addition, the individual processing-condition filters were analysed in various ways for the purpose of the correlational analysis.

A/B test

The main outcome of the A/B testing is presented in Fig. 1, which shows the normalised BTL scores for each processing condition together with estimated 95% confidence intervals, based on the data from all test subjects and all sound samples. A high BTL score indicates that the condition is more likely to be preferred in a comparison against another randomly selected condition, and non-overlapping confidence intervals are taken as an indication of a statistically significant difference.

The main result is that $HA0_{REIG}$ was preferred over both $HA0_{REAR1}$ and $HA0_{REAR2}$, which suggests that in consideration of sound quality a REIG-target approach should be preferred over a REAR-target approach. In addition, it is seen that the best preference rating was given to the low-resolution $HA0_{REIGlowres}$. This result was unexpected and is further investigated below.

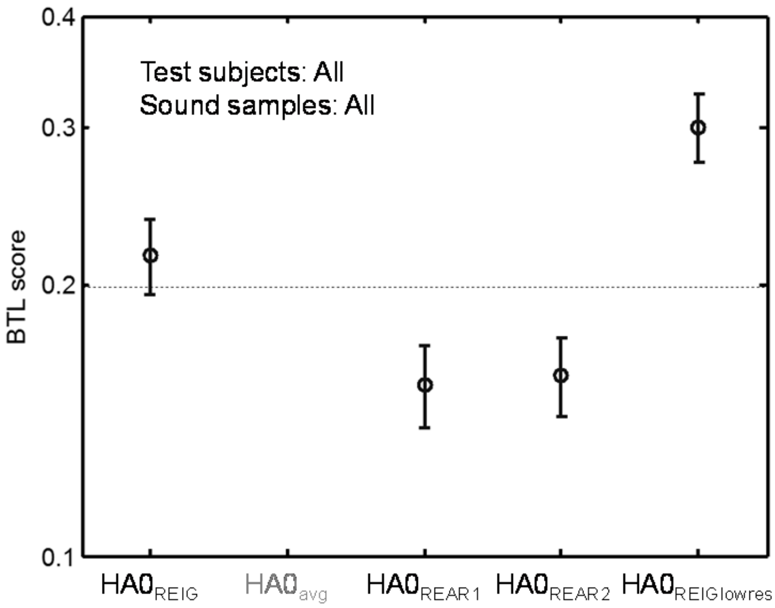


Fig. 1: Overall BTL scores for the five processing conditions with 95% confidence intervals indicated. The dashed line indicates chance level.

A more detailed analysis indicates that the results in Fig. 1 are more pronounced if data obtained only with the music samples are considered, as shown in Fig. 2(a). In contrast, the pattern of results is somewhat different for the speech samples, as indicated in Fig. 2(b). This observation agrees well with the comments from several test subjects, who stated that they applied different criteria for the music and speech samples, mainly because speech intelligibility – rather than sound quality – became the prevailing criterion for the samples including speech.

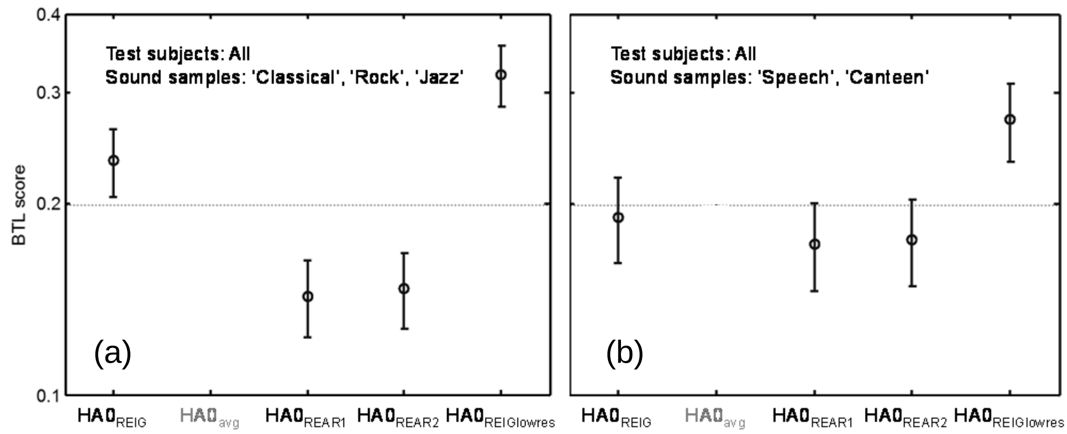


Fig. 2: BTL scores and confidence intervals for (a) the music samples alone, and (b) the samples containing speech.

Furthermore, it should be noted that the individual patterns of preference for some test subjects deviate considerably from the all-test-subjects patterns shown in Figs. 1 and 2. This is not surprising, due to the individual nature of the ATs, which means that the contrasts among the different processing conditions were not the same for all test subjects.

Correlational analysis

The most surprising result from the above BTL analysis was that $HA0_{REIGlowres}$ was preferred over $HA0_{REIG}$. Therefore, it was investigated whether this preference could be related to specific characteristics of the respective processing-condition filter responses. Figure 3(a) shows the mean 'colouration responses' for the two relevant conditions. The colouration responses are similar to the processing-condition filter responses except that the individual insert-phone compensation was removed. The results in Fig. 3(a) show slight systematic differences between the $HA0_{REIG}$ and the $HA0_{REIGlowres}$ conditions (unintended artefacts of the lowres procedure), e.g., a 2-dB boost above 6 kHz and an attenuation in other frequency ranges. The colouration-response differences were averaged across the 6-8 kHz frequency band and across the two ears of each test subject to describe the amount of high-frequency boost. Similar quantities were computed for the attenuation bands. These measures were then used as predictors of the percentage of comparisons where $HA0_{REIG}$ was preferred over $HA0_{REIGlowres}$. None of these predictors turned out to be significant on a 5% level. Nevertheless, the most striking relation (involving the 6-8 kHz high-frequency boost) is illustrated in Fig. 3(b). Although the correlation ($r = -0.40$) is not significant ($p = 0.10$), a larger high-frequency boost ($HA0_{REIGlowres}$ over $HA0_{REIG}$) seems to be associated with stronger preference for $HA0_{REIGlowres}$ (low % values in Fig. 3(b)).

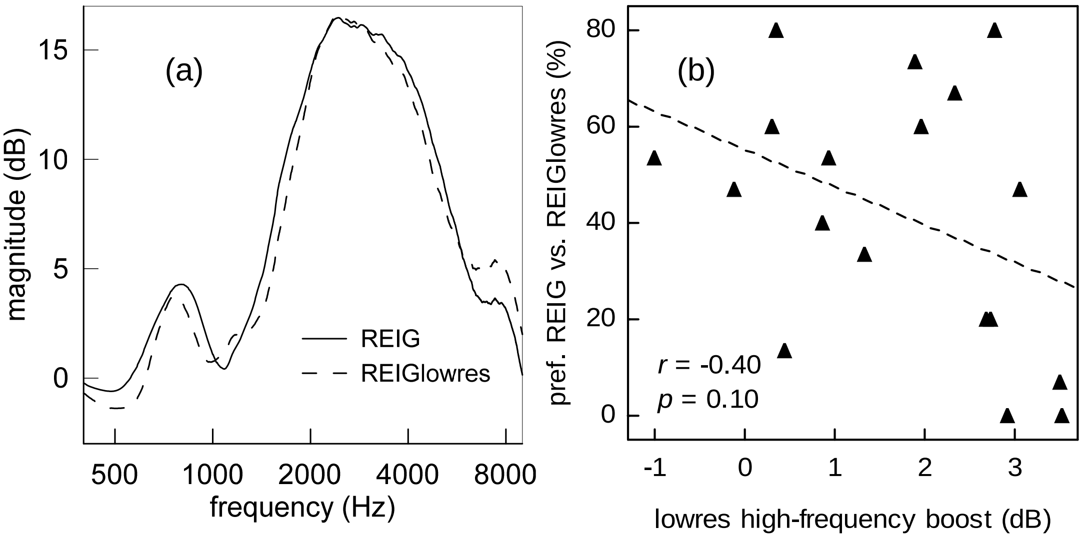


Fig. 3: (a) Mean colouration responses, as indicated. (b) Relation between the lowres high-frequency boost predictor variable and the preference for $HA0_{REIG}$ over $HA0_{REIGlowres}$. The preference data from the 18 test subjects were averaged across the five sound samples and three repetitions.

DISCUSSION

This study investigated the potential sound-quality benefit from fitting hearing aids according to individualised acoustical transforms instead of average transforms.

The key result was that (simulated) hearing aids individualised according to a real-ear insertion gain (REIG) target were preferred over hearing aids individualised according to a real-ear aided response (REAR) target. An earlier study (Abrams *et al.*, 2012) found benefits from individualising to a REAR target relative to a non-individualised approach. However, the main outcome measure used in the Abrams study (the APHAB questionnaire) assesses different benefit domains than sound quality – The APHAB's predefined sub-scales are: Ease of Communication, Reverberation, Background Noise, and Aversiveness of Sounds.

In addition, representing the individualised transforms in lower frequency resolution was preferred over the representation in fine spectral detail. The analysis suggests that this may be because of an artefact of the low-resolution representation which added a slight boost in the 6-8 kHz frequency range. Recall that the test subjects in this study had normal hearing, and therefore might appreciate the brighter timbre brought about by the lowres high-frequency boost. A similar outcome is not expected for test subjects with high-frequency hearing loss, who are less likely to appreciate additional amplification in the frequency range above 6 kHz (Dillon, 2012). The present study was limited to normal-hearing test subjects and simulated hearing-aid processing. Hence, the next step is to perform a similar investigation with hearing-impaired test subjects listening through real hearing aids.

REFERENCES

Aazh, H., Moore, B.C., and Prasher, D. (**2012**). "Real ear measurement methods for open fit hearing aids: Modified pressure concurrent equalization (MPCE) versus modified pressure stored equalization (MPSE)," Int. J. Audiol., **51**, 103-107.

Abrams, H.B., Chisolm, T.H., McManus, M., and McArdle, R. (**2012**). "Initial-fit approach versus verified prescription: Comparing self-perceived hearing aid benefit," J. Acoust. Soc. Am., **23**, 768-778.

Bradley, R.A. and Terry, M.E. (**1952**). "Rank analysis of incomplete block designs: I. The method of paired comparisons," Biometrika, **39**, 324-345.

Bramsløw, L. (**2010**). "Preferred signal path delay and high-pass cut-off in open fittings," Int. J. Audiol., **49**, 634-644.

Dillon, H. (**2012**). *Hearing Aids*. 2nd ed., Stuttgart, Germany: Thieme.

Dillon, H. and Keidser G. (**2003**). "Is probe-mic measurement of HA gain-frequency response best practice?," Hear. J., **56**, 28-30.

Humes, L.E. (**2012**). "Verification and validation: The chasm between protocol and practice," Hear. J., **65**, 8-10.

Luce, R.D. (**1959**). *Individual Choice Behavior: A Theoretical Analysis.* New York: Wiley.

Moore, B.C.J. and Tan, C.-T. (**2003**). "Perceived naturalness of spectrally distorted speech and music," J. Acoust. Soc. Am., **114**, 408-419.

Mueller, H.G. (**2014**). "20Q: Real-ear probe-microphone measures – 30 years of progress?" Audiology Online, article 12410.

Mueller, H.G. and Picou, E.M. (**2010**). "Survey examines popularity of real-ear probe-microphone measures," Hear. J., **63**, 27-28.

Nelson, S.R. (**2013**). "The impact of using real ear measures to calculate prescriptive targets on hearing aid follow-up visits," Retrieved from:
http://digitalcommons.wustl.edu/pacs_capstones/669/

Saunders, G.H. and Morgan, D.E. (**2003**). "Impact on hearing aid targets of measuring thresholds in dB HL versus dB SPL," Int. J. Audiol., **42**, 319-326.

Schärer, Z. and Lindau, A. (**2009**). "Evaluation of equalization methods for binaural signals," Proc. 126th Audio Engineering Society Convention, **1**, 15-31.

Seewald, R., Moodie, S., Scollie, S., and Bagatto, M. (**2005**). "The DSL method for pediatric hearing instrument fitting: Historical perspective and current issues," Trends Ampl., **9**, 145-157.

van Buuren, R.A., Festen, J.M., and Houtgast, T. (**1996**). "Peaks in the frequency response of hearing aids: Evaluation of the effects on speech intelligibility and sound quality," J. Speech Lang. Hear. Res., **39**, 239-250.

Wickelmaier, F. and Schmid, C. (**2004**). "A Matlab function to estimate choice model parameters from paired-comparison data," Behav. Res. Meth. Instr. Comp., **36**, 29-40.

A profiling system for the assessment of individual needs for rehabilitation with hearing aids

WOUTER DRESCHLER* AND INGE BRONS

Academic Medical Center, Department of Clinical & Experimental Audiology, Amsterdam, The Netherlands

Despite the huge number of hearing aids and the different options in terms of functionality, there is lack of a systematic approach how to select specific hearing aid models, or at least functionalities that may contribute to an optimal compensation of the hearing loss. If we can design such a systematic approach, this can not only be supportive for hearing aid selection, but also for a well-structured evaluation of the hearing aid benefits. If applied in a large-scale approach, this will yield practice-based evidence that will compensate for the lack of evidence-based practice in hearing aid selection.

INTRODUCTION

Although diagnostic data from pure tone audiometry and speech audiometry are essential for fitting a selected hearing aid, their role in the selection of a hearing aid itself is limited. We propose to design a systematic approach for hearing aid selection with focus on signal-processing functionalities rather than on features and operational issues like volume controls, connectivity, or options for tinnitus masking and (bi)CROS-units. For this purpose, additional information is required about the limitations experienced by the hearing-impaired client in daily life without or with their old hearing aids (pre-fitting), and how this changes with new hearing aid use (post-fitting). Therefore, it is useful to draw up an inventory of both the disabilities experienced by the hearing-impaired listener and the individual objectives for rehabilitation (fitting targets).

Kramer *et al.* (1995) developed a questionnaire to assess hearing impairment in daily life, the Amsterdam Inventory for Auditory Disability and Handicap (AIADH), which was shown to have good reliability and validity (Meijer *et al.*, 2003). For the disability profile used in this study, the AIADH was slightly adapted to the AVAB (in Dutch: Amsterdam Questionnaire for Auditory Disabilities): We only used the disability-related questions, added three questions, and rearranged the questions into six dimensions or factors: detection of sounds, speech in quiet, speech in noise, auditory localization, focus, and noise tolerance (see Table 1). Such a profile might be useful in tailoring a hearing aid to the specific needs of a patient, as well as in evaluating the benefit of a hearing aid for an individual with respect to the six different aspects of auditory functioning (see also Fuente *et al.*, 2012). An important

*Corresponding author: w.a.dreschler@amc.uva.nl

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

disadvantage of a questionnaire like the AVAB is that it evaluates a fixed list of common listening situations, which are not by definition situations that are relevant for the patient. As an alternative for questionnaires with fixed situations, Dillon *et al.* (1997) proposed the Client Oriented Scale of Improvement (COSI) for the evaluation of hearing aids, in which patients are asked to define their own targets for rehabilitation. Although the COSI is very useful for individual patients, the major disadvantage is that the individualization complicates the comparison of needs or benefits for groups of patients. This makes the COSI still useful for clinical practice, but of low value for research purposes and evaluation of hearing aid types or functions over groups of patients. In order to improve comparability between patients, Dillon *et al.* (1997) proposed to categorize each individual target into a set of sixteen predefined categories. Zelski (2000) concluded that the intra-observer agreement was high, but that the number of categories could be reduced. On the other hand, there is for instance no category for ‘localization of sounds’, despite the potential importance of this aspect in hearing aid selection and fitting.

As an alternative for the sixteen categories proposed by Dillon *et al.* (1997), the six dimensions of the AVAB might be useful. If this approach proves to be applicable, individual hearing disabilities and individual compensation targets can be compared along the same dimensions and can be taken together in a six-dimensional human-related-intended-use profile. These dimensions cover a broad range of important auditory functionalities and might be related to hearing aid functions. An advantage of using the same dimensions for AVAB and COSI is that, when using both AVAB and COSI, the COSI can help the interpretation and weighting of the AVAB results. If categorization of the COSI targets can be done in a reproducible way, COSI is a valuable tool in the hearing aid prescription and evaluation process, both for clinical practice and research purposes, by being individual and general at the same time.

The goal of this study was to determine whether the six categories defined by the AVAB disability profile are appropriate to also categorize individual COSI targets. The main two aspects of this question are (1) whether the inter-observer agreement between clinicians is sufficiently high, and (2) whether categories are regarded as missing or superfluous.

METHODS

Fitting targets from hearing aid candidates and hearing aid users

The COSI targets used in this study were administered during regular clinical practice in the Academic Medical Center. A total number of 533 COSI targets were collected from 151 consecutive patients who visited the clinic in fall 2014 and early 2015 for hearing aid fitting. During the first visit pure-tone audiometry, speech audiometry, AVAB questionnaire, and COSI questionnaire were all administered and documented. 103 patients were new hearing aid users and 48 patients already had a hearing aid. Data were gathered retrospectively from the database, thus patients and clinicians were not aware of the purpose of this study during administration of the targets. Personal information was removed to make the data anonymous.

Observers and procedure

Eight professional audiologists (six clinical physicists in audiology and two hearing-aid dispensers) participated in this study. There was a wide range in experience administering the AVAB and COSI. For the purpose of this study, this was regarded as an advantage. If inter-observer correspondence is not dependent on the level of experience, we may assume that the categorization of COSI targets is robust.

Participating audiologists received a file with the 533 COSI targets and a user interface for categorization, accompanied by written instructions. In order to make sure that they understood what was meant by the six categories (Table 1), they first got the possibility to read all AVAB questions sorted by category. After they confirmed that they understood the categories, they started the categorization procedure.

A user interface showed one target at the time and presented 3 questions to be answered for each of the targets:

1. The first question was which AVAB category describes the COSI target best. Only one category could be assigned, and observers were forced to make a choice. However, apart from the six categories, there was an option ‘not possible to categorize’ for targets that did not fit well in one of the categories.
2. The second question was if additional categories were required to describe the COSI-target. Observers were allowed here to add one or more categories, if this was judged to be relevant for the categorization of the COSI target.
3. The third question was whether the COSI target was formulated in a sufficiently specific way. Possible answers were yes or no.

Audiologists were allowed to stop at each moment and continue at a later moment from the point they stopped. After categorizing all 533 targets, the audiologist had the possibility to indicate whether they found the classification feasible, or whether they missed categories or perceived categories as superfluous. Finally, they had the possibility to give additional remarks.

AVAB: Profile of “general” disabilities	
Det	Detection
SiQ	Speech in Quiet
SiN	Speech in Noise
Tol	Noise Tolerance
Foc	Focus / Discrimination
Loc	Localization / spatial hearing

Table 1: List of dimensions that are derived from the AVAB questionnaire to inventory “general” disabilities.

RESULTS

The primary dimension

Figure 1 indicates the distribution of all judgments (8 observers times 533 COSI targets) regarding the primary dimension. Speech perception in noise and in quiet and detection were the dimensions mostly used as primary dimensions. In about 15% of the cases the audiologists chose for the option 'not possible to categorize' (indicated as "rest"). Some examples of COSI targets that did not match the six dimensions were: "To reduce the hinder from my tinnitus" or "Less problems with feedback". COSI targets could also be categorized as "rest", if the target was not specified in enough detail, e.g., "Communication with others", "Safety in my job", or "Less miscommunication at home". Figure 1 also indicates that the distributions for new users (gray bars) and experienced users (black bars) are very similar.

We analysed the numbers of COSI targets that were classified identically. In 389 out of 533 COSI targets, the primary dimension was the same for 8 audiologists (55%) or 7 audiologists (18%). This indicates a good agreement between observers. We also calculated Cohen's kappa as a metric for inter-observer correspondence (Cohen, 1960). If we include all dimensions into the analysis, Cohen's Kappa was 0.81. This may be considered as a substantial (or even almost perfect) agreement (Landis and Koch, 1977). Other measures for inter-observer agreement (Fleiss' kappa and Gwet's Agreement Coefficient 1) gave comparable results, both for the analysis of all dimensions and for sub-analyses for separate dimensions. The analyses of the individual dimensions revealed that the correspondence between audiologists in the categorization of focus/discrimination is only weak to fair (kappa value of 0.3).

The use of additional dimensions

As indicated, additional dimensions could be used to categorize the COSI target. Figure 2 shows that some of the audiologists used only the primary dimensions in the majority of cases (e.g., observers 2, 3, 4, and 8), while others frequently used 2, 3, or even more dimensions. Further analysis indicated that the combinations of dimensions that occurred most frequently were:

- Speech in quiet and speech in noise (for 38% of the cases where speech in quiet was chosen as primary dimension, speech in noise was chosen as additional dimension, and for 35% of the cases where speech in noise was chosen primarily, speech in noise was added as secondary dimension).
- Detection and localization (for 25% of the cases where detection was chosen as primary dimension, localization was chosen as additional dimension, and for 38% of the cases where localization was chosen primarily, detection was added as secondary dimension).
- Detection and focus (for 15% of the cases where detection was chosen as primary dimension, focus was chosen as additional dimension, and for 38% of the cases where detection was chosen primarily, focus was added as secondary dimension).

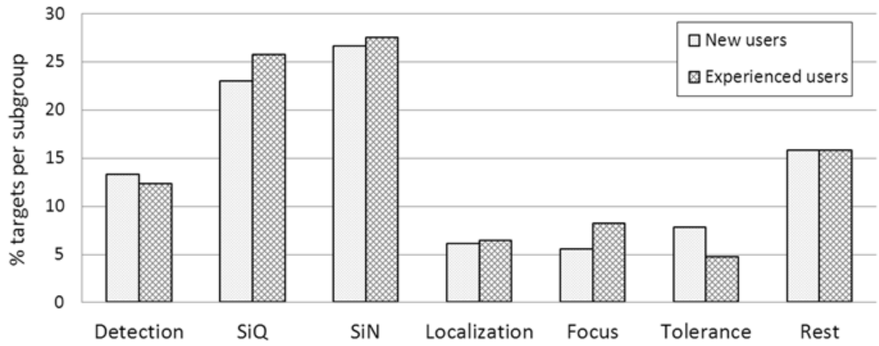


Fig. 1: Distribution of categories for the primary dimension, split for new users (gray bars; n=103) and experienced users (black bars; n=48). COSI targets that didn't match one of the 6 dimensions were categorized as "rest".

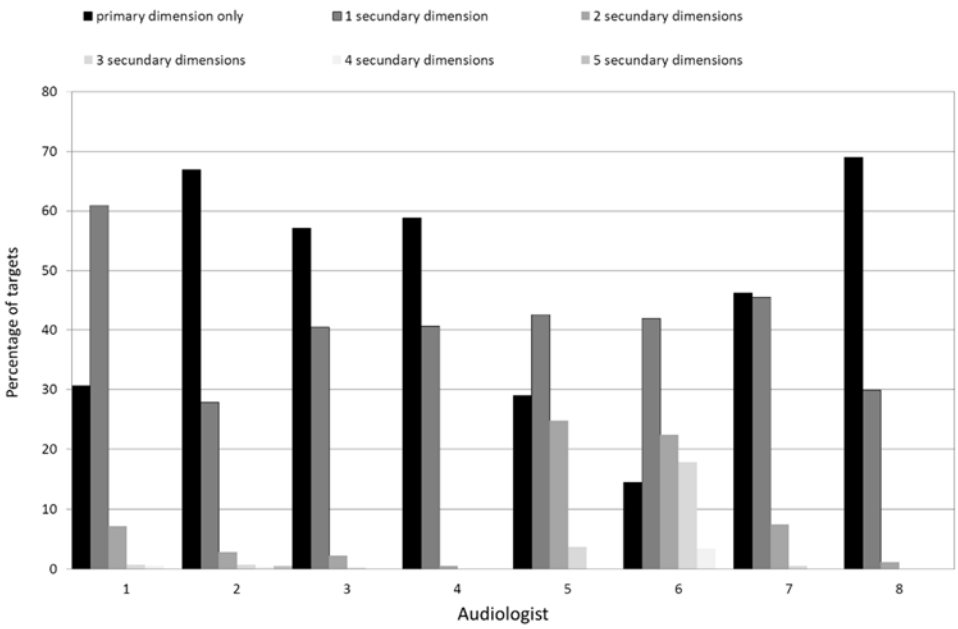


Fig. 2: Distributions of the number of dimensions used to categorize the 533 COSI targets by the 8 audiologists.

Missing dimensions

At the end of the session the audiologists answered some overall questions. The classification in six dimensions was regarded as feasible, but some categories were indicated as missing. Tinnitus was mentioned as a missing dimension by 5 out of 8 audiologists. Other dimensions that were missing were related to speech from a distance, listening effort, music and sound quality, and the perception of loud sounds. On the other hand, focus/discrimination was indicated to be more or less superfluous and was not often used.

DISCUSSION

In this study, we found a good agreement between eight audiologists in the categorization of COSI goals into the six AVAB dimensions. The agreement was very high given the fact that the observers reported that some targets are not specific enough, some targets do not fit well into the six dimensions, and some targets can easily be categorized differently, for instance: "To hear someone coming from behind, speaking". The study yielded some suggestions to combine dimensions and add new ones. Future research is needed to design a new set of "optimal" dimensions. But, despite the possible improvements in the future, this study indicates that COSI targets can be expressed reasonably well along the same dimensions as the disability profile defined by AVAB. This allows the following two steps:

From categorized COSI targets to a target profile

COSI can now be used to define a target profile in the same six dimensions as used in the AVAB-based profile of disabilities. The purpose of the hearing aid selection and fitting is then to improve the AVAB results by using a hearing aid until the clients meet the target profile resulting from COSI. As a starting point, for instance, we use a score of 3 for each of the six dimensions, thus 18 points overall. Based on the dimensions chosen for the COSI goals, these points can be re-divided over the dimensions, with more weight for dimensions for which COSI goals were formulated than for the other dimensions.

This can be done in different ways. But as a starting point, we implemented the following weighting:

- Each fitting target was assigned to one or more dimensions. We decided not to discriminate between primary and secondary/tertiary dimensions.
- The subject's priority was an important component of the weighting. The assignments were weighted according to the priority of the fitting target.
- The sum scores of the weighted assignments determined the relative importance of each dimension.

This way, the "18-points" are distributed according to their relative importance, indicated by the user, into an individually shaped target profile. For some subjects the resulting pattern was rather general (the points were more or less equally distributed across the six dimensions). In other subjects focused profiles were found, in which specific dimensions were much more important than others.

Combining the profiles into a profile of compensation needs

Given the finding that auditory disabilities and fitting targets can be expressed reasonably well along the same dimensions, they can be combined in a compensation profile. Of course, this can be implemented in different forms, but this section illustrates the choices that have been made in the BRIDGE program, that has been introduced recently in the Netherlands.

Figure 3 illustrates the way that disability profiles (assessed with the AVAB questionnaire) and target profiles (assessed with the COSI approach discussed above) can be combined in one profile for "compensation needs". The profile of disability is the starting point. For severe cases, the scores are closer to the centre (a severe disability gives a low score). The target profile is around 3, but may be focused as discussed above. Figure 3 shows a hypothetical but unusual example that speech in quiet is the top priority target. The difference between these two profiles is indicative of the compensation needs for an individual user and may be applied in the selection process for a hearing aid model/type and/or specific hearing aid features.

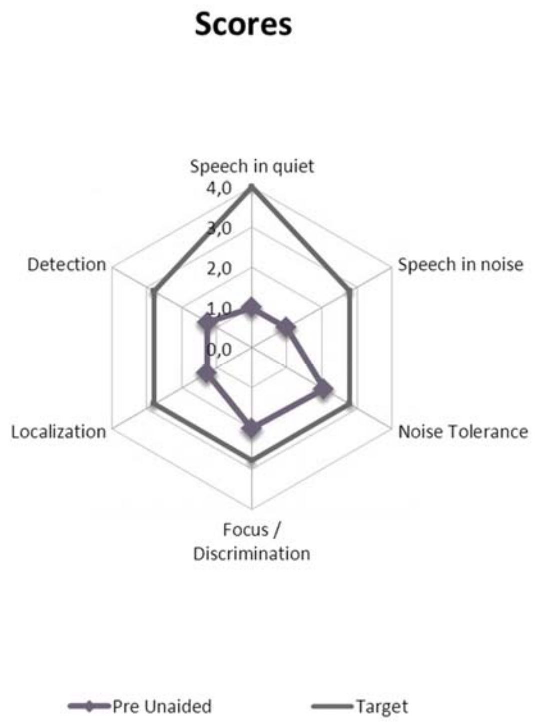


Fig. 3: Illustration of the combination of profiles. The compensation profile is composed of the difference between profile of disabilities (the starting point from AVAB) and the target profile (the user needs from COSI).

APPLICATIONS

There are three major areas where our approach of a disability profile and a target profile along the same dimensions, combined in a compensation profile, can be supportive:

1. The compensation profile is a means that can be helpful in the selection of hearing aids and/or hearing aid functionalities. The overall degree of compensation needs should be related to the minimum level of technology that is required for an adequate compensation. In addition, the profile of the compensation needs along the six dimensions indicates which aspects deserve special attention in the selection process. A possible outcome can be that a person with relatively modest compensation needs nevertheless should be fitted with a high-end hearing aid, because his/her main problems are in the "focus" dimension that is hard to compensate with low-end hearing aids.

2. The profiles provides a well-structured basis for the evaluation of the post-fitting situation. The AVAB questionnaire can be used for pre-post comparisons and the post-fitting results should – in principle – meet the targets along each of the dimensions, because these were defined as the fitting targets. It should be realized that this is not always feasible (e.g., restoration of localization in one-sided deaf subjects), but in that case clear argumentation is required that helps to interpret the post-fitting outcome measures. In addition the COSI questions about the "degree of change" and the "final ability" will be used. Both components of evaluation form a good combination by being individual (COSI) and general (AVAB) at the same time.
3. If applied on a large scale, a system like BRIDGE is able to collect knowledge for better hearing aid selection. The system can be used to collect practice-based evidence and this data can be used to learn the clinicians more about reference values and the potential for beneficial effects of hearing aids. This knowledge can partly be used to update the system in the future.

CONCLUSIONS

This study shows a way to translate individual patterns of user needs into more general dimensions derived from a disability questionnaire. Now we are able to calculate a qualitative indication of the compensation power required in six dimensions, based on the degree of disability and the individual user needs. This categorization is a starting point in hearing aid selection. Also this approach offers a systematic approach for the evaluation of the post-fitting results. Finally, the approach is able to collect practice-based evidence, if applied on a large scale.

REFERENCES

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

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. Ac. Audiol., **8**, 27-43.

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.

Kramer, S.E., Kapteyn, T.S., Festen, J.M., and Tobi, H. (**1995**). "Factors in subjective hearing disability," Int. J. Audiol., **34**, 311-320.

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

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.

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 Theses and Dissertations, University of South Florida.

Predicting individual hearing-aid preference in the field using laboratory paired comparisons

MARTIN DAHLQUIST*, JOSEFINA LARSSON, SOFIA HERTZMAN, FLORIAN WOLTERS, AND KAROLINA SMEDS

ORCA Europe, Widex A/S, Sweden

Two gain settings were compared in two hearing-aid programs. Twenty participants with impaired hearing evaluated the settings during a two-week field trial period using a double-blind design. During the field test, the participants used a diary to report which program they preferred in various self-selected situations. After the field trial, the participants stated their overall preferred setting in an interview and answered questions about their preferred settings in various predefined sound scenarios. In the laboratory, the participants made paired comparisons of preference, speech intelligibility, comfort, and loudness. The analysis focused on whether the laboratory test could predict the results obtained in the field. On a group level, it looked as if the results from the diary and questionnaire (data from the field) agreed well with the laboratory paired comparisons. However, on an individual level, the laboratory paired comparisons were not effective in predicting real-life preference. Potential reasons for this result and the consequences of the result are discussed.

BACKGROUND

Currently, there is a certain focus in the hearing-device research community on how more realistic laboratory tests should be designed. Another question is how we can collect data that is more sensitive to small signal-processing differences from the test participants' real life.

Testing using paired comparisons (PCs) is often advocated as a sensitive measure when small differences in for instance signal processing are studied (Amlani and Schafer, 2009; Kuk, 2002). However, the correlation between PCs performed in the laboratory and the preference experienced in the field is not commonly documented. The purpose of the current study was to see if laboratory PCs could predict the preference experienced in the field.

METHOD

Twenty participants compared two hearing-aid gain settings in the field and in the laboratory using a double-blind design. In a two-week field trial, the participants compared the two settings in two hearing-aid programs in a balanced design. The following outcome measures were used: An interview that focused on the preferred

*Corresponding author: martin.dahlquist@orca-eu.info

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

hearing-aid program, a diary for paired comparisons in the field, a questionnaire answered after the field test, and the hearing aid log data. In the laboratory, the participants made paired comparisons of preference, speech intelligibility, comfort, and loudness.

Participants

Twenty experienced hearing-aid users, 8 females and 12 males (average age: 74 years) were recruited from the ORCA Europe database. They all had symmetrical hearing losses and all had experience with hearing aids of other brands than Widex. Measured pure-tone thresholds are found in Fig. 1.

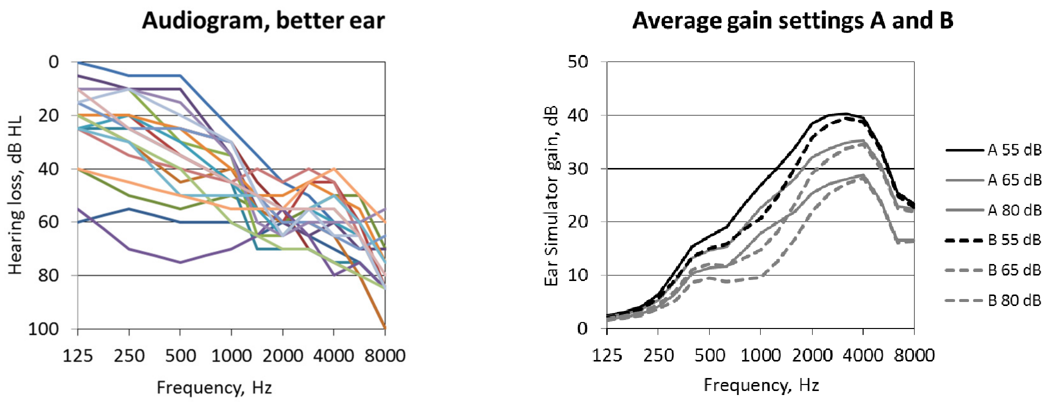


Fig. 1: Left: Better ear audiometric thresholds for all participants. Right: Average hearing-aid gain (Ear simulator) for the two prescriptions A and B measured using speech at three input levels (55, 65, and 80 dB SPL).

Hearing aids and prescriptions

A research receiver-in-the-ear (RITE) hearing aid was fitted using custom earmoulds (15 participants) or standard domes (5 participants) with varying degree of ventilation (selected based on audiogram configuration and type of currently used hearing aids). A directional microphone system and an SII-based noise reduction were switched on. One hearing-aid program was programmed using Widex' general prescription, and the other prescription differed from the Widex prescription by reduction of the gain in a fairly broad region around 1 kHz (Fig. 1, right panel). The hearing-aid fittings were verified and documented using real-ear measurements (Interacoustics Equinox REM440). The hearing-aid settings were also documented using box measurements (Interacoustics Equinox HIT440, equipped with test box TBS25 and an ear simulator G.R.A.S. RA0045).

Field – Diary

The purpose of the diary was to give the participants an opportunity to make direct paired comparisons of the two hearing-aid programs in real-life situations. Seven sound scenario categories were described in the booklet that constituted the diary. Short descriptions can be found in Fig. 2. These sound scenario categories focused on activities and the participants' intent rather than purely on the acoustical environment. Prior to the field test, the participants suggested at least one example of a situation from their everyday life matching each scenario category, and the participants were encouraged to make the evaluation in this specific situation as well as in other situations.

The participants were asked to categorize experienced sound situations into one of the seven pre-defined scenarios, to make a paired comparison of the two programs and write a comment if they wanted to describe why a particular program was chosen.

Field – Interview

When the participants returned to ORCA Europe after the field trial, a structured interview was performed. Overall preference for one of the hearing-aid programs was the main outcome of the interview.

Field – Sound scenario questionnaire

Together with the test leader, the participants also filled out a questionnaire with questions about preference in a number of presented scenarios. For each of the seven main sound scenario categories used for the diary (except category 4), one or two more specific examples, assumed to be encountered by a majority of the participants, were presented. The participants assessed the occurrence of each sound scenario example, which hearing-aid program they preferred in the scenario, and the strength or certainty of their choice. The participants also had the possibility to add four own sound scenario examples that were not covered by the ten pre-defined scenarios. Short descriptions of the scenarios can be found in Fig. 3.

Field – Data log

The hearing aid logged data (active program, sound level, and volume control setting) in 24.4-min intervals during 104 hours.

Lab – Paired comparisons (PC)

In the laboratory, sound field paired comparisons were made for four attributes (preference, speech intelligibility, comfort, and loudness) for a number of stimuli (Table 1). Six loudspeakers (KRK R6, powered by two Rotel RMB-1075 amplifiers) were placed at 1.0 m distance from the reference point (in the middle of the listener's head) at 0, 45, 90, 180, 270, 315 degrees azimuth in a sound-treated test booth.

During the test, the hearing aids were connected to a computer via a USB link. The volume control was set to default (the fitted gain). A Matlab script controlled playback of the test stimuli, the program switching, and the storage of the responses. The sound files were looped and played back continuously, while the participants used a graphical user interface to control which hearing-aid program was active and to indicate their choice of program. They also indicated the magnitude of difference between the two programs (small/moderate/large difference).

For each sound example, the order of the hearing-aid programs was randomized. A pre-conditioning time of 15 seconds was used at the beginning of each new sound stimulus in order for the hearing aids to stabilize their performance. During this time, the program selection buttons were locked. Generally, one round of comparison was made for each rating attribute and sound stimulus, but preference was also assessed at the end of the visit in order to collect retest data.

Stimulus	Level, dB	# Loud-speakers	Prefe-rence	Speech Intell.	Com-fort	Loud-ness
Outdoors with birds	51	2	X			
Speech in Quiet	55	1	X	X		X
Speech in Quiet	65	1	X	X		X
Speech in Cafeteria noise	75/71	1+5	X	X	X	X
Music,string quartet	75	2	X			X
Music, piano	75	2	X			
Soccer chant	85	4	X		X	

Table 1: Paired comparisons: Sound stimuli, presentation levels, loudspeaker setup, and rating attributes.

Relationships between measures

Individual results from field and laboratory tests were collected and processed to allow for correlation analysis. From the field test, the interview gave 1 variable, the diary 7 variables, and the questionnaire 10 variables. From the laboratory test, there were 16 variables from the PCs. The outcomes were transformed into difference measures. Both Spearman's rank correlation and a binomial method were applied.

RESULTS

Field – Interview

During the interview 9 participants stated that they preferred setting A and 11 that they preferred setting B, with varying degree of confidence.

Field – Diary

The participants made paired comparisons in a variety of relevant sound environments, performing various activities. In median, the participants had 27 entries (range 4-80) and these entries were often described in detail. For each participant and each sound scenario category, the preference for A and B was calculated as percentages. Then an average was calculated across all participants. Fig. 2 shows close to equal preference, but setting A (providing more gain) was slightly preferred for live focused listening, whereas B (providing less gain) was slightly preferred for sound monitoring and passive listening.

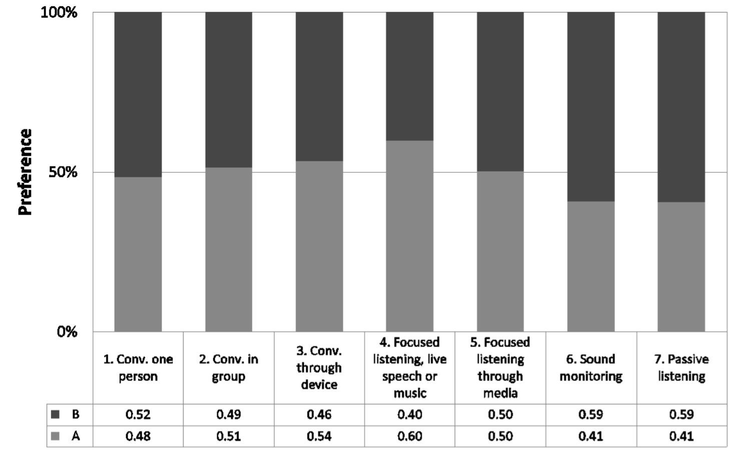


Fig. 2: Diary. Average preference for setting A (light grey) and B (dark grey) for the seven sound scenario categories used in the diary.

Field – Questionnaire

For all participants who had experienced the various predefined scenarios described in the questionnaire, the preference is indicated in Fig. 3. This shows close to equal preference, but setting A (providing more gain) was slightly preferred for speech communication scenarios whereas setting B (providing less gain) was slightly preferred for passive listening scenarios. Only the difference for the last scenario ("Resting on a train") was statistically significant (p<0.05, sign test).

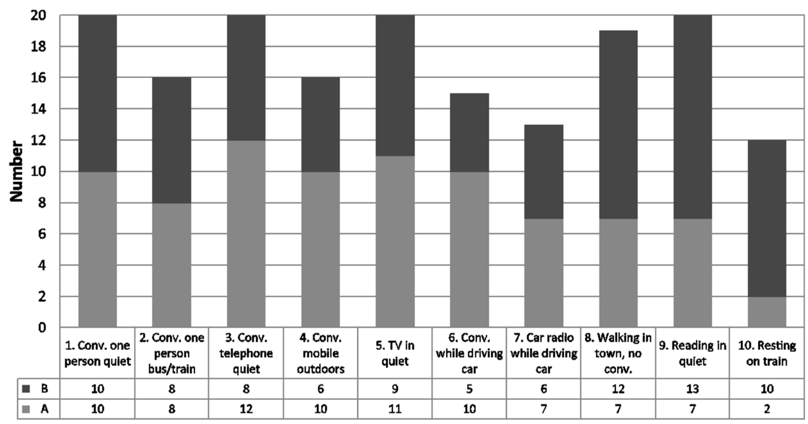


Fig. 3: Number of participants who preferred setting A (light grey) and setting B (dark grey) in the ten sound scenarios described in the questionnaire.

Field – Data log

The data log showed that the hearing-aids were used 11.4 hours per day in median. The preferred setting was generally used more than the non-preferred. The volume control was on average changed equally often up and down for both programs, but the fitted gain was used on average 70% of the time.

Lab – Paired comparisons (PC)

When laboratory PC data for all sound stimuli were pooled (Fig. 4), there was about equal preference for settings A and B. Setting A was preferred for speech intelligibility and setting B for comfort, and A was judged as louder than B (these differences were statistically significant). Differences in the preference pattern were seen for the various stimuli (not shown in the figure): For soft speech there was a preference for A, while B was preferred for speech in cafeteria noise (both differences statistically significant). The interpretation was that preference in the latter scenario was more related to comfort than to speech intelligibility. Statistical testing was done using Wilcoxon signed ranks test with $p<0.05$.

Relationships between measures

Correlation analyses were made with two methods, whose results agreed well. Generally, the results from the correlation analyses showed that a large number of the field outcome variables correlated with each other. Specifically, there were statistically significant correlations ($p<0.05$) with the interview question about preferred setting after the field trial for 6 out of 7 diary sound scenario categories and for 6 out of 10 questionnaire sound scenarios. On the other hand, none of the laboratory PC results correlated with the result of this interview question about overall preference.

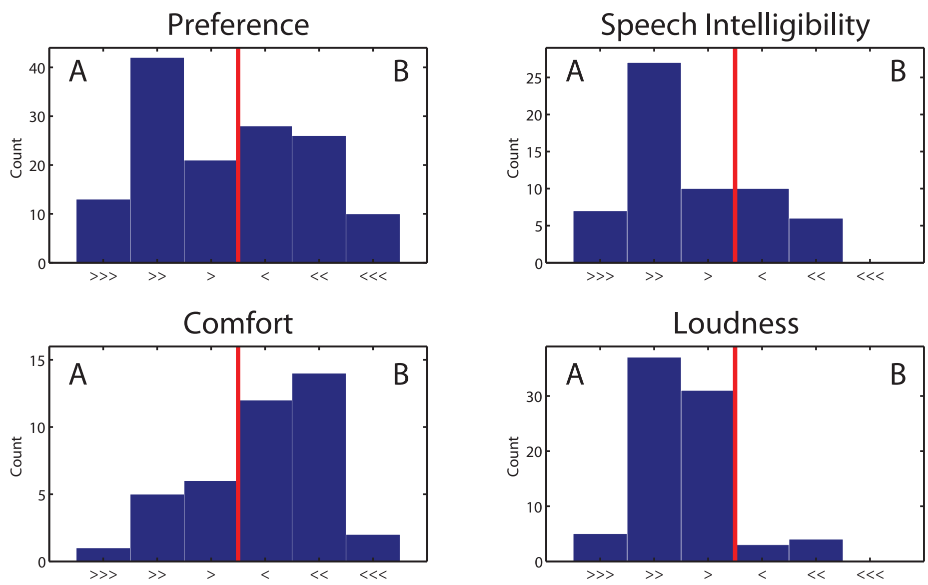


Fig. 4: Laboratory PC results for the attributes when all sound stimuli were pooled. Bars to the left of the vertical line indicate that A was selected more often, bars to the right that B was selected more often. The y-axes represent the total number of ratings. The symbols on the x-axes indicate the magnitude of difference between the two programs.

DISCUSSION AND CONCLUSIONS

The laboratory PC group data seemed to agree with the group data from the diary and the questionnaire. The clear difference between the settings found in the laboratory for speech intelligibility and comfort (Fig. 4) were mirrored by similar tendencies in the field (Figs. 2 and 3). But, on an individual level, the laboratory PC data only correlated with very few of the field test outcomes. Specifically, the laboratory PC data could not predict the overall preference in the field. Potential explanations for this prediction mismatch will be discussed.

If the difference between compared settings is so small that the participants have difficulties detecting the difference, the results found in this study would be expected. But, that did not seem to be the case here. The audiologists who performed the testing reported that the participants did not seem to have any difficulties hearing the difference between the two programs, neither during the field test, nor during the laboratory testing. The ratings done in connection to both field and laboratory paired comparisons also showed that the participants often rated the difference to be at least “moderate”.

Laboratory paired comparisons are sensitive when small differences are evaluated, but there are a number of problems associated with these tests when the results are compared to real-life performance.

One limitation with the traditional laboratory PC setup used in this study is the artificial situation and task. In particular, the participants only listened to speech, instead of participating in a dialogue. This means that potentially important aspects of the signal processing might have been lost. These aspects could be related to the sound quality of the participants’ own voice and to the changes in signal processing perceived when the speech levels change during a dialogue. This could create a difference between the field and the laboratory. At ORCA Europe, we have subsequently tried a dialogue-based paired comparison task. Own voice is included and a more complex activity is created. Initial testing has indicated that this might be a possible way forward.

Another aspect of the difference between the laboratory and the field is the selection of sound stimuli. For the laboratory PC, a fairly large range of presentation levels and sound types were selected (Table 1). These stimuli might not be representative of the situations the participants encountered in real life. This listening situation mismatch could perhaps be overcome by selecting stimuli in the laboratory test based on the situations encountered in the field (reported in the diary). However, a separate analysis of the current laboratory PC data, only including the most commonly experienced stimuli, did not show a better correlation between laboratory and field outcomes.

Further, the loudness difference between the two settings probably play a larger role in the laboratory PCs than in the field, where the volume control could be used. It is also possible that the participants in the laboratory PC focused on some easily identified details, perhaps specific to the recording or talker, in a way that is not done in the field.

In addition to these general shortcomings of the laboratory PCs, it turned out that the test-retest reliability for the laboratory PCs was poor for the speech stimuli, but acceptable for the music stimuli. The number of repetitions was too limited, but there also seemed to be some other methodological difficulties with the test when speech was used. Especially when the speech was easily understood, the participants seemed to find it difficult to judge preference and speech intelligibility.

Some participants also indicated a confusion when using the preference attribute. First, PCs for preference was measured, then followed the three other attributes before a retest for the preference was performed. During the first preference measurements, the participants did not seem to find the task difficult, whereas some of them commented things like “Do you want me to concentrate on speech intelligibility or comfort?” when they were asked to rate preference the second time. That “attribute confusion” did not take place for the music stimuli, for which only preference and loudness were measured, and the preference judgments seemed less complex.

Development of more realistic laboratory tests is one way of improving the evaluation of hearing-aid characteristics. Another strategy will probably be to perform more controlled and sensitive field tests. Advanced logging of field preference using smartphones (e.g., Kissner *et al.*, 2015) and semi-controlled field tests, for instance inspired by “Soundwalks” (e.g., Adams *et al.*, 2008), seem promising.

In conclusion, the current laboratory paired comparisons could not predict outcomes in real life. Suggestions for improving the laboratory paired comparisons (including both basic methodological questions to improve test-retest reliability, and more substantial changes to the task included in the comparisons) have been presented and alternative methods for collecting real-life sensitive data have been mentioned.

REFERENCES

Adams, M.D., Bruce, N.S., Davies, W.J., Cain, R., Jennings, P., Carlyle, A., Cusack, P, Hume, K., and Plack, C. (**2008**). “Soundwalking as a methodology for understanding soundscapes,” Proc. Ins. Ac., **30**.

Amlani, A.M. and Schafer, E.C. (**2009**). “Application of paired-comparison methods to hearing aids,” Trends Amplif., **13**, 241-259.

Kissner, S., Holube, I., and Bitzer, J. (**2015**). “A smartphone-based, privacy-aware recording system for the assessment of everyday listening situations,” In Proc. ISAAR 2015: *Individual Hearing Loss – Characterization, Modelling, Compensation Strategies*. 5th International Symposium on Auditory and Audiological Research, Nyborg, Denmark. Eds. S. Santurette, T. Dau, J.C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

Kuk, F.K. (**2002**). “Paired-comparisons as a fine-tuning tool in hearing aid fitting,” in *Strategies for selecting and verifying hearing aid fittings*, 2nd ed. Ed. M. Valente, New York, NY, USA: Thieme medical publishing, pp. 125-150.

Relating hearing aid users' preferred noise reduction setting to different measures of noise tolerance and distortion sensitivity

TOBIAS NEHER[1,3,*], KIRSTEN C. WAGENER[2,3], MARKUS MEIS[2,3], AND ROSA-LINDE FISCHER[4]

[1] *Medizinische Physik, Oldenburg University, Oldenburg, Germany*

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

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

[4] *Sivantos GmbH, Erlangen, Germany*

Recently, there has been growing interest in the personalisation of hearing aid fittings. In two previous studies, we investigated preference for different types of noise reduction (NR) processing and found that we could partly explain individual differences based on audiometric and cognitive factors. In the current study, we explored a number of psychoacoustic and self-report measures in terms of their ability to help explain these results. Groups of hearing aid users with clear preferences for either weak ($N = 13$) or strong ($N = 14$) NR participated. Candidate measures included maximally acceptable background noise levels, detection thresholds for speech distortions caused by NR processing, and self-reported 'sound personality' traits. Participants also adjusted the strength of the binaural coherence-based NR algorithm to their preferred level. Analyses confirmed the basic group difference concerning preferred NR strength. Furthermore, detection thresholds for speech distortions were higher for 'NR lovers' than for 'NR haters'. In terms of maximally acceptable noise levels, there was a tendency for NR lovers to be less tolerant towards background noise than NR haters. Group differences were generally absent in the self-report data. Altogether, these results suggest that differences in preferred NR setting are partly related to individual sensitivity to background noise and speech distortions.

INTRODUCTION

Digital hearing aids (HAs) are typically equipped with a large range of signal processing algorithms such as noise reduction (NR) and directional processing (e.g., Dillon, 2012). Because individual users are known to respond very differently to such algorithms it is of interest to find ways for their individualisation. In two recent studies, we therefore investigated individual speech recognition with, and preference for, different types of NR (Neher, 2014; Neher *et al.*, 2015). In short, we observed

*Corresponding author: tobias.neher@uni-oldenburg.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

large variability in outcome, which we could partly explain based on differences in pure-tone average hearing loss (PTA) and cognitive function. That is, we found participants with larger PTA and worse cognitive function to have weaker preferences for inactive NR and stronger preferences for strong NR than participants with smaller PTA and better cognitive function. These results could suggest that the former types of participants are more affected by noise and thus favour greater noise removal even at the cost of added speech distortions, whereas the latter do not.

The main purpose of the current study was to find out if the results summarised above are related to individual differences in noise tolerance and distortion sensitivity. In particular, it had the following three aims:

1. To confirm the previously observed differences in preferred NR setting using a method of self-adjustment;
2. To investigate if performance on two psychoacoustic measures of noise tolerance and distortion sensitivity can account for preferred NR setting;
3. To explore a novel 'sound personality' questionnaire in terms of its ability to reveal differences in preferred NR setting.

METHODS

Participants

For the current study, we recruited 27 participants aged 61-81 yr. All of them had taken part in our earlier studies and were experienced HA users with symmetrical sensorineural hearing impairments. Furthermore, all of them were screened for a number of sensory and neuropsychological deficits (cf. Neher, 2014). Eligibility for the current study was determined based on their overall preference for NR processing. Using the data from our previous studies, we computed an aggregate preference score per participant for 'inactive', 'moderate', and 'strong' NR (see below). From the 60 available participants, we then chose those 13 participants with the clearest preference for inactive NR ("NR haters") and those 14 participants with the clearest preference for strong NR ("NR lovers"). The two resultant groups did not differ in terms of age (73 vs. 70 yr, $p > 0.17$), PTA across 500 Hz to 4 kHz (45 vs. 47 dB HL, $p > 0.5$) or reading span (40 vs. 40% correctly recalled target words, $p > 0.9$; cf. Carroll *et al.*, 2015).

Test setup and HA signal processing

All measurements were carried out in a soundproof booth. They were controlled from a personal computer (PC) running the measurement software. This PC was connected to another PC via a digital audio interface. The other PC was running a real-time HA simulation (implemented on the Master Hearing Aid research platform; Grimm *et al.*, 2006), which was controlled via the measurement PC.

The HA processing closely resembled that we had used previously (cf. Neher, 2014). It included binaural coherence-based NR (Grimm *et al.*, 2009), NAL-RP amplification (Byrne *et al.*, 1991), and equalisation of the magnitude spectrum of the headphones used for stimulus presentation (Sennheiser HDA 200). Concerning the

NR processing, the algorithmic parameter that we varied was the processing strength indexed by the parameter α (cf. Grimm *et al*., 2009). Setting α to 0, 0.75, or 2 resulted in the inactive, moderate, and strong NR settings tested previously.

Speech stimuli

The stimuli closely resembled those we had used previously. They were based on recordings from the Oldenburg sentence test (Wagener *et al*., 1999; Wagener and Brand, 2005). To simulate a realistic complex listening situation we convolved these recordings with pairs of head-related impulse responses measured in a reverberant cafeteria using a head-and-torso simulator equipped with two behind-the-ear HA dummies (Kayser *et al*., 2009). Specifically, we used the measurements made with the front microphones of each HA dummy and a frontal source at a distance of 1 m from, and at the same height as, the head-and-torso simulator. For the interfering signal, we used a recording made in the same cafeteria with the same setup during a busy lunch hour. During the measurements, we presented this signal at a nominal sound pressure level of 65 dB and mixed it with the target sentences, the level of which we adjusted to produce a given signal-to-noise ratio (SNR).

Self-adjusted NR strength

To confirm the basic group difference concerning NR preference we asked our participants to adjust the strength of the NR algorithm such that they would be willing to listen to the stimuli for a prolonged time. Participants could make these adjustments in real-time using a slider on a graphical user interface displayed on a touch screen. Measurements were performed at two input SNRs: 0 and 4 dB. They started with two training runs (one per input SNR), followed by six test runs (three per input SNR) in randomised order. For the analyses, we used the median of the three self-adjusted NR strengths per input SNR and participant.

Acceptable noise level

To assess noise tolerance we performed measurements based on the acceptable noise level (ANL) test (Nabelek *et al*., 1991). Using a graphical user interface, participants had to adjust the level of the cafeteria noise three times in a row: (1) so they no longer could follow the target speaker, (2) so they could follow the target speaker very easily, and (3) so they would just about be able to tolerate the noise while trying to follow the target speaker for a prolonged time (the 'maximally acceptable noise level'). Unlike in the original ANL procedure, we presented the target speech at a fixed, nominal level of 65 dB SPL. We then obtained our ANL estimates by taking the difference between the nominal speech level and the maximally acceptable noise level. Note that, as in the original ANL procedure, a lower value therefore indicates more tolerance towards noise.

We measured ANLs for the inactive, moderate, and strong NR settings tested previously. The measurements with inactive NR served as estimates of general noise tolerance ('baseline ANL'). The measurements with moderate and strong NR served to verify the expected benefit from active NR with respect to (greater) noise tolerance. We started with six training runs (two per NR setting), followed by nine test runs (three per NR setting) in randomised order. For the analyses, we then used the median of the three ANL estimates per NR setting and participant.

Distortion sensitivity

To assess distortion sensitivity we used an adaptive 3-interval 2-alternative forced-choice paradigm coupled with a 1-up 3-down rule to estimate the 79.4% detection threshold (Levitt, 1971) for the distortions imposed onto the target speech (cf. Brons *et al*., 2014). The task of the participants was to choose which of two sound samples was different from a reference sound sample. The reference sound sample, which was always presented in the first interval, was an unprocessed speech signal. The target sound sample was the same speech signal processed with the NR gains computed for the speech-in-noise mixture. Before presentation, we equated the target and reference sound samples in terms of their root-mean-square levels. We then applied level roving of up to ±2 dB during the second and third intervals to prevent our participants from relying on any potentially remaining loudness differences, and also instructed them to concentrate on differences other than loudness to complete the task. There was one training run, followed by two test runs. Feedback was provided throughout. As our detection threshold estimate, we took the median of the last eight reversal points per test run and participant.

Self-reported sound personality

To assess self-reported traits related to noise tolerance and distortion sensitivity we used a new sound personality questionnaire intended to predict usage of, and preference for, different types of HA technology (Meis *et al*., 2015). This questionnaire consists of 46 items that were derived based on expert interviews as well as focus groups and in-depth interviews with normal-hearing and hearing-impaired listeners. In analysing the data from 622 predominantly older participants with different degrees of hearing loss, Meis *et al*. uncovered seven underlying factors: (F1) annoyance/distraction by background noise, (F2) importance of sound quality, (F3) noise sensitivity, (F4) avoidance of unpredictable sounds, (F5) openness towards loud/new sounds, (F6) preference for warm sounds, and (F7) detail in environmental sounds/music. In the current study, we explored the predictive power of these factors with respect to NR preference. For the analyses, we calculated the mean score across the items belonging to a given factor.

RESULTS

Self-adjusted NR strength

To analyse the self-adjusted NR strength data we performed a repeated-measures analysis of variance (ANOVA) with SNR as within-subject factor and listener group as between-subject factor. We found significant effects of SNR ($F_{1,25} = 12.5$, $p < 0.01$, $\eta_p^2 = 0.33$) and listener group ($F_{1,25} = 11.4$, $p < 0.01$, $\eta_p^2 = 0.31$) and a non-significant interaction of these two factors ($p > 0.5$). Consistent with our expectations, the NR haters preferred weaker NR processing than the NR lovers (mean α: 0.8 vs. 1.5; see Fig. 1). Also consistent with our expectations, both groups preferred stronger NR at 4 dB SNR than at 0 dB SNR (mean α: 1.3 vs. 0.9).

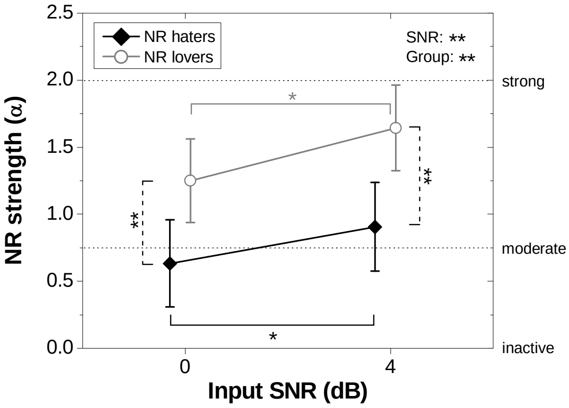


Fig. 1: Self-adjusted NR strength. Means and 95% confidence intervals for the two listener groups and input SNRs. α-values corresponding to inactive, moderate and strong NR are also indicated. * $p < 0.05$, ** $p < 0.01$.

Acceptable noise level

Despite several training runs, one participant was unable to perform the ANL test reliably and was thus excluded from the analyses. Performing a repeated-measures ANOVA with NR setting as within-subject factor and listener group as between-subject factor on the other data revealed a significant effect of NR setting ($F_{2,48} = 15.3$, $p < 0.00001$, $\eta_p^2 = 0.39$), a non-significant effect of listener group ($p > 0.7$), and a NR setting × listener group interaction that exceeded the 5% significance level slightly ($F_{2,48} = 3.0$, $p = 0.058$, $\eta_p^2 = 0.11$). For the NR lovers, noise tolerance increased by 3.7 and 4.5 dB with moderate and strong NR, respectively (see Fig. 2); for the NR haters, no statistically significant ANL changes were observed. This was due to the baseline ANLs (with inactive NR) of the NR lovers being about 2 dB higher (poorer) than those of the NR haters.

Distortion sensitivity

Since one (out of the 54) distortion sensitivity thresholds that we obtained was classified as an outlier we excluded it from the analyses. Performing a repeated-measures ANOVA on the remaining data with test run as within-subject factor and listener group as between-subject factor revealed a significant effect of listener group ($F_{1,23} = 5.7$, $p = 0.026$, $\eta_p^2 = 0.20$) and non-significant effects of test run ($p > 0.09$) and listener group × test run ($p > 0.8$). Consistent with our expectations, the NR lovers were less sensitive to the speech distortions than the NR haters (α-value at threshold: 0.44 vs. 0.31; see Fig. 3).

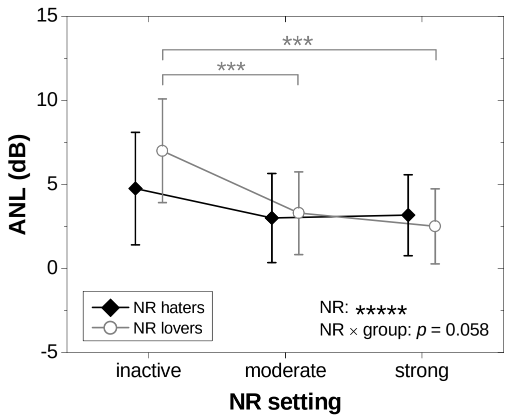


Fig. 2: ANL. Means and 95% confidence intervals for the two listener groups and three NR settings. *** $p < 0.001$, ***** $p < 0.00001$.

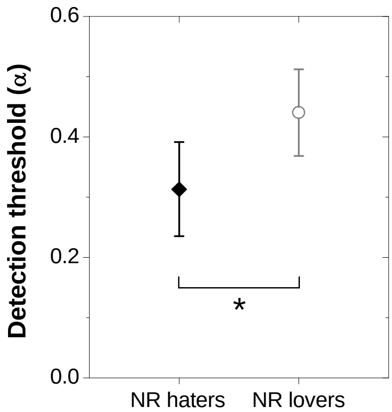


Fig. 3: Distortion sensitivity. Means and 95% confidence intervals for the two listener groups. * $p < 0.05$.

Self-reported sound personality

Figure 4 shows boxplots of the scores for the seven sound personality factors separated by listener group.

To check for any significant group differences we performed a series of two-tailed Mann-Whitney *U*-tests. However, none of these tests led to a significant result (all $p > 0.05$). Differences among the two groups were most apparent for F4 ('avoidance of unpredictable sounds'; $U = 1.8$, $p = 0.065$), followed by F6 ('preference for warm sounds'; $U = 1.3$, $p = 0.21$) and F3 ('noise sensitivity'; $U = 1.2$, $p = 0.24$).

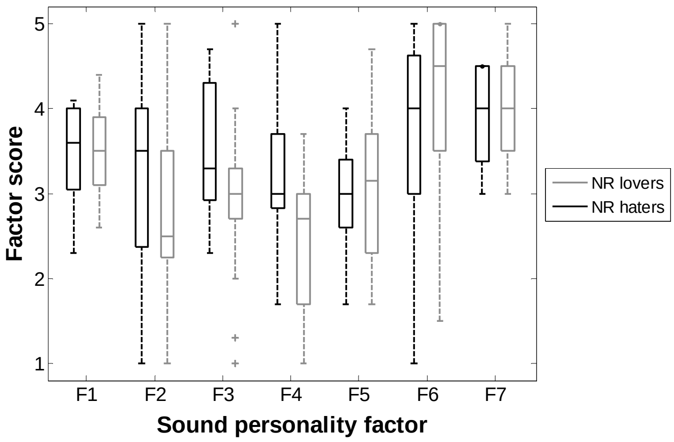


Fig. 4: Self-reported sound personality. Boxplots of scores for the seven factors.

SUMMARY

With respect to the three research aims outlined above, the results of the current study can be summarised as follows:

1. NR lovers set the strength of the algorithm tested here to almost twice the value chosen by NR haters, thereby confirming the group differences regarding preferred NR strength found previously with pre-selected (inactive, moderate, and strong) NR settings.
2. NR lovers obtained higher detection thresholds for speech distortions caused by the algorithm tested here than NR haters, indicating reduced sensitivity to such processing artefacts. Also, there was a (non-significant) tendency for NR lovers to have higher baseline ANLs than NR haters, indicating less tolerance towards background noise.
3. For the NR conditions considered here, the sound personality questionnaire did not reveal any clear differences among NR lovers and NR haters.

Altogether, these results provide a conceptual framework for factors seemingly involved in preference for NR processing (i.e., noise tolerance and distortion sensitivity). Future research should (i) confirm the putative link between preferred NR strength and baseline ANL, (ii) consider other types of NR algorithms, and (iii) apply the sound personality questionnaire to a wider range of HA conditions with a broader range of acoustical and perceptual effects.

ACKNOWLEDGEMENTS

This research was funded by the DFG Cluster of Excellence EXC 1077/1 Hearing4all and by Siemens Audiologische Technik (now Sivantos) GmbH.

REFERENCES

Brons, I., Dreschler, W.A., and Houben, R. (**2014**). "Detection threshold for sound distortion resulting from noise reduction in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am., **136**, 1375-1384.

Byrne, D., Parkinson, A., and Newall, P. (**1991**). "Modified hearing aid selection procedures for severe/profound hearing losses," in *The Vanderbilt Hearing Aid Report II*. Eds. G.A. Studebaker, F.H. Bess, and L.B. Beck (York Press, Parkton, NC), pp. 295-300.

Carroll, R., Meis, M., Schulte, M., Vormann, M., Kießling, J., and Meistere, 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.

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.

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 Trans. Audio Speech, **17**, 1408-1419.

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., 298605.

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

Meis, M., Huber, R., Fischer, R.L., Schulte, M., and Meister, H. (**2015**). "Develop-ment of a questionnaire for the assessment of sound preferences and hearing habits of people with different degrees of hearing impairment," 12th Congress Europ. Fed. Audiol. Soc. (EFAS), Istanbul, Turkey, May 27-30.

Nabelek, A.K., Tucker, F.M., and Letowski, T.R. (**1991**). "Toleration of background noises: Relationship with patterns of hearing aid use by elderly persons," J. Speech Hear. Res., **34**, 679-685.

Neher, T. (**2014**). "Relating hearing loss and executive functions to hearing aid users' preference for, and speech recognition with, different combinations of binaural noise reduction and microphone directionality", Front. Neurosci., **8**, 391.

Neher, T., Wagener, K.C., and Fischer, R.L. (**2015**). "Directional processing and noise reduction in hearing aids: Individual and situational influences on preferred setting," under review.

Wagener, K.C., Brand, T., and Kollmeier, B. (**1999**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache. I-III: Design, Optimierung und Evaluation des Oldenburger Satztests (Development and evaluation of a sentence test for the German language. I-III: Design, optimization and evaluation of the Oldenburg sentence test)", Zeitschrift für Audiologie (Audiological Acoustics), **38**, 4-15, 44-56, 86-95.

Wagener, K.C. and Brand, T. (**2005**). "Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters", Int. J. Audiol., **44**, 144-157.

For hearing aid noise reduction, babble is not just babble

HELEN CONNOR SØRENSEN* AND CHARLOTTE T. JESPERSEN

Global Audiology, GN ReSound A/S, Ballerup, Denmark

Most modern hearing aids provide single-microphone noise reduction without specifying how they work. The current study investigates how noise reduction is applied to babble noise in current premium hearing aids. Coupler gain measurements were performed in an acoustic test chamber. The signals used were standardized test signals, as well as babble noises compiled with different numbers of speakers (2, 4, 6, 8, and 10 speakers). The output of the hearing aid was measured with the noise reduction off and the strongest setting available. The gain reduction was calculated as the difference between the two settings. The results showed that, for an unmodulated test signal, the noise reduction algorithms applied quite different amounts of gain reduction across frequency. For the babble noise, some of the algorithms reduced gain very little, even for the 10-person babble. Other algorithms applied a graduated response, i.e., most gain reduction for 10-person babble, and the least amount of noise reduction for 2-person babble. Along with previous studies, this study highlights the need to have a standardized benchmarking procedure to define not only how noise reduction works in hearing aids but also which listening situations in which the noise reduction is active.

INTRODUCTION

For many hearing aid users, listening to speech in noisy situations is an important goal. For this reason, most modern hearing aids have single-microphone noise reduction. The general aim of hearing aid noise reduction is to reduce background noise while preserving speech information and sound quality. This is usually done by detecting in which frequency regions noise is more intense than speech and reducing gain in these regions. While there is only limited evidence that noise reduction improves speech intelligibility in noise, there is evidence of other benefits including improved sound quality and listening comfort, reduced noise annoyance, as well as possible improvements in listening effort and cognitive load (see Brons *et al.*, 2013 for recent discussion).

Previous studies have found large differences in how noise reduction works in commercial hearing aids (Bentler and Chiou, 2006; Brons *et al.*, 2013; 2014; Hoetink *et al.*, 2009; Smeds *et al.*, 2010). Quantitatively, there are more than 10-dB differences in how gain reduction is applied in a given frequency region. These differences can be heard by normally-hearing and hearing-impaired listeners

*Corresponding author: hconnor@gnresound.com

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

(Brons *et al.*, 2013; 2014). There are also differences in which signals activate the noise reduction (Hoetlink *et al.*, 2009) and the signal-to-noise ratio required to activate noise reduction (Smeds *et al.*, 2010; Brons *et al.*, 2013; 2014). This demonstrates that commercial implementations of noise reduction can behave differently in differerent listening situations.

This experiment is part of a larger study to investigate how different noise reduction algorithms in hearing aids are active in various listening situations. During preliminary measurements, it was observed that sometimes babble noise would result in a large gain reduction. For other measurements using the same hearing aids and a different babble signal, no gain reduction was applied by the same noise reduction setting. Given that listening situations with babble as background noise are highly relevant for hearing aid users, we decided to investigate this further. The purpose of this experiment is to investigate the effect of varying the number of babble speakers on how noise reduction is applied in current premium hearing aids.

METHOD

Five premium receiver-in-ear (RIE) hearing aids were programmed linearly to a mild, sloping hearing loss. Except for the noise reduction algorithm, all other advanced signal processing strategies were turned off. Recordings from the hearing aids were performed in an ear simulator in an anechoic test box. The amount of gain reduction applied by the noise reduction was calculated by comparing the output of the hearing aids with the noise reduction i) off, and ii) on with the strongest noise reduction setting available.

Hearing aids, programming, and verification

The hearing aids included were the latest premium hearing aids from five manufacturers, as of June 2015. They were all RIE (Receiver-In-the-Ear) form factor. The receiver was the lowest power level available for each aid. The hearing aids were programmed linearly with all other adaptive features off, including directional microphones and automatic program changes. If possible, expansion was switched off and the maximum power output was set to its maximum value. In the respective fitting softwares, occluded earpieces were selected. Each hearing aid was programmed with two listening programs: one with noise reduction off and the other with noise reduction on, with the strongest setting available.

Using the NAL-NL stand-alone software (v1.927), coupler targets were generated for the standardized N2 hearing loss (Bisgaard *et al.*, 2010), which is a mild, sloping hearing loss (Fig. 1). The targets were calculated using the NAL-NL2 rationale (Keidser *et al.*, 2011) specified for a 65-dB speech input. The audiological input variables used were: hearing thresholds measured using supra-aural headphones with default adult acoustic transforms. The user's sex was unspecified. The fitting variables were: bilateral, behind-the-ear fitting with RIE tubing with an occluded earmold. The compressor was assumed to be 18-channel with an intermediate compressor speed.

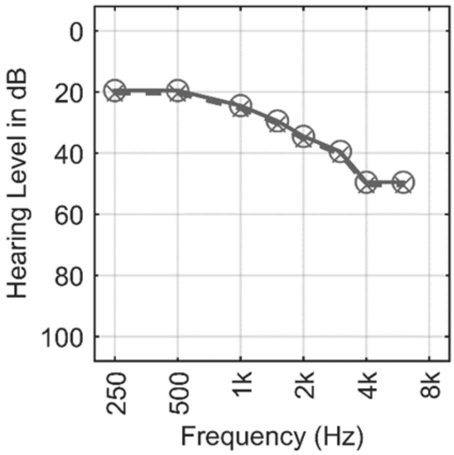


Fig. 1: The standardized N2 audiogram used to generate coupler targets for the hearing aids (Bisgaard *et al.*, 2010).

Fitting verification was performed in a 2-cc coupler in an Aurical HIT box with OTOsuite software (v. 4.75.00). For each of the hearing aids, coupler gain was matched to the NAL-NL2 targets within ±3 dB between 500 to 4000 Hz using the International Speech Test Signal (ISTS, Holube *et al.*, 2010). To check linearity, the ISTS was presented between 50 to 75 dB SPL at 5-dB intervals. In addition, it was verified that the coupler gain for the two hearing aid programs (noise reduction on and off) for each hearing aid were equal using the ISTS at 65 dB SPL.

Equipment

Recordings from the hearing aids were performed in an anechoic test box (Brüel and Kjær type 4232) using a PC with a Fireface UFX sound card. The recording software was Adobe Audition 3.0 (build 7283). The hearing aids were coupled to an IEC 60318-4 ear simulator (Brüel and Kjær type 4157) using a type DS 0540 earmold holder and sealed with adhesive gum. The recordings were performed with no earmoulds to optimise the seal. The microphones and amplifiers were Brüel and Kjær type 4192 measurement and reference microphone, type 2669 pre-amplifiers, and type 2692-C Nexus charge amplifier for very high input.

Signals

The three different types of signal are listed below. All signals were 60-seconds long and presented at 67 dB SPL. The frequency and modulation spectrum of the signals are plotted in Fig. 2 below.

1. The ISTS (Holube *et al.*, 2010) was included.
2. Wave files with babble noises consisting of varying amounts of speakers (2, 4, 6, 8, and 10-speakers) were created based on the ISTS. To do this, the pauses in the ISTS utterances were found. Then nine new sound tracks were

created, each with a start point at a randomly assigned pause and then looped back to the start. Then the new tracks were superimposed on the original ISTS to create wave files with the required number of babble speakers. Finally, the overall level of the wave files was adjusted to match the RMS level of the original ISTS wave file re: max.

3. An unmodulated speech-shaped noise was created by spectrally shaping the ANSI speech noise (ANSI S3.42, 1992) to match a real female speech signal (Cox *et al.*, 1987), which resembles the ISTS. The shaping was performed in 1/6-octave bands using an FIR filter with 2048 taps at sample rate 44.1 kHz.

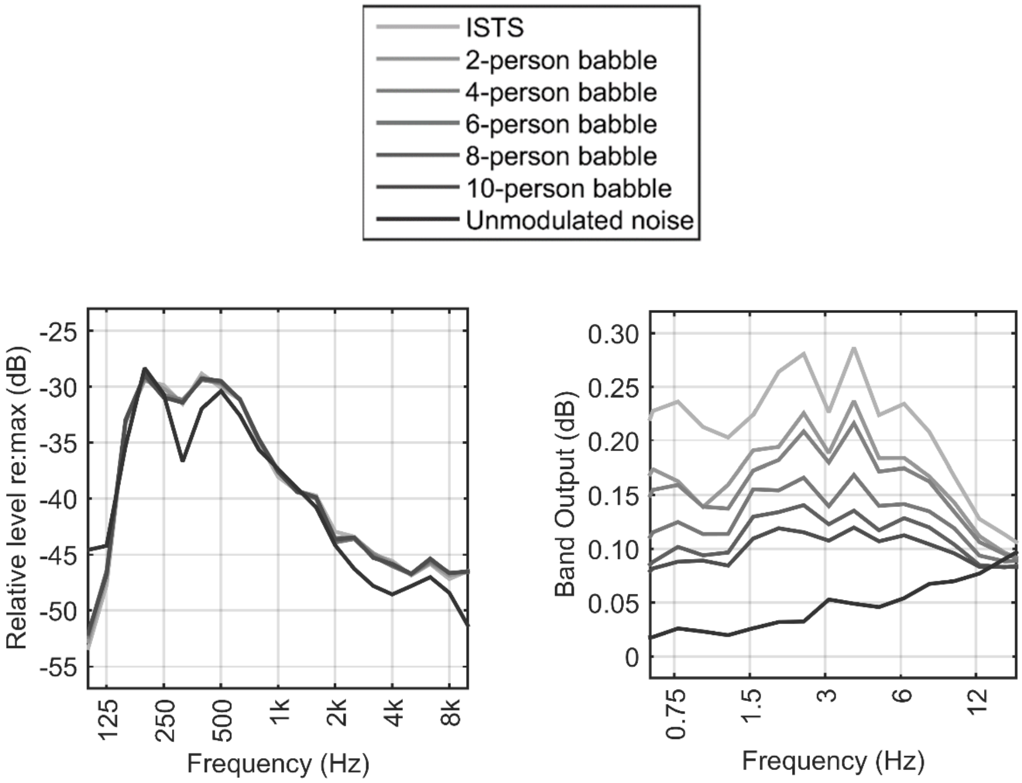


Fig. 2: The long-term average frequency (left panel) and modulation spectrum (right panel) of the signals.

Procedures

For each combination of signal, hearing aid, and program (noise reduction on and off), recordings of the hearing aid output were performed in the ear simulator. For each combination, the output was calculated in 1/3-octave bands between 250 to 8000 Hz using a 125-ms analysis window. Gain reduction was calculated as the difference in the output for the two programs, after a 35-second pre-conditioning time.

RESULTS

Recordings were made for each of the five hearing aids with noise reduction i) off, and ii) on, with the strongest setting of noise reduction. The difference in output for the noise reduction switched on and off were used to calculate the long-term average gain reduction due to the noise reduction (Fig. 3). The results presented in this manuscript were carefully cross-verified using similar signals on different software and hardware (the noise reduction measurement functionality of the OTOsuite software with the Aurical HIT box).

As expected, none of the noise reduction algorithms applied gain reduction to the ISTS signal. All the noise reduction systems applied the most gain reduction to the unmodulated noise. The maximum amount of gain reduction in a given frequency area varied from 7 dB to more than 15 dB.

Two of the hearing aids (A and D) applied a graduated response and applied most gain reduction to 10-person babble and least noise reduction to 2-person babble. The remaining three hearing aids seemed to apply the same graduated response, but only became active when the babble consisted of 8 or 10 speakers. (B, C, and E).

DISCUSSION

For the standardized test signals, none of the hearing aids reduced gain for the ISTS, suggesting that all noise reduction systems could appropriately detect speech, at least in quiet. All hearing aids applied the most gain reduction for the unmodulated noise. There were large differences in how much gain reduction was applied across frequency, with some only applying 7 dB in certain frequencies and others applying more than 15 dB gain reductions across the whole frequency spectrum. The range of differences is consistent with previous studies (Brons *et al.*, 2013; 2014; Hoetlink *et al.*, 2009; Smeds *et al.*, 2010).

For the babble noise signals, some of the hearing aids applied very little gain reduction (< 5 dB), even for a 10-person babble. Two of the hearing aids (A and D) applied a graduated approach and began to reduce gain for the 2-person babble and gradually increase the amount of noise reduction as the number of speakers increased. The other three hearing aids (B, C, and D) did not become active until there were 8 or 10 persons present in the babble.

The matter of how much gain reduction is appropriate has not been established. In addition, potential confounders include the level of the signal across frequency, and the hearing aid user’s hearing thresholds, potential cognitive factors, as well as how the algorithm is implemented (Arehart *et al.*, 2015). Generally speaking if a noise reduction algorithm applies too much gain reduction then it may reduce audibility for speech. If too little gain reduction is applied, then the effect of noise reduction will not be audible to the user, and some of the positive benefits of noise reduction, such as improved listening comfort and reduced annoyance, will not be as optimal as they could be. There is little published evidence of how much gain reduction is appropriate, but Brons *et al.* (2013; 2014) have demonstrated that the differences in how noise reduction is applied can impact speech intelligibility, noise annoyance and listener preference.

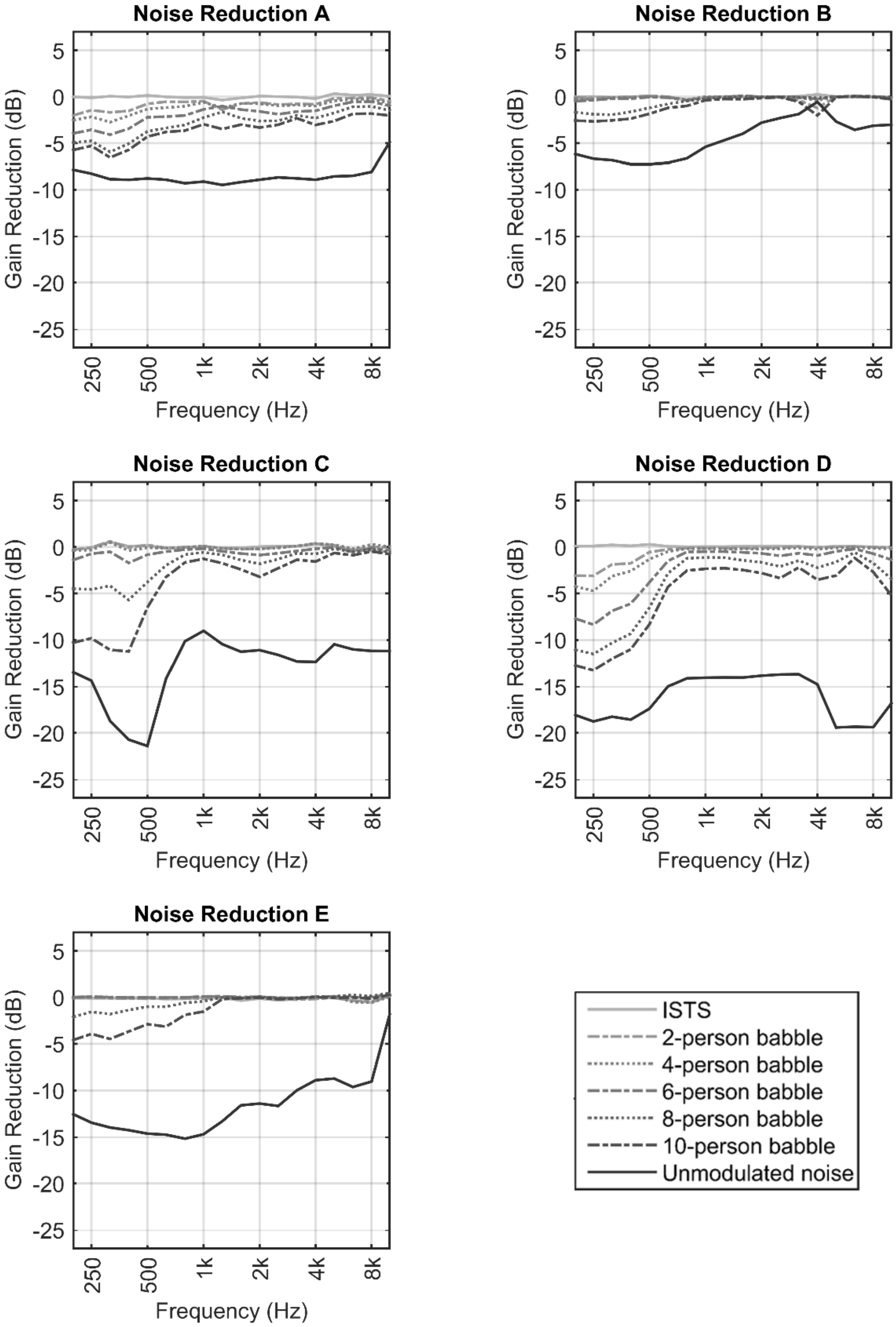


Fig. 3: Long-term average gain reduction in one-third octave bands for each of the seven test signals used. Every line represents the difference between noise reduction on and off averaged over 25 seconds after a 35-second pre-conditioning time.

This study demonstrates that there are differences among hearing aids in terms of which listening situations that the noise reduction activates. The range of listening situations explored by this experiment was quite broad, varying from a small group (two speakers) to a fairly large group (ten people speaking at the same time). There were considerable differences in how active the noise reduction systems were in these situations. Listening in groups is an important need for hearing aid users (Kochkin, 2010), yet a noise reduction algorithm may not be active in the situations that the user or audiologist expect it to be.

Other authors have raised the need for standardized measurements to describe how these noise reduction systems work (Bentler and Chiou, 2006; Brons *et al.*, 2013; 2014; Hoetink *et al.*, 2009; Smeds *et al.*, 2010). We suggest that such a test battery should include a range of realistic, yet well-described test signals. This would help the audiologists discern when the noise reduction systems are active in order to help them select an appropriate hearing aid for the listening needs of the user, and possibly to help fine-tune the hearing aid. If babble noise is included in a test battery, it is important to specify how it is compiled.

ACKNOWLEDGMENTS

Thanks to research engineer Martin Rung from GN Netcom for use of his Matlab script to frequency-match signals, and Ellen Raben Pedersen (Ph.D. candidate, University of Southern Denmark) for her Matlab script to plot frequency and modulation spectra. Thanks to Professor Bert de Vries (GN ReSound and University of Technology, Eindhoven) for helpful discussions during the formulation of the abstract.

REFERENCES

ANSI S3.42 (**1992**). *American National Standard: Testing Hearing Aids with a Broad-Band Noise Signal*. New York: American National Standards Institute.

Arehart, K., Souza, P., Kates, J., Lunner, T., and Pedersen, M.S. (**2015**). “Relationship among signal fidelity, hearing loss, and working memory for digital noise suppression,” Ear Hearing, **36**, 505-516.

Bisgaard, N., Vlaming, M.S.M.G., and Dahlquist, M. (**2010**). “Standard audiograms for the IEC 60118-15 measurement procedure,” Trends Amplif., **14**, 113-120.

Bentler, R. and Chiou, L.-K. (**2009**). “Digital noise reduction: An overview,” Trends Amplif., **10**, 67-82.

Brons, I., Houben, R., and Dreschler, W.A. (**2013**). “Perceptual effects of noise reduction with respect to personal preference, speech intelligibility, and listening effort,” Ear. Hearing, **34**, 29-41.

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**, 1-10.

Cox, R.M., Alexander, G.C., and Gilmore, C. (**1987**). “Development of the Connected Speech Test (CST),” Ear. Hearing, **8**, 199S-126S.

Hoetink, A.E., Körössy, L., and Dreschler, W.A. (**2009**). "Classification of steady state gain reduction produced by amplitude modulation based noise reduction in digital hearing aids," Int. J. Audiol., **48**, 444-455.

Holube, I., Fredelake, S., and Vlamming, M. (**2010**). "Development and analysis of an international speech test signal (ISTS)," Int. J. Audiol., **49**, 891-903.

Keidser, G., Dillon, H., Flax, M., Ching, T., and Brewer, C. (**2011**). "The NAL-NL2 prescription procedure," Audiol. Res., **1S**, e24.

Kochkin S. (**2010**). "MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing," Hearing J., **63**, 19-32.

Smeds, K., Bergman, N., Hertzman, S., and Nyman, T. (**2010**). "Noise reduction in modern hearing aids – long-term average gain measurements using speech," in Proceedings of ISAAR 2009: *Binaural Processing and Spatial Hearing*. 2nd International Symposium on Auditory and Audiological Research. Elsinore, Denmark. Eds. J.M. Buchholz, T. Dau, J. Christensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2 (The Danavox Jubilee Foundation, Copenhagen), pp. 289-298.

Promoting off-axis listening and preserving spatial cues with Binaural Directionality II

Tobias Piechowiak[1,*], Jesper Udesen[1], Kimi Moeller[1], Fredrik Gran[1], and Andrew Dittberner[2]

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

[2] *GN ReSound North America, Glenview, IL, USA*

Hearing in complex acoustic scenes is a challenge for hearing-impaired persons that often persists after amplification is applied even when fitted bilaterally. From a hearing aid (HA) processing point of view there can be several reasons for this. First, directional filters in a symmetric fitting can help increase signal-to-noise ratio for on-axis signals-of-interest. However, they also can render off-axis signals inaudible. Second, HA microphone location can degrade spatial cues that are important for localization and thus listening in complex acoustic scenes. Third, amplification itself, when applied independently at both ears, can affect spatial cues, mainly interaural-level-differences. Finally, changing acoustic scenes might require changing processing. In order to overcome some of these challenges we propose a bilateral fitting scheme that can be symmetric or asymmetric depending on the acoustic scene. The respective HA processing modes can be (a) omnidirectional, (b) directional, or (c) directional with preservation of spatial cues. In this study it was shown that asymmetric fitting helps improve off-axis audibility when prioritized while it provides natural sound and decreases listening effort for symmetric fitting in situations when audibility is not the main focus.

INTRODUCTION

Hearing aids' primary objective is to restore audibility of desired target signals. This is usually done by applying frequency dependent gains on the input signal. However, it does not solve the problem hearing aid users have in background noise (Kochkin, 2000). Directional filters can help alleviate some of the problems by supressing distracting sounds from a certain direction. However, there are some drawbacks to this approach. Generally, benefits of directional filters are most salient in laboratory investigations and studies have failed to establish comparable benefits in "real-world environments" (e.g., Walden *et al.*, 2000; Nielsen, 1973). There might be several reasons for this. First, directional benefit is largest in anechoic environments but dimishes relatively quickly when reverberation is added **?**. This is also the case when multiple sound sources are presented creating a more diffuse sound field that decreases the performance of a directional filter. Second, directional filters create worse "off-axis" listening. Sounds from the sides and rear are reduced in amplification, users may report feelings of

*Corresponding author: tpiechowiak@gnresound.com

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

"tunnel hearing" being cut off from their surroundings and not being able to re-orientate their focus to other, new target signals. This is especially true when target/interferer sources are moving as is usually the case in real acoustic environments. Third, and very related to the previous point, is that directional filters and microphone placement alone on the device itself can deteriorate spatial cues and can lead to a decrease in loclalization performance (Van den Bogaert *et al.*, 2006; Keidser *et al.*, 2006). It has also been shown that the usage of spatial cues are the main key to spatial unmasking which constitutes an important part of speech understanding in complex environments (Edmonds and Culling, 2005, 2006). Another reason, more of psychological nature, is the fact that only a small fraction (30%) of hearing aid users with switchable hearing aids use this feature (Cord *et al.*, 2004) although it could be beneficial. The reasons for this neglect is that in most situations the user does not know when to switch or does not know in which situations the respective programs could be of benefit. In order to alleviate some of the listed challenges above we introduced a asymmetric fitting scheme that adaptively switches between different microphone processing modes which in the following will be called Binaural Directionality II (BDII). Its ultimate rationale is to increase audibility in challenging situations when the target signal is well defined while maximizing spatial awareness in quieter situations and when the target signal is not clearly defined.

CONCEPT

There are four mircophone processing modes for BDII:

Omnidirectional (Omni) : The front microphone is used as input.

Fixed Directionality (Dir) is a fixed 2-mic beamformer that maximizes DI.

Pinna Restoration (PR) mimics the open-ear response of a KEMAR ear at the coupler by a fixed 2-mic beamformer. Its rationale is to deteriorate spatial cues less than with classical microphone modes like omnidirectional mode and to provide better natural sound.

Bilateral Compression (BC) aims to preserve natural interaural-level-differences and thus mimics the head shadow effect. This is done with the help of ear-to-ear synchronization and the exchange of envelope power estimates across ears.

The functionality of the different microphone modes is illustrated in the intensity plots in Fig. 1. The open ear response shows the characteristic "pinna valley" around 120° and between $5-7$ kHz. In this frequency range scattering from the pinna mainly takes place. Pinna Restoration is a 2-mic beamformer that optimizes towards a KEMAR's coupler response. This is shown in the plot where the "pinna valley" is clearly indicated. In contrast the omnidirectional mode is listening "sideways" having its direction of maximum sensitivity around 100° across all frequencies. Fixed directionality has a hypercardioid shape with maximum sensitivity at 40° and its "null" at around 120°.

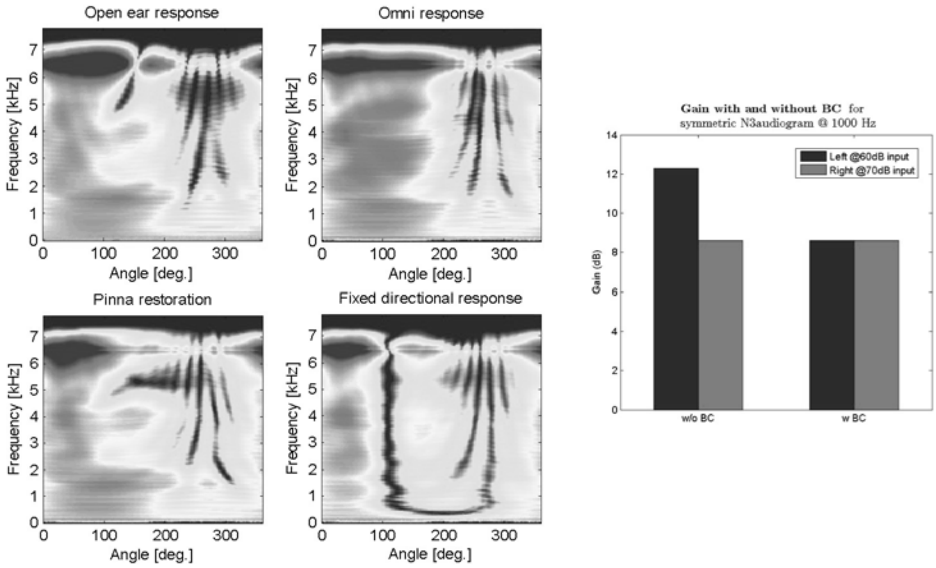


Fig. 1: Left panel: Intensity plots showing the open ear, omnidirectional, Pinna Restoration (PR) and Fixed Directionality (Dir) responses. The hearing aid was placed on KEMAR's left ear. Right panel: Gain, given with and without Bilateral Compression for a typical audiogram

The right panel shows a working example of BC. Insertion gain is shown when BC is switched on and when it is switched off. Hearing loss at 1000 Hz was symmetric and difference in input power (ILD) was 10 dB. Without BC the ILD will *decrease* from 10 dB to 6 dB while the original ILD is preserved when BC is switched on. For Binaural Directionality II, PR and BC are only applied jointly. PR serves as a front-end for BC in order to achieve a reliable ILD estimate (close to the ear-drum). As stated before in the Binaural Directionality concept, microphone modes are changed according to the acoustic environment. Figure 2 shows the switching strategy of BDI and BDII.

Concept	Acoustic scene	Left device		Right device	
		BD I	BD II	BD I	BD II
Spatial Cue Preservation	Quiet w/o speech	Omni	PR + BC	Omni	PR + BC
	Quiet w/ speech	Omni	PR + BC	Omni	PR + BC
Better-ear effect	Noise w/o speech	Omni		Dir	
	Noise w/ speech	Omni		Dir	
Speech enhancement	Noise w/ speech dominance in front of listener	Dir		Dir	
	Noise w/ speech dominance at Right device	Dir		Omni	

Fig. 2: Switching strategy for Binaural Directionality I/II depending on the respective acoustic scene.

There are three different concepts that are applied here. Spatial cue preservation is the new concept in Binaural Directionality II. It is applied when speech is present in a relative quiet environment. In BDI omnidirectional mode is used instead.

EXPERIMENTS

Subjects and fitting

The same 11 hearing-impaired subjects participated in all the experiments. They were fitted with open domes of various sizes. The mean audiograms of the subjects are shown in Fig. 3.

Corresponding gains were obtained by using GN Resound's *Audiogram+* fitting rule. After applying prescribed gains, fine tuning was performed on each subject. No real-ear measurement (REM) was performed.

Localization

Subjects were placed onto a chair inside the array of 12 speakers facing the speaker at 0° as shown in Fig. 5. The stimulus was a 1-sec broadband telephone signal in quiet. It was presented at 70 dB SPL at the position of the head. The task was to identify the loudspeaker from which sound emanated by naming the number of the respective speaker. With each subject two sessions were held two weeks apart. This was done for testing test-retest reliability and also to assess if learning took place for localization during the time interval between visits (see section 3.4). Front-back confusions and the root-mean-square (RMS) error were used for data analysis.

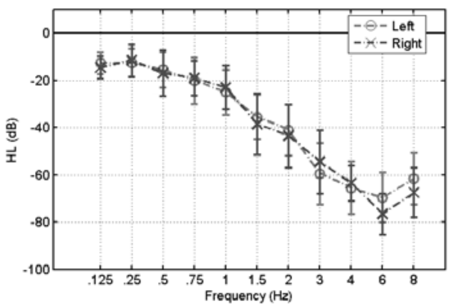


Fig. 3: Audiogram of the participating subjects.

Speech intelligibility

The setup for speech intelligibility measurements is shown in Fig. 4. It was inspired by the work of Hornsby and Ricketts (2007). The rationale was that, depending on the listening situations, switching between different microphone modes would yield the best performance compared to a single symmetric mode. In this study symmetric Omni mode (Omni/Omni) served as reference. HINT sentence material was used to determine speech reception thresholds (SRTs) with an adaptive 2-up 1-down procedure.

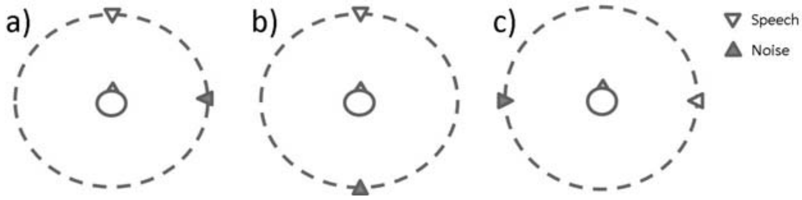


Fig. 4: The three different acoustic environments which are used in the intelligibility test: a) speech from front, noise from side; b)speech from front, noise from behind; c) speech from side, noise from side.

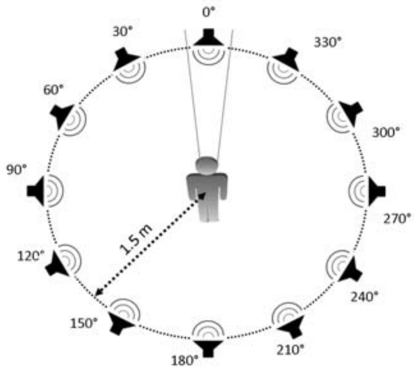


Fig. 5: Setup of the localization experiment.

Field trial

In order to evaluate efficiency of BDII ("real-world benefit") a simple crossover design was used. Five of the 11 subjects were initially fitted with BDI activated. The remaining subjects were fitted with BDII activated. The SSQ questionnaire (Gatehouse and Noble, 2004) was used to evaluate naturalness of sound, listening effort and orientation ability:

Q1: You are talking with one other person and there is a TV in the same room. Can you follow what the person you are talking to says? (Anchors: Not at all - Perfectly)

Q2: You are in a group of about five people, sitting around a table. It is an otherwise quiet place. You can see everyone else in the group. Can you follow the conversation?

Q3: You are in a group of about five people in a busy restaurant. You can see everyone else in the group. Can you follow the conversation? (Anchors: Not at all - Perfectly)

Q4: You are outside. A dog barks loudly. Can you tell immediately where it is, without having to look? (Anchors: Not at all - Perfectly)

Q5: Can you tell how far away a bus or truck is, from the sound? (Anchors: Not at all - Perfectly)

Q6: Can you tell from the sound whether a bus or truck is coming towards you or going away? (Anchors: Not at all - Perfectly)

Q7: Do everyday sounds that you can hear easily seem clear to you (not blurred)? (Anchors: Not at all - Perfectly)

Q8: Do you have to concentrate very much when listening to someone or something? (Anchors: Concentrate Hard - Not need to concentrate)

Q9: When you are the driver in a car can you easily hear what someone is saying who is sitting alongside you? (Anchors: Not at all - Perfectly)

Q10: When you are a passenger can you easily hear what the driver is saying sitting alongside you? (Anchors: Not at all - Perfectly)

After a two-week use, devices were re-programmed to the alternative setting and again were worn for a period of another two-weeks. At each visit subjects were instructed to complete the SSQ questionnaire. Individual subject audiograms were used to define *Audiogram+* target gains. Only one program was made available at a time (either BDI or BDII) so that subjects could not unintendedly change the settings.

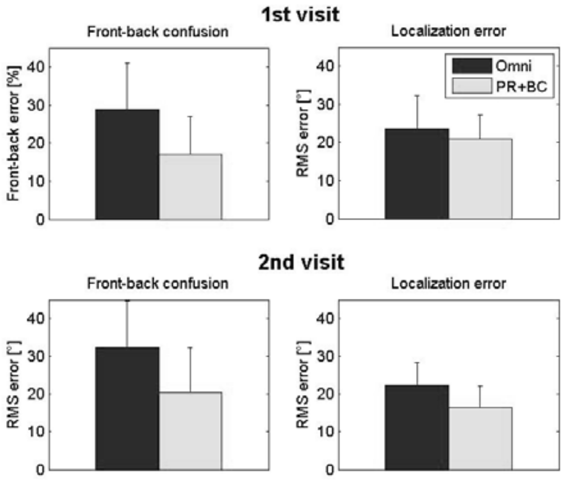


Fig. 6: Upper panel: Front-back confusion and RMS error at first visit. Lower panels: Front-back confusion and RMS error at second visit.

RESULTS

Localization

Results for the localization experiment are shown in Fig. 6. A two-tailed *t*-test revealed that front-back confusions were significantly lower for BDII than for BDI at both visits (first visit: $t(10) = 2.5, p < 0.05$, second visit: $t(10) = 2.9, p < 0.05$). However, there was no significant difference between front-back confusions between first and second visit for either program. No difference in RMS errors was found, neither between programs nor across visits.

Speech intelligibility

Speech intelligibility results are shown in Fig. 7. Directional benefit was defined as the difference between SRTs in the respective microphone modes and the omnidirectional mode which served as reference. Generally, for speech coming from the front, all directional microphone modes provided higher SRT than symmetric omnidirectional mode. The situation was different when speech was presented from the side. Here, Dir/Dir was significantly worse than the reference while Dir/Omni performed significantly better. The microphone modes that provided benefit in all three conditions were those that are automatically chosen by Binaural Directionality in the respective environment.

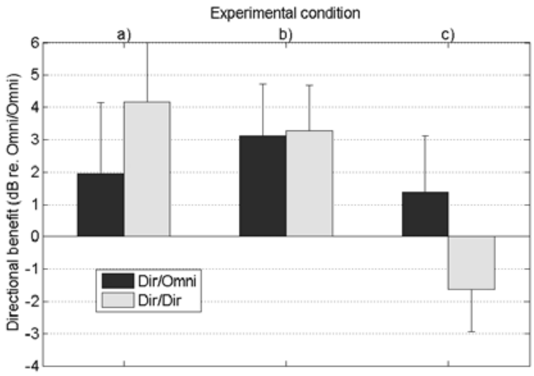


Fig. 7: Speech reception thresholds for the three experimental conditions shown in section 3.3

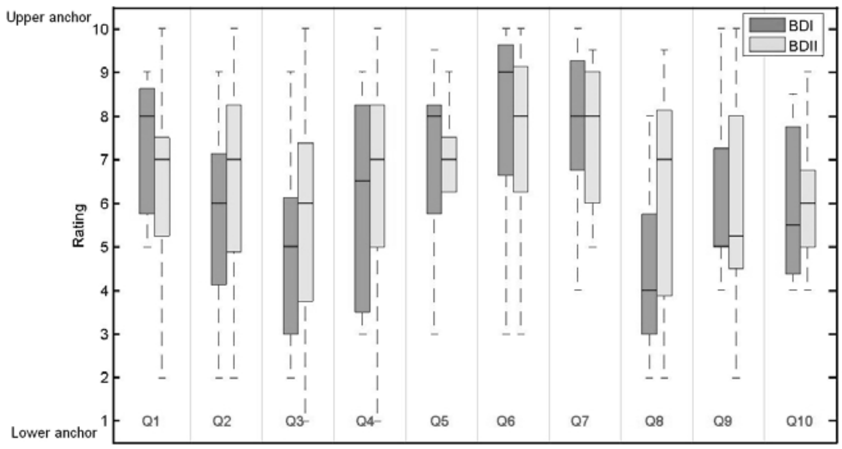


Fig. 8: Results from the field trial are shown for the two program modes BDI and BDII. The box plots show the median and the 25% and 75% whiskers. Questions(Q1 - Q10) are indicated on the *x*-axis below each pair of data.

Field trial

Results from the field trial are shown in Fig. 8. Questions are indicated on the *x*-axis. The *y*-axis shows the 10-point rating scale having the low anchor at the bottom (e.g., *Not at all*) and the high anchor at the top (e.g., *Perfectly*). The rating in question 8 (Q8) was shown to be significantly different for BDI and BDII (two-tailed *t*-test:

$t(10) = 2.5, p < 0.05$). Thus, listening is perceived as more effortless in BDII than in BDI.

CONCLUSIONS

Binaural Directionality II is a concept that switches hearing aid microphone input modes depending on the acoustic scene. It also tries to preserves spatial cues in less challenging (meaning good SNR) situations. It was found that BDII (a) provided higher off-axis intelligibility, (b) gave better localization, (c) required less effort listening in real-world acoustic situations, and (d) tended to be preferred in multi-talker situations.

REFERENCES

Cord, M.T., Walden, R.K., Walden, B.E., and Dyrlund, O. (**2004**). “Relationship between laboratory measures of directional advantage and everyday success with directional microphone hearing aids,” J. Am. Acad. Audiol., **15**, 353-364.

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.

Edmonds, B.A. and Culling, J.F. (**2006**). “The spatial unmasking of speech: Evidence for better-ear listening,” J. Acoust. Soc. Am., **120**, 1539-1545.

Gatehouse, S. and Noble, W. (**2004**). “The Speech, Spatial and Qualities of Hearing Scale (SSQ),” Int. J. Audiol., **43**, 85-99.

Hoffmann, H.W., Trine, T.D., Buckley, K.M., and van Tasell, D.J. (**1994**). “Robust adaptive microphone array processing for hearing aids: Realistic speech enhancement,” J. Acoust. Soc. Am., **96**, 759-770.

Hornsby, B.W.Y. and Ricketts, T.A. (**2007**). “Effects of noise source configuration on directional benefit using symmetric and asymmetric directional hearing aid fittings,” Ear Hearing, **28**, 177-186.

Keidser, G., Rohrseitz, K., Dillon, H., Hamacher, V., Carter, L., Rass, U., and Convery, E. (**2006**). “The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers,” Int. J. Audiol., **45**, 563-579.

Kochkin, S. (**2000**). “Why my hearing aids are in the drawer: The consumer’s perspective.” Hear. J., **53**, 34-42.

Nielsen, H.B. (**1973**). “A comparison between hearing aids with directional microphone and hearing aids with conventional microphone,” Scand. Audiol., **2**, 173-176.

Van den Bogaert T., Klasen, J.K., Van Deun, L., and Wouters, J. (**2006**). “Horizontal localization with bilateral hearing aids: Without is better than with,” J. Acoust. Soc. Am., **119**, 515-526.

Walden, B.E., Surr, R.K., Cord, M.T., Edwards, B.E., and Olson, L. (**2000**). “Comparison of benefits provided by different hearing aid technologies,” J. Am. Acad. Audiol., **11**, 540-560.

Effects of dynamic-range compression on temporal acuity

Alan Wiinberg[1,*], Morten Løve Jepsen[2], Bastian Epp[1], and Torsten Dau[1]

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

[2] *Hearing Loss Compensation, Department of Electronics and Audiology, Widex A/S, Lynge, Denmark*

Some of the challenges that hearing-aid listeners experience with speech perception in complex acoustic environments may originate from limitations in the temporal processing of sounds. To systematically investigate the influence of hearing impairment and hearing-aid signal processing on temporal processing, temporal modulation transfer functions (TMTFs) and "supra-threshold" modulation-depth discrimination (MDD) thresholds were obtained in normal-hearing (NH) and hearing-impaired (HI) listeners with and without wide-dynamic range compression (WDRC). The TMTFs were obtained using tonal carriers of 1 and 5 kHz and modulation frequencies from 8 to 256 Hz. MDD thresholds were obtained using a reference modulation depth of −15 dB. A compression ratio of 2:1 was chosen. The attack and release time constants were 10 and 60 ms, respectively. For both carrier frequencies the TMTF thresholds decreased with the physical compression of the modulation depth due to the WDRC. Indications of reduced temporal resolution in the HI listeners were observed in the TMTF patterns for the 5-kHz carrier. Significantly higher MDD thresholds were found for the HI group relative to the NH group. No relationship was found between the MDD thresholds and the TMTF threshold. These findings indicate that the two measures may represent different aspects of temporal processing.

INTRODUCTION

Modern hearing aids use wide-dynamic range compression (WDRC) as an amplification strategy to compensate for loudness recruitment in sensorineural hearing-impaired (HI) listeners (Ricketts and Bentler, 1996). This is typically accomplished by providing a higher gain for low-level sounds than for high-level sounds (Jenstad *et al*., 2000). Besides restoring audibility of a wide dynamic range of sound levels, fast-acting WDRC reduces the depth of low-frequency amplitude modulations and distorts the temporal envelope waveform. The degree of the temporal distortion and reduction in modulation depth depends on the settings of WDRC system (Stone and Moore, 1992).

Temporal changes in the envelope (amplitude modulations) of speech convey linguistic information about consonant manner and voicing as well as prosodic cues

*Corresponding author: alwiin@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

(e.g., Rosen, 1992). Amplitude modulations have been shown to contribute significantly to high speech recognition (Shannon *et al.*, 1995; Stone *et al.*, 2011). Hence, modulation depth reduction and temporal distortion of speech by fast-acting WDRC may impair speech recognition (Stone and Moore 2003).

Temporal modulation transfer functions (TMTFs) and "supra-threshold" modulation-depth discrimination (MDD) have previously been used as measures of temporal processing (Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001; Lee and Bacon, 1997). In a TMTF experiment, amplitude modulation detection thresholds are measured as a function of modulation frequency. In an MDD experiment the just-noticeable increase in modulation depth from a (supra-threshold) standard modulation depth is measured as a function of modulation frequency.

In the present study, the influence of hearing impairment and WDRC on temporal envelope encoding was investigated. TMTF and MDD thresholds were obtained in normal-hearing (NH) and hearing-impaired (HI) listeners with and without WDRC. Tonal carriers were used. Compared to noise carriers, tonal carriers have the advantage that they do not contain any intrinsic modulations which may mask, and thereby limit, the detectability of the imposed modulation (Dau *et al.*, 1997; 1999). However, the disadvantage is that the imposed modulation introduces spectral sidebands which may be perceived as separate tones, if the sidebands are sufficiently far in frequency from the carrier frequency (Kohlrausch *et al.*, 2000). This is not a problem for broadband carriers as the modulation sidebands generally fall within the spectrum of the carrier and are therefore masked. Hence, results obtained with tonal carriers may provide a better measure of temporal resolution of the auditory system than modulation results using noise carriers, provided that the modulation frequency is within the range where spectral resolution does not play a major role.

METHOD

Listeners

Nine adult listeners (5 males and 4 females) with normal hearing were tested, with ages ranging from 21 to 50 years. All had absolute thresholds better than 20 dB hearing level at all audiometric frequencies. Seven adult listeners (4 males and 3 females) with mild to moderate/severe sensorineural hearing loss were tested, with ages ranging from 50 to 73 years. Their absolute thresholds for the test ear, measured using conventional audiometry, are shown in Fig. 1.

Procedure

The TMTF was measured using an adaptive three-alternative forced-choice 3-down 1-up procedure tracking the 79.4% point on the psychometric function. The gated carrier was unmodulated in two of the intervals and modulated in the other; the listeners had to select the interval containing the modulated carrier. The step size was 5 dB down to the reversal and 2 dB thereafter. Each run was terminated after seven reversals, and the threshold estimate for that run was computed as the mean value of 20 log(m) at the last six reversals (where m is the modulation index). Three runs were obtained for each condition.

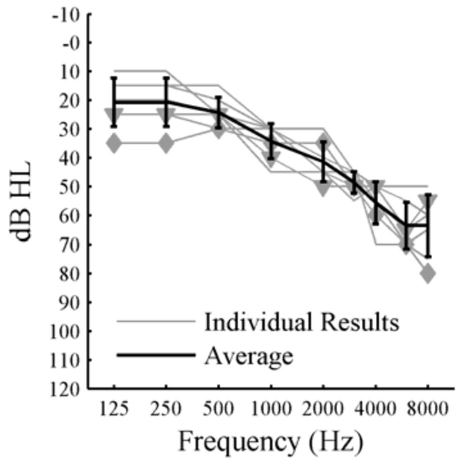


Fig. 1: Individual and average absolute thresholds for the tested ear of the hearing-impaired (HI) listeners, measured using conventional manual audiometry, and expressed in dB HL. The error bars represents ±1 standard deviation (SD). The thresholds for two of the listeners are displayed with symbols as their TMTF results at 1 kHz differ from the others.

The procedure was the same for measuring the MDD as the TMTF procedure, except that the carrier was modulated with a constant standard modulation depth (m_s) in two of the intervals and modulated with a higher modulation depth ($m = m_s + \Delta m$) in the other; the listeners had to select the interval containing the carrier with the highest modulation depth. The modulation depth of the target was adjusted in steps of 2 dB, and thereafter 1 dB.

Stimuli

In both experiments, tonal carriers and modulators were used. The frequency of the carrier was either 1 kHz or 5 kHz. The modulation frequencies were 8, 16, 32, 64, 128, and 256 Hz in the unprocessed condition and 8, 16, 32 Hz in the WDRC condition. The presentation intervals were 600 ms in duration with 30-ms rise and fall times. The pause between presentations within a trial was 500 ms. The overall level of the presentations was the same, regardless of modulation depth and WDRC processing. For the NH listeners, the presentation level was 65 dB SPL. For the HI listeners, the presentation level was increased according to a NAL-R(P) frequency dependent prescription gain that depends on their individual audiometric thresholds (Byrne *et al.*, 1990). The standard modulation depth m_s was 0.18 (−15 dB). The frequency response of the headphones at the eardrum was estimated using an ear simulator.

All signals were generated digitally with a sampling rate of 44.1 kHz and were presented to the listeners via an RME soundcard and DT 770 PRO Beyerdynamic headphones. The listeners were seated in sound-attenuating booth and the sound was played monaurally to the audiometric best ear.

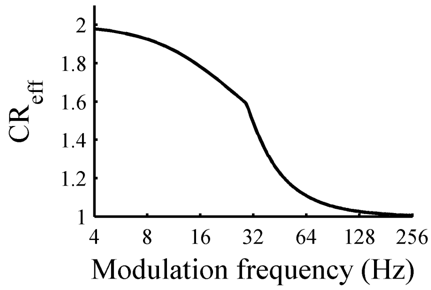


Fig. 2: The effective compression ratio (CR_{eff}) for the used WDRC system as function of modulation frequency (e.g., Stone and Moore, 1992).

WDRC system

The WDRC system was implemented in Matlab. The envelope of the incoming signal was extracted using a Hilbert transform and smoothed using single-pole low-pass filters. The smoothed envelope was converted to dB SPL and was input to a broken-stick gain function (linear gain below the compression kneepoint and constant compression above). The resulting gain was applied to the input. The static compression ratio was set to 2:1. The attack and release time constants were 10 and 60 ms, respectively. The compression kneepoint was set to 20 dB SPL. The effective compression ratio for this WDRC system as a function of modulation frequency is shown in Fig. 2.

RESULTS

The TMTF thresholds for the NH listeners are shown in Fig. 3. The average data (without compression) for both carrier frequencies are consistent with earlier work (Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001). For the 1-kHz carrier (top panel), the threshold decreases with increasing modulation frequency above 64 Hz. This reflects the detection of spectral sidebands, as shown in Kohlrausch *et al.* (2000). For the 5-kHz carrier (bottom panel), the threshold tends to increase slightly as the modulation frequency is increased from 128 Hz to 256 Hz. For both carrier frequencies, the modulation detection thresholds are increased when compression is applied.

The TMTF thresholds for the HI listeners are shown in Fig. 4. The data (without compression) for both carrier frequencies are consistent with earlier work on HI listeners (Moore and Glasberg, 2001). For the 1-kHz carrier (top panel), the threshold decreases with increasing modulation frequency for five of the seven listeners above 128 Hz. The remaining two listeners (marked by symbols) showed no corresponding decrease in threshold at the high modulation frequencies, probably because of reduced frequency selectivity. For the 5-kHz carrier, the thresholds tend to increase slightly with increasing modulation frequencies above 64 Hz. A two-way analysis of variance (ANOVA) indicated significantly higher thresholds for the NH listeners relative to the HI listeners for the 5-kHz carrier [$F(1,60) = 4.7$, $p = 0.03$]. The 95% confidence interval ranges from −0.2 dB to −3.6 dB. However, no significant differences in the TMTF thresholds were observed for the 1-kHz carrier.

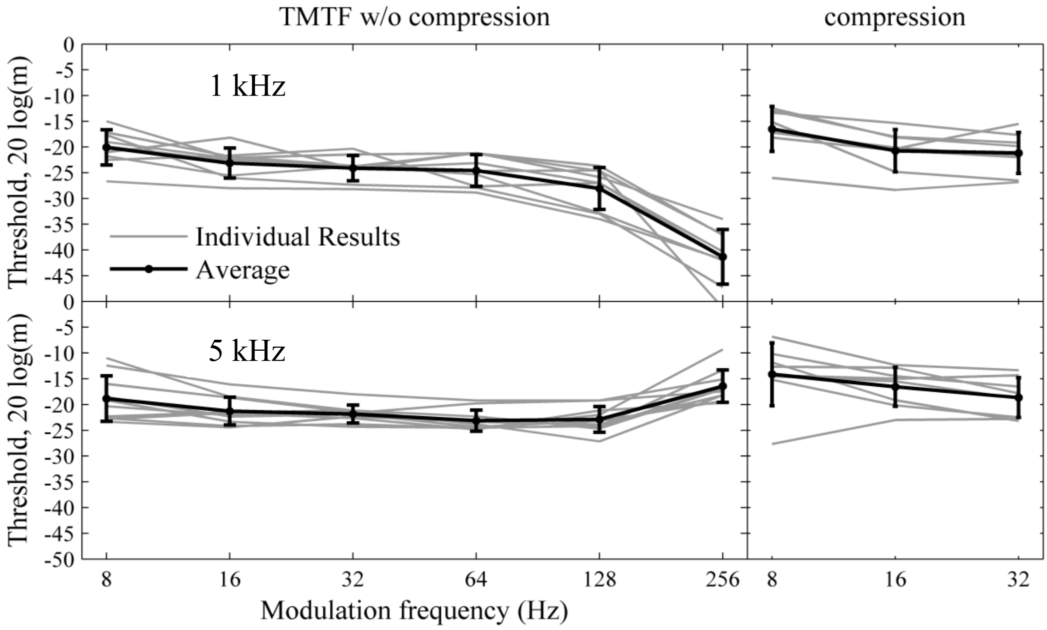


Fig. 3: Individual and average TMTF results for the NH listeners. The error bars represents ±1 SD. The modulation detection threshold ($20 \log m$) is shown as a function of modulation frequency. The upper panels show the results for the 1-kHz carrier, and the lower panels show the results for the 5-kHz carrier. The left panels show the results obtained without WDRC, and the right panel show the results obtained with WDRC.

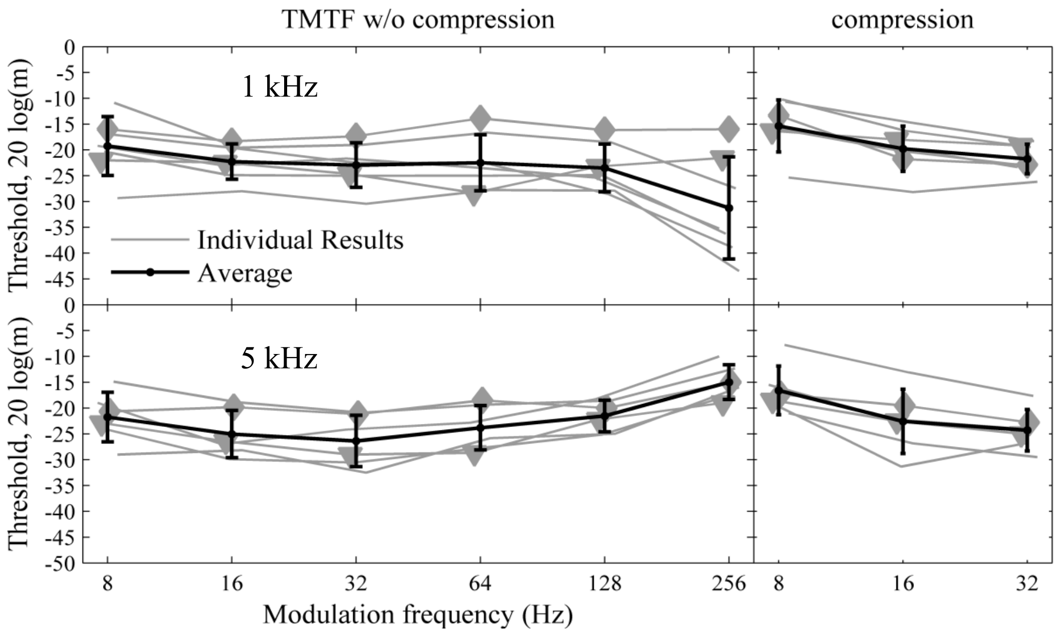


Fig. 4: Individual and average TMTF results for the HI listeners. Otherwise as in Fig. 3.

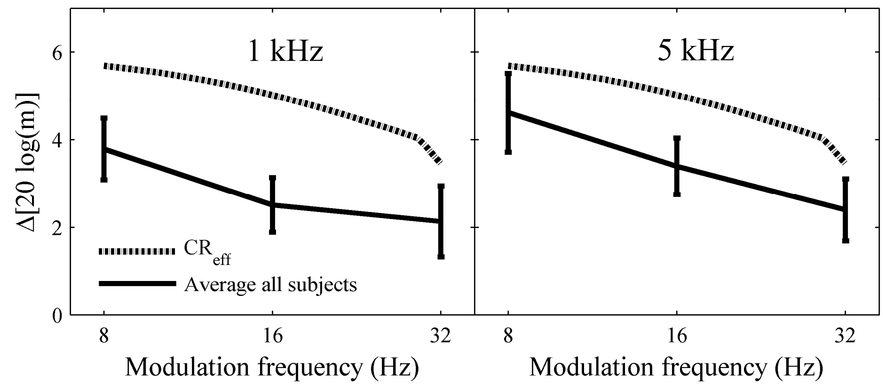


Fig. 5: The effect of WDRC on the modulation detection thresholds. The left panel is the result for the 1-kHz carrier and the right panel is the result for the 5-kHz carrier. For comparison, the black dotted curves display the physical reduction of the modulation depth from Fig. 2.

The effect of WDRC on the modulation detection thresholds is shown in Fig. 5. The effect is computed as the threshold for the non-compressed condition, in dB, subtracted from the threshold for the compressed condition, in dB (e.g., Edwards, 2004). It can be seen that the change in the modulation detection threshold due to WDRC is consistent with the physical reduction of the modulation depth.

The MDD thresholds for the NH listeners are shown in Fig. 6. The thresholds obtained without WDRC are consistent with those found by previous researchers (Lee and Bacon, 1997). The MDD threshold is rather insensitive to modulation frequency. The performance is not affected by compression. Hence, the reduction of the standard modulation depth due to compression does not seem to affect the discrimination performance.

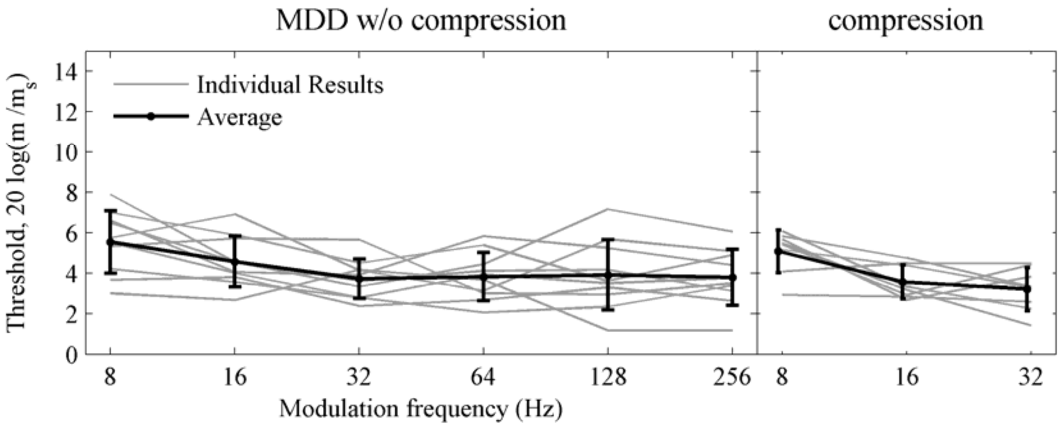


Fig. 6: Individual and average MDD results for the NH listeners for the 1-kHz carrier. The error bars represents ±1 SD. The modulation discrimination threshold (20 log (m/m_s)) is plotted as function of modula-tion frequency. The left panels show the results obtained without WDRC, and the right panel are the results obtained with WDRC.

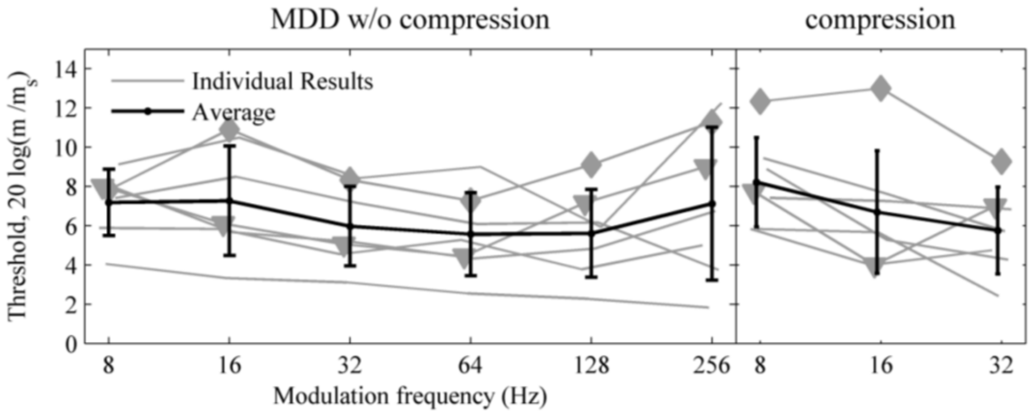


Fig. 7: Individual and average MDD results for the HI listeners for the 1-kHz carrier. Otherwise as in Fig. 6.

The MDD thresholds for the HI listeners are shown in Fig. 7. Larger individual differences across the HI listeners can be observed relative to the NH listeners. A two-way ANOVA indicated significantly lower MDD thresholds for the NH listeners than for the HI listeners [$F(1,60) = 18.2$, $p < 0.0001$]. The 95% confidence interval ranges from 1.0 dB to 2.7 dB.

DISCUSSION

The results obtained in the HI group revealed larger individual differences across listeners in the ability to discriminate amplitude changes in the envelope as well as in terms of spectral resolvability of the sidebands despite similar pure-tone sensitivity. The modulation detection thresholds were, on average, significantly lower for the HI listeners at the 5-kHz carrier in the "flat region" of the TMTF pattern. Thus, the HI listeners seem to have an improved ability to detect amplitude modulations. In contrast, higher MDD thresholds were observed in the HI group relative to the NH group. Hence, the ability to process amplitude changes of the envelope for a given modulation frequency seems to be reduced in the HI listeners. No significant correlation between the MDD thresholds and the TMTF thresholds was found, indicating that the two measures may represent different aspects of temporal processing.

Temporal resolution derived from TMTFs is often characterized by a time constant, τ, defined as $(2\pi f_c)^{-1}$, where f_c is the frequency at which sensitivity has declined by 3 dB relative to that measured for low modulation frequencies. A decline in sensitivity is thought to reflect a limitation in resolving fast amplitude modulations in the auditory system (Kohlrausch *et al.*, 2000). Such a measure cannot be applied to the data for the 1-kHz carrier due to resolved spectral sidebands at high modulation frequencies in this condition. For the 5-kHz carrier, a decreased sensitivity was observed at high modulation frequencies for both NH and HI listeners. For the average NH data, the value of f_c was 160 Hz ($\tau \approx 1.0$ ms), while the value for the HI listeners was around 93 Hz ($\tau \approx 1.7$ ms). Thus, there is some indication of reduced temporal resolution in the hearing-impaired listeners.

CONCLUSION

The encoding of temporal envelope fluctuations in the auditory system seems to be affected by sensorineural hearing impairment: The ability to detect slow and moderate envelope fluctuations can be superior in the hearing-impaired listeners at high carrier frequencies. In contrast, the ability to discriminate amplitude changes in the envelope and temporal resolution seems to be reduced. No indication of a relationship between modulation detection and modulation discrimination thresholds was found. Fast-acting WDRC was found to reduce the ability to detect slow envelope fluctuations.

REFERENCES

Byrne, D., Parkinson, A, and Newall, P. (**1990**). "Hearing aid gain and frequency response requirements of the severely/profoundly hearing-impaired," Ear Hearing, **11**, 40-49.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Modulation detection and masking with narrowband carriers," J. Acoust. Soc. Am., **102**, 2892-2905.

Dau, T., Verhey, J., and Kohlrausch, A. (**1999**). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrowband noise carriers," J. Acoust. Soc. Am., **106**, 2752-2760.

Edwards B. (**2004**). "Hearing aids and hearing impairment," in *Speech Processing in the Auditory System.* Eds. A. Greenberg, W.A. Ainsworth, A.N. Popper, and R.R. Fay. New York, NY: Springer, pp. 339-421.

Jenstad L.M., Pumford, J., Seewald, R.C., and Cornelisse, L.E. (**2000**). "Comparison of linear gain and wide dynamic range compression hearing aid circuits II: Aided loudness measures," Ear Hearing, **21**, 32-44.

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am., **108**, 723-734.

Lee, J. and Bacon, S.P. (**1997**). "Amplitude modulation depth discrimination of a sinu-soidal carrier: Effect of stimulus duration," J. Acoust. Soc. Am., **101**, 3688-3693.

Moore, B.C.J. and Glasberg, B.R. (**2001**). "Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am., **110**, 1067-1073.

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

Ricketts, T.A. and Bentler, R.A. (**1996**). "The effect of test signal type and band-width on the categorical scaling of loudness," J. Acoust. Soc. Am., **99**, 2281-2287.

Shannon R.V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science, **270**, 37-39.

Stone, M.A. and Moore, B.C.J. (**1992**). "Syllabic compression: Effective com-pression ratios for signals modulated at different rates," Br. J. of Aud., **26**, 351-361.

Stone, M.A., and Moore, B.C.J. (**2003**). "Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task," J. Acoust. Soc. Am., **114**, 1023-1034.

Stone, M.A., Füllgrabe, C., Mackinnon, R.C., and Moore, B.C.J. (**2011**). "The im-portance of speech intelligibility of random fluctuations in 'steady' background noise," J. Acoust. Soc. Am., **130**, 2874-2881.

Effects of time of day and cognitive load on aided performance

SHARON A. SANDRIDGE[1,*], LIA M. SANTIAGO[1], CRAIG W. NEWMAN[1], AND THOMAS BEHRENS[2]

[1] *Cleveland Clinic, Cleveland, Ohio, USA*

[2] *Centre for Applied Audiology Research, Oticon A/S, Smoerum, Denmark*

A link among hearing loss, fatigue, listening effort, and cognitive drain has been suggested to impact benefit from amplification. Hornsby (2013) investigated the effects of hearing aid (HA) use on effort and fatigue for complex listening, suggesting that these negative consequences can be reduced by using well-fit HAs. To probe into this, an experiment was designed where 14 HA users were tested aided in complex listening tasks on late Friday afternoon, Saturday morning, and late Saturday afternoon. In between the two Saturday tests participants were taken on a tour, designed to span a range of challenging listening tasks. This was done twice, using two different levels of hearing technology. Single and dual task versions of the hearing in noise test (HINT) were used to test listening abilities. Self-report probed into fatigue and vigor, different aspects of perceived listening, and characterized participants as morning, intermediate, or evening types. In addition to audiometric measures, the reading span was used to assess cognitive status. Results showed that aided listening changed over the course of a day, performance in the morning was not the best despite most participants being morning types, and well-rested and speech understanding was better in the afternoon despite self-perceived fatigue being increased. Higher technology level did positively affect some objective and subjective listening abilities.

PURPOSE

The purpose of this study was to evaluate fatigue and cognitive effort using two different levels of HA signal processing technology at three different time points associated with a day of active listening activities.

METHODS

Participants

Fourteen experienced adult HA users (7 male; 7 female) ranging in age from 55-83 years (mean = 70; SD = 8.9) participated. Subjects met the following inclusion criteria:

*Corresponding author: sandridges@ccf.org

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

- ≥ 12 months of previous hearing aid use;
- Bilaterally symmetric sensorineural hearing loss within the fitting range of the test HAs (Fig. 1);
- Passed a vision screening assessment (ability to read aloud and comprehend a short passage printed with 12-point font); and
- Passed the Montreal Cognitive Assessment (Nasreddine *et al.*, 2005).

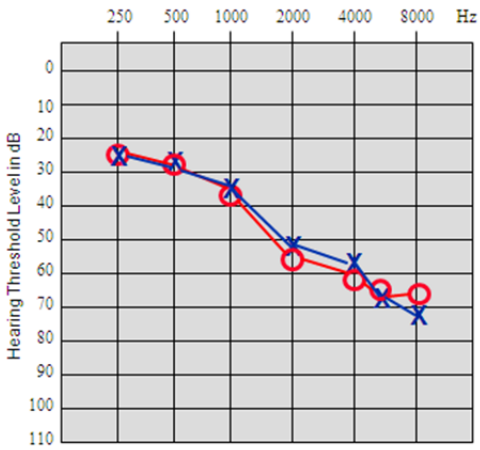


Fig. 1: Mean pure tone air conduction thresholds of the participants (n=14).

Laboratory measures: Objective testing

Hearing in Noise Test (HINT; Nilsson *et al.*, 1994): Listeners repeated sentences spoken by a male speaker in the presence of a fixed (70 dBA) competing noise in the sound field using the standard adaptive technique to determine the speech reception threshold (SRT). The HINT was administered under two conditions:

- HINT Single Task: The HINT was administered as an auditory task solely.
- HINT Dual Task: The HINT was administered along with a simultaneous visual task – the Pattern Completion Test (PCT; Pittman and Petersen, 2011) to assess cognitive effort.
- PCT: Various geometric symbols occurred in a row in a pattern of 2, 3, or 4 shapes and were presented on a computer monitor. A total of 11 symbols were presented and the subject was required to select which of 4 possible symbols would be the next symbol in the pattern.

Self-Report Questionnaires: Subjective Testing

Morningness-Eveningness Questionnaire (MEQ; Horn and Ostberg, 1976): 19-item questionnaire designed to assess whether a subject is more alert in the morning or evening (e.g., *How alert do you feel during the first half hour you are awake in the morning?*). Scores range from 16 – 86 points:

< 41 points	evening types
42-58 points	intermediate types
> 59 points	morning types

Profile of Mood States (POMS; McNair *et al.*, 1971): 15-item questionnaire used to verify fatigue/vigor. Participants rated on a 5-point scale (0 = "not at all" to 4 = "extremely") how well the item related to his or her feelings at that time.

Effort Questionnaire: Six questions examining perceived listening effort, willingness or ability to compensate for the various listening environments, any changes in listening strategy during the day, and perceived physical fatigue were responded to by the participants using a 11-point scale with anchors for 0 and 10 as indicated in Table 1.

Question no.	Wording used	Anchors
1	How often did you find it hard to hear during the activity you just completed?	Never/All the time
2	Did you ever stop trying to hear?	Never/All the time
3	How often do you participate in an activity like this one?	Never/All the time
4	Do you feel tired from the effort you had to make to hear?	Not at all/Completely
5	Do you feel tired from the physical effort associated with the activity?	Not at all/Completely
6	Was the activity enjoyable?	Not at all/Completely

Table 1: Overview of questionnaire used to assess aspects of listening effort in the morning, in the afternoon and in the evening.

Test Devices

- Device HA 1: Oticon Alta Pro; premium level device
- Device HA 2: Oticon Nera Pro; midlevel device

PROCEDURES

The participant was fit with the test device approximately 7 days prior to the weekend activities allowing one-week acclimatization period. Participants were fit binaurally with each set of devices counterbalancing which set was tested first. Programming of the devices followed standard clinical procedures. Verification of fit was performed using real ear measurement.

Each participant was tested three times within a 24-hour period per set of devices. These time-of-day (TOD) sessions occurred on Friday afternoon, Saturday morning, and Saturday afternoon. At each session, the POMS, HINT Single Task, and HINT Dual Task were administered. In between the Saturday morning and Saturday late afternoon experimental tests sessions, participants were taken as a group of 4-6 participants and spouses on a listening tour of local community sites/events designed to span a range of challenging listening situations. These included talking to other participants (previously unknown to each other) on the bus, at a busy mall and restaurant, during a tour at a museum, and other environments. These listening situations included various background noise and acoustical conditions, yet the situations were controlled across the participants. During the day's activities, the participants were asked to complete the Effort Questionnaire at three different time intervals (morning, noon, and afternoon).

The above test paradigm was repeated the following weekend for the other set of devices. To verify that both weekends offered very similar listening environments, a dosimeter was used to monitor each event. All weekends were found to be comparable.

RESULTS

The Morningness-Eveningness Questionnaire revealed that the majority of the participants were morning type as shown in Fig. 2.

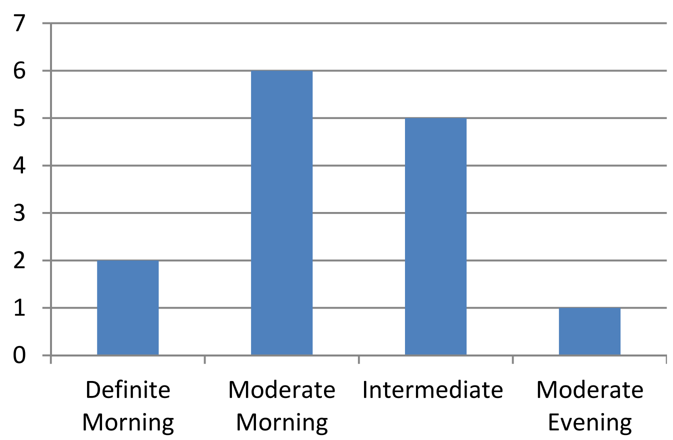


Fig. 2: Number of subjects who were categorized as a morning, intermediate, or evening type (n=14).

The HINT Single and Dual Tests performance was slightly better for the Saturday PM TOD. However, statistical significance was not reached as seen in Fig. 3.

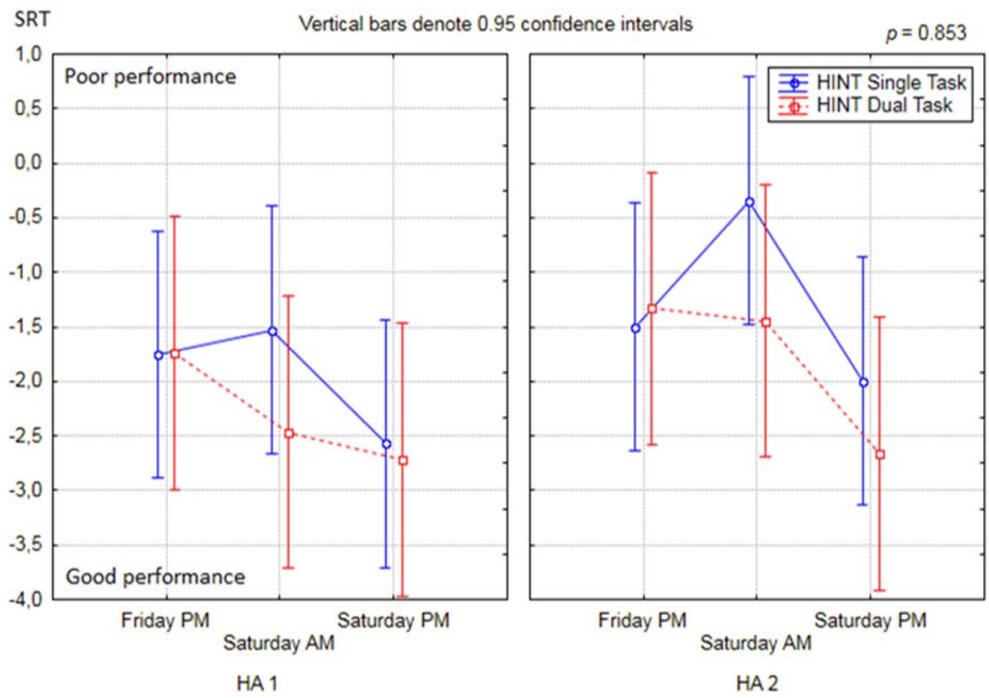


Fig. 3: Mean SRT for the HINT Tests across HA Level and Time of Day. Morning test time was abbreviated AM and afternoon test time was abbreviated PM.

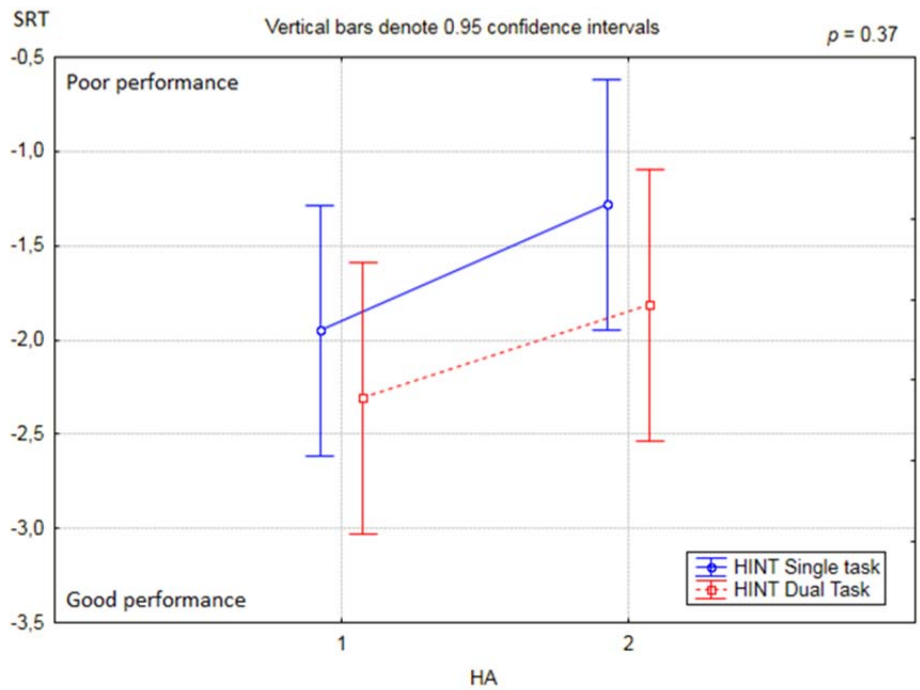


Fig. 4: The mean HINT SRTs for the Single and Dual Tasks across HA Level collapsing time of day.

Figure 4 illustrates the means for the HINT Single and Dual Tasks across the HA levels. Note that although the results are not statistically different, there is a trend showing performance on the Dual Task was better than the Single Task and that HA 1 was better than HA 2. Further, Killion *et al.* (2004) suggested that for every 1 dB of change in signal-to-noise ratio on the SRT there is an 11% change in speech intelligibility, so there may be clinical significance of these findings.

The POMS Scale Fatigue and Vigor results suggested that the participants maintained low fatigue and high vigor across the TOD as seen in Fig. 5.

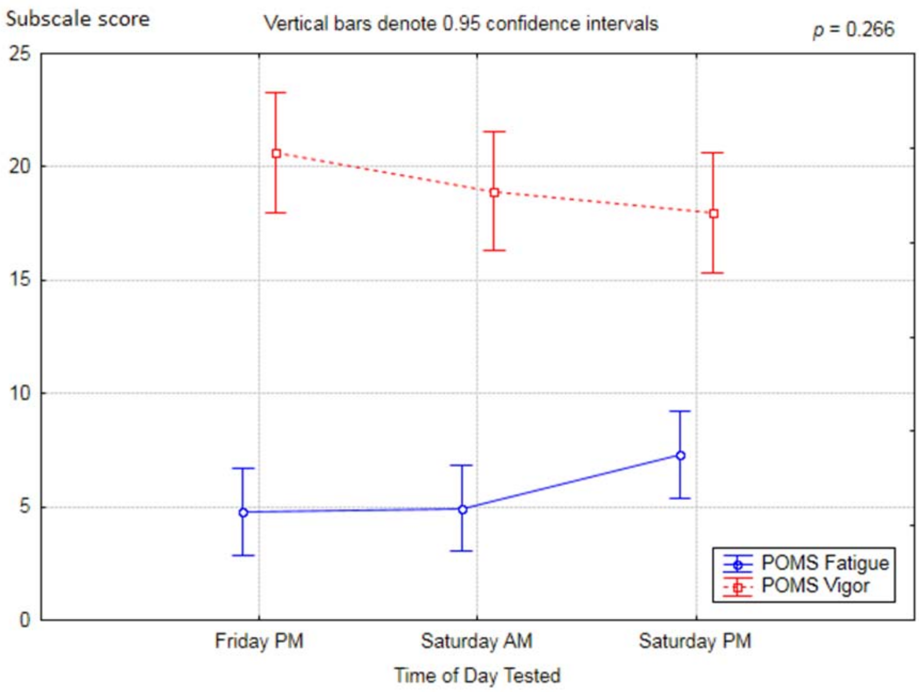


Fig. 5: Mean ratings for 2 subscales of the POMS. Total possible scores for the *Vigor* subscale is 32 points and for the *Fatigue* subscale is 28 points (n=14). Morning test time was abbreviated AM and afternoon test time was abbreviated PM.

Figure 6 illustrates the mean ratings for selected items from the Effort Questionnaire. Results suggest that overall, participants found listening effort, difficulty hearing, and the physical tiredness to be minimal while finding the day quite enjoyable.

Post hoc (Tukey HSD test, $p < 0.01$) revealed that:

- Within HA 1 – participants had less tiredness, effort and difficulty in the Saturday morning and Saturday afternoon compared to the Saturday noon testing;

- Within HA 2 – no statistical differences were found across the 3 Saturday assessment times (morning, noon, and afternoon); and
- Between HA/s – results from HA 1 were statistically better than the results from HA 2 for Saturday morning and Saturday afternoon.

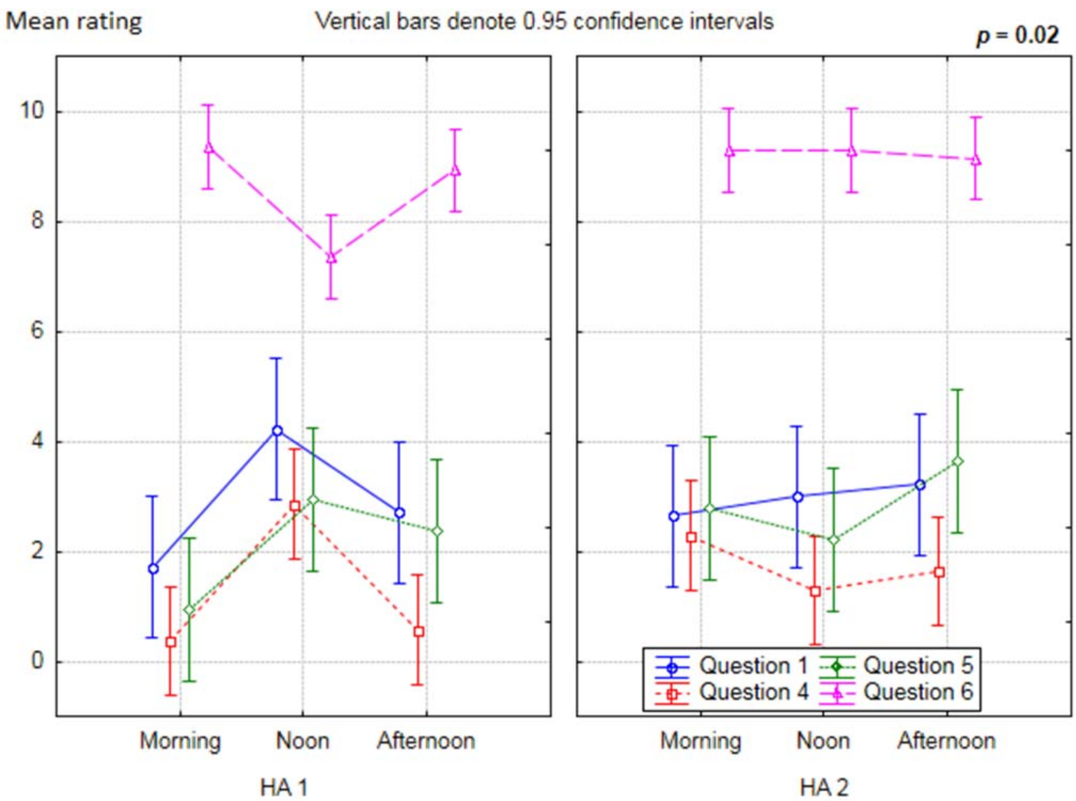


Fig. 6: Mean ratings for items on the Effort Questionnaires.

SUMMARY/CONCLUSIONS

The majority of the participants indicated that they were more 'morning' types. Yet, a clear pattern of better performance for the Saturday morning testing was not evident. In fact, results on the HINT Single Task SRTs were the poorest (largest SRTs) for Saturday morning compared to the other time of day test sessions.

Performance on the HINT for both the Single and Dual tasks showed acceptable SRTs for HA use (ranging from −1.3 to −2.7) and overall, participants demonstrated greater SRTs for the Dual Task compared to the Single Task. It may be speculated that the HA technology decreased the cognitive load allowing greater resources for processing the auditory stimuli.

As the day progressed, it was expected that the participants would demonstrate greater fatigue and less vigor. While there was a slight trend consistent with that expectation, the results were not significant. In addition, the HINT scores were actually the best (lowest SRTs) for the Saturday afternoon testing for both sets of devices. These two findings suggest that amplification may reduce the listening effort lessening the fatigue from having to work hard at listening.

The results of the Effort Questionnaire are also consistent with the findings of the POMS and HINT. Mean scores for Questions 1, 4, and 5 were all below 3 on a 11-point scale with responses closer to the *Never* or *Not At All* (0) anchor than the *All the Time* or *Completely* (10). Accordingly, participants reported minimal difficulty hearing, minimal effort in listening, and less overall physical tiredness from the day's activities while reporting that they very much enjoyed the day (mean rating for Question 6 was > 9 pts).

These results are promising showing that amplification may negate, to some degree, the negative impact of the interaction of hearing loss, fatigue, and cognitive load. Further research is clearly needed to investigate these relationships more in depth while using a larger sample size.

ACKNOWLEDGEMENTS

This study was funded by Oticon A/S.

REFERENCES

Horn, J. and Ostberg, O. (**1976**). "A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms," Int. J. Chronobiol., **4**, 97-110.

Hornsby, B.W.Y. (**2013**). "The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands," Ear Hearing, **34**, 523-534.

Killion, M.C., Niquette, P.A., Gudmundsen, G.I., Revit, L.J., and Banerjee, S. (**2004**). "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am., **116**, 2395-2405.

McNair, D., Lorr, M., Heuchert, J., and Droppleman, L.F. (**1971**). *Manual: Profile of Mood States.* San Diego, CA: Educational and Industrial Testing Service.

Nasreddine, Z., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., and Chertkow, H. (**2005**). "The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment," J. Am. Geriatr. Soc., **53**, 696-699.

Nilsson, M., Soli, S., and Sullivan, J. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Ac. Soc. Am., **95**, 1085-1099.

Pittman, A. and Petersen, K. (**2011**). *Pattern Completion Task.* Unpublished Behavioral Measurement, Arizona State University.

Compensating for impaired prosody perception in cochlear implant recipients: A novel approach using speech preprocessing

FELIX KUHNKE[1,2,*], LORENZ JUNG[1,2], AND TAMÁS HARCZOS[1,2,3]

[1] *Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany*

[2] *Institute for Media Technology, Faculty of Electrical Engineering and Information Technology, Ilmenau University of Technology, Ilmenau, Germany*

[3] *Cochlear-Implant Rehabilitationszentrum Thüringen, Erfurt, Germany*

Due to inherent device limitations, cochlear implant (CI) recipients are provided with greatly reduced pitch information. However, detecting changes in pitch is necessary to perceive intonation, a main feature of prosody. Therefore, CI recipients' ability to perceive prosody is typically below that of normal-hearing subjects. We propose a novel preprocessing algorithm to enhance intonation perception by broadening the range of pitch changes in speech signals. To proof this concept, we have developed the pitch range extension (PREX) algorithm. PREX is capable of low-delay pitch modifications to speech signals. In addition, it provides automatic and intonation based amplification of pitch movements. In an evaluation with 23 CI recipients, the proposed algorithm significantly improved intonation perception in a question vs. statement experiment. However, the improved performance of CI subjects was still inferior to the performance of normal-hearing subjects. The results support the idea that preprocessing algorithms can improve the perception of prosodic speech features. Furthermore, we suggest utilizing the PREX algorithm for individualized treatment and rehabilitation.

INTRODUCTION

Over the last decades, speech recognition rates with cochlear implants steadily improved, with the main focus of the cochlear implant (CI) treatment being to improve the perception of words and sentences. However, CI recipients still perform very poorly on pitch-related tasks such as melody recognition (e.g., Wang *et al*., 2011) and the perception of voice pitch information (e.g., Meister *et al*., 2009). The perception of variation of pitch in speech is crucial to perceive intonation, a main aspect of prosody. As a consequence of poor pitch perception, cochlear implantees may not perceive the emotions expressed by a speaker or whether a sentence is meant as a question or a statement.

---

*Corresponding author: felix.kuhnke@gmail.com

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

With the steady improvement of cochlear implants, CIs have become powerful computing platforms. Modern CIs employ sophisticated signal processing algorithms that react to the incoming signal and may change their processing parameters to improve sound perception. Apart from that, CI recipients can change their device settings according to the sound environment, applying different sound processing technology at different times. Thus, CIs provide the platform and possibilities for sound-specific and user-specific signal processing algorithms.

We propose to use preprocessing algorithms to enhance features of speech signals that are difficult to perceive for CI recipients, such as pitch. As a proof of concept we developed a method to enhance intonation perception in CI recipients by broadening the range of pitch changes made by speakers.

PITCH RANGE EXTENSION (PREX) ALGORITHM

As already stated, the algorithm should first be implemented as a preprocessing algorithm, meaning that we preprocess the audio signal before it enters the usual CI processing chain. Even though one could imagine modifying the stimulation pattern of the electrodes (speech-processing strategy) directly, the use of a preprocessing algorithm has several advantages. First, the algorithm is independent of implant type and electrode design, which allows usage across different devices. Further, it can be used in devices such as hearing aids, which is advantageous for bimodal fitted patients. Finally, the output quality of the algorithm can easily be evaluated with normal-hearing subjects. However, as a final step the algorithm should be embedded inside the speech-processing strategy for improved performance and lower delays.

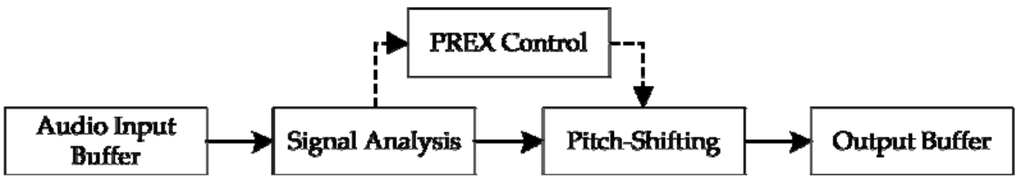


Fig. 1: Overview of the algorithm processing chain.

Figure 1 shows the algorithm overview. The PREX algorithm analyses the incoming speech signal and automatically determines a new pitch value and synthesises the according pitch-shifted signal. In the following we explain the different modules:

Audio samples that have been recorded by the microphone are stored in the Audio Input Buffer. The Signal Analysis module reads a predefined amount of samples from the Audio Input Buffer and estimates fundamental frequency (F0), root-mean-square energy, and zero crossing rate. In the first step these features are used to detect voiced and unvoiced segments of the signal. In the following step the extracted features and the durations between voiced segments are used by the PREX Control module to detect intonational structures. For the PREX protoype, we used heuristics to classify the change between intonational structures. For example, a new intonational structure was detected after 200 ms of silence, as humans are only able to perceive unconnected F0 contours as one, until the gap between them exceeds 200 ms (Nooteboom, 1997). Based on the intonational structures, the PREX Control module computes the pitch shift scale factors accordingly (see the next section for the calculation of new pitch values). The pitch-shifting is done by a customized implementation of the PSOLA algorithm (Hamon *et al.*, 1989). We enhanced the classical PSOLA for lower delays. Instead of using segments of the size of two (or more) pitch periods we used only a single period of the voiced signal as synthesis segment. By applying an adaptive window for every period we removed signal discontinuities that would otherwise arise during overlap-add synthesis.

The algorithmic delay is dependent on the lowest frequency the algorithm should process. If the F0 value is below this frequency, no pitch-shifting is performed. For a minimum supported frequency of 62.5Hz we get an algorithmic delay of 18.75ms and for 100Hz, 11.7ms, respectively.

Calculating new pitch values (PREX Control)

Based on the detected F0 of the speech signal (f_{in}) the new pitch values (f_{out}) for the synthesized signal are calculated according to Eq. 1:

$$f_{out} = f_{in} + f_{in} \times \log_2\left(\frac{f_{in}}{f_{start}}\right) \times PRSF \qquad \text{(Eq. 1)}$$

PRSF is the pitch range scale factor that allows to modify the global amount of range extension. We found that different factors for up and down pitch range extension are necessary to produce natural sounding results. This can easily be accomplished using separate pitch range scale factors for upward and downward extension. The perception and production of voice pitch is generally not on a linear scale (Nolan, 2003). Therefore, we use a psycho-acoustic logarithmic pitch scale to uniformly describe intonation across different speakers. To preserve natural intonation we use the first F0 of every intonational structure (f_{start}) as reference frequency for pitch range extension. This approach will not alter the pitch register and will produce a more natural sounding result. Finally, the new pitch values are used to compute the corresponding pitch shift scale factor for the pitch shifting.

Figure 2 shows the result of PREX preprocessing on a sentence uttered as a question. The processed F0 curve (dashed line) shows a much higher pitch range.

EVALUATION

We use an intonation hearing experiment based on the question vs. statement paradigm: A number of recorded sentences (stimuli) are presented to the subject. For every stimulus the subject has to decide whether it was spoken as a statement or a question.

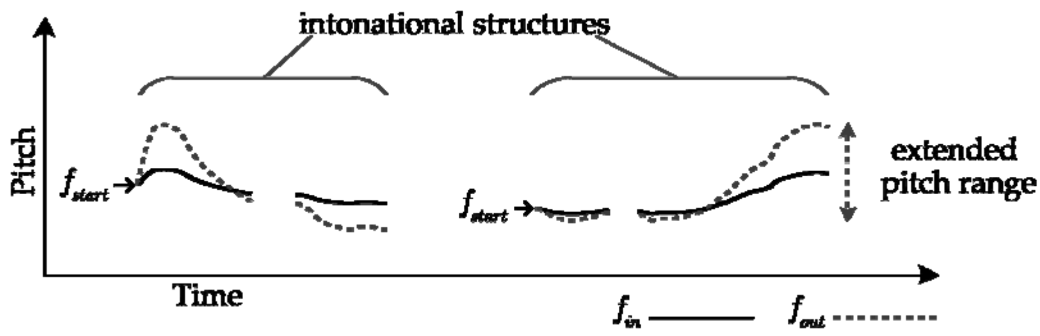


Fig. 2: Pitch range extension based on intonational structures. F0 contour of a german sentence uttered as question, before (solid line) and after PREX preprocessing (dashed line).

Stimuli

We recorded 36 sentences from 3 female and 3 male speakers, once as statement and once as question. To provide no lexical cues, the sentences were the same for questions and statements, e.g., "She will arrive at ten o'clock." vs. "She will arrive at ten o'clock?". This produced 72 test stimuli. In addition, all 72 stimuli were processed with the PREX algorithm, resulting in a total of 144 stimuli with a total length of 5 minutes and 24 seconds.

Subjects

23 CI recipients from the Cochlear-Implant Rehabilitationszentrum Thüringen were asked to perform the test. The subjects had the following characteristics: Subjects were aged 17 to 77 years, with a mean of 54 years. The duration of the subjects' CI-experience was ranging from 1 month to over 11 years. 12 female and 11 male subjects participated. Furthermore, subjects were using Cochlear and MED-EL implants.

Procedure

A single loudspeaker (YAMAHA Monitor Speaker MS101 II) was positioned in front of the subject. The maximum presentation level at the subject's position, about 50 cm in front of the loudspeaker, was set to 70 dB(A) (measured using stimuli of the first male speaker). All stimuli were played from a laptop. Furthermore, a small program was developed to play the stimuli in random order.

The tests were carried out with one CI subject at a time. First, the task was explained. At this point, unilaterally implanted subjects were asked to put on a single-sided headphone to mask the contralateral ear with noise. We used uncompressed OLSA noise (Wagener *et al.*, 1999) at 81.9 dB(A). Afterwards, every subject completed a training phase, where feedback was given for every response.

The length and intensity of the training phase was dependent on the subject. This was required, because some subjects easily identified the different sentences and were quickly ready for the main test, while others needed repetitive presentation of stimuli and even visual cues to learn what they had to listen for. However, all subjects received the same training stimuli, which were not used in the main test. The main test usually took between 25 to 35 minutes. Every stimulus was presented only once and no feedback was provided during testing. The testing was conducted in an anechoic room at the Cochlear-Implant Rehabilitationszentrum Thüringen. Afterwards, the percentage of correct question/statement identifications (score) was measured for every subject. Furthermore, the subjects' responses were sorted to provide scores for natural stimuli and PREX modified stimuli.

Test verification with normal-hearing subjects

A verification test was conducted to assess whether or not the PREX preprocessing harms the recognition of intonation and that the natural stimuli can be correctly identified by normal-hearing (NH) subjects. Five NH subjects participated in the test. The average identification performance for both types of stimuli was 99.7%. Surprisingly, one subject achieved 98.6% (142 of 144 correct), whereas all others reached 100%. The erroneous identifications could be caused by a lack of concentration as every stimulus was only played once. However, the results showed that NH subjects can perform the test with near perfect results.

RESULTS

Because of the small sample size, scores were not assumed to be normally distributed. The box plot in Fig. 3 shows the scores for both stimulus groups. It can be seen that the identification of questions and statements for natural stimuli and modified stimuli was worse compared to the near perfect score achieved by NH subjects. Furthermore, the median of the PREX stimuli score is about 10% higher than the corresponding natural stimuli score median. To analyse the results in more detail, a scatter plot was used.

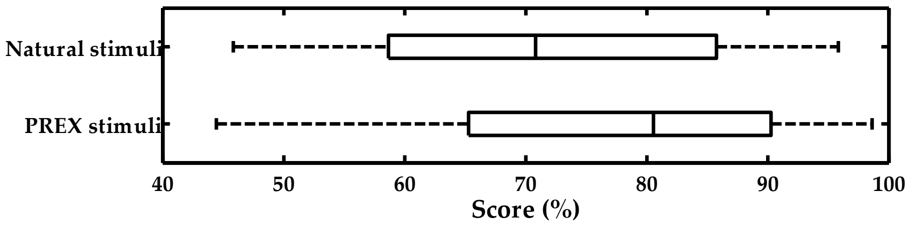


Fig. 3: Box plot of the results (percent correct scores) for the two stimulus groups.

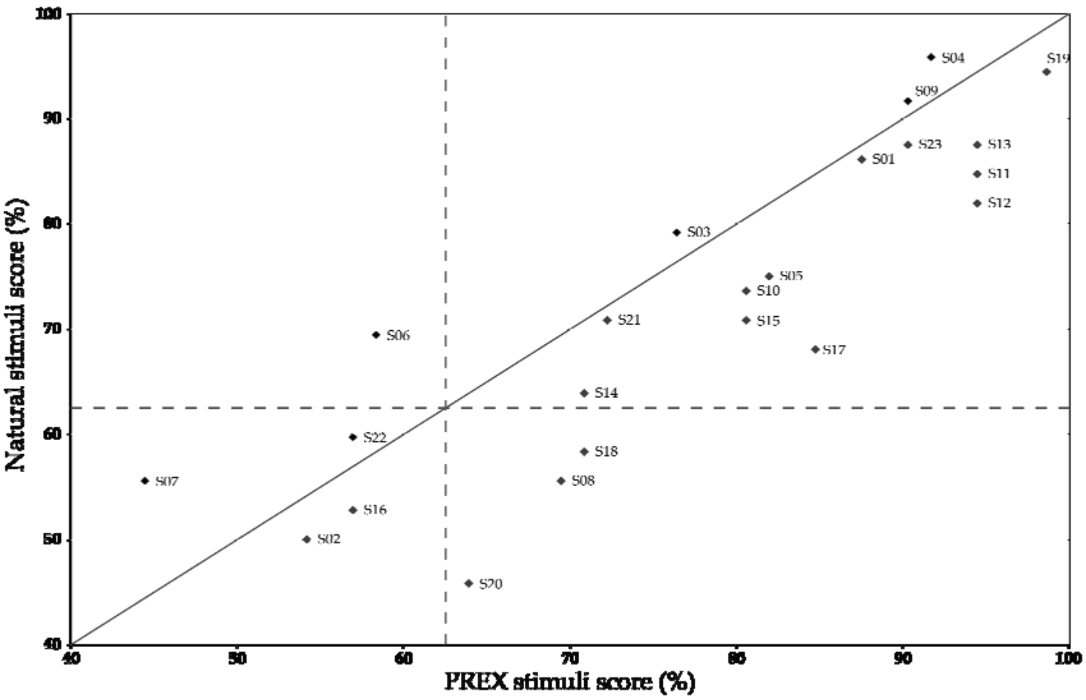


Fig. 4: Percentages of correct question/statement identifications (score) for all CI subjects. Performance for the natural stimuli is plotted against the performance for the PREX processed stimuli. The diagonal line represents equal performance for both types of stimuli. Subjects had to achieve more than 62.5% (shown with a dashed line) for each condition to perform better than chance.

The scatter plot (Fig. 4) uncovers the scores of every subject. The plot shows that some CI subjects (S02, S07, S16, and S22) had huge problems in perceiving the difference between question and statement stimuli. The binomial test revealed that these subjects did not perform significantly better than chance (p = 0.5), which would require at least 45 of 72 (62.5%) correct identifications for either case (p = 0.0444), two sided binomial test). Interestingly, subjects S06, S08, S18, and S20 performed only above chance level for one stimuli group. The other subjects were above chance level but identification scores were widely distributed and subjects showed large inter-individual scatter. However, scores were mostly close to the diagonal, indicating only small differences between the stimuli groups.

Visual inspection of Fig. 4 also suggests a trend towards the lower right side of the scatter plot, as points are more often found below the diagonal. This finding indicates a better performance with the PREX stimuli. The non-parametric Wilcoxon matched-pairs signed rank test (two sided) was employed to test for a significant difference of medians between PREX stimuli scores and natural stimuli scores. In the indicated case, the test rejected the null hypothesis of equal medians at the 5%

significance level with $p = 0.0103$. Therefore, the score for PREX stimuli is significantly higher than the score for natural stimuli.

In addition, we analysed for relationships between scores and subject characteristics. Astonishingly, no significant relationship was found for experience, processing strategy, age, or residual hearing. This surprising finding may be associated with the small number of subjects.

DISCUSSION

Based on the results, it can be concluded that the PREX preprocessing significantly improves question vs. statement identification in CI recipients. The increased ability to identify sentences as questions or statements suggests that PREX preprocessing improves the overall perception of intonation and prosody. However, these results must be interpreted with caution, as the question vs. statement test cannot be considered representative for all forms of intonation perception.

While a significant difference was found, it seems to be very small when the median values are taken as references. On the other hand, the subjects heard PREX processed stimuli in the test situation for the first time and they had no time to get accustomed to the new stimuli. Using PREX on a daily basis might reveal additional improvements.

The main weakness of the evaluation is that speech intelligibility of PREX stimuli was not measured. Even though NH subjects did not report any difficulties to understand the processed sentences, the same cannot be assumed for CI recipients. However, CI subjects did not report any problems with intelligibility. Often, they stated that they did not have any problem understanding the sentences but did not know whether it was spoken as a question or a statement.

CONCLUSION AND OUTLOOK

We presented a preprocessing algorithm that enhances intonation perception by broadening the range of pitch changes in speech signals. It provides automatic and intonation based amplification of pitch movements. In an evaluation with 23 CI recipients, the proposed algorithm significantly improved intonation recognition, likely caused by the fact that pitch movements became more easily identifiable by CI subjects. Based on these findings, it would be very interesting to see if PREX processing could improve speech intelligibility for tonal languages such as Mandarin.

The results support the idea that the perception of a variety of speech features that are difficult to perceive for CI recipients or hearing aid users can be improved by speech preprocessing algorithms. These additional speech features include loudness, vowel quality, and duration. Speech preprocessing methods could be used in CI rehabilitation to specifically exercise and improve individual weaknesses. The areas of application range from speech perception to speech production. Similar to music students learning musical pieces at lower tempi, CI recipients could use time stretching to learn voice recognition at a lower speech tempo. Furthermore, PREX preprocessing could be used while training speech production. Implantees may achieve an improved recognition of their own sound production and subsequently improve their prosody production. Finally, signal modifications do not need to be fixed to a certain intensity, but could be set to meet individual needs.

REFERENCES

Hamon, C., Mouline, E., and Charpentier, F. (**1989**), “A diphone synthesis system based on time-domain prosodic modifications of speech,” Proc. International Conference on Acoustics, Speech, and Signal Processing, 1989, 238-241.

Meister, H., Landwehr, M., Pyschny, V., Walger, M., and Wedel, H.v. (**2009**), “The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients,” Int. J. Audiol., **48**, 38-48.

Nolan, F. (**2003**), “Intonational equivalence: An experimental evaluation of pitch scales,” Proc. 15th International Congress of Phonetic Sciences, 771-774.

Nooteboom, S. (**1997**), “The prosody of speech: Melody and rhythm,” in *The Handbook of Phonetic Sciences*. Eds. Hardcastle, W.J. and Laver, J. (Oxford, UK: Blackwell), pp. 640-673.

Wagener, K., Kühnel, V., and Kollmeier, B. (**1999**). “Entwicklung und Evaluation eines Satztests in deutscher Sprache I: Design des Oldenburger Satztests,” Zeitschrift für Audiologie/Audiological Acoustics, **38**, 4-15.

Wang, W., Zhou, N., and Xu, L. (**2011**), “Musical pitch and lexical tone perception with cochlear implants,” Int. J. Audiol., **50**, 270-278.

Frequency selectivity improvements in individual cochlear implant users with a biologically-inspired preprocessing algorithm

FLORIAN LANGNER* AND TIM JÜRGENS

Medizinische Physik and Cluster of Excellence “Hearing4all”, Carl von Ossietzky University, Oldenburg, Germany

The ability to distinguish between two sounds of different frequency is known as frequency selectivity, which can be quantified using psychoacoustic tuning curves (PTCs). Normal-hearing (NH) listeners show level- and frequency-dependent sharp PTCs, whereas frequency selectivity is strongly reduced in cochlear implant (CI) users. This study aims at (i) assessing the individual shapes of PTCs measured psycho-acoustically in CI users, (ii) comparing these shapes to those of simulated CI listeners, and (iii) improving the sharpness of PTCs using a biologically-inspired preprocessing algorithm. A 3-alternative-forced-choice forward masking technique was used to assess PTCs in eight CI users (with their own speech processor) and 11 NH listeners (with and without listening to a vocoder to simulate electric hearing). CI users showed large inter-individual variability in sharpness, whereas simulated CI listeners had shallow, but homogeneous PTCs. Furthermore, a biologically-inspired dynamic compression algorithm was used to process the stimuli before entering the CI users’ speech processor or the vocoder simulation. This algorithm was able to partially restore frequency selectivity in both groups, meaning significantly sharper PTCs than unprocessed.

INTRODUCTION

Frequency selectivity is an important characteristic of the individual listener’s ability to perceive sounds. Psychoacoustic tuning curves (PTCs) can be used to estimate frequency selectivity in normal-hearing (NH) or hearing-impaired (HI) listeners. PTCs display the masking threshold – i.e., the level of a pure-tone masker that is necessary to render a specific target tone inaudible – as a function of different masker frequencies. NH listeners show sharp PTCs with slightly lower masking thresholds at the low-frequency tail (due to upward spread of masking, cf. Moore, 1978; Oxenham and Plack, 1998). HI listeners show broader PTCs than NH listeners. Their PTC shape can be considerably sharpened using a dynamic compression algorithm (Jürgens *et al.*, 2014), as used in hearing aids. In cochlear implant (CI) users, “spatial tuning curves” (Nelson *et al.*, 2011) can be used to assess the spatial selectivity of electric stimulation on single electrodes using a

*Corresponding author: langner.florian@mh-hannover.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

similar paradigm to PTCs in acoustic hearing. These spatial tuning curves are not being measured using the CI user's personal speech processor, but using a research interface, which allows controlled stimulation of single electrodes. Spatial tuning curve shapes were found to be highly individual across CI users (Nelson *et al.*, 2011). A comparison to PTCs in NH and HI listeners is difficult, because such a comparison would require exact mappings of electric current to acoustic level and mappings of electrode location to acoustic frequency. Furthermore, spatial tuning curves (measured using a research interface) do not necessarily reflect the frequency selectivity of the CI user in their everyday life, because their speech processor and sound coding strategy are not used.

For performing such a PTC comparison, experiment 1 of this study measured PTCs in simulated CI users using a vocoder. In experiment 2, the same psychoacoustic measurement is then performed with individual CI users, which means that PTCs are measured with acoustic stimuli presented via the CI user's own speech processor. This allows a direct comparison across individual CI users, but also comparisons to NH listeners and simulated CI users. Finally, the hypothesis is tested whether improvements of the PTC shape due to preprocessing with a multi-channel dynamic compression algorithm (Meddis *et al.*, 2013) are possible in both simulated and actual CI users.

METHODS

Subjects and procedure

Eleven NH subjects (22–30 years, average age of 26 years) acted as the simulated CI listeners and were measured using Sennheiser HDA200 headphones listening through a software-implemented vocoder (adapted from Bräcker *et al.*, 2009, see below). Eight actual CI listeners (seven postlingual and one prelingual deafened, see Table 1, average age of 42 years) participated in the study. These CI users were presented with acoustic sounds using an audio cable connected directly from the sound card to the input of their sound processor.

ID	Age	Sex	Etiology	Duration of deafness (y)	CI usage (y)	Device
CI1	25	M	Ototoxic	17	8	Freedom Hybr.
CI2	23	F	Acute hearing loss	0.5	3	CP810
CI3	45	M	Lack of oxygen	44	0.6	CP910
CI4	19	F	Short hair cells	8	12	OPUS 2
CI5	64	M	Meningitis	49	1	CP810
CI6	46	F	Acute hearing loss	6	0.5	CP910
CI7	53	M	Since birth	7	6	CP910
CI8	63	M	Acute hearing loss	10	4	CP810

Table 1: Details about all participating CI listeners.

A 3-interval forced-choice (3-IFC) 1-up-2-down forward masking paradigm was used to determine the individual masking thresholds for pure-tone maskers with 7 frequencies relative to (0.5, 0.7, 0.9, 1.0, 1.1, 1.3, and 1.7 times) the fixed frequency of the 2-kHz pure-tone target. The target level was fixed at 10 dB sensation level (SL) and was determined for each listener beforehand using the same 3IFC method. The 106-ms masker was followed by 10 ms of silence and the 16-ms target tone. Three repetitions were averaged to obtain one masking threshold.

BioAid processing

BioAid (Meddis *et al.*, 2013) is a multi-channel dynamic compression algorithm that mimics two essential mechanisms in the healthy auditory system. The signal processing flow is shown in Fig. 1. Nine different frequency channels with half-octave-wide Butterworth filters at half-octave spacing were used. The first mechanism of BioAid is the instantaneous compression of the basilar membrane which is technically realized by an instantaneous ‘broken-stick’ compression. The second mechanism is the reflex of the medial olivocochlear complex which is realized by a slow and time-delayed feedback loop using a time constant of 50 ms. The latter process is called delayed feedback attenuation control (DFAC) in the algorithm and controls the attenuation adaptively in each channel.

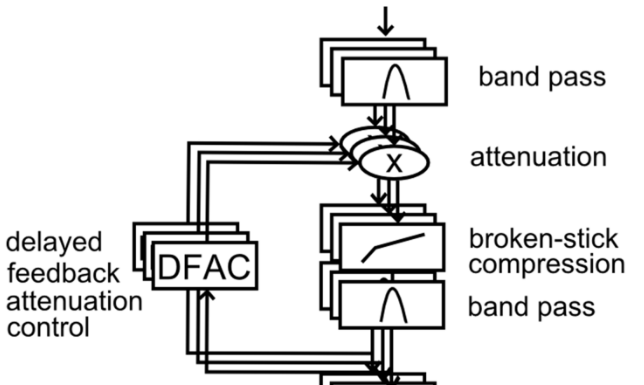


Fig. 1: Signal processing structure in the BioAid algorithm, different layers symbolize different frequency channels.

Both physiological mechanisms are missing in CI listeners, which is why their imitation might improve or even restore frequency selectivity. Both the DFAC and instantaneous compression consist of an activation threshold that is personalized for each listener.

CI simulation and measurement conditions

A vocoder mimicking details of the signal processing and the physiology of CI users (Bräcker *et al.*, 2009) was used for simulating CI users with NH listeners. This vocoder was structured to resemble the implant type of a Cochlear Contour Advance electrode array with 22 electrodes. PTCs were measured for simulated and actual CI users in three conditions: unprocessed (i.e., vocoded-only for simulated CI listeners), BioAid without and BioAid with instantaneous compression. In addition, NH listener's PTCs were measured without vocoder and BioAid as a reference.

RESULTS

Figures2 and 3 show PTCs as masker threshold levels in dB SL, which means that the zero line indicates the absolute threshold of the target tone. Circles indicate the averages over all three measurement repetitions for this masker frequency, while error bars indicate one standard deviation. PTCs were fitted using a 2[nd] order rounded exponential (ROEX) fit (Patterson *et al.*, 1982).

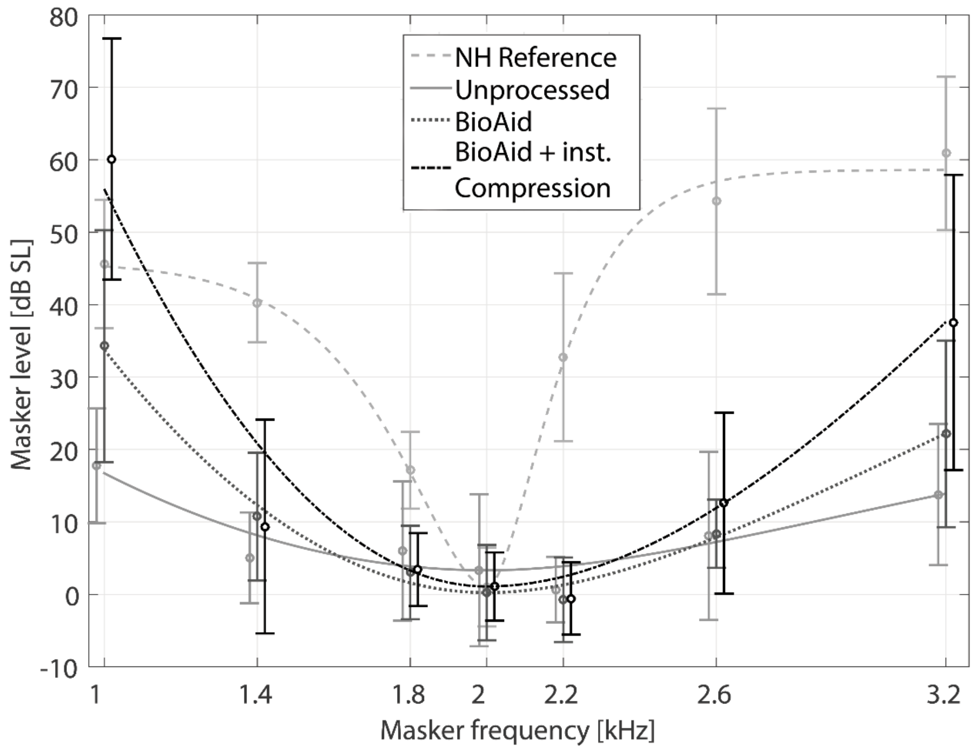


Fig. 2: Averaged group PTCs of all 11 simulated CI listeners: Masker level in dB SL as a function of masker frequency in kHz.

Experiment 1 – Simulated CI listeners

Figure 2 shows average masking thresholds and resulting PTCs averaged across simulated CI listeners. The NH reference PTC (gray dashed line) is relatively sharp in agreement with studies from the literature (e.g., Moore, 1978). The unprocessed

CI simulation (realized using the vocoder only, gray continuous line) shows a very flat shape. BioAid without instantaneous compression (black dotted line) shows small improvements in frequency selectivity in terms of a sharper PTC curve and a higher masking threshold at the outer-most masker frequencies (1 and 3.2 kHz). The improvement is stronger using BioAid with instantaneous compression (dashed-dotted line).

Experiment 2 – Actual CI listeners

Individual PTCs for actual CI listeners showed high variability among subjects (Fig. 3). In most cases, the unprocessed condition (grey continuous lines) resulted in a relatively flat PTC shape (similar to the unprocessed PTC of simulated CI listeners, see Fig. 2). High variability in PTC shape can also be observed regarding the effect of preprocessing the stimuli with BioAid. BioAid without instantaneous compression resulted in slightly sharper PTCs for some CI listeners (CI2, 3, 4, 6, and 8), while others showed no change (CI1, 5, and 7). For BioAid with instantaneous compression (dashed-dotted line), the PTC shape was strongly (CI2, 3, and 4), modestly (CI1, 5, 6, and 8) or not at all (CI7) affected by the algorithm. Thus, BioAid had a much stronger frequency selectivity restoration effect with than without instantaneous compression, especially in terms of higher masking thresholds at outlying masker frequencies (1, 1.4, 2.6, and 3.2 kHz).

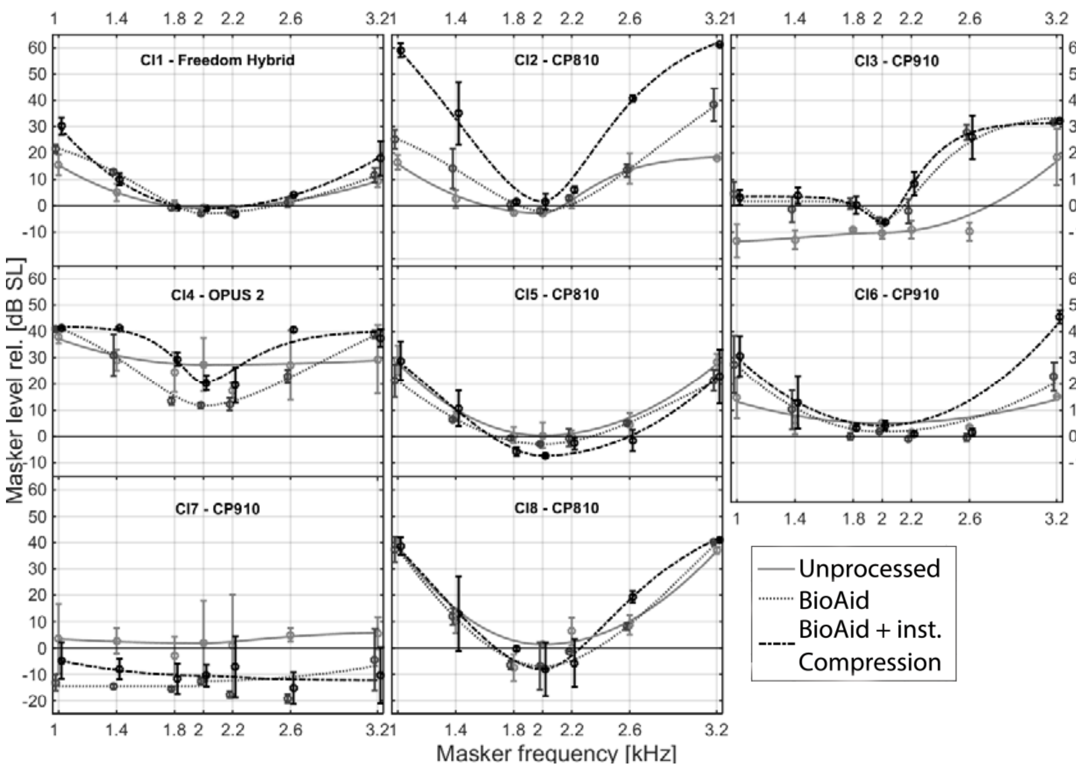


Fig. 3: Individual PTCs of 8 actual CI listeners in three conditions: relative masker level in dB SL as a function of masker frequency in kHz.

Statistical comparisons

Figure 4 shows two measures to quantify the sharpness of PTC shape: D_{PTC} (cf. Lecluyse *et al.*, 2013) and Q_{10dB}. D_{PTC} is a depth measure, in dB, and is the difference between the mean of all four outlying masker frequencies and the mean of the three centre masker frequencies. This measure is suitable for capturing the large-scale shape of the PTC. Q_{10dB} is the ratio between the centre frequency and the bandwidth 10 dB above the tip of the curve. For relatively flat PTCs, it captures variations only near the centre frequencies and is therefore a small-scale measure for PTC shape comparisons. The Friedman-test was used for statistical comparisons.

For the simulated CI listeners, D_{PTC}s were significantly different between the unprocessed condition and both BioAid conditions ($p < 0.01$), as well as significantly different between both BioAid conditions ($p < 0.05$). No significant difference was found between BioAid with instantaneous compression and the NH reference ($p > 0.1$), implying that the PTC sharpness (as measured using D_{PTC}) was fully restored. Significant differences in Q_{10dB} were found between the unprocessed and both BioAid conditions ($p < 0.01$). However, the NH reference condition showed a highly significant difference to all other conditions ($p < 0.01$). For the actual CI listeners, a significant difference between the unprocessed and BioAid with instantaneous compression condition was found both regarding D_{PTC} and Q_{10dB} ($p < 0.05$). There was no statistical difference in D_{PTC} or Q_{10dB} between unprocessed and BioAid without instantaneous compression in actual CI listeners.

DISCUSSION

Similar PTC shapes were observed across simulated CI listeners, in contrast to the very individual PTC shapes across the actual CI listeners. This large variability is in line with CI listeners' spatial tuning curves reported in Nelson *et al.* (2011). Different physiological factors, such as spatial spread of the electric field, number and distribution of auditory nerve fibers and the individual electric dynamic range, may have contributed to this high degree of individuality. However, also different signal processing schemes (four different devices from two different manufacturers) may have contributed as well. These factors can, in principle, be implemented also in the vocoder being used in this study (Bräcker *et al.*, 2009) for a systematic investigation of how strong the influence of these factors is on the PTC shape.

In line with earlier findings in HI listeners (Jürgens *et al.*, 2014), the PTC shape was sharpened in all simulated CI users and in 7 out of 8 actual CI users due to the algorithm BioAid. This highlights that frequency selectivity can be improved independently of CI manufacturer and device. The introduction of the D_{PTC} measure revealed that masking threshold increases were mainly present at remote masker frequencies. Frequency selectivity changes at nearby masker frequencies were limited, as the Q_{10dB} measure showed. Two different mechanisms in BioAid are responsible for the improvements in frequency selectivity, which can be separated by the two BioAid processing conditions tested in this study. The frequency-selective DFAC attenuates the masker, but leaves the target tone almost unchanged

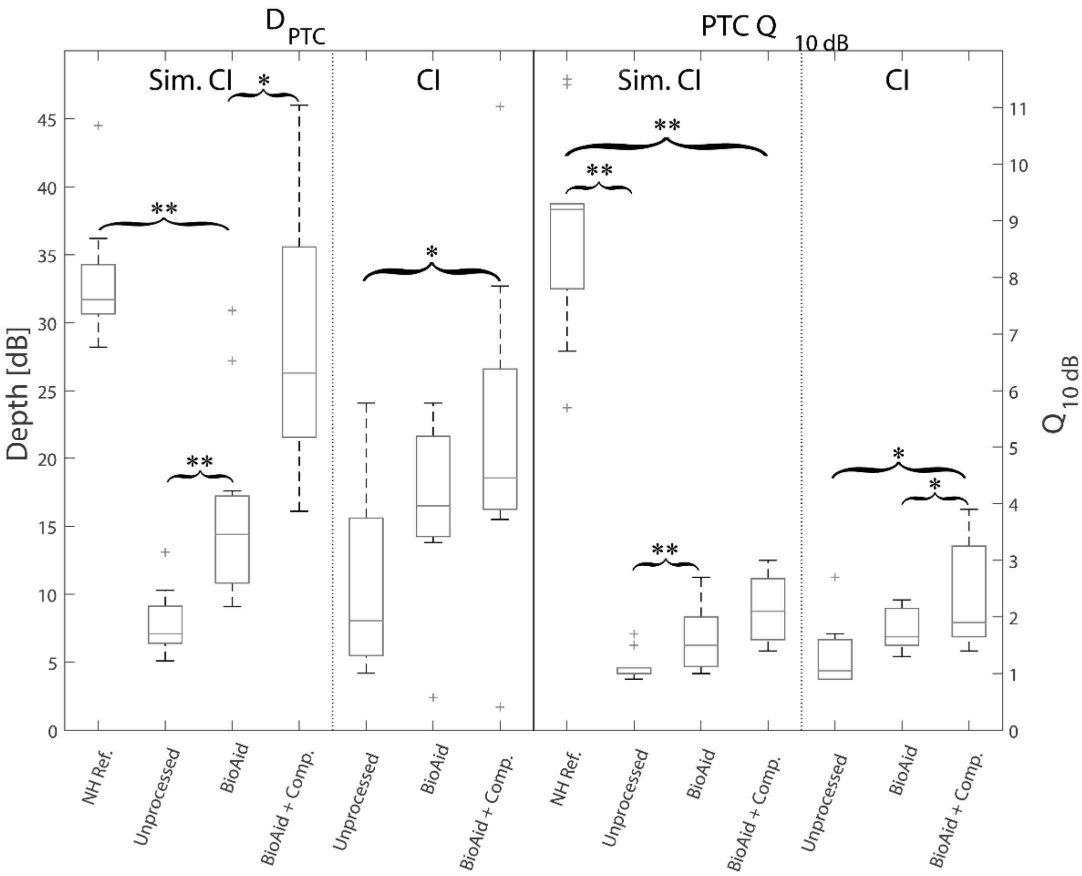


Fig. 4: Boxplots for simulated and actual CI listeners: the horizontal line within the box indicates the median; edges are the 25th and 75th percentiles, whiskers the most extreme data points and outliers are shown as plus signs. Significance symbols indicate $p < 0.05$ with * and $p < 0.01$ with **.

in amplitude if target and masker fall in different frequency channels (i.e., for remote masker frequencies). Thus, the masking effect for maskers with remote frequencies is diminished allowing higher masker levels at threshold. If masker and target fall in the same frequency channel both are attenuated due to the DFAC and masker thresholds are virtually unchanged. Enabling instantaneous compression in addition to the DFAC (in BioAid + instantaneous compression) diminishes the masking effect for remote-frequency maskers further, because the masker (higher in level) is being compressed, whereas the target tone is not.

It is important to consider the different compression stages in both the simulated and the actual CI listeners. While simulated CI listeners use only one compression stage in BioAid with instantaneous compression (in addition to their healthy basilar membrane compression), actual CI listeners use up to three compression stages (BioAid with instantaneous compression, a broadband automatic gain control or adaptive dynamic range optimization (ADRO) preceding, and instantaneous compression within their sound coding strategy). These differences are most likely responsible for the smaller frequency selectivity improvements for actual CI listeners than for simulated CI listeners.

It is conceivable that the direct implementation of BioAid's mechanisms into a CI coding strategy could enlarge their effect on frequency selectivity even further. This might especially prove useful for programs that are fitted for listening to music.

CONCLUSIONS

PTCs of simulated CI users were found to be broader than those obtained with the NH reference group. PTCs of the actual CI users were also broader, but varied strongly across users. In both groups, the multi-channel dynamic compression algorithm BioAid was able to partially restore the sharpness of PTCs, except for one CI user (CI7). This indicates that frequency selectivity can be improved using a compressive processing preceding the CI speech processor. Future research should investigate the implementation of BioAid's algorithm structure into a music coding strategy for cochlear implants.

ACKNOWLEDGEMENTS

This study was supported by DFG cluster of Excellence "Hearing4all" and the Scholarship for ISAAR. Special thanks to Ray Meddis for fruitful discussion and Torsten Dau for helpful comments on an earlier version of the manuscript.

REFERENCES

Bräcker, T., Hohmann, V., Kollmeier, B., and Schulte, M. (**2009**). "Simulation und Vergleich von Sprachkodierungsstrategien in Cochlea-Implantaten," Zeitschrift der Audiologie/Audiological Acoustics, **48**, 158-169.

Jürgens, T., Clark, N.R., Lecluyse, W., and Meddis, R. (**2014**). "The function of the basilar membrane and the MOC reflex mimicked in a hearing aid algorithm," J. Acoust. Soc. Am., **135**, 2385.

Lecluyse, W., Tan, C.M., McFerran, D., and Meddis, R. (**2013**). "Acquisition of auditory profiles for good and impaired hearing," Int. J. Audiol., **52**, 596–605.

Meddis, R., Clark, N.R., Lecluyse, W., and Jürgens, T. (**2013**). "BioAid – ein biologisch inspiriertes Hörgerät," Zeitschrift der Audiologie/Audiological Acoustics, **52**, 148-152.

Moore, B.C.J. (**1978**). "Psychophysical tuning curves measured in simultaneous and forward masking," J. Acoust. Soc. Am., **63**, 524-532.

Nelson, D., Kreft, H.A., Anderson, E.S., and Donaldson, G.S. (**2011**). "Spatial tuning curves from apical, middle, and basal electrodes in cochlear implant users," J. Acoust. Soc. Am., **129**, 3916-3933.

Oxenham, A.J. and Plack, C. (**1998**). "Suppression and the upward spread of masking," J. Acoust. Soc. Am., **104**, 3500-3501.

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.

Simple spectral subtraction method enhances speech intelligibility in noise for cochlear implant listeners

MATTHIAS LEIMEISTER[1,2], CSANÁD EGERVÁRI[1,3], FELIX KUHNKE[1,2], ANJA CHILIAN[1,4,5], CHARLOTT VOIGT[1], AND TAMAS HARCZOS[1,2,5,*]

[1] *Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany*

[2] *Institute for Media Technology, Faculty of Electrical Engineering and Information Technology, Ilmenau University of Technology, Ilmenau, Germany*

[3] *Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary*

[4] *Institute of Biomedical Engineering and Informatics, Faculty of Computer Science and Automation, Ilmenau University of Technology, Ilmenau, Germany*

[5] *Cochlear-Implant Rehabilitationszentrum Thüringen, Erfurt, Germany*

It has been demonstrated that while clean speech is well intelligible by most cochlear implant (CI) listeners, noise quickly degrades speech intelligibility. To remedy the situation, CI manufacturers integrate noise reduction (NR) algorithms (often using multiple microphones) in their CI processors, and they report that CI users benefit from this measure. We have implemented a single-microphone NR scheme based on spectral subtraction with minimum statistics to see if such a simple algorithm can also effectively increase speech intelligibility in noise. We measured speech reception thresholds using both speech-shaped and car noise in 5 CI users and 23 normal-hearing listeners. For the latter group, CI hearing was acoustically simulated. In case of the CI users, the performance of the proposed NR algorithm was also compared to that of the CI processor's built-in one. Our NR algorithm enhances intelligibility greatly in combination with the acoustic simulation regardless of the noise type; these effects are highly significant. For the CI users, trends agree with the above finding (for both the proposed and the built-in NR algorithms), however, due to low sample number, these differences did not reach statistical significance. We conclude that simple spectral subtraction can enhance speech intelligibility in noise for CI listeners and may even keep up with proprietary NR algorithms.

INTRODUCTION

Signal processing chains of modern cochlear implant (CI) processors (like the Nucleus® CP910 from Cochlear™ or the Naída CI *Q70* from Advanced Bionics) include noise reduction (NR) methods to enhance speech perception in noise. However, for studies involving novel speech processing strategies, the elements of

*Corresponding author: tamas.harczos@gmail.com

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

the CI's built-in signal processing chain are typically not available. Since our research plans include the testing of novel strategies combined with noise reduction, we created our own plain NR implementation, which we will abbreviate as PNR all through this document. This paper describes the functional principle of PNR and elaborates on the study we did to test PNR with CI users and normal-hearing (NH) listeners.

METHODS

Noise reduction

PNR is based on a single-microphone spectral subtraction algorithm that was proposed by Martin (1994). The first algorithm of that kind was introduced by Boll (1979) and many variations are widely used in communications and audio processing. Given a speech signal that is corrupted by additive noise, spectral subtraction aims at estimating the magnitude power of the noise spectrum. By applying one of several subtraction rules on the frames of a short time Fourier transform (STFT), this noise estimate is subtracted from the mixture leading to an approximation of the clean sound. The resulting time domain signal is obtained by applying the overlap-add technique. The noise is commonly only estimated in the magnitude or power domain while the original phase values are not modified for the reconstruction. This is due to the observation that estimating the phase of the clean signal is not crucial for the intelligibility of the resulting output (Loizou, 2007).

Most variants of spectral subtraction use a speech activity detector in order to estimate the noise spectrum in speech pauses. However, this can be a source of error. If the detection does not work correctly, parts of the speech might contribute to the noise estimate and would be attenuated by the subtraction rule. The extension that is used in this work (Martin, 1994) circumvents this by estimating the noise spectrum at the minima of a smoothed power spectrum. Under the assumption that the noise can be observed in isolation within a certain search window, one arrives at a steadily updated noise floor.

More detailed, the subband signal power P_x is computed from an STFT that is smoothed along the time axis by a first order low-pass. The estimated minimum power P_{min} is computed as the minimum within a given search window. By multiplying with a correction parameter *omin* that accounts for bias in the minimum estimate one arrives at the estimated noise power $P_n = omin \cdot P_{min}$ (for details see Martin, 1994). Given the STFT $X(t,k)$ of the noisy signal with time index t and subband index k, the output $Y(t,k)$ is computed as

$$|Y(t,k)| = \begin{cases} \sqrt{subf \cdot P_n(t,k)} & \text{if } |X(t,k)| \cdot Q(t,k) \leq \sqrt{subf \cdot P_n(t,k)} \\ |X(t,k)| \cdot Q(t,k) & \text{else} \end{cases} \qquad \text{(Eq. 1)}$$

where the spectral weighting factor Q is given as

$$Q(t,k) = 1 - \sqrt{osub \cdot \frac{P_n(t,k)}{|X(t,k)|^2}} \qquad \text{(Eq. 2)}$$

To improve the quality of the reconstructed signal, some more parameters have been introduced. Because the noise cannot be estimated perfectly, $Y(t,k)$ contains spectral peaks that change rapidly between frames. In the reconstruction, this leads to audible tonal artefacts with fast changing frequencies that are known as musical noise. To reduce those peaks, the noise power is over-estimated by the factor *osub*. As this can lead to very small and even negative values, the reconstructed signal is bounded from below by a noise floor that can be adjusted by the factor *subf*. For our experiments, the following parameters are used: *subf*=0.001, *osub*=5.5, and *omin*=0.4. An overview of the proposed NR system is shown in Fig 1.

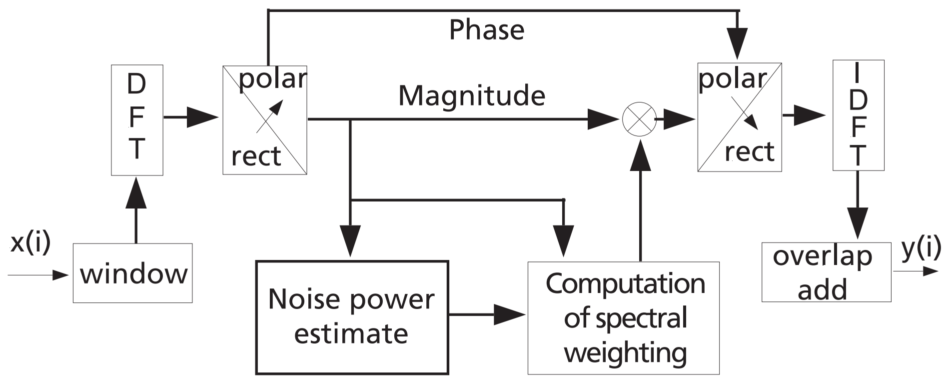


Fig 1: System overview of spectral subtraction algorithm (Martin, 1994).

Noise types

We used two noise types for our tests: a fluctuating speech-shaped noise (abbreviated *ols*) from the Oldenburg sentence test (OLSA, see Wagener *et al.*, 1999) and interior noise of a car driving steadily (abbreviated *car*). The two types of noise were normalized so that their A-weighted sound pressure level was the same (measured with a Phonic PAA3 handheld audio analyzer). The spectrograms are shown in Fig. 2.

Subjects

All subjects in this study were speaking German at the level of a native speaker. All CI users were fitted with a CP910 or CP920 processor using the ACE (Advanced Combination Encoder) strategy. Further details are listed in

Table **1**.

With the hearing subjects (age min=21, max=52.6, Md=27.9 years) we performed bilateral pure-tone audiometry at 500, 1000, and 2000 Hz, and calculated the pure-tone average (PTA). Based on the results (PTA min=5, max=18.3, Md=8.3 dB HL), all subjects could be considered normal-hearing at the time of the study.

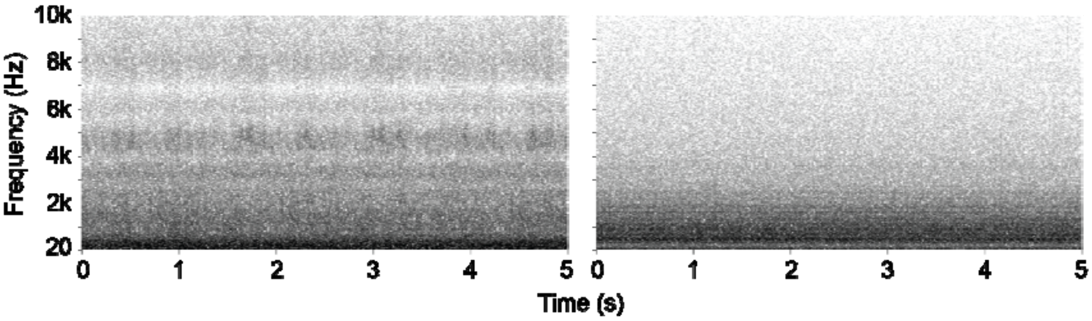


Fig. 2: Spectrograms of the two noise types: *ols* (left) and *car* (right).

Subject	Age (years)	CI (months)	Nucleus implant type	Lateralization	Note
S1	69	11	CI422	Bimodal	Ménière's disease
S2	42	86	CI24RE (CA)	Bilateral	Congenital
S3	13	87	CI24RE (CA)	Bilateral	Congenital
S4	63	14	CI422	Bimodal	Cause unknown
S5	22	15	CI24RE	Bilateral	Meningitis

Table 1: Demographics of cochlear-implanted subjects of the study.

Acoustic simulation of cochlear implant hearing

For normal-hearing listeners, we simulated cochlear implant hearing using the ACE strategy (channel stimulation rate of 900 pps with N=8 selected channels) as described in Chilian *et al.* (2011). Chilian *et al.* extended the signal synthesis of the general vocoder approach by combining two different carrier signals. As a result, both place and rate pitch mechanisms could be simulated. The algorithm also includes models of the electrode-tissue-interface and loudness perception.

In this study, we used the following parameters for the acoustic simulation: λ=8 mm (range of current spread), s=0.25 (synchronisation factor), PLL=300 Hz (phase-locking limit), α_p=0.75 mm, and α_s=4.5 mm (pass-band and stop-band filter bandwidths, respectively, as measured along the cochlea).

Test environment

Listening tests were performed in a soundproof booth (in accordance with the guidelines of ITU-R BS.1116) using a pair of Tapco S5 studio monitors (frequency response flatness: ±3 dB for 64 to 20000 Hz, according to the specifications) with an approximate loudspeakers-to-ears distance of 1 meter, driven by a Creative Sound

Blaster Live! 24-bit external (USB) sound card having excellent frequency response (within ±0.2 dB for 20 to 20000 Hz, measured with RightMark Audio Analyzer 5.5 using an external loopback, at 48-kHz sampling rate and 24-bit resolution).

Test procedure

During the listening tests, we captured the speech reception threshold (SRT) using the Oldenburg sentence test. However, we applied some modifications to the original test procedure, as follows. First, we embedded the sentences in either the original noise (*ols*) or the *car* noise. Second, speech and noise were not spatially separated, but mixed and played back from both loudspeakers. Third, the volume of each processed sentence was set so that the sound pressure level at the ears reached but did not exceed 70 dB SPL(A) during the playback (based on 100-ms measurement windows). Processing steps are shown in Fig. 3.

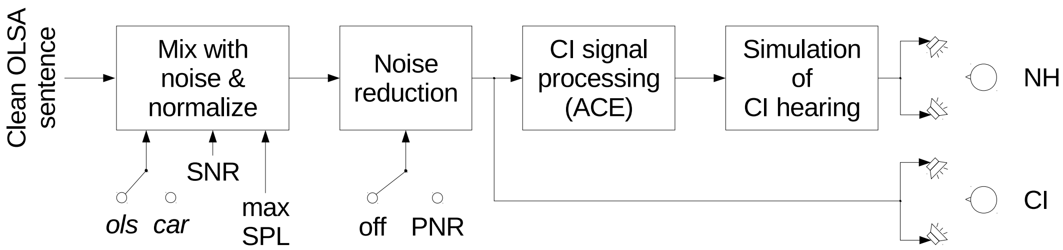


Fig. 3: System overview. NH and CI abbreviates normal-hearing and cochlear-implanted listener, respectively.

Before the listening tests, all subjects were made comfortable with OLSA: After the explanation of the test procedure, subjects examined a table showing all possible words of the OLSA sentences for 3 minutes, which was then followed by a warm-up list of 30 sentences (with feedback). The results of the warm-up list were excluded from any evaluation. During the subsequent actual tests, no feedback was provided. Between lists of 30 sentences, subjects could choose to have a short break for refreshments.

For each CI user, two variants of their everyday CI setting (map) were created: one with built-in NR disabled and one (otherwise identical copy) with built-in NR enabled. The CI processor's program was then switched between the OLSA sentence lists according to the desired test condition.

RESULTS

The results of the listening tests are displayed in Fig. 4. For the NH listeners, the evaluation of the speech reception threshold, which was measured using acoustic simulation of CI hearing with ACE, shows statistically significant benefit with PNR over non-processed noise corrupted speech (for both noise types; statistical test used: paired-sample Wilcoxon signed rank test). The improvements in SRT (median differences) with PNR were 2.45 dB for *car* noise and 1.025 dB for *ols* noise. Speech intelligibility in the presence of *car* noise was better for both unprocessed and processed signals, which we will further elaborate on in the discussion section.

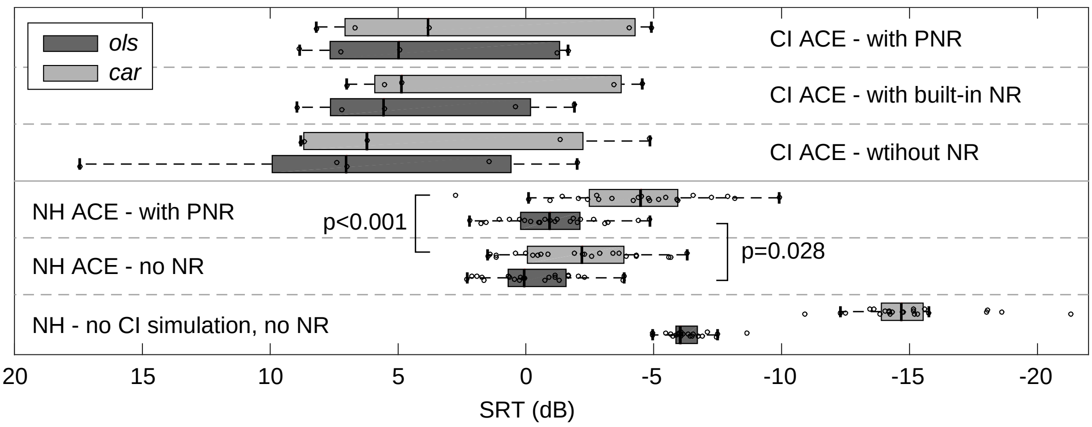


Fig. 4: Overview of the main results.

For CI users, a similar trend can be observed when comparing SRTs for speech corrupted with noise (CI ACE – without NR) and that for PNR applied to the signal before playback (CI ACE – with PNR). However, no statistical significance could be shown, which was likely due to the small sample number. The measured median improvements with PNR in SRT were 2.6 dB for *car* noise and 2.4 dB for *ols* noise, respectively. *Car* noise always allowed for better intelligibility than *ols* noise.

Finally, when comparing the CI's built-in noise reduction stage with PNR, the evaluation showed that both approaches result in improved SRT and that PNR seems on par with the built-in algorithm. The built-in method achieved median SRT improvements of 1.7 dB for *car* noise and 1.3 dB for *ols* noise. Again, to establish statistically significant results, a higher number of test subjects would be desirable.

DISCUSSION

Comparison of OLSA and car noise

Both for CI users and NH subjects, better intelligibility could be observed in the case of *car* noise. Further analysis of the disturbed signals and intermediate stages of the PNR algorithm showed that the spectral shape of the used *car* noise is less destructive to the clean speech signal than that of the speech-shaped noise. The energy of the *car* noise is more stable over time than in the case of *ols* noise so that spectral subtraction can distinguish better between the noise floor and the signal of interest. Furthermore, it is concentrated mostly outside of the individual bands that are important for speech intelligibility, whereas the *ols* noise is concentrated in exactly this region of the spectrum. Fig. 5 illustrates the effect of the two noise types.

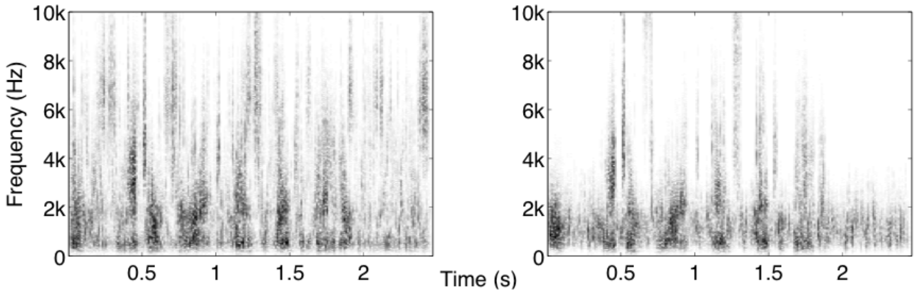


Fig. 5: Spectrograms of noisy speech after noise reduction and CI processing for *ols* noise (left) and *car* noise (right).

Influence of processing order

Commercial CI processors typically implement a pre-emphasis filter, which behaves like a high-pass filter on the audio input of the CI, just before further processing and filtering steps. One issue that needs further analysis is the order of pre-emphasis filtering and noise reduction. Because, for the sake of this study, it was not feasible to implement the PNR algorithm within the processing chain of commercial CI processors, the algorithm was applied to the audio signals before playing them back to the CI users. The built-in pre-emphasis filter of the CI was therefore applied after PNR processing. In a real-world scenario the noise reduction would run within the CI processing after the pre-emphasis.

Because PNR involves taking the maximum between the noise-subtracted spectrum and zero, the operation is non-linear and cannot be exchanged with the pre-emphasis filter, as in the case of linear time-invariant filters. However, in an informal analysis, the signals showed only minor differences when the two processing steps were swapped. Visual inspection of the resulting spectrograms suggests even a slightly improved noise reduction for the order of pre-emphasis followed by noise reduction followed by the CI processing, as can be seen in Fig. 6. Given this, it would be interesting to implement the proposed noise reduction stage within the CI hardware for further analysis.

Future directions

There are several possible extensions to the basic PNR algorithm that could improve its performance, such as multi-band processing and psycho-acoustically motivated spectral subtraction techniques (Zoghlami *et al.*, 2010). Among new approaches in the field of speech enhancement, deep neural networks become more and more popular. In recent studies they showed superior performance to classic methods as well as matrix factorization approaches (e.g., Liu *et al.*, 2014). In addition to using them as a pre-processing stage, such machine learning methods might provide the possibility to work well directly in the coded domain of the CI, which can be an interesting topic for future research.

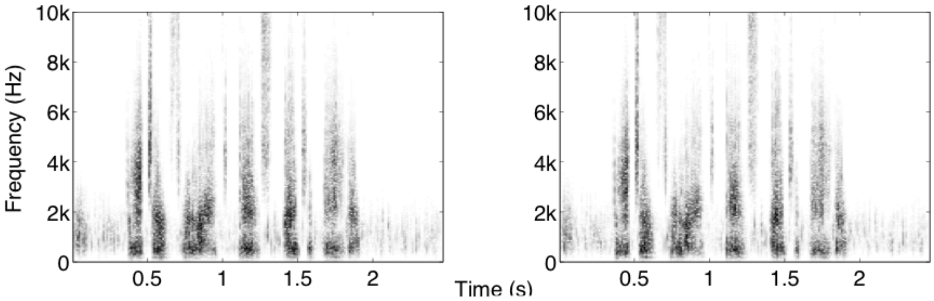


Fig. 6: Comparison of processing orders. Left: noise reduction – pre-emphasis – CI processing. Right: pre-emphasis – noise reduction – CI processing.

REFERENCES

Boll, S. (**1979**). "Suppression of acoustic noise in speech using spectral subtraction," IEEE T. Acoust. Speech, **27**, 113-120.

Chilian A., Braun E., and Harczos T. (**2011**). "Acoustic simulation of cochlear implant hearing," Proc. of 3rd International Symposium on Auditory and Audiological Research (ISAAR 2011) – *Speech Perception and Auditory Disorders*, Nyborg, Denmark, pp. 425-432.

Liu, D., Smaragdis P., and Kim M. (**2014**). "Experiments on deep learning for speech denoising," Proc Annual Conference of the International Speech Communication Association (INTERSPEECH), Singapore.

Loizou, P.C. (**2007**). *Speech Enhancement: Theory and Practice.* CRC Press: Boca Raton, FL, USA.

Martin, R. (**1994**). "Spectral subtraction based on minimum statistics," Proc. European Signal Processing Conference (EUSIPCO) '94, Edinburgh, Scotland, nr. 1, pp. 1182-1185.

Wagener, K., Kühnel, V., and Kollmeier, B. (**1999**). "Entwicklung und Evaluation eines Satztests in deutscher Sprache I: Design des Oldenburger Satztests (Development and evaluation of a sentence test in German language I: Design of the Oldenburg sentence test)," Z. Audiol, **38**, 4-15.

Zoghlami, N., Lachiri Z., and Ellouze N. (**2010**). "Perceptually motivated generalized spectral subtraction for speech enhancement," Advances in Nonlinear Speech Processing, Lecture Notes Artif. Int., **5933**, 136-143.

Perceptual space induced by cochlear implant all-polar stimulation mode

JEREMY MAROZEAU[1,*] AND COLETTE MCKAY[2]

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

[2] *Bionics Institute, East Melbourne, Australia*

It has often been argued that a main limitation of the cochlear implant is the spread of current induced by each electrode, which activates an inappropriately large range of sensory neurons. In order to reduce this spread, a new stimulation mode, the all-polar mode, was tested with 5 participants. It was designed to activate all the electrodes simultaneously with appropriate current levels and polarities to recruit narrower regions of auditory nerves in the region of specific intra-cochlear electrode positions (denoted all-polar electrodes). In this study, the all-polar mode was compared to the current commercial stimulation mode: the monopolar mode. The participants were asked to judge the sound dissimilarity between pairs of 2-electrode stimuli that differed in the electrode positions and were presented in either monopolar or all-polar mode. The dissimilarity ratings were analysed using a multidimensional scaling technique and a three-dimensional stimulus perceptual space was produced. For both modes, the first perceptual dimension was highly correlated with the average position of the electrical stimulation and the second dimension moderately correlated with the distance between the two electrodes. The monopolar and all-polar stimuli were separated by a third dimension, which may indicate that all-polar stimuli have a perceptual quality that differs from monopolar stimuli.

INTRODUCTION

The cochlear implant (CI) is a biomedical device that can restore functional hearing for a large portion of people with severe to profound hearing loss (Blamey *et al.*, 2013). Despite this great success the sound quality produced by the device needs to be improved to help CI users to better understand speech in noise and to enjoy music. In the most common setup (for example, a Cochlear® device with the monopolar ACE strategy), the input signal is band-pass filtered. Then the envelope of the output of each filter is extracted to modulate a fixed-rate electric pulse train that activates specific electrodes. In order to avoid uncontrolled current interaction only one electrode is activated at a time (sequential interleaved stimulation). In the monopolar (MP) mode, each singly-activated intra-cochlear electrode is paired with an extra-cochlear return electrode.

*Corresponding author: jemaroz@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

Recently a new stimulation mode has been designed to better control the current interaction and to create independent and focused places of electrical stimulation along the cochlea. In this mode, called all-polar (AP), all the electrodes are activated simultaneously. The current levels and polarities on each electrode are set in order to create a sum of all potentials that will confine the current field to specific and independent places within the cochlea. Recent behavioral experiments (Marozeau *et al.*, 2015) have shown that AP mode produces less current summation when 2 electrodes are simultaneously activated compared to MP mode. However, when the stimuli were matched in loudness, no significant advantage in terms of spread of neural excitation was found for the AP mode.

This study aimed to investigate the difference between MP and AP modes in the perceptual space generated by dual-electrode stimuli using a multidimensional scaling technique.

METHOD

Participants

Five CI users participated to the experiment (including 3 women). Their age ranged between 44 and 82 years old (mean: 67.2; std: 17) with a duration of deafness before first implantation ranging from 9 to 31 years (mean: 19; std: 8.5). All the participants were unilateral CI users who had received a second research implant on the contralateral side that could be connected to an external stimulator via a percutaneous connector. During an 18-month period, they participated in a number of experiments (for example Marozeau *et al.*, 2015). While not participating in experiments, the participants connected their research implant to a standard sound processor programmed with the ACE strategy (McDermott *et al.*, 1992; Vandali *et al.*, 2000) via a wearable adaptor (van den Honert and Kelsall, 2007). After the research period, participants were explanted and re-implanted with a standard commercial cochlear implant. This project conformed to The Code of Ethics of the World Medical Association (Declaration of Helsinki), and was approved by the Royal Victorian Eye and Ear Hospital Human Research Ethics Committee (Project 11-993H). Recruitment was conducted through the Cochlear Implant Clinic at the Royal Victorian Eye and Ear Hospital and the Hearing CRC.

Stimuli

The stimuli were generated by an experimental stimulator that was able to activate all 22 electrodes simultaneously to produce MP or AP stimuli. In this study we will refer to the electrode around which a focused current field is created in AP mode, by activating all the electrodes simultaneously, as an “AP electrode”. Likewise, the term “MP electrode” designates the single intra-cochlear electrode activated in MP mode. AP electrodes were created by first measuring the impedances between all possible pairs of electrodes. Then a weight matrix that defined the relative current amplitudes across the array predicted to produce the focused current field at each electrode position was derived by inverting the impedances matrix (van den Honert and Kelsall, 2007; Marozeau *et al.*, 2015).

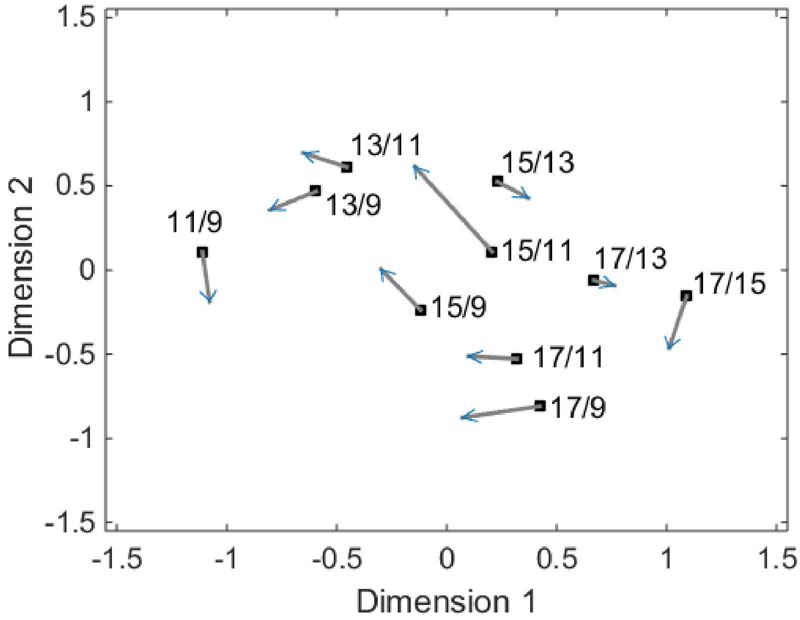
A set of 20 dual-electrode stimuli were created: 10 in AP mode and 10 in MP mode. Each MP stimulus was a 500-ms-duration pulse train, with two biphasic pulses per period of 10 ms. The two pulses were presented sequentially with an onset to onset delay of 232 μs to two different MP electrodes. Each biphasic pulse had a phase width of 100 μs and an interphase gap of 20 μs. The current levels of each electrode were adjusted so that each electrode contributed equally to the overall loudness, and all the dual-electrode stimuli were adjusted to have an equal comfortable loudness (using a loudness balance method described in Marozeau *et al.*, 2015). The MP electrodes were selected in order produce different electrode separations and different average electrode positions: 17/15, 17/13, 17/11, 17/9, 15/13, 15/11, 15/9, 13/11, 13/9, and 11/9[1,2]. Stimuli presented in AP mode were similar in all aspects other than the mode, and were loudness balanced to the MP stimuli.

Task

Participants were presented, first, with each of the 20 stimuli in random order to acquaint them with the range of perceptual differences in the set of stimuli. They were allowed to hear them as many times as they wanted. Then, they were informed that the goal of the experiment was to estimate the similarity in sound quality between pairs of sounds. Remaining small differences of loudness were to be ignored. They were presented with every possible pair of the 20 stimuli in random order, totalling 380 pairs (excluding pairs with repeated stimuli). In each trial, the participants were instructed to judge how similar the pairs were, and to respond by moving a cursor on a slider bar labelled from “most similar” to “least similar”. Participants could listen to the pair as many times as they wanted, by pressing a “listen again” button. When they were satisfied with their judgment, they pressed a “validate” button, and the next trial began.

RESULTS

An MDS solution was derived based on the dissimilarity scores averaged across the five participants. In order to reduce space distortion due to a bound dissimilarity scale, dissimilarity scores were transformed with a hyperbolic arctangent transformation (as in Marozeau and de Cheveigné, 2007). The scores were then analysed using the MDSCAL procedure, implemented according to the SMACOFF algorithm (Borg and Groenen, 1997). A three-dimensional solution was selected because higher-dimensional solutions did not significantly decrease the stress of the model. As the MDSCAL solution is rotationally undetermined, the solution was rotated with a procrustean procedure in order to maximize the correlation between the MDS dimensions and some physical descriptors (described below).



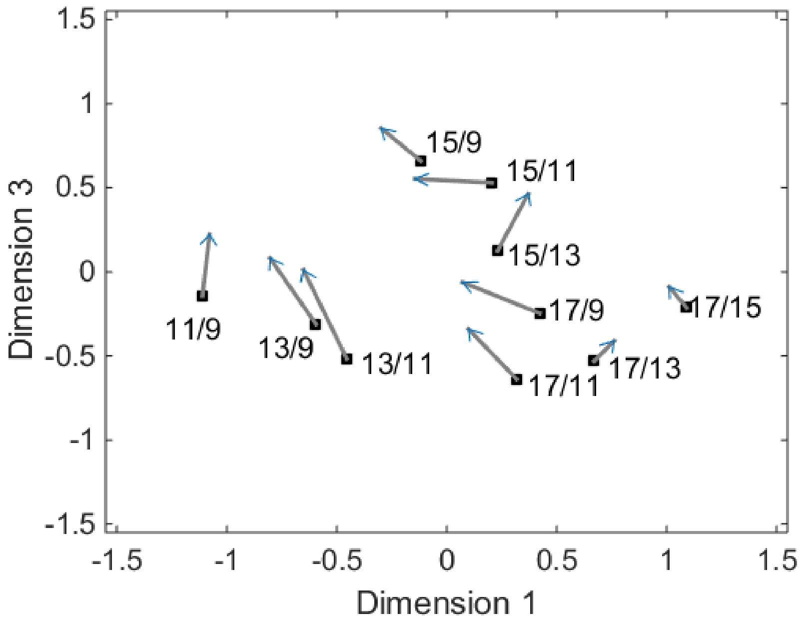


Fig. 1: MDS solution. Each MP stimuli is represented by a square and each AP stimuli is represented by the end of the arrow. The two numbers next to each stimulus indicate the “AP” and “MP” electrodes activated. Each MP and AP stimulus that shared the same activated electrodes are linked by an arrow.

Figure 1 shows the 3-dimensional solution. Each MP stimulus is represented by a square and each AP stimulus is represented by the end of the arrow. The two numbers next to each stimulus indicate the AP and MP activated electrodes. The MP and AP stimuli that shared the same activated electrodes are linked by an arrow. The figure shows that the stimuli are grouped into clusters based on the most apical electrode. The projection on the first dimension is highly correlated with the average activated electrode position [$R^2 = 0.93$, $df = 19$, $p < 0.0001$]. The second dimension is significantly correlated with the distance between the two activated electrodes [$R^2 = 0.44$, $df = 19$, $p = 0.001$]. Two features can be observed on the third dimension: first, the stimuli with electrode 15 as the most apical (15/13, 15/11, and 15/9) are separated in that dimension from the other stimuli; secondly, the AP stimuli are consistently separated from the MP stimuli (i.e., the arrows are always pointing upward).

Figure 2 shows the average difference of the projection on each dimension between the position of the MP stimuli and their AP counterparts. On average, in the first dimension, AP stimuli are located on the left of the MP stimuli [$t(9) = 2.42$, $p = 0.0389$], and upward on the third dimension [$t(9) = -5.3008$, $p < 0.0001$]. No significant difference can be observed on the second dimension [$t(9) = 0.0780$, $p = 0.9395$].

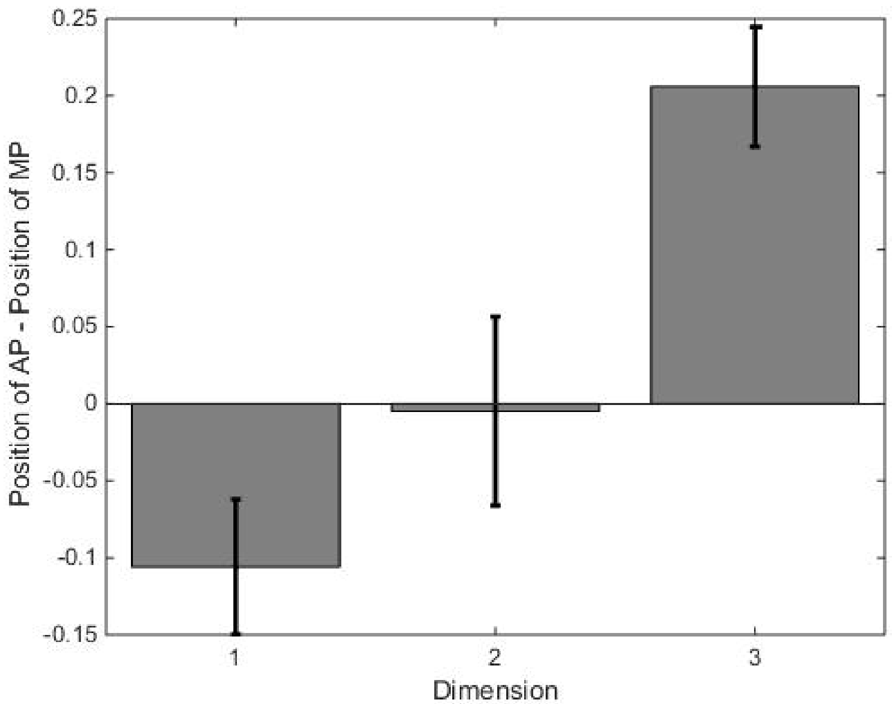


Fig. 2: Average difference of the projection on each dimension between the position of the MP stimuli and their AP counterparts.

DISCUSSION

The main aim of the experiment was to study the effect of stimulation mode (AP vs MP) on a perceptual space. Overall, the AP stimuli and their MP counterparts are always located closely to each other on the three dimensional space. However, two clear effects can be observed. First, as showed in Fig. 2, on average AP stimuli are shifted toward the left of first dimension compared to the MP stimuli. As this dimension is clearly correlated with the average activated electrode position, it can be interpreted as a dimension linked to the perception of pitch, ranging from high pitch stimuli on the left to lower pitch on the right. This will indicate that AP stimuli are perceived as higher in pitch than MP stimuli. Second the effect of mode can be seen on the third dimension as an upward shift. It is unclear why on that dimension the stimuli 15/13, 15/11 and 15/9 are also shifted upward in both modes compared to the other electrode positions. However, it is possible that the 3-D solution is composed of bended 2-D plans like a half cylinder (or a horse saddle). This kind of distortion is often found in MDS studies, where a 1-D solution is represented as a horse shoe in a 2-D solution (for example McKay *et al.*, 1996). If this distortion is ignored, then the third dimension clearly separated the MP and AP stimuli. This result would indicate that the AP mode differed from the MP mode along a perceptual dimension that was independent of electrode position and separation.

The first two dimensions of the 3-D solution can be strongly correlated with simple physical descriptors. Those descriptors are the CI equivalent of common acoustical descriptors of timbre: the spectral centroid and the spectral spread (see Marozeau *et al.*, 2003 for a complete description). This indicates that the perception of those dimensions might be similar to the perception of timbre by normal hearing listeners (Kong *et al.*, 2009; Kong *et al.*, 2012).

Similar results were previously found by McKay *et al.* (1996). They asked four CI participants to rate the dissimilarity between pairs of dual-electrode bipolar stimuli that varied in electrode separation and overall position. The bipolar stimulation widths were also varied with two distances between the active and return electrodes of the bipolar pair in order to test the effect of current spread. They found that for most CI participants a two dimensional solution related to the average activated electrode position and the activated electrode separations. They also found similar MDS solutions with the two bipolar stimulation widths. However, as this parameter was not varied within the same session, it was not possible to assess whether the width of the bipolar stimuli produced an isometric shift along a specific dimension as observed in the current experiment.

ACKNOWLEDGMENTS

The authors would like to thank the five participants who selflessly dedicated 18 months of their life to this study. We would like also to thank Brett Swanson and Cochlear Ltd for technical support, and the Hearing CRC for their clinical support. This project was funded by The Garnett Passe and Rodney Williams Memorial Foundation and supported by Cochlear Ltd. The Bionics Institute acknowledges the support it receives from the Victorian Government through its Operational Infrastructure Support Program.

ENDNOTES

[1] The electrode will be identified by the Cochlear Ltd convention in which electrode 22 is the most apical electrode, and electrode 1 the most basal.

[2] The stimuli were shifted basally by 2 electrode positions for one participant in order to avoid high AP threshold regions. The relative electrode positions were identical to those of the other subjects.

REFERENCES

Blamey, P., Artieres, F., Baskent, D., Bergeron, F., Beynon, A., *et al.* (**2013**). "Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients," Audiol. Neurootol., **18**, 36-47.

Borg, I. and Groenen, P.J.F. (**1997**). *Modern Multidimensional Scaling: Theory and Applications.* Springer, New York.

Kong, Y.Y., Tansman, P., Marozeau, J., and Epstein, M. (**2009**). "Perceptual dimensions for musical timbre in cochlear-implant users," Am. Audit. Soc. Annu. Meet.

Kong, Y.Y., Mullangi, A., and Marozeau, J. (**2012**). "Timbre and speech perception in bimodal and bilateral cochlear-implant listeners. Ear Hearing, **33**, 645-659.

Marozeau, J., de Cheveigné, A., McAdams, S., and Winsberg, S. (**2003**). "The dependency of timbre on fundamental frequency," J. Acoust. Soc. Am., **114**, 2946-2957.

Marozeau, J. and de Cheveigné, A. (**2007**). "The effect of fundamental frequency on the brightness dimension of timbre," J. Acoust. Soc. Am., **121**, 383-387.

Marozeau, J., McDermott, H.J., Swanson, B., and McKay, C.M. (**2015**). "Perceptual interactions between electrodes using focused and monopolar cochlear stimulation," J. Assoc. Res. Otolaryngol., **16**, 401-412.

McDermott, H.J., McKay, C.M., Vandali, A.E. (**1992**). "A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant," J. Acoust. Soc. Am., **91**, 3367-3371.

McKay, C.M., McDermott, H.J., and Clark, G.M. (**1996**). "The perceptual dimensions of single-electrode and nonsimultaneous dual-electrode stimuli in cochlear implantees," J. Acoust. Soc. Am., **99**, 1079-1090.

Vandali, A.E., Whitford, L.A., Plant, K.L., and Clark, G.M. (**2000**). "Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system," Ear Hearing, **21**, 608-624.

van den Honert, C. and Kelsall, D.C. (**2007**). "Focused intracochlear electric stimulation with phased array channels," J. Acoust. Soc. Am., **121**, 3703-3716.

Coding of interaural phase differences in BiCI users

STEFAN ZIRN*, SUSAN ARNDT, AND THOMAS WESARG

Department of Oto-Rhino-Laryngology of the Medical Center, University of Freiburg, Germany

The ability to detect a signal masked by noise is improved in normal-hearing (NH) listeners when interaural phase differences (IPD) between the ear signals exist either in the masker or the signal. We determined the impact of different coding strategies in bilaterally implanted cochlear implant (BiCI) users with and without fine-structure coding (FSC) on masking level differences. First, binaural intelligibility level differences (BILD) were determined in NH listeners and BiCI users using their clinical speech processors. NH subjects (n=8) showed a significant mean BILD of 7.5 dB. In contrast, BiCI users (n=9) without FSC as well as with FSC revealed a barely significant mean BILD (0.4 dB respectively 0.6 dB). Second, IPD thresholds were measured in BiCI users using either their speech processors with FS4 or direct stimulation with FSC. With the latter approach, synchronized stimulation providing an interaural accuracy of stimulation timing of 1.67 μs was realized on pitch matched electrode pairs. The resulting individual IPD threshold was lower in most of the subjects with direct stimulation than with their speech processors. These outcomes indicate that some BiCI users can benefit from increased temporal precision of interaural FSC and adjusted interaural frequency-place mapping presumably resulting in improved BILD.

INTRODUCTION

Interaural timing cues are important for normal-hearing (NH) listeners for sound source localization and for binaural unmasking of speech in the presence of spatially separated interfering sounds (e.g., Moore 2012; Colburn *et al.*, 2006).

In bilaterally implanted cochlear implant (BiCI) users sound source localization as well as binaural unmasking of speech are impaired compared to NH listeners. The main limitation may arise from limited availability of interaural timing cues when using their clinical devices, whereas interaural level differences can be perceived with a considerably higher precision (Kerber *et al.*, 2012; Seeber *et al.*, 2008; van Hoesel *et al.*, 2008).

On the other hand, BiCI users show considerable sensitivity to interaural time differences (ITD) in bilaterally synchronized electric pulse trains with and without on-/offset differences. Especially Laback *et al.* (2007) showed that even ongoing ITDs are perceivable to selected BiCI users at low pulse rates.

*Corresponding author: stefan.zirn@uniklinik-freiburg.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

Furthermore, Goupell *et al.* (2015) measured binaural masking level differences (BMLDs) in constant high-rate pulse trains. They found average BMLDs up to 11 dB.

These outcomes on single electrode pairs are promising for real-life benefits with multi-electrode stimulation. However, if and how single electrode experiments in this context translate to multi-electrode scenarios is unknown. With the fine-structure coding strategy 'FS4' from MED-EL, a coding strategy that processes interaural phase differences (IPDs) is clinically available.

Aim of this work

With this work we investigated if the fine-structure coding strategy FS4 enables binaural unmasking of speech in terms of measurable binaural intelligibility level differences (BILD). As a reference BILD was evaluated in the same BiCI users programmed with HDCIS and in NH listeners.

Furthermore, IPD thresholds with signals processed by the clinical CI processors programmed with FS4 or HDCIS were compared with the same signals processed with synchronized bilateral direct stimulation obtained with a research platform.

METHODS

Stimuli

For measuring the influence of IPD in broad-band signals, the binaural intelligibility level difference (BILD) has been measured similar to the approach first described by Licklider (1948). Speech material was taken from the German Oldenburger Sentence Test (OLSA, 2011). The masker was OLnoise, a steady-state noise with a speech shaped spectrum. The BILD was determined as the difference in speech reception threshold (SRT) between two listening conditions. First, a diotic condition (speech in noise on both ears, no difference between ear signals) and a dichotic condition (speech in noise on both ears, the phase of the speech signal was inverted on one ear).

With speech in noise two interaural cues are available to the listeners: interaural envelope differences and interaural fine-structure differences. To examine the effect of interaural fine-structure differences exclusively, IPD sensitivity was measured using narrow-band signals. For this purpose, a 150 Hz pure tone was ramped up and down with hann windows. On the right ear, the pure tone was presented without any phase variations (Eq. 1).

$$y(t) = A_c \sin(\omega_c t) \qquad (1)$$

The phase of the left pure tone, however, was modulated according to Eq. 2.

$$y(t) = A_c \sin(\omega_c t + m(t) + \varphi_c) \qquad (2)$$

where $\omega_c = 2 * \pi * f_c$, f_c is the carrier frequency (150 Hz), $m(t)$ is the sinusoidal modulation signal, and $\varphi_c = 0°$.

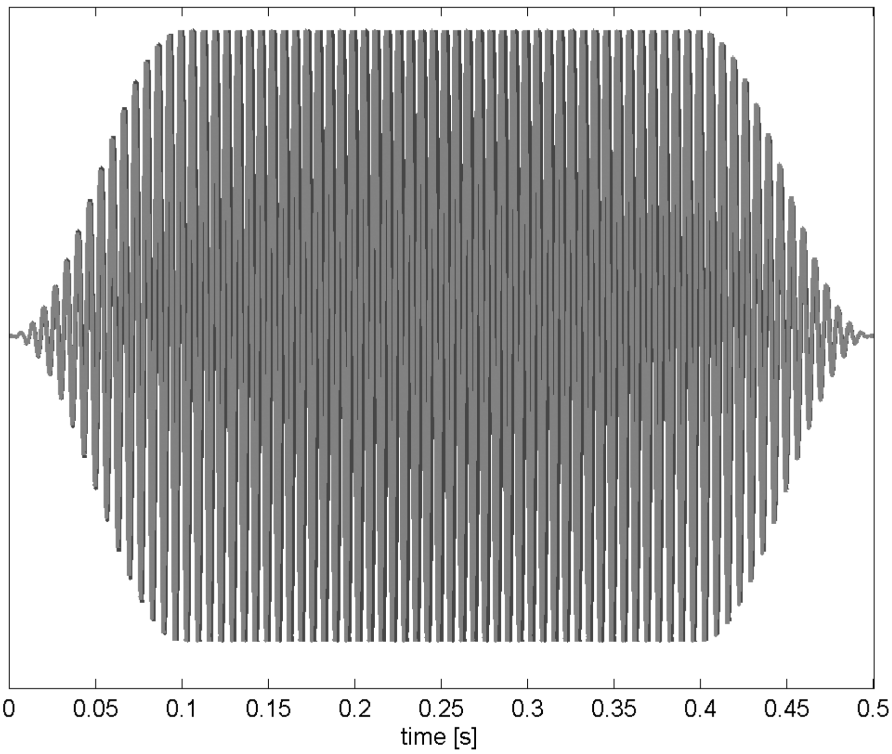


Fig. 1: IPD stimuli – two 150 Hz sinusoidal tones. The left ear signal (dark grey) was phase modulated which then lags behind the right ear signal (light grey). Note that no envelope differences between the ear signals exist.

150 Hz corresponds to the center frequency of CI channel 1 in the typical FS4 map with all 12 CI channels activated.

Procedure

For the IPD test, a 3-alternative forced-choice paradigm for masked threshold determination was implemented. For every two correct answers one after one other, the IPD was lowered, one false response lead to an increase of the IPD (2-down 1-up procedure).

For the BILD effect speech understanding of 5 word OLSA sentences was investigated with an adaptive procedure as well. For every sentence with more than two words correct, the signal-to-noise ratio was lowered. For every response with two or less than two words correct, the signal-to-noise ratio was increased. The procedure was similar to that described in the OLSA documentation (OLSA, 2011). In contrast to the BILD procedure described in this documentation, the signals were not presented in free field, but using the auxiliary inputs of the speech processors in BiCI users and headphones in NH listeners.

The BILD was defined similarly to Licklider (1948) and to Goverts and Houtgast (2010): the difference between the speech-reception threshold in noise in diotic presentation mode (SRT N_0S_0) and in dichotic presentation mode with antiphasic speech (SRT N_0S_π).

Participants, cochlear implants, and stimulation hardware

Nine BiCI users and eight NH listeners participated in the broad-band BILD experiment. Six BiCI users and three NH listeners participated in the narrow-band IPD experiment. All BiCI users had two cochlear implants from MED-EL types PULSAR, SONATA, CONCERTO, or SYNCHRONY with electrode arrays of either 31.5 or 28 mm length. All of them used two OPUS 2 processors.

Acoustic stimuli were generated using a PC, MATLAB, and soundcard type RME Fireface UC with 44.1 kHz sampling frequency and 16 bit quantization depth. The stimuli were presented to the BiCI users using a y-audio cable and the auxiliary inputs of the OPUS 2 processors.

Electric stimuli (biphasic current pulse trains) were generated using the RIB2 direct stimulation platform manufactured at the University of Innsbruck, Austria with custom made MATLAB code. The coding strategy implemented for direct stimulation with the RIB2 was orientated at the FS4 strategy. The major difference was that stimulation was synchronized across ears, which is possible with the RIB2. Furthermore, an increased sampling frequency relative to the 6000-12000 Hz applied with FS4 was implemented. With direct stimulation we used 1 MHz. This led to a higher temporal precision of zero-crossing determination. It was then limited by the RIB2 with a temporal precision of 1.67 µs. In the following we call this coding strategy "Fine HighPrecision".

RESULTS

BILD

The BILD results of NH listeners and BiCI users are shown in Fig. 2.

The SRT of NH listeners in the diotic condition was −7.1 ± 0.8 dB SNR (mean ± standard deviation), in the dichotic condition −14.6 ± 1.6 dB SNR. The BILD was considered as the difference of these SRTs. In this group of NH listeners, the mean BILD was 7.5 dB. Statistical analysis with the Wilcoxon signed-rank test revealed that this BILD was significant (p=0.008).

In BiCI users programmed with FS4, the SRT in the diotic condition was −2.1 ± 1.8 dB SNR, in the dichotic condition −2.6 ± 1.9 dB SNR. In this group of BiCI users, the mean BILD was 0.5 dB with FS4. This BILD was statistically significant (p=0.05).

The same BiCI users programmed with HDCIS reached an SRT in the diotic condition of −1.4 ± 1.9 dB SNR, in the dichotic condition −2.0 ± 2.0 dB SNR. The resulting mean BILD was 0.6 dB with HDCIS. This BILD was also statistically significant (p=0.02).

No significant difference between the SRTs using either HDCIS or FS4 occurred. This held for the diotic and the dichotic condition.

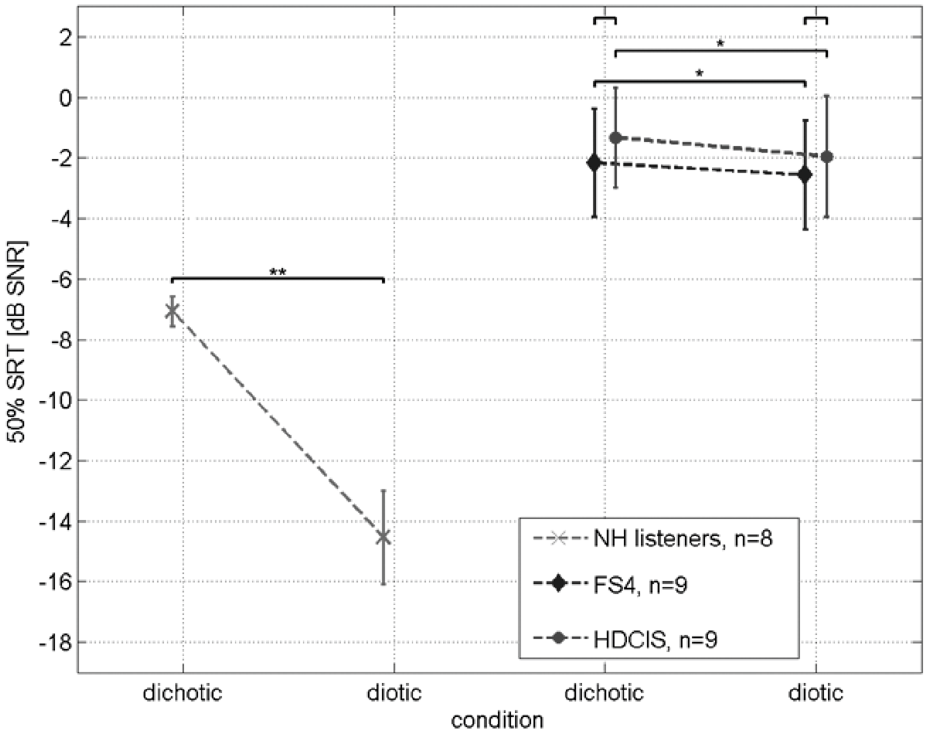


Fig. 2: Results of the BILD experiment. The data of NH listeners is shown on the left. The results of BiCI users tested with two different coding strategies (HDCIS and FS4) are shown on the right. An adaptation phase of three months was applied in each BiCI user to familiarize to changes in the coding strategy.

IPD threshold measurements

Figure 3 shows the IPD thresholds obtained in NH listeners and BiCI users.

The three NH listeners reached a mean IPD threshold of 25° which corresponds to 0.46 ms.

The results of the 6 BiCI users included in this experiment were very inhomogeneous and dependent on the coding strategy. The constant rate coding strategy HDCIS led to the worst IPD threshold of 180° or 3.3 ms, which was defined as the upper limit of the test. The same BiCI users programmed with FS4 reached lower IPD thresholds except two. The lowest IPD thresholds were achieved with the Fine HighPrecision coding strategy in every BiCI user.

Two out of the six BiCI users (CI1 and CI6) reached IPD thresholds close to those of NH listeners.

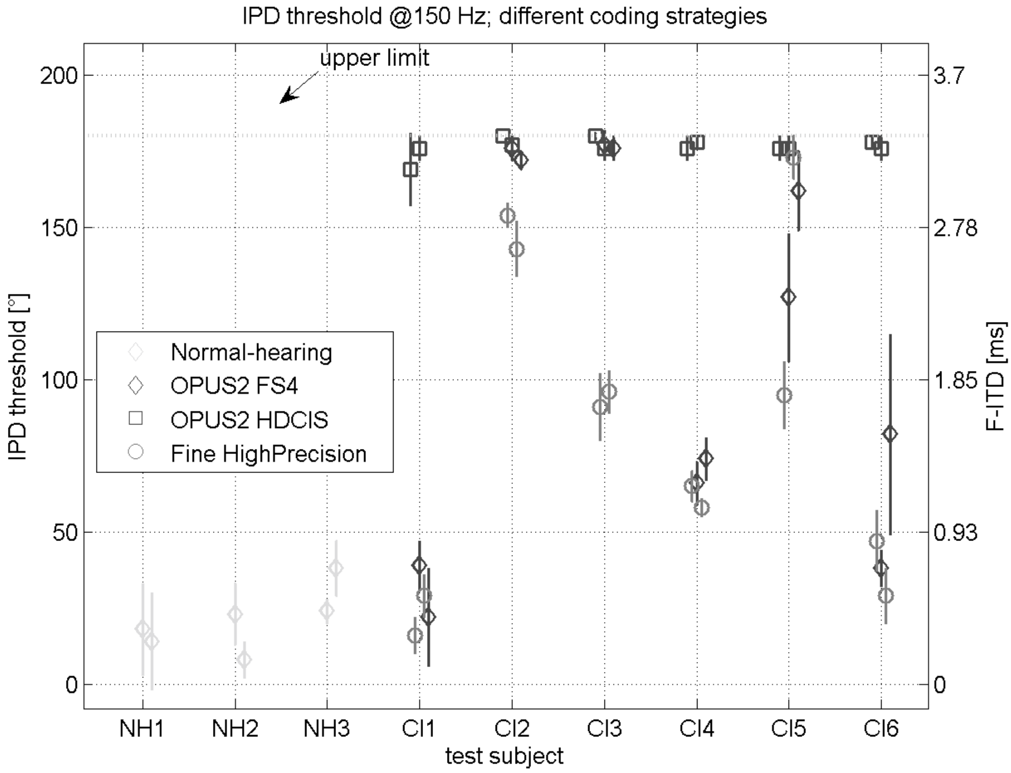


Fig. 3: Results of the phase modulation experiment. The data of NH listeners are shown on the left. The results of BiCI users are shown on the right.

DISCUSSION

The broad-band experiment showed that SRTs as well as BILD were not dependent on the coding strategy in our group of BiCI users. With both FS4 and HDCIS, the included BiCI users reached significantly lower SRTs in the dichotic condition, i.e., with interaural phase differences. Thus, envelope and fine-structure information coded with FS4 was as effective as coding of the envelope only in this group of BiCI users. An explanation might be that with HDCIS the representation of the envelope is more precise than with FS4 in the apical CI channels.

The narrow-band experiments were designed in a way that no interaural envelope differences occurred. According to this, the IPD thresholds with HDCIS were at the upper limit of the test (180° corresponding to 3.3 ms). Thus 0% of the BiCI users showed IPD sensitivity with HDCIS. With FS4, four out of six BiCI users (66%) reached better IPD thresholds than with HDCIS, whereas all six BiCI users (100%) showed better IPD sensitivity with Fine HighPrecision.

Consequently, some BiCI users can benefit from increased temporal precision of interaural fine-structure coding and adjusted interaural frequency-place mapping. With such a high precision fine-structure coding strategy an improved BILD might be achieved provided that the effect is still present in a multi-channel stimulation strategy.

However, the ideal way of interaural frequency-place mapping is still a topic of discussion. Hu and Dietz (2015) pointed out recently that the optimal interaural electrode pairing method might not be pitch matching as done in this work. They compared three such methods namely pitch matching, ITD sensitivity, and/or binaural interaction potentials (BIC). Another study by Kan *et al.* (2013) supports this approach. They showed that lateralization in BiCI users was still possible with up to 3 mm of interaural mismatch determined by pitch matching. But they also pointed out that mismatched inputs might not be ideal since it leads to a distorted auditory spatial map. On the other hand, the auditory system is adaptive. Therefore, it is still a topic of discussion how to optimize the frequency-place mapping for bilateral CI stimulation in order to achieve an optimized binaural multi-electrode stimulation strategy.

ACKNOWLEDGEMENTS

This work was supported by MED-EL Elektromedizinische Geraete Gesellschaft m.b.H., Innsbruck, Austria.

REFERENCES

Goupell, M.J. and Litovsky, R.Y. (**2015**). "Sensitivity to interaural envelope correlation changes in bilateral cochlear-implant users," J. Acoust. Soc. Am., **137**, 335-349.

Goverts, S.T. and Houtgast, T. (**2010**). "The binaural intelligibility level difference in hearing-impaired listeners: the role of supra-threshold deficits," J. Acoust. Soc. Am., **127**, 3073-3084.

Hu, H. and Dietz, M. (**2015**). "Comparison of interaural electrode pairing methods: Pitch matching, ITD sensitivity and binaural interaction component," Conference on Implantable Auditory Prostheses, Lake Tahoe, California.

Kan, A., Stoelb, C., Litovsky, R.Y., and Goupell, M.J. (**2013**). "Effect of mis-matched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users," J. Acoust. Soc. Am., **134**, 2923-2936.

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

Laback, B., Majdak, P., and Baumgartner, W.D. (**2007**). "Lateralization discrimination of interaural time delays in four-pulse sequences in electric and acoustic hearing," J. Acoust. Soc. Am., **121**, 2182-2191.

Licklider, J. (**1948**). "The influence of interaural phase relations upon the masking speech by white noise," J. Acoust. Soc. Am., **20**, 150-159.

Moore, B.C.J. (**2012**). *An Introduction to the Psychology of Hearing.* 6th Ed., Brill.

OLSA (**2011**). *Oldenburger Satztest. Bedienungsanleitung für den manuellen Test auf Audio-CD.* http://www.hoertech.de/web/dateien/HT.OLSA_Handbuch_Rev01.0_mitUmschlag.pdf

Seeber, B.U. and Fastl, H. (**2008**). "Localization cues with bilateral cochlear implants," J. Acoust. Soc. Am., **123**, 1030-1042.

Colburn, H.S., Shinn-Cunningham, B., Kidd, G., Jr., and Durlach. N. (**2006**). "The perceptual consequences of binaural hearing," Int. J. Audiol. Suppl., **45**, S34-44.

van Hoesel, R., Bohm, M., Pesch, J., Vandali, A., Battmer, R.D., and Lenarz, T. (**2008**). "Binaural speech unmasking and localization in noise with bilateral cochlear implants using envelope and fine-timing based strategies," J. Acoust. Soc. Am., **123**, 2249-2263.

Long-term changes in music perception in Korean cochlear implant listeners

GWANGHUI RYU AND YANG-SUN CHO*

Department of Otorhinolaryngology-Head and Neck Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea

The purpose of this study was to assess long-term post-implant changes in music perception in cochlear implant (CI) listeners using the Korean version of the Clinical Assessment of Music Perception (K-CAMP) test and questionnaire for music listening. Twenty-seven patients, including 5 men and 22 women participated in this prospective study. Their music perception ability was evaluated with the K-CAMP test which consists of pitch discrimination, melody, and timbre identification. Also, a questionnaire was used to quantify listening habits, and level of musical experience. Median postoperative durations of the first and second test were 12.8 and 30.9 months. Participants were divided into two groups: good or poor performance in the first test with reference to the average of each performance. Among the demographic factors, the good performance group was younger than the poor performance group at the time of the test, and the ability of pitch discrimination decreased with aging at 262 Hz for the first test and at 391 Hz for the second test. Pitch discrimination in the second test in the good performance group showed no difference with the first test, but in the poor performance group, the pitch discrimination score significantly improved. Similarly, timbre test results significantly improved in the poor performance group. In the melody identification test, the two groups showed no change at the second test. Scores for listening habit and level of musical experience significantly decreased postoperatively and did not recover during the follow-up period. The pitch discrimination and timbre identification ability improved in the CI listeners who had poor ability shortly after surgery. However, the ability of melody identification showed no difference in both groups after the lapse of time. Age was related to pitch discrimination and younger people showed good performance. Listening habits and level of musical experience decreased after CI surgery without time-dependent improvement.

INTRODUCTION

While cochlear implants (CIs) are remarkably effective in speech perception, they are less adequate for listening to music. However, music appreciation and perception is quite important in the daily life of CI recipients. Thus, music perception is challenging for CI listeners and the majority of CI users reported music to sound strange, noisy, unnatural, and mechanical. As music is connected with the everyday environment and

*Corresponding author: yscho@skku.edu

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

emotional communication, perceptibility of music for CI recipients is related to quality of life and social integration (Gfeller and Knutson, 2003). Efforts to develop technology for music perception are underway, and further researches are warranted to better define the music perception and appraisal in CI users.

Most studies compared the music perception ability of CI listeners with normal-hearing listeners. Adult CI recipients showed poorer results than normal-hearing persons or hearing aid users on recognition of pitch, melody, and timbre except simple rhythmic patterns (Kong *et al.*, 2004; Gfeller *et al.*, 2008; Looi *et al.*, 2008). Regarding rhythmic perception, CI users can perceive simple rhythm patterns as well as normal-hearing listeners (McDermott, 2004; Looi *et al.*, 2012) as their temporal resolution skills are similar to normal listeners. However, CI listeners are definitely poorer on pitch and spectral-based tasks than normal-hearing listeners (Limb and Rubinstein, 2012).

The Korean version of the Clinical Assessment of Music Perception (K-CAMP) test (Jung *et al.*, 2010) is a modified version of the University of Washington's Clinical Assessment of Music Perception (UW-CAMP) test (Nimmons *et al.*, 2008), which consists of pitch discrimination, melody identification, and timbre identification.

A number of studies were conducted to establish the effectiveness of the CAMP test with post-lingually deafened patients. Kang *et al.* (2009) reported a validity and test-retest reliability of the CAMP test. The CAMP test is feasible in a clinical setting because of relatively short test time of about 30-40 minutes and excellent test-retest reliability.

Various literatures proved significant improvement in speech perception ability during the post-implantation follow-up (Krueger *et al.*, 2008; Lenarz *et al.*, 2012). However, there are few evidences which proved the improvement of music perception according to the post-operative periods. In a longitudinal cohort study, Gfeller et al Gfeller *et al.* (2010) revealed modest improvement in familiar melody recognition and recognition of melody excerpts with lyrics (MERT-L) from year 1 to year 2. The suggested predictors of improvement were hearing aid use, bilateral CI use, and musical training experience.

The present study aimed to investigate long-term changes of music perception in post-lingually deafened adult CI listeners using the K-CAMP test after implantation. Also authors evaluated musical listening patterns with questionnaires for listening habits and level of musical experience.

MATERIAL AND METHODS

Patients

A total of 27 post-lingually deafened adults (mean age 49 years at the time of surgery; range from 19.2 to 69.9 years; SD = 12.7) who underwent CI surgery unilaterally at a tertiary referral center between October 2001 and January 2013 were enrolled in this study. CI devices from 3 manufacturers (Cochlear®, Advanced Bionics®, and Med-El®) were implanted. After cochlear implantation, 8 patients were bimodal users. The

K-CAMP test was scheduled to be conducted at annual follow-up visits, and the patients were evaluated at least two times after cochlear implantation. They were also asked to complete a questionnaire to quantify listening habits and level of musical experience. The Institutional Review Board of the Samsung Medical Center approved this study (IRB No. 2010-01-080).

K-CAMP test

Three subtests are comprised in the protocol of the CAMP test, which were pitch discrimination, melody, and timbre identification (Nimmons *et al.*, 2008).

The pitch discrimination test was performed with a one-up, one-down procedure (Levitt, 1971) and two-alternative forced-choice using three base frequencies (262, 330, 391 Hz). The tones were presented at 65 dBA and lasted 760 ms (Nimmons *et al.*, 2008; Drennan *et al.*, 2015) and the test started with a 12-semitone interval. The patient chose the higher pitch between two buttons on computer screen. The threshold was calculated with the mean interval size for 3 adaptive tracks.

The melody identification test used 12 isochronous melodies in a close-set task and the test was finished with 36 presentations. In the K-CAMP test, 10 of 12 melodies were changed into familiar melodies for the Korean population (Jung *et al.*, 2010). Melodies were played with 500-ms duration in an 8-note pattern and a tempo of 60 beats per minute.

The timbre identification test consisted of 24 presentations with 8 musical instruments: piano, violin, cello, acoustic guitar, trumpet, flute, clarinet, and saxophone. Melody and timbre identification were estimated with total percent correct score.

In bimodal listeners, the contralateral hearing aid was removed and the test was performed under a CI-only condition.

Questionnaires

Questionnaires to quantify listening habits and level of musical experience were completed at the same time as the K-CAMP test. Music listening habit inquiries were comprised of 4 items with interest in music before hearing loss and after implantation, mean time of music listening hours per week before hearing loss and after implantation. Answers of music listening habit were graded from 1 to 4 using Likert-type scales. Level of musical experience consisted of 5 yes/no questions prior to hearing loss and after implantation. Responders were asked status prior to hearing loss, including enjoying listening to music, enjoying listening to the radio, visiting indoor or outdoor concerts, attending musical ensembles (e.g., band, choir, or orchestra), and participation in musical lessons. Two different items were inquired instead of the second and third questions, such as difficulty in communicating with background music and difficulty in appreciating unfamiliar music. The answer was to score a ‘1’ if that factor improved musical experience.

Variables		Total (N=27)	Good (N=16)	Poor (N=11)	p-value
Age at the operation	Mean (SD), years	49.0 (12.7)	45.3 (11.3)	54.2 (13.1)	0.072*
Age at the 1^{st} test	Median (IQR), years	52.5 (15.3)	48.9 (15.4)	60.9 (20.2)	**0.038**†
Age at the 2^{nd} test	Median (IQR), years	53.5 (15.0)	51.7 (14.8)	63.6 (19.9)	**0.030**†
POD at the 1^{st} test	Median (IQR), months	12.8 (16.4)	12.3 (17.3)	13.3 (12.3)	0.790†
POD at the 2^{nd} test	Median (IQR), months	30.9 (26.0)	33.6 (26.5)	26.7 (26.0)	0.942†
Interval of tests (1^{st} to 2^{nd} test)	Median (IQR), months	18.0 (20.2)	18.3 (19.2)	17.2 (22.0)	0.767†
Gender	Male vs. Female, No. (%)	5 : 22 (18.5 : 81.5)	3 : 13 (18.8 : 81.2)	2 : 9 (18.2 : 81.8)	1.000‡
Final education	Elementary, No. (%)	3 (11.1)	2 (12.5)	1 (9.1)	
	Middle school	4 (14.8)	1 (6.2)	3 (27.3)	0.323‡
	High2 school	14 (51.9)	8 (50.0)	6 (54.5)	
	College	6 (22.2)	5 (31.2)	1 (9.1)	
Duration of deafness	Median (IQR), years	10.0 (15)	7.5 (19)	10.0 (15)	0.618†
CI manufacturer	Cochlear, No. (%)	11 (40.8)	4 (25.0)	7 (63.6)	
	Med-El	6 (22.2)	5 (31.2)	1 (9.1)	
	Advanced Bionics	10 (37.0)	7 (43.8)	3 (27.3)	0.116‡
Year of implantation	Before 2008, No. (%)	8 (29.6)	5 (31.2)	3 (27.3)	1.000‡
	From 2008	19 (70.4)	11 (68.8)	8 (72.7)	

Abbreviations. CI: cochlear implant, IQR: interquartile range, POD: post-operative day, SD: standard deviation

* Two-sample t-test, † Mann-Whitney test, ‡ Fisher's exact test

Table 1: Demographic data of patients. Univariate analysis of the good and poor performance group based on the mean of pitch discrimination.

Statistical analysis

Data are expressed as mean ± SD unless otherwise stated. Results were analyzed using SPSS 22 (SPSS, Inc., Chicago, Illinois) using a two sample t-test, Wilcoxon's rank sum test or Fisher's exact test, and Spearman correlation analysis with statistical significance set at 0.05.

RESULTS

Demographics

Twenty-seven patients, including 5 men and 22 women participated in this prospective study. Demographic data are shown in Table 1. Median postoperative durations of the first and second test were 12.8 and 30.9 months. Half of the patients (13 patients) completed their first test within one year after surgery. Participants were divided into two groups: good or poor performer in the first test with reference to the average of each performance. Univariate analysis of demographic factors between the good and poor performance group based on the mean of pitch discrimination demonstrated no difference in gender, final education, duration of deafness, CI manufacturer, surgeon, and year of implantation (Table 1). However, the good performance group was younger than the poor performance group at the time of the test (p = 0.038 for the first test, p = 0.030 for the second test). The ability of pitch discrimination decreased with aging at 262 Hz for the first test (p = 0.042, ρ = 0.765) and at 391 Hz for the second test (p = 0.013, ρ = 0.473 by Spearman correlation analysis).

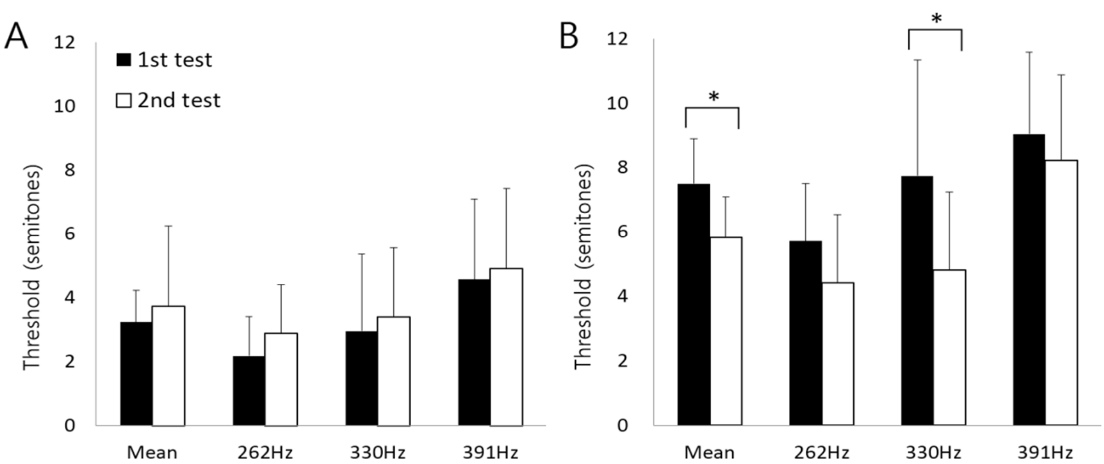


Fig. 1: Average of pitch discrimination: (A) good performance group and (B) poor performance group. * Significance at $p < 0.05$.

K-CAMP scores

Thresholds of pitch discrimination for the first test were 3.61±2.28, 4.90±3.75, 6.38±3.34, and 4.97±2.43 semitones at base frequencies of 262 Hz, 330 Hz, 391 Hz, and on average, respectively. In the second test, thresholds were 3.51±2.23, 3.97±2.53, 6.26±3.02, and 4.58±1.73 semitones at base frequencies of 262 Hz, 330 Hz, 391 Hz, and on average, respectively. Patients were divided into two groups: better or worse than the average performance (4.97 semitones) for analysis. As a result, pitch discrimination of the second test in the good performance group showed no difference with the first test (p = 0.468) (Fig. 1A), but in the poor performance group, the pitch discrimination score significantly improved (p = 0.005) (Fig. 1B). However, the second test result of the good performance group was 3.73±1.50 semitones which was still better than the poor performance group with 5.81±1.27 semitones (p = 0.001).

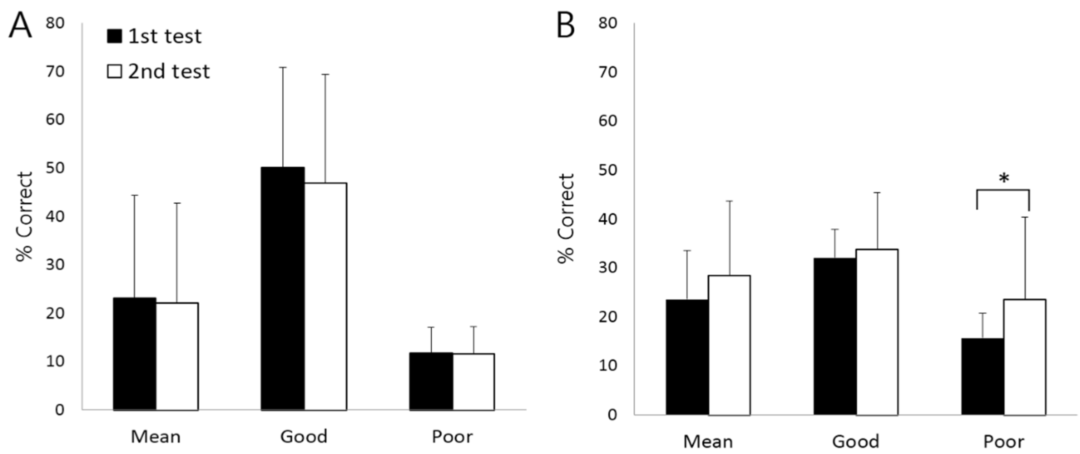


Fig. 2: K-CAMP subtest scores: (A) melody identification and (B) timbre identification. * Significance at $p < 0.05$.

The timbre identification scores of the first and second test were 23.6±9.9% and 28.4±15.2% correct, respectively. Patients were also divided into two groups: better or worse than the average performance (23.55%). Similar to the pitch test, the performance in the timbre test in the poor performance group significantly improved from 15.7±5.1% to 23.5±16.9% correct in the second test (p = 0.029) (Fig. 2B).

In the melody identification test, the average of correct identification was 23.1±21.3% for the first test and 22.0±20.7% for the second test. The patients were also divided into two groups based on the mean performance of the first test (23.1%): good performance group (50.1±20.6%) and poor performance group (11.7±5.3%). In the second test, the melody identification performance changed to 46.9±22.5% and 11.6±5.7%, respectively (Fig. 2A), and the changes were insignificant (p = 0.596).

Questionnaires

Scores for listening habit and level of musical experience significantly decreased post-operatively (p = 0.06 and $p < 0.001$ respectively). In the first and second follow-up tests, these scores did not recover during follow-up periods as shown in Fig. 3.

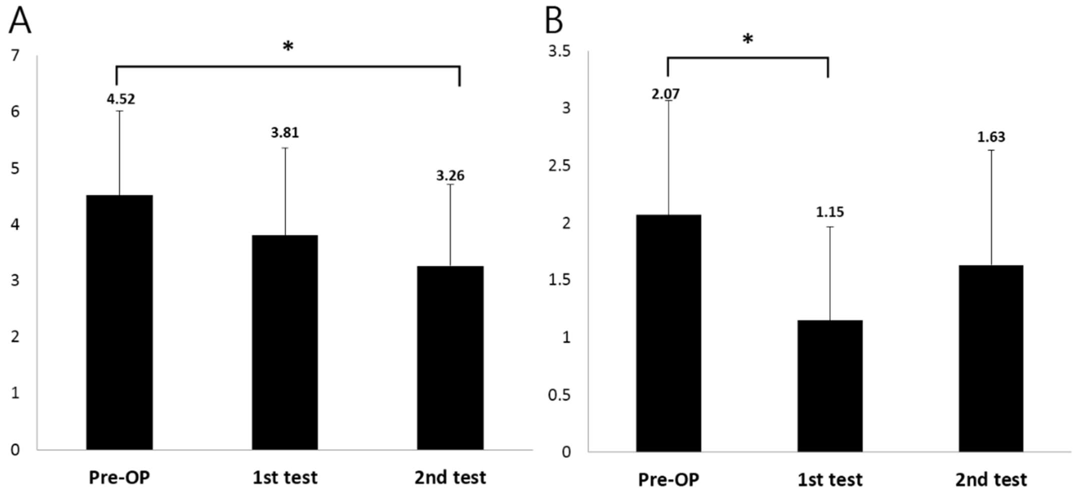


Fig. 3: Questionnaires: (A) listening habits before hearing loss and after implantation and (B) level of musical experience before hearing loss and after implantation. * Significance at $p < 0.05$.

CONCLUSIONS

The pitch discrimination and timbre identification ability improved in the CI listeners who had poor ability shortly after surgery. However, the ability of melody identification showed no difference in both groups after the lapse of time. Age at the test was related to pitch discrimination and younger people showed good performance. Listening habits and level of musical experience decreased after CI surgery without time-dependent improvement.

REFERENCES

Drennan, W.R., Oleson, J.J., Gfeller, K., Crosson, J., Driscoll, V.D., Won, J.H., Anderson, E.S., and Rubinstein, J.T. (**2015**). “Clinical evaluation of music perception, appraisal and experience in cochlear implant users,” Int. J. Audiol., **54**, 114-123.

Gfeller, K. and Knutson, J.F. (**2003**). “Music to the impaired or implanted ear: Psychosocial implications for aural rehabilitation,” ASHA Leader, **8**, 1-15.

Gfeller, K., Oleson, J., Knutson, J.F., Breheny, P., Driscoll, V., and Olszewski, C. (**2008**). “Multivariate predictors of music perception and appraisal by adult cochlear implant users,” J. Am. Acad. Audiol., **19**, 120-134.

Gfeller, K., Jiang, D., Oleson, J.J., Driscoll, V., and Knutson J.F. (**2010**). “Temporal stability of music perception and appraisal scores of adult cochlear implant recipients,” J. Am. Acad. Audiol., **21**, 28-34.

Jung, K.H., Cho, Y.S., Cho, J.K., Park, G.Y., Kim, E.Y., Hong, S.H., Chung, W.H., Won, J.H., and Rubinstein, J.T. (**2010**). “Clinical assessment of music perception in Korean cochlear implant listeners,” Acta Otolaryngol., **130**, 716-723.

Kang, R., Nimmons, G.L., Drennan, W., Longnion, J., Ruffin, C., Nie, K., Won, J.H., Worman, T., Yueh, B., and Rubinstein, J. (**2009**). "Development and validation of the University of Washington Clinical Assessment of Music Perception test," Ear Hearing, **30**, 411-418.

Kong, Y.Y., Cruz, R., Jones, J.A., and Zeng, F.G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear Hearing, **25**, 173-185.

Krueger, B., Joseph, G., Rost, U., Strauss-Schier, A., Lenarz, T., and Buechner, A. (**2008**). "Performance groups in adult cochlear implant users: Speech perception results from 1984 until today," Otol. Neurotol., **29**, 509-512.

Lenarz, M., Sönmez, H., Joseph, G., Büchner, A., and Lenarz, T. (**2012**). "Long-term performance of cochlear implants in postlingually deafened adults," Otolaryngol. Head Neck Surg., **147**, 112-118.

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

Limb, C.J. and Rubinstein, J.T. (**2012**). "Current research on music perception in cochlear implant users," Otolaryngol. Clin. North Am., **45**, 129-140.

Looi, V., McDermott, H., McKay, C., and Hickson, L. (**2008**). "Music perception of cochlear implant users compared with that of hearing aid users," Ear Hearing, **29**, 421-434.

Looi, V., Gfeller, K., and Driscoll, V. (**2012**). "Music appreciation and training for cochlear implant recipients: A review," Semin. Hear., **33**, 307-334.

McDermott, H.J. (**2004**). "Music perception with cochlear implants: A review," Trends Amplif., **8**, 49-82.

Nimmons, G.L., Kang, R.S., Drennan, W.R., Longnion, J., Ruffin, C., Worman, T., Yueh, B., and Rubinstein JT. (**2008**). "Clinical assessment of music perception in cochlear implant listeners," Otol. Neurotol., **29**, 149-155.

Best application of head-related transfer functions for competing voices speech recognition in hearing-impaired listeners

LARS BRAMSLØW*, MARIANNA VATTI, RENSKJE K. HIETKAMP, AND NIELS HENRIK PONTOPPIDAN

Eriksholm Research Centre, Snekkersten, Denmark

When presenting separated speech sources over hearing aids, should the normal physical spatial cues be restored? The answer was sought by presenting speech sources to a listener via headphones, either directly or after application of generic head-related-transfer functions (HRTF) in different modes to simulate free-field listening. For the presentation of two competing voices, we have measured the relative monaural and binaural contributions to speech intelligibility using a previously developed competing voices test. Two consecutive tests, using 13 and 10 hearing-impaired listeners with moderate, sloping hearing losses were conducted, combining different HRTF modes and horizontal plane angles. We found that neither the monaural HRTF gain nor the binaural cues imposed through crosstalk do affect the speech recognition. The only factor improving the competing voices scores is a large spatial separation, with as little mixing of the two voices as possible.

INTRODUCTION

Many situations require listeners to attend to two equally important voices, e.g., a dinner situation, or answering questions while watching TV. In some cases, the two voices are available separately, e.g. streaming from two phone lines simultaneously. In a normal, physical acoustic situation, the two voices will always be mixed, but one might imagine a perfect separation algorithm. In this latter case, the question again comes up: Is the application of generic head-related transfer functions (HRTF) beneficial? Moreover, which HRTF contributions are important: the monaural open-ear gain (see Fig. 1) and/or the binaural cross-talk that provides natural interaural level cues (interaural level and time differences)?

According to Brungart and Simpson (2005), with normal-hearing listeners there is a advantage of roughly 5% by going from separate (dichotic) to binaural HRTF (termed ‘3D’ by the authors), in a word-based test known as Coordinate Response Measure (CRM; Bolia *et al.*, 2000).

The severity of the problem is typically larger for hearing-impaired (HI) listeners, and this study only concerns this group for hearing aid applications.

*Corresponding author: labw@eriksholm.com

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

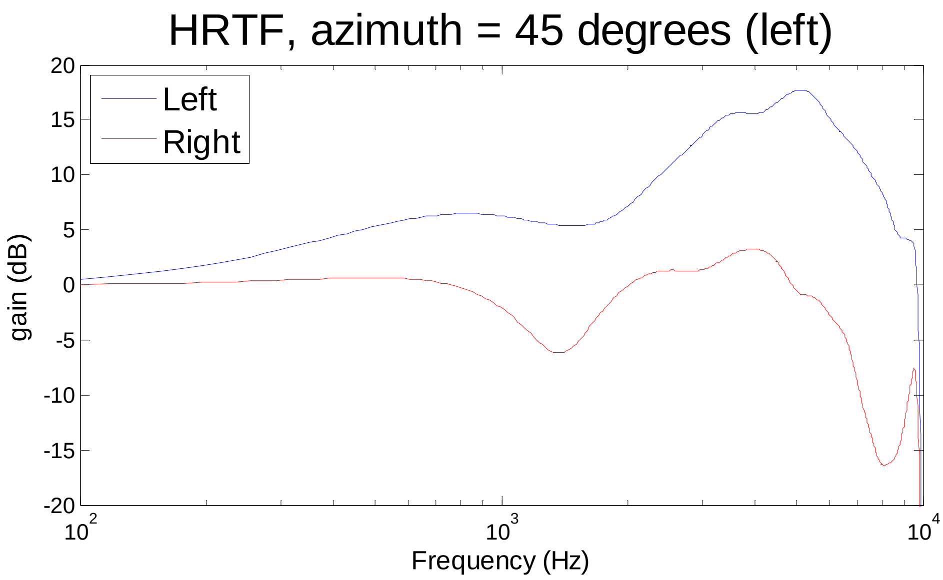


Fig. 1: Example of head-related transfer function (HRTF) from free-field frontal incidence (45 degrees azimuth) to blocked ear canal.

Aim of the study

To investigate the effect on competing voice situations from these components:

- The monaural component – due to the large gain applied by the open ear gain (OEG) contained in the HRTF. See Fig. 1 and, e.g., Fig. 3 (left pane).
- The binaural component – due to the interaural cues provided by the crosstalk from a sound source to the contralateral ear. See, e.g., Fig. 4 (right pane).
- The spatial separation effect from a co-located to a left-right configuration. See, e.g., Fig. 4 (right and left panes).

METHODS AND MATERIAL

Test method

The competing voices scenario was evaluated by using the Competing Voices Test (CVT), where two Danish Hearing In Noise Test (HINT; Nilsson *et al.*, 1994) sentences are played simultaneously in a spatial configuration. The Competing Voices Test was developed for these types of scenarios (Bramsløw *et al.* 2014; 2015). The task of the listener is to repeat sentences spoken either by the male or the female as prompted randomly ($p = 0.5$) by a sign on a PC monitor. The outcome measure was percent correct score, which was rau-transformed to provide better 'normal' distribution of the data (Studebaker, 1985).

Tests conducted

Two separate tests were conducted, which covered different aspects of the overall aim:

1. Test 1 evaluated the effect of applying generic binaural HRTF vs no HRTF in a spatially separated configuration. The HRTFs were measured on a Brüel & Kjær 4128 HATS manikin at the entrance of the blocked ear canal. The data shown here is the relevant subset from a larger experiment.
2. After test 1, it was speculated that normal binaural HRTF application (with crosstalk) was not the optimal. Therefore, test 2 evaluated the same contrast as test 1 plus the separate contributions from binaural HRTF (with crosstalk) and monaural HRTF (without crosstalk).

Spatial configurations via headphones

The possible spatial configurations are shown in Figs. 2, 3, 4 below.

Fig. 2: HRTF 'Off' spatial modes.

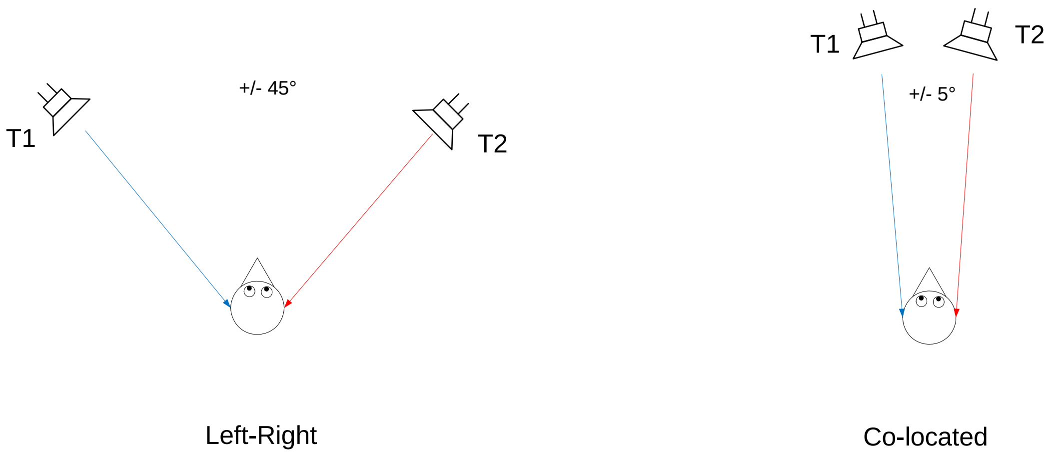


Fig. 3: HRTF 'No crosstalk' spatial modes.

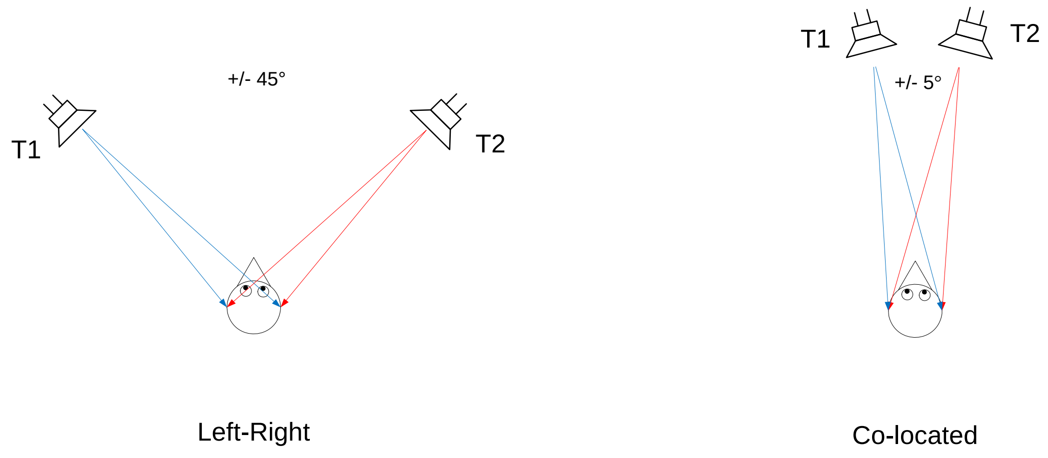


Fig. 4: HRTF 'Crosstalk' spatial modes. This corresponds to a normal physical loudspeaker setup.

The two tests covered these six possible configurations as listed in Table 1.

Conditions	Test 1		Test 2		
HRTF mode	**Off**	**Crosstalk**	**Off**	**No crosstalk**	**Crosstalk**
Left-Right	Separate	+/- 45 º	Separate	+/- 45 º	+/- 45 º
Co-located		+/- 5 º	Sum	+/- 5 º	+/- 5 º

Table 1: Overview of spatial modes and HRTF modes employed in the two tests.

Listeners

Both tests used elderly hearing-impaired listeners with moderate, sloping sensorineural hearing losses. Test 1 used 13 listeners and test 2 used 10 listeners, with an average age of 70 years. The hearing losses were compensated linearly according to the CAMEQ linear gain rule (Moore and Glasberg, 1998) and the listening level was set to most comfortable level during the initial training phase of the test.

RESULTS

The mean value scores are summarised in Table 2.

% CVT scores	Test 1		Test 2		
HRTF mode	**Off**	**Crosstalk**	**Off**	**No crosstalk**	**Crosstalk**
Left-Right	66.7 %	63.2 %	72.3 %	76.6 %	70.6 %
Co-located		50.9 %	54.1 %	76.7 %	54.6 %

Table 2: Overview of CVT scores from the two tests. The significantly different values in the two tests are underlined.

Both test 1 and test 2 were analysed using repeated-measures analysis of variance. In both cases, the following significant effects were found:

- Test person (test 1 and test 2). The values spread from app. 20 to 80% across test persons (not shown).
- Spatial mode (test 1 and test 2).
- HRTF mode (test 2).
- Spatial mode * HRTF mode (test 2). Note that test 1 was not a complete design of spatial mode and HRTF mode.

For test 2, the interaction of spatial mode is shown in Fig. 5. A post-hoc Tukey HSD test showed that:

Left-right: The scores in this spatial mode are not significantly different across the HRTF modes. Neither HRTF gain, nor crosstalk on/off affect the scores.

Co-located: ‘No crosstalk’ is significantly better than the two other conditions, because this does not have the large contralateral contribution as the other two HRTF modes do. The differences between ‘Off’ and ‘Crosstalk’ is the large gain from HRTF to both ears; however this does not affect the scores.

CONCLUSION

The best scores were obtained in the left-right (spatially separated) mode, as expected. In this spatial mode, there is no effect of HRTF, neither with or without HRTF. Thus, the HRTF (‘3D’ spatialised) advantage found in Brungart and Simpson (2005) was not replicated here for a hearing-impaired group.

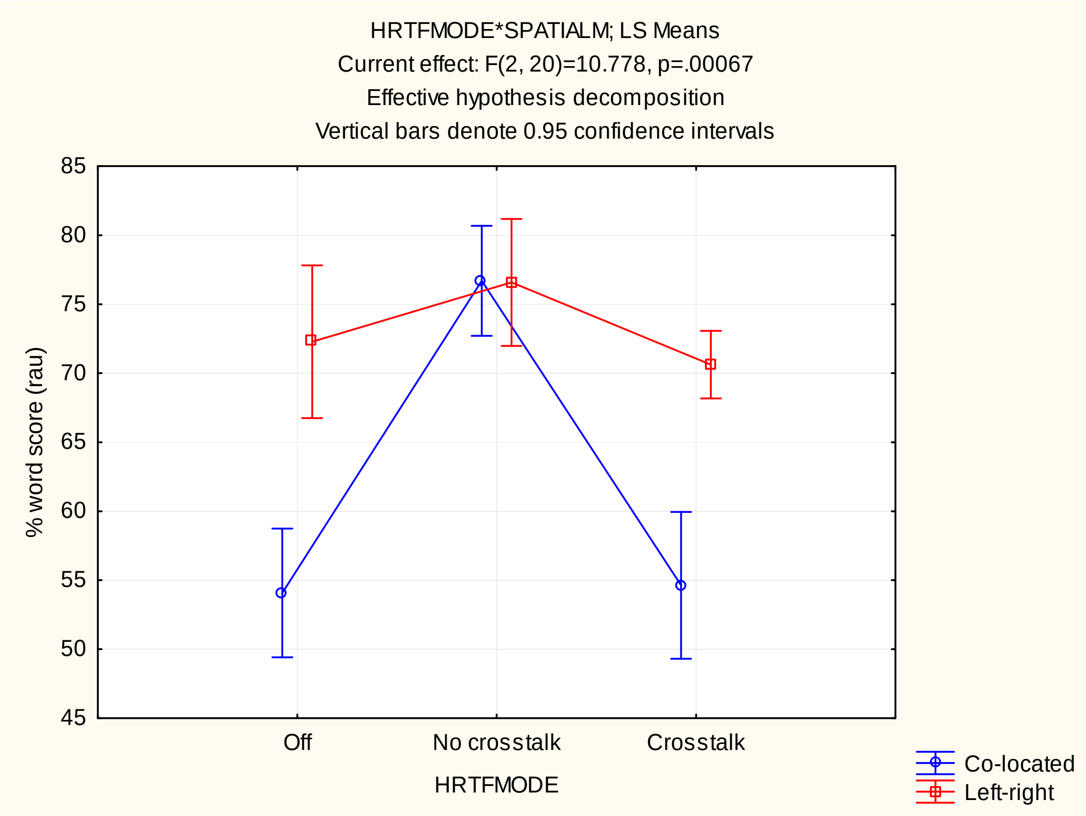


Fig. 5: Test 2 – Plot of the HRTF mode and spatial mode interaction.

Likewise, in the co-located mode, the effect of HRTF was due to removal of crosstalk, rather than due to the substantial gain difference (see Figure 1) added by the HRTF. The mode with no crosstalk is essentially separated regardless of angles.

The best presentation mode is thus spatially separated, without crosstalk. The normal transformation from free field to eardrum applied in hearing aids will satisfy this requirement.

REFERENCES

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

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) 2014, Lake Tahoe, CA, USA. Available at: http://www.eriksholm.com/about-us/news/IHCON_2014.

Bramsløw, L., Vatti, M., Hietkamp, R.K., and Pontoppidan, N.H. (**2015**). "Binaural speech recognition for normal-hearing and hearing-impaired listeners in a competing voice test. Poster presented at Speech in Noise Workshop 2015, Copenhagen. Available at: http://www.eriksholm.com/about-us/news/2015/SPIN_2015.

Brungart, D.S. and Simpson, B.D. (**2005**). "Improving multitalker speech communication with advanced audio displays," In *New Directions for Improving Audio Effectiveness.* Neuilly-Sur-Seine, France, pp. 30/1-30/18.

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

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.

Studebaker, G.A. (**1985**). "A "rationalized" arcsine transform," J. Speech Hear. Res., **28**, 455-462.

Effects of auditory acclimatization to bilateral amplification on audio-visual sentence-in-noise processing and speech-evoked potentials

JULIA HABICHT[1,3,*], MAREIKE FINKE[2,3], BIRGER KOLLMEIER[1,3], AND TOBIAS NEHER[1,3]

[1] *Medizinische Physik, Oldenburg University, Oldenburg, Germany*
[2] *Hannover Medical School, Hannover, Germany*
[3] *Cluster of Excellence Hearing4all, Oldenburg & Hannover, Germany*

Recently, Wendt *et al.* (2014) developed an eye-tracking paradigm for estimating *how quickly* a participant can grasp the meaning of audio-visual sentence-in-noise stimuli (the 'processing time'). Using this paradigm, Wendt *et al.* (2015) and Habicht *et al.* (2015) found that hearing-impaired listeners with prior hearing aid (HA) experience performed faster on this task than hearing-impaired listeners without any HA experience, despite comparable speech recognition performance. To better understand this finding the current study investigated the effects of auditory acclimatization to bilateral amplification on this task using a longitudinal study design. Groups of novice and experienced HA users took part. The novice users were tested before and after 12 weeks of acclimatization to bilateral HAs. The experienced users were tested with their own devices over the same time period. In addition to the processing time measurements, speech-evoked potentials were measured. Initial results show a tendency for shorter processing times for linguistically complex sentences and no changes in speech-evoked potentials. Additional analyses based on a set of measurements collected after another 12 weeks of acclimatization will make it possible to scrutinize the variables of interest further.

INTRODUCTION

Although a number of studies have investigated the effects of auditory acclimatization to hearing aids (HAs; see reviews by Turner *et al.*, 1996; Palmer *et al.*, 1998; Munro, 2008), the results of these studies are not consistent. Some studies found acclimatization effects (e.g., Munro and Lutman, 2003) while others did not (e.g., Humes and Wilson, 2003). Furthermore, these studies often used outcome measures which are not necessarily indicative of real-world communication abilities (e.g., loudness perception). Therefore, we wanted to investigate the potential effects of HA use on speech comprehension in complex listening situations. To that end, we used a recently developed audio-visual test paradigm for the assessment of speech comprehension in noise (Wendt *et al.*, 2014). This paradigm allows estimating *how quickly* a participant can grasp the meaning of sentence-in-noise stimuli (the

*Corresponding author: julia.habicht@uni-oldenburg.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

'processing time'). Using this method, we previously obtained results suggesting that HA experience leads to better performance on this task, irrespective of the amplification characteristics (Habicht *et al.*, 2015). Because these results were obtained using an across-group design, it is unclear whether they were due to a lack of auditory stimulation for the hearing-impaired listeners without prior HA experience. Here, we therefore investigated the effects of auditory acclimatization to bilateral amplification on processing times using a longitudinal study design. Furthermore, we measured speech-evoked potentials using a test paradigm of Finke *et al.* (2014) that allowed us to explore any higher-level neurophysiological changes due to HA provision.

Our hypotheses were as follows:

1. Acclimatization to bilateral amplification will lead to improved (i.e., shorter) processing times.
2. Bilateral amplification will also result in larger amplitudes and shorter latencies of late auditory potentials.
3. For experienced users, no such changes will be apparent.

METHODS

Participants

We recruited 15 habitual HA users with at least one year of HA experience ('eHA group') and 18 novice HA users ('nHA group'). The nHA users were acclimatized to bilateral HAs for 12 weeks. The eHA users continued to wear their own HAs for the same period. Inclusion criteria were (1) age from 60 to 80 yr, (2) bilateral, sloping, sensorineural hearing loss ranging from 40 to 80 dB hearing level (HL) at 3-8 kHz, and (3) self-reported normal or corrected-to-normal vision. The two groups were matched closely (see Table 1) in terms of age and pure-tone average hearing loss from 500 Hz to 4 kHz (PTA). Furthermore, their speech reception thresholds corresponding to 80%-correct speech intelligibility for the speech stimuli used here (SRT_{80}) were very similar. All participants were required to wear their HAs for at least six hours per day. In this contribution, we show results from 22 participants who at the time of writing had completed all measurements.

Hearing aids and amplification

At the beginning of the study, the nHA users were fitted with Sivantos pure micon 7mi receiver-in-the-canal devices. These HAs are equipped with 20-channel dynamic range compression and active noise management. Acoustic coupling was achieved via standard double click domes or, if ear canals were too small, closed click domes. The HAs were fitted according to NAL-NL1 prescription targets (Byrne *et al.*, 2001). Target gains were verified with real-ear insertion gain measurements. The nHA users were given up to three days to get used to their devices, and gains were adjusted only if participants felt that they could not tolerate the prescribed amplification for the duration of the study. Two nHA users were satisfied with the prescribed amplification, whereas for the other 16 nHA users gains had to be reduced for frequencies above 4 kHz. Following fine-tuning, no further adjustments were made, and the participants were not able to alter the amplification themselves. The eHA participants, who were all users of receiver-in-the-canal devices (various brands), were tested with their own HA fittings. Figure 1 shows mean prescription targets and user gains for a 65 dB input signal level for the two groups of participants.

	eHA group	**nHA group**
N	10	12
Age (yr)	73.7 (3.7)	73.2 (5.0)
PTA (dB HL)	42.4 (4.2)	38.2 (6.0)
SRT_{80} (dB SNR)	−1.6 (1.6)	−1.6 (0.8)
HA use (hr/day)	11.1 (4.5)	8.1 (3.5)

Table 1: Means (and standard deviations) for the age, PTA, SRT_{80} and HA use data for the two groups of participants.

During the measurements (see below), all stimuli were amplified in accordance with the measured individual insertion gains using the Master Hearing Aid research platform (Grimm *et al.*, 2006).

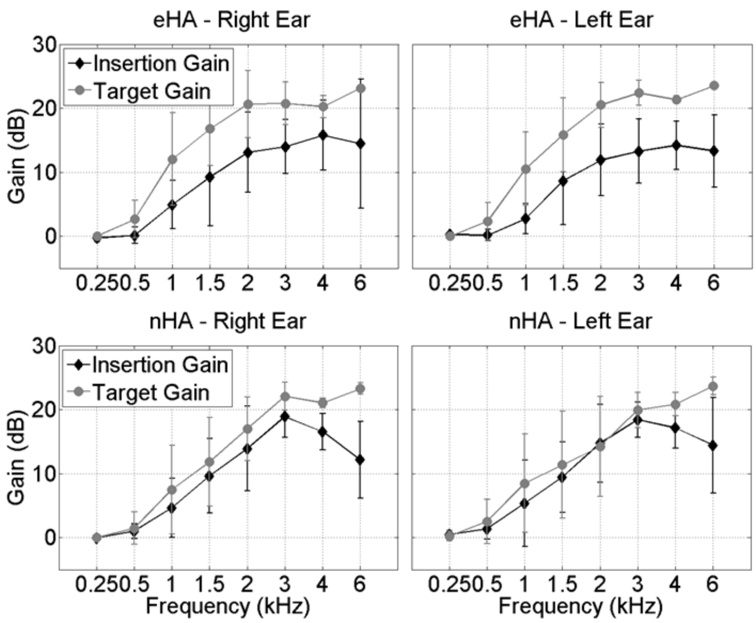


Fig. 1: Mean insertion gains and target gains (based on NAL-NL1) for the eHA (top, N = 10) and nHA (bottom, N = 12) groups.

Eye-tracking measurements

The eye-tracking measurements were based on sentences from the “Oldenburg corpus of Linguistically and Audiologically Controlled Sentences” (Uslar *et al.*, 2013). This corpus consists of (grammatically correct) sentence structures that vary in linguistic complexity. For our measurements, we used two sentence structures with either low or high linguistic complexity. In the German language, the linguistic complexity of these sentences is determined by relatively subtle grammatical or acoustic cues (see 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).

Sentences were presented in stationary speech-shaped noise at individual SRT_{80}’s via closed headphones (Sennheiser HDA 200). On each trial, two similar pictures (one target, one competitor) were displayed on a monitor positioned in front of the participant. In the target picture, the subject and object matched those conveyed by the corresponding acoustic sentence; in the competitor picture, the roles of the subject and object were interchanged so that there was a cross-modal mismatch. The task of the participant was to select the picture that matched the sentence by pressing a button on a hardware controller as quickly 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 (target) picture. In the following, the time elapsed for this to occur will be referred to as the “processing time”.

A total of four test blocks were performed per participant and visit. Within a test block there were 30 trials based on 15 sentences with low linguistic complexity and 15 sentences with high linguistic complexity plus seven catch trials. The different blocks were presented in randomized order across the different participants.

<table>
<tr><td rowspan="2">Low</td><td>Der$_{\text{nom}}$</td><td>müde$_{\text{nom}}$</td><td>Drache</td><td>fesselt</td><td>den$_{\text{acc}}$</td><td>großen$_{\text{acc}}$</td><td>Panda.</td></tr>
<tr><td colspan="7">Meaning: “The tired dragon ties up the big panda.”</td></tr>
<tr><td rowspan="2">High</td><td>Den$_{\text{acc}}$</td><td>müden$_{\text{acc}}$</td><td>Drachen</td><td>fesselt</td><td>der$_{\text{nom}}$</td><td>große$_{\text{nom}}$</td><td>Panda.</td></tr>
<tr><td colspan="7">Meaning: “The big panda ties up the tired dragon.”</td></tr>
</table>

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 and high). In each case, the grammatically salient ***word endings*** and corresponding cases (nom = nominative; acc = accusative) are indicated, as are the English meanings.

Event-related potentials

In addition to the eye-tracking measurements, we measured event-related potentials (ERPs) to also investigate potential acclimatization effects based on the latencies and amplitudes of the P3 response, which is known to reflect post-perceptual processing. For that purpose, we used an active oddball paradigm of Finke *et al.* (2014) with stimuli from Rufener *et al.* (2014). Standards were spoken words describing non-living objects (e.g., invoice or window). Deviants described living beings (e.g., mother or eagle). The participants were seated in a comfortable chair in an electrically shielded booth and looked at a visual marker. Their task was to press a button whenever they heard a deviant. The stimuli were presented diotically in quiet via insert earphones (Etymotic EAR 3A). The length of all stimuli was 800 ms with an inter-stimulus interval of 1.5 s and a jitter of maximally 50 ms. We presented 350 trials (270 standards and 80 deviants) in three blocks (1 × 140 trials, 2 × 105 trials). The block order was randomized across participants. At least two standard stimuli were presented in-between two deviant stimuli. The duration of the blocks was 4 to 5 min. The ERPs were recorded from 66 active scalp electrodes according to the International 10-20 system. Additionally, we placed two reference electrodes at the earlobes. To analyze P3 amplitudes and latencies we averaged the ERPs from the electrodes Pz, P1, P2, P3, and P4.

Test protocol

Each nHA participant attended four visits. At the first visit, the HAs were fitted. At the second visit, individual insertion gains and SRT_{80}'s were measured. At the third and fourth visit, the eye-tracking and ERP measurements were carried out. Between the second and third visit, participants used their HAs for about 12 weeks. The first and second visit took 1 hr each, whereas the third and fourth visit took 2 hr each. The eHA participants only attended visits 2 to 4.

RESULTS

Eye-tracking measurements

Mean processing times with 95% confidence intervals are shown in Fig. 2. To analyze these data we performed a mixed-model analysis of variance (ANOVA) with listener group (eHA, nHA) as between-subject factor and linguistic complexity (low, high) and time point (baseline, 12 weeks) as within-subject factors. This revealed significant effects of linguistic complexity ($F_{1,20}$ = 35.4, p < 0.00001, η_p^2 = 0.64) and listener group [$F_{1,20}$ = 15.6, p < 0.001, η_p^2 = 0.44], but not of time point [p > 0.2]. Interestingly, however, there was a tendency for the processing times for sentences with high linguistic complexity to decrease following 12 weeks of acclimatization.

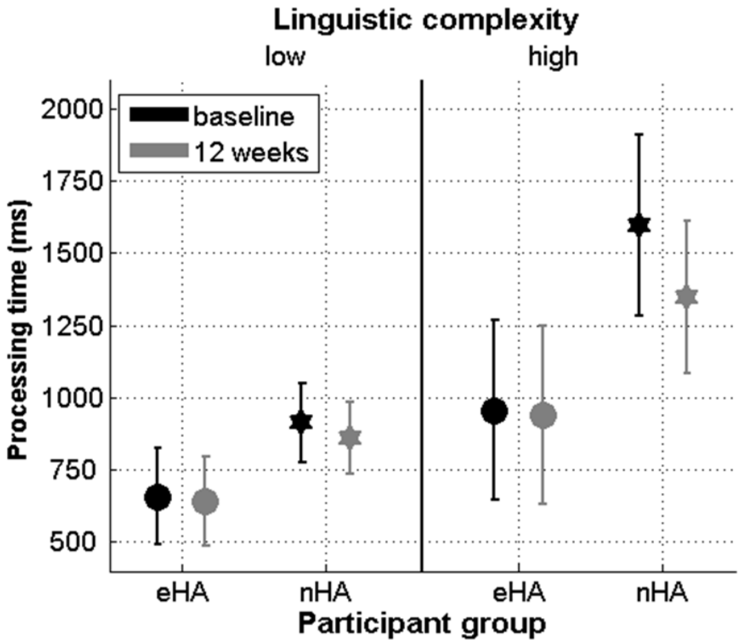


Fig. 2: Mean processing times for the eHA (circles, N = 10) and nHA (stars, N = 12) groups before ('baseline', black) and after 12 weeks (gray) of HA use for sentences with low (left) and high (right) linguistic complexity.

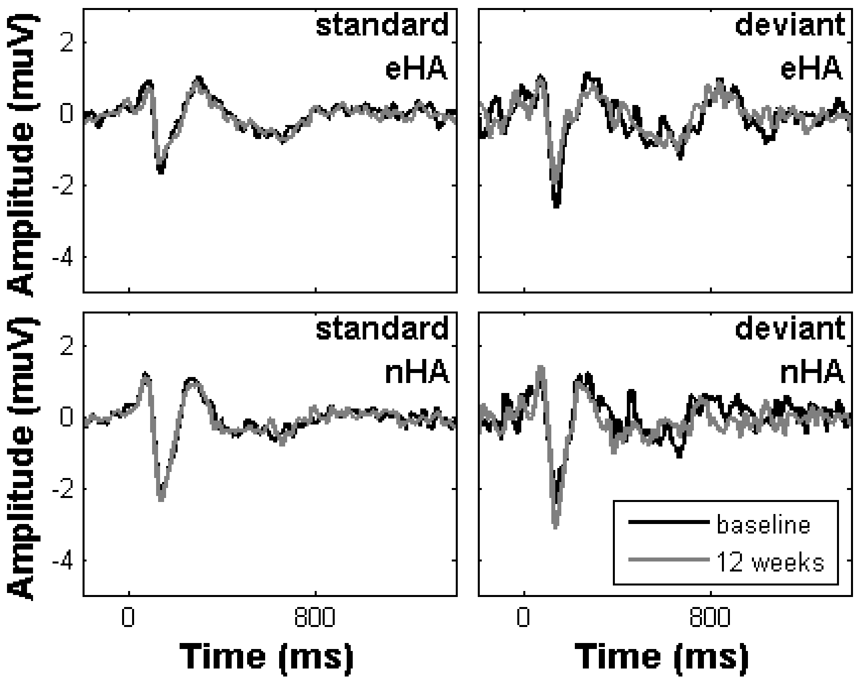


Fig. 3: Averaged speech-evoked potentials with standard (left) and deviant (right) stimuli before ('baseline', black) and after 12 weeks (gray) of HA use for the eHA (top, N = 10) and nHA (bottom, N = 12) participants.

Event-related potentials

Figure 3 shows averaged speech-evoked potentials for the two groups of participants, the two stimulus types, and the two time points. To analyze the ERPs we performed two mixed-model ANOVAs, one on the latencies and one on the amplitudes of the P3 response (which occurs around 800 ms after stimulus onset; see Finke *et al.*, 2014), with listener group (eHA, nHA) as between-subject factor, and stimulus type (standard, deviant) and time point (baseline, 12 weeks) as within-subject factors. The ANOVA performed on the amplitude data revealed significant effects for stimulus type [$F_{1,17} = 48.6$, $p < 0.00001$, $\eta_p^2 = 0.74$], but neither for listener group nor time point [both $p > 0.4$]. The ANOVA performed on the latency data revealed no significant effects [all $p > 0.05$].

SUMMARY AND CONCLUSIONS

In this contribution, we presented initial data from a longitudinal investigation into the effects of auditory acclimatization to bilateral amplification on audio-visual sentence-in-noise processing times and speech-evoked potentials. For this, we acclimatized a group of hearing-impaired listeners without HA experience to bilateral amplification for 12 weeks. In addition, we tested a group of experienced users with their own HA fittings over the same time period. As expected, the analysis of the processing time data showed significant effects of linguistic complexity and listener group. However, the effect of acclimatization was non-significant. Nevertheless, we observed a tendency for shorter processing times for sentences with high linguistic complexity following 12 weeks of acclimatization. Preliminary analyses of the measured P3 responses revealed larger amplitudes for deviant stimuli, but no effects of acclimatization.

Follow-up analyses based on the data from a total of 30 participants and an additional set of processing time and ERP measurements following 24 weeks of HA use will allow for more comprehensive analyses of the effects of HA use on the (neuro)physiological outcomes investigated here.

ACKNOWLEDGMENTS

We thank Katharina Simone Rufener for the provision of some of the stimulus material and Sivantos GmbH for supplying the hearing aids. This research was funded by the DFG Cluster of Excellence EXC 1077/1 “Hearing4all”.

REFERENCES

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.

Finke, M., Büchner, A., Ruigendijk, E., Meyer, M., and Sandmann, P. (**2014**). “Electrophysiological correlates of listening effort on speech understanding in cochlear implant users,” 8th International Symposium on Objective Measures in Auditory Implants, Toronto, Canada.

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. (**2015**). “Untersuchung der Effekte von Hörgeräteversorgungen bei Sprache im Störgeräusch anhand einer audio-visuellen Messmethode,” Proc. Annual Meeting of the German Audiological Society, Bochum, Germany.

Humes, L.E. and Wilson, D.L. (**2003**). “An examination of changes in hearing-aid performance and benefit in the elderly over a 3-year period of hearing-aid use,” J. Speech Lang. Hear. Res., **46**, 137-145.

Munro, K.J. and Lutman, M.E. (**2003**). “The effect of speech presentation level on measurement of auditory acclimatization to amplified speech,” J. Acoust. Soc. Am., **114**, 484-495.

Munro, K.J. (**2008**). “Reorganization of the adult auditory system: Perceptual and physiological evidence from monaural fitting of hearing aids,” Trends Amplif., **12**, 254-271.

Palmer, C.V., Nelson, C.T., and Lindley IV, G.A. (**1998**). “The functionally and physiologically plastic adult auditory system,” J. Acoust. Soc. Am., **103**, 1705-1721.

Rufener, K.S., Liem, F., and Meyer, M. (**2014**). “Age-related differences in auditory evoked potentials as a function of task modulation during speech–nonspeech processing,” Brain Behav., **4**, 21-28.

Turner, C.W., Humes, L.E., Bentler, R.A., and Cox, R.M. (**1996**). “A review of past research on changes in hearing aid benefit over time,” Ear Hearing, **17**, 14S-25S.

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.

Wendt, D., Kollmeier, B., and Brand, T. (**2015**). How hearing impairment affects sentence comprehension: Using eye fixations to investigate the duration of speech processing,” Trends Hear., **19**, 2331216515584149.

Impact of background noise and sentence complexity on cognitive processing demands

DOROTHEA WENDT[1,2,*], TORSTEN DAU[1], AND JENS HJORTKJÆR[1,3]

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

[2] *Eriksholm Research Centre, Snekkersten, Denmark*

[3] *Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital Hvidovre, Denmark*

Speech comprehension in adverse listening conditions requires cognitive processing demands. Processing demands can increase with acoustically degraded speech but also depend on linguistic aspects of the speech signal, such as syntactic complexity. In the present study, pupil dilations were recorded in 19 normal-hearing participants while processing sentences that were either syntactically simple or complex and presented in either high- or low-level background noise. Furthermore, the participants were asked to rate the subjectively perceived difficulty of sentence comprehension. The results showed that increasing noise levels had a greater impact on the perceived difficulty than sentence complexity. In contrast, the processing of complex sentences resulted in greater and more prolonged pupil dilations. The results suggest that while pupil dilations may correlate with cognitive processing demands, acoustic noise has a greater impact on the subjective perception of difficulty.

INTRODUCTION

Everyday listening situations usually take place at high signal-to-noise ratios (SNR), ranging from +5 to +15 dB (Smeds *et al.*, 2015). Nevertheless, in some situations, listeners may experience considerable difficulties with listening to speech even though intelligibility is at 100%. The processing demands might be high in such situations and comprehension may be experienced as effortful. The listening difficulties may arise from the acoustic disturbance of the speech source due to the background noise, or caused by a hearing impairment, but may further result from purely endogenous factors, such as the complexity of the speech signal that is being processed. While both acoustic and cognitive factors may challenge the processing load, it is still unknown whether they interact in the experience of listening effort. Different measures have been used in order to investigate effortful listening, ranging from subjective measures, such as subjectively rated effort, to more objective or physiological measures, such as task-evoked pupil dilation as an indicator of increased cognitive processing demands (McGarrigle *et al.*, 2014).

*Corresponding author: wendt@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

Several studies have demonstrated a correlation between pupil dilations and task demands (Beatty, 1982; Kahneman and Beatty, 1966). Thus, pupillometry has been increasingly used to examine processing load (sometimes termed 'listening effort') during speech recognition in difficult listening environments (Zekveld *et al.*, 2010; 2011). For example, Zekveld *et al.* (2010) examined the pupil response of listeners with normal hearing who listened to sentences presented in noise at several signal-to-noise ratios. They reported that mean pupil dilation and peak pupil dilation increased with increasing noise level indicating higher processing demands. Besides background noise, a few studies indicated that linguistic aspects of the speech signal, such as syntactic complexity, affect speech processing. More complex sentence structures can lead to a decrease in speech intelligibility (Uslar *et al.*, 2013) or an increase in processing duration (Wendt *et al.*, 2014; 2015). Piquado *et al.* (2010) used the pupillary response in younger and older adults to test cognitive processing demands due to syntactically complex sentences and sentence length. They found that the pupil response correlated with the length of the sentences, especially for elderly people. These studies indicated that processing demands are substantially higher when processing linguistically complex sentences in noise than when processing sentences with a simple linguistic structure.

The relationship between these subjective and objective measures of processing demands is still not well established (e.g., McGarrigle *et al.*, 2014). Although both measures have been employed, it seems that perceived effort and pupil dilation are not necessarily correlated (see, e.g., Zekveld *et al.*, 2011). In the present study, we examined the effects of syntactic complexity and noise level on processing demand using both a subjectively rated difficulty measure and pupil dilation. The rationale behind combining different measures in an audio-visual picture-matching task was to better understand the relationship between subjectively perceived difficulty and a physiological measure of effort.

MATERIAL AND METHODS

Participants

Eleven female and eight male participants with normal hearing carried out the experiment, with an average age of 23 years (ranging from 19 to 36 years). The participants had pure-tone hearing thresholds of 15 dB hearing level (HL) or better at the standard audiometric frequencies in the range from 125 to 8000 Hz. All participants had normal or corrected-to-normal vision.

Material

Speech material. Two different sentence types were recorded by translating 39 items from the German OLACS corpus (see Uslar *et al.*, 2013). All sentences contained a transitive full verb, an auxiliary verb (*vil* – 'will'), a subject noun phrase (SNP) and an object noun phrase (ONP). Two different types of sentence structures were realized by varying the word order to either subject-verb-object structure (SVO) or to object-verb-subject structure (OVS). For each sentence structure, two different propositions were realized (*bjørn* – 'bear' as agent vs. *robot* – 'robot' as agent; see SVO I and OVS II in Table 1). The word order (the position of the main verb, e.g., *vække* – 'wake up') was the only cue to understanding *who* (the agent, i.e., the entity that carries out the action) *did what to whom* (the patient, i.e., the entity that is affected by the action). Both sentence structures (SVO and OVS) were locally ambiguous with respect to their meaning as well as to the grammatical role of the involved entities (e.g., *bjørn* and *robot* in Table 1) until after the auxiliary verb (*vil*).

Sentence type	**Example**							
	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word8
SVO I	Den	flinke	bjørn	vil	$\text{vække}_{\text{PTD}}$	den	rare	robot.
	The agile bear will wake up the nice robot.							
SVO II	Den	rare	robot	vil	$\text{vække}_{\text{PTD}}$	den	flinke	bjørn.
	The nice robot will wake up the agile bear.							
OVS I	Den	flinke	bjørn	vil	den_{PTD}	rare	robot	vække.
	The agile bear, the nice robot will wake up.							
OVS II	Den	rare	robot	vil	den_{PTD}	flinke	bjørn	vække.
	The nice robot, agile bear will wake up.							

Table 1: Examples of the two different sentence structures that were used in the current study, i.e. subject-verb-object structure (SVO) and object-verb-subject structure (OVS). PTD indicates the point of target disambiguation.

In both structures, the disambiguating word, which is the word that allows a thematic role assignment of the agent and the patient – *who* (agent) *did what to whom* (patient) – is the auxiliary verb (see word 5 in Table 1). For instance, the position of the verb *vække* of the SVO structure (see SVO I and II in Table 1) disambiguates the sentence in a way that enables the participants to relate the spoken sentence to the target picture. For the OVS structure, the lack of a main verb in front of the article *den* (Table 1) informs the participants about the object role of the first noun in the sentence. Therefore, the onset of word 5 was defined as the "point of target disambiguation" (PTD). The SVO structure is considered syntactically simple and easy to process. Written and spoken OVS clauses in Danish, however, are typically more difficult to process (see Kristensen, 2013).

Visual stimuli. For each spoken sentence, a single picture was shown, which was either a target picture or competitor picture. The target picture illustrated the situation described by the spoken sentence (see right picture in Fig. 1). In the competitor picture, the roles of the agent and the object were interchanged (left picture in Fig. 1).

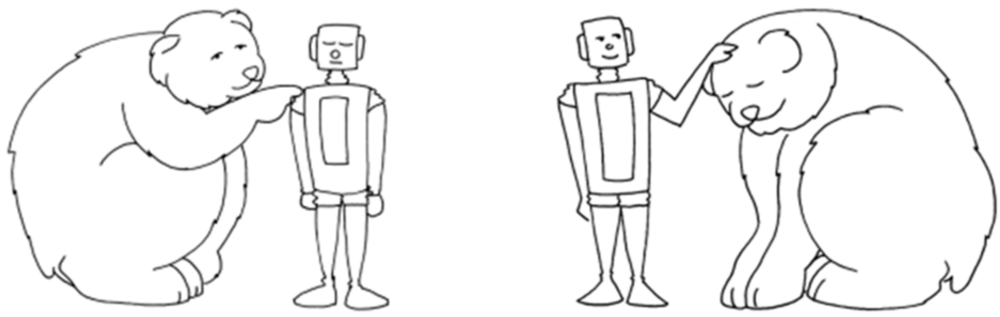


Fig. 1: Example of a target picture (right) and a competitor picture (left) for the Danish versions of the sentence "*The nice robot will wake up the agil bear*" or "*The agile bear, the nice robot will wake up.*"

PROCEDURE AND DATA ANALYSIS

All participants performed an audio-visual matching paradigm (see Fig. 2). First, a picture was shown on the screen for 2000 ms. The participants then heard the sentence (e.g., *The nice robot will wake up the agile bear*) while presented with a fixation cross. The sentences began 3000 ms after the picture offset and were presented in background noise. The background noise started 3000 ms before the sentence onset and ended 3000 ms after the sentence. After the noise offset, the participants' task was to decide whether the sentence matched the picture or not. 12 filler trials were included that did not contain a target or competitor picture, but showed an unrelated picture depicting different characters or different actions. After the comprehension question, the participants were asked to rate the perceived difficulty on a rating scale, i.e., how difficult they perceived the sentence comprehension to be. First, the participants performed one training block, which contained 10 trials. Afterwards, each participant listened to 159 sentences, divided into two blocks. The sentences were presented either at a low-noise level (+12 dB SNR) or at a high-noise level (−6 dB SNR). The noise masker was a stationary speech-shaped noise with the long-term frequency spectrum matching that of the speech. Changes in pupil size were measured for each participant from the onset of the noise until the comprehension task. An eye-tracker system (EyeLink 1000 desktop system, SR Research Ltd.) was used with a sampling rate of 1000 Hz to record pupil dilations.

Pupil data analysis. The pupil data were analysed using a similar procedure as described in Piquado *et al.* (2010) and Zekveld *et al.* (2010; 2011). First, the pupil data were cleaned for eye-blinks by classifying samples as an eye-blink for which the pupil value was below 3 standard deviations of the mean. Eye-blinks were removed and linearly interpolated, starting ten samples before and ending twenty samples after a blink. Trials for which more than 20% of the data required an interpolation were removed from further data analysis. The data of the de-blinked

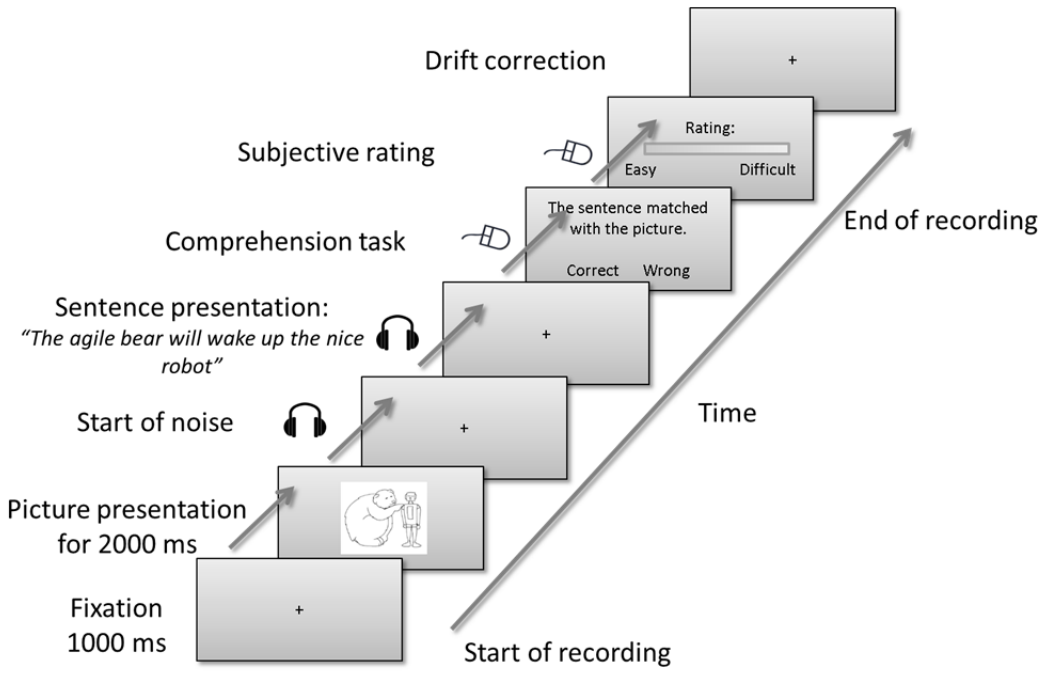


Fig. 2: Audio-visual picture-matching paradigm used for recording pupil dilation and subjectively perceived difficulty. Participants were presented a picture, followed by a spoken sentence. Their task was to decide whether the sentence either did or did not match with the picture. The comprehension question was followed by a subjective rating of the difficulty of the task.

trials were smoothed by a four-point moving average filter and then averaged for each condition and participant. In order to control for individual differences in pupil range, the minimum pupil value of the entire trial time series (from trial onset to the comprehension task) was subtracted from each trial data point. Afterwards, the pupil data were divided by the range of the pupil size within the entire trial. This method was applied to ensure consistent scaling of the range of the pupil value between 0 and 1 within each trial. Finally, the pupil data were normalized by subtracting a baseline-value which was defined as the averaged pupil value across 1.5 seconds before sentence presentation (when listening to noise alone). The maximum pupil dilation and time-averaged pupil dilation was calculated for the time interval between the sentence onset until the comprehension question.

RESULTS

A two-way repeated measures analysis of variance (ANOVA) was applied on the pupil data (on both mean pupil dilation and maximum pupil dilation) and the rated effort separately using SPSS 20, with *complexity* and *noise* as within-subjects factors. Significant effects were followed up with pairwise comparisons using post-hoc tests (applying a Bonferroni correction).

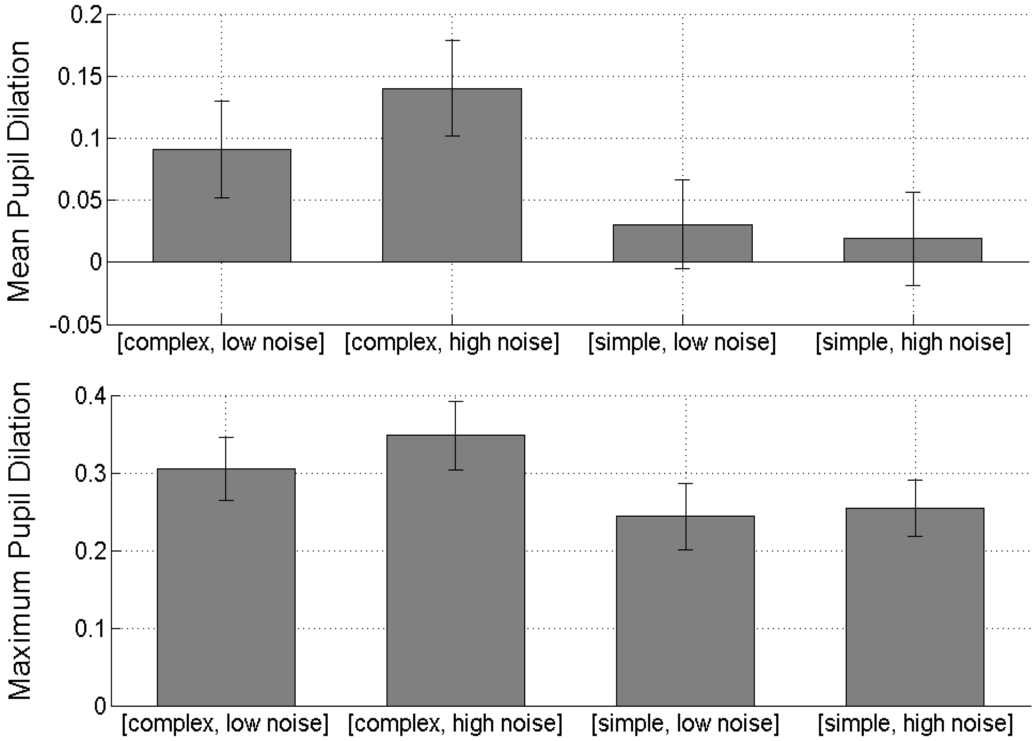


Fig. 3: Normalized pupil dilation averaged across all participants for four different conditions. The mean pupil dilation was calculated from the onset of sentence presentation until about 7500 ms after the sentence onset (3000 ms after sentence offset). Error bars indicate standard deviations.

Mean pupil dilation. A significant main effect was observed for the factor *complexity*. Post-hoc tests revealed significant differences in mean pupil dilation when processing syntactically simple and complex sentences [$F(1,18) = 22.0$, $p < 0.001$]. However, no interaction between noise and complexity was found.

Maximum pupil dilation. The ANOVA revealed a main effect of *complexity* [$F(1,18) = 13.0$, $p = 0.002$] indicating higher pupil dilation when processing syntactically complex sentences.

Difficulty rating. A significant main effect was found for the factor *noise*. Post-hoc tests revealed significant differences in pupil response in low and in high noise levels [$F(1,18) = 16.0$, $p < 0.001$]. This indicates that participants perceived processing sentences as more effortful within higher noise levels. The effect of sentence complexity, however, was rather small.

DISCUSSION

This study investigated subjective and physiological effects of linguistic complexity and background noise on sentence processing using an audio-visual picture-matching paradigm. Sentence processing demands were tested at two different levels

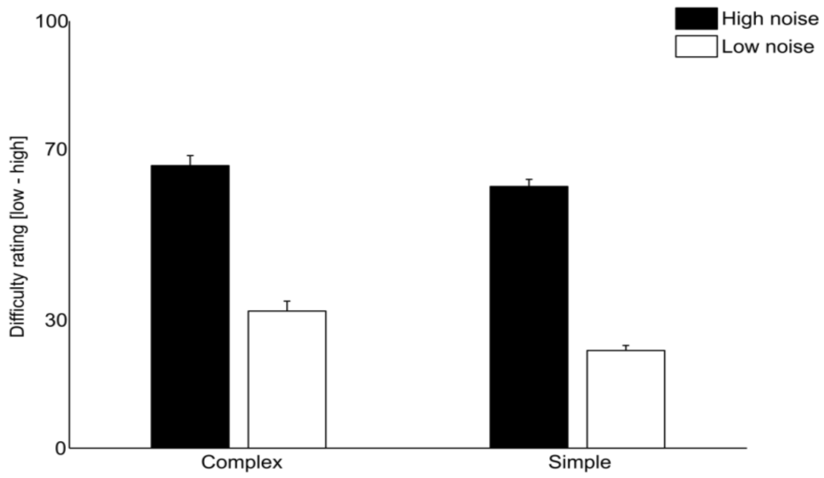


Fig. 4: Rated difficulty after the audio visual task averaged across all participants when OVS sentences (complex) or SVO (Simple) sentences were presented in low noise level (white) and high noise levels (black). Error bars show standard errors.

of background noise, whereby speech intelligibility was always relatively high. The results suggest that noise level and syntactic complexity relate in different ways to subjectively perceived effort and physiological markers of speech processing. The syntactic complexity was found to increase pupil dilation while processing sentences in background noise. However, the effect of the background noise level on the mean pupil dilation was rather small. An increase of the processing demand due to higher a noise level was only reflected in the subjective ratings. In other words, the poorer acoustical speech signal (due to the presence of background noise) led to a higher perceived demand on sentence processing. However, the interaction between background noise level and sentence structure was rather small. No combined effect of complexity and noise on either task-evoked pupil response or on perceived difficulties was found.

Our data indicate that both noise-induced and speech-induced processing demands can be found in listening situations that reflect everyday communication situations when speech intelligibility is still high. Moreover, our results demonstrate that the subjectively perceived effort is not directly reflected by the pupil dilation. The subjectively rated effort was found to be sensitive to changes in the noise level and, therefore, may reflect sensory processing difficulties that occur at early stages of speech processing (bottom-up processes). In contrast, the pupil response, which is often used as an indicator of the listening effort (Zekveld *et al.*, 2010), seems to be more sensitive to syntactic complexity and, thus, may reflect demands associated with cognitive processes that are required for sentence comprehension (top-down processes).

ACKNOWLEDGMENTS

This project was funded by the Oticon Centre of Excellence for Hearing and Speech Sciences (CHeSS). Thanks go to Line Burholt Kristensen for helping with the development of the sentence material. We thank the AULIN project for using their picture material.

REFERENCES

Beatty, J. (**1982**). “Task-evoked pupillary responses, processing load, and the structure of processing resources,” Psychol. Bull., **91**, 276-292.

Kahneman, D. and Beatty, J. (**1966**). “Pupil diameter and load on memory.” Science, **154**, 1583-1585.

Kristensen, L.B. (**2013**) *Context, You Need: Experimental Approaches to Information Structure Processing.* PhD thesis, University of Copenhagen.

McGarrigle, R., Munro, K.J., Dawes, P., Stewart, A.J., Moore, D.R., Barry, J.G., and Amitay, S. (**2014**). “Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group ‘white paper’,” Int. J. Audiol., **53**, 433-445.

Piquado, T., Isaacowitz, D., and Wingfield, A. (**2010**). “Pupillometry as a measure of cognitive effort in younger and older adults,” Psychophysiology, **47**, 560-569.

Smeds, K., Wolters, F., and Rung, M. (**2015**). “Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios.” J. Am. Acad. Audiol., **26**, 183-196.

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., **124**, 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.

Wendt, D., Kollmeier, B., and Brand, T. (**2015**). “How hearing impairment affects sentence comprehension: Using eye fixations to investigate the duration of speech processing,” Trends Hear., **19**, 1-18.

Zekveld, A.A., Kramer, S.E., and Festen, J.M. (**2010**). “Pupil response as an indication of effortful listening: The influence of sentence intelligibility,” Ear Hearing, **31**, 480-490.

Zekveld, A.A., Kramer, S.E., and Festen, J.M. (**2011**). “Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response,” Ear Hearing, **32**, 498-510.

Subjective listening effort and electrodermal activity in listening situations with reverberation and noise

INGA HOLUBE[1,*] KRISTINA HAEDER[1], CHRISTINA IMBERY[1], RALF HEINDORF[1], AND REINHARD WEBER[2]

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

[2] *Department of Medical Physics and Acoustics, University of Oldenburg, Oldenburg, Germany*

Disturbing factors like reverberation or ambient noise can obstruct speech recognition and raise the listening effort needed for communication in daily life. Situations with high listening effort are considered to incur an increased stress for the listener. The aim of this study was to assess listening effort in situations with background noise and reverberation. For this purpose, a subjective scaling of the listening effort, together with the electrodermal activity (EDA) as a measure of the autonomic stress reaction, was used. Ten young normal-hearing (NH) and 17 elderly hearing-impaired (HI) participants listened to sentences from the Oldenburg sentence test in stationary background noise and reverberation. Four listening situations were generated, an easy and a hard one for each of the two disturbing factors, which were related to each other by the Speech Transmission Index (STI). The results of the subjective scaling showed significant differences between the easy and the hard listening situations in both subject groups. However, various analyses of the EDA values indicate differences between the results of the groups. For the NH listeners, similar tendencies were observed both in the subjective results and the physiological EDA data. For the HI listeners, these effects in the EDA data were less pronounced.

INTRODUCTION

In this study, listening effort is regarded as the mental load needed to reach maximum speech recognition. Disturbing factors, such as reverberation or ambient noise, can obstruct speech recognition and increase listening effort. Situations with high listening effort are considered to imply an increased stress for the listener (Mackersie and Cones, 2011). When exposed to stress, the human body reacts with a change in many physiological parameters via the autonomic nervous system (Gramann and Schandry, 2009; Goldstein and Kopin, 2007). One of these physiological measures is the electrodermal activity (EDA), also known as skin conductance. The EDA describes the electrical conductance and potential changes of the skin (Schandry, 1989). It is influenced by the innervation of eccrine sweat

*Corresponding author: inga.holube@jade-hs.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

glands, which are stimulated sympathetically (Critchley, 2002). The sympathetic nervous system is that part of the nervous system that stimulates the fight-or-flight response (Goldstein and Kopin, 2007) of the body. The goal of the experiment in the present study was to examine the relationship between the subjectively-rated listening effort and the physiological EDA measure in easy and in hard listening situations in the presence of noise and reverberation for young normal-hearing and elderly hearing-impaired participants.

METHODS

Participants

Ten NH and 17 HI subjects participated in the experiment. The age of the NH subjects was 19 to 28 (average: 23 yrs) and the HI subjects were 52 to 85 (average: 73 yrs). NH was defined as a hearing threshold of ≤ 20 dB HL at all audiometer frequencies in the range from 250 Hz to 8 kHz. The HI subjects exhibited a mild-to-moderate hearing loss of 23.1 to 53.1 dB HL (average of 0.5, 1, 2, and 4 kHz, PTA_4). All participants received a compensation (12 Euro/h) for their expenses. The experiment was approved by the ethics commission of the Carl von Ossietzky University in Oldenburg, Germany.

Stimuli and test conditions

The Oldenburg Sentence Test (OLSA, Wagener *et al.*, 1999), with 30 sentences per list, was used as speech signal. The speech stimuli were either mixed with speech-simulating stationary noise ("Olnoise") or convoluted with impulse responses of real rooms, to add reverberation. For both situations, noise and reverberation, an easy and a hard hearing condition were generated. Previous experiments by Rennies *et al.* (2014) and Schepker *et al.* (2015), using the same stimuli, showed a similar subjective listening-effort rating for NH and HI both for at SNRs of −6 dB and −2 dB (as hard conditions), as well as 6 dB and 10 dB SNR (as easy conditions). Therefore, these values were also chosen in the current study. The room impulse responses characterized by their reverberation time T_{60} were chosen to provide approx. the same Speech Transmission Index (STI, Houtgast and Steeneken, 1985; Schepker *et al.*, 2015), i.e., approx. 4 s (NH) and 2 s (HI) for the hard, and 0.5 s (NH) and 0.3 s (HI) for the easy condition. The level of the speech signals was adjusted to a sound pressure level (SPL) of 55 dB for the NH subjects and to the same individual subjective loudness in categorical units, using loudness scaling, for the HI subjects (Rennies *et al.*, 2013). This resulted in an average presentation level of 69 dB SPL (STD 4.7 dB) for the HI subjects.

Measurement procedure

To minimize any muscle activities due to body movements, the participants were located in a relaxed position on a couch in a sound-isolated test booth and wore headphones (Sennheiser HD650). The experiment started with a relaxation time of approx. 10 min, followed by two training lists with the OLSA. Then, the first of the four randomly-presented test conditions, started by a recovery time of approx. 5 min and was followed by one test list of the OLSA during which the participants repeated the words they recognized. After each test list, the participants were interviewed and asked to subjectively rate their listening effort. Subsequently, the next test condition started with another recovery time of about 5 min. During the whole duration of the experiment, including test and recovery phases, the EDA was recorded as amplitude in µS using Nexus 10-MKII via electrodes and using a low, constant current on the middle phalanx of the index finger and the middle finger of the non-dominant hand.

Subjective listening effort was rated using a categorical scale showing seven labeled categories and six intermediate steps (Luts *et al.*, 2010). Effort scale categorical units (ESCU) were assigned to the categories. They were labeled from no effort (German "mühelos"; 1 ESCU), very little effort ("sehr wenig anstrengend"; 3 ESCU), little effort ("wenig anstrengend"; 5 ESCU), moderate effort ("mittelgradig anstrengend"; 7 ESCU), considerable effort ("deutlich anstrengend"; 9 ESCU), very much effort ("sehr anstrengend"; 11 ESCU), to extreme effort ("extrem anstrengend"; 13 ESCU). The values in ESCU were not visible to the subjects.

RESULTS

Subjective rating

The results of the subjective listening effort scaling obtained with the NH and HI subjects in each of the four conditions are shown in Fig. 1. For both subject groups, significant differences between the conditions were observed (Friedman test $p<0.001$ for NH and SH). Post-hoc Wilcoxon tests revealed significant differences between easy and hard conditions in noise and in reverberation (noise: $p=0.005$ for NH and $p<0.001$ for SH; reverberation: $p<0.001$ for NH and $p=0.005$ for SH). The ratings for the easy noise and the easy reverberation condition were similar in both subject groups, but the two hard conditions were significantly different in both groups ($p=0.010$ for NH and $p=0.002$ for SH). The hard conditions required a higher listening effort in reverberation than noise.

EDA

An example of the time course of the EDA during the whole experiment for one subject is given in Fig. 2. The EDA typically decreases during the recovery phases between the different conditions. At the beginning of each test list, the EDA typically showed an onset followed by a decay during its duration, but also exhibited several maxima and minima during other test lists. During the interview and the listening-effort rating directly after each test list, the EDA showed high amplitudes and substantial variations that are mainly based on motor activities of the body during this phase.

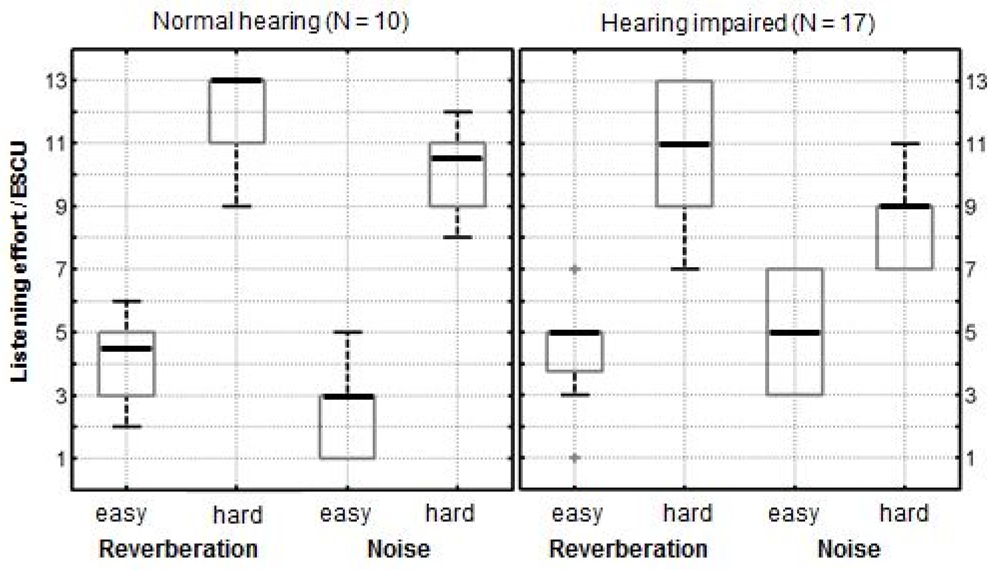


Fig. 1: Results of subjective listening effort ratings for normal-hearing (left) and hearing-impaired subjects (right) in the easy and hard reverberant and noise condition.

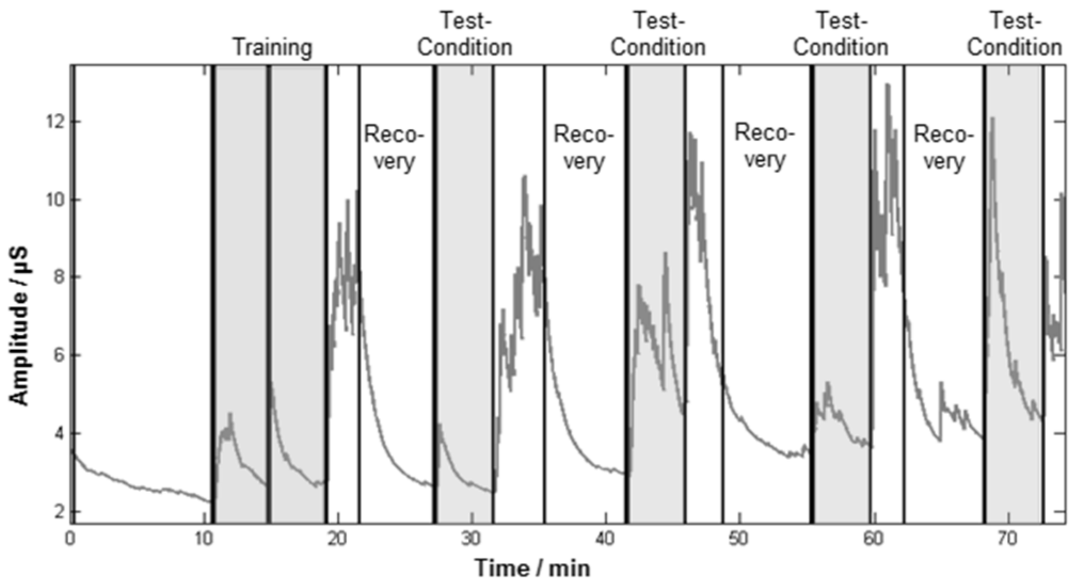


Fig. 2: Example of the time course of the EDA for one subject. The first grey area indicates the two training lists whereas the other four grey areas indicate the four test lists with different test conditions.

The EDA amplitudes during the test lists were compared across the four test conditions in terms of their averaged *z*-values (Mackersie and Cones, 2011; Mackersie *et al.*, 2015) and their relative peak rate. For the *z*-values, the average EDA amplitude during the last second before the start of each test list was regarded as the baseline for each subject and each condition. This baseline was subtracted from the EDA amplitudes averaged over the whole duration of the following test list, respectively. The results for each subject and each test condition were converted to *z*-values by subtracting the average of all subjects and test conditions and dividing by the respective standard deviation. The results in Fig. 3 show for NH subjects (left panel) the same tendency as for the subjective ratings, but non-significant differences (Friedman test, p=0.073). For HI subjects (right panel), the same tendency was observed for reverberation only, but not for the noise conditions (Friedman test, p=0.153).

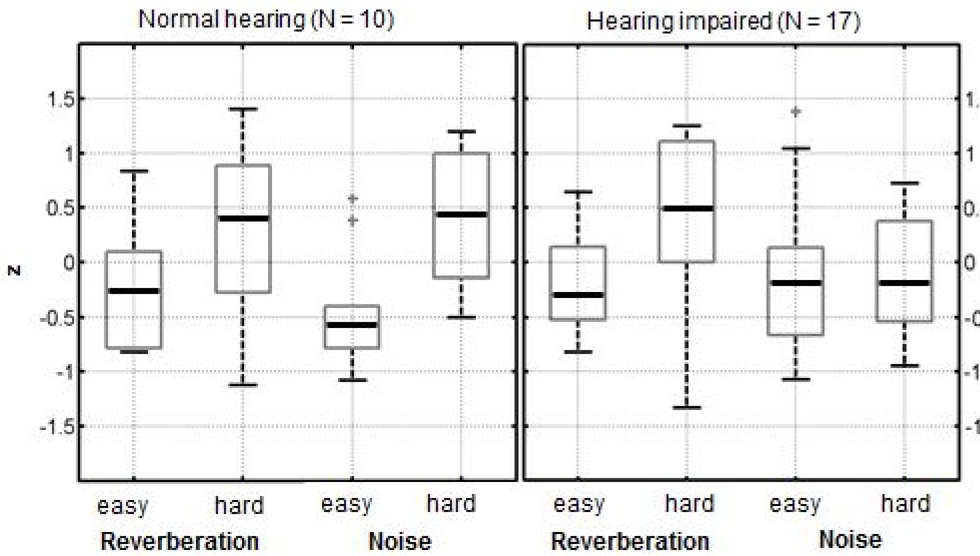


Fig. 3: Results of the *z*-values of the EDA for normal-hearing (left) and hearing-impaired subjects (right) in the easy and hard reverberant and noise condition.

The second measure, the relative peak rate, indicating the sympathetical level of excitation (Bruns and Praun, 2002), was calculated by counting the peaks of the EDA within the last three minutes of every recovery phase and within each test phase and dividing it by the duration of the respective recording periods. The relative peak rate of each test condition was given by the individual difference of the fluctuations/min in the test phase and in the previous recovery phase. Figure 4 shows the results for both groups. For the NH subjects (left panel), the results for the four test conditions were significantly different (Friedman test, p=0.008). However, a post-hoc Wilcoxon paired comparison test with Bonferroni correction did not show a significant difference. For the HI subjects, no significant differences were found between the test conditions.

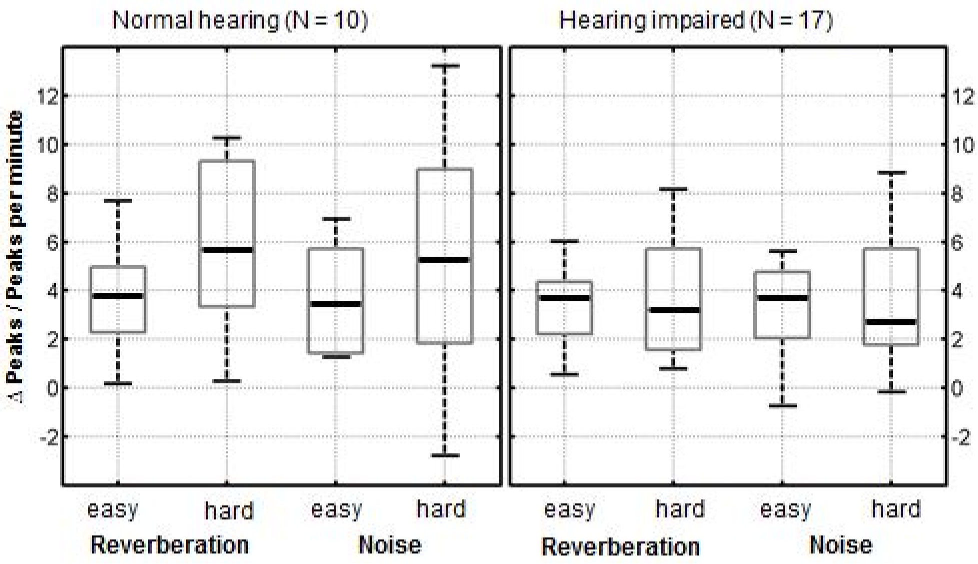


Fig. 4: Results of relative peak rate of the EDA for normal-hearing (left) and hearing-impaired subjects (right) in the easy and hard reverberant and noise condition.

The results of the EDA in *z*-values and in relative peak rate for all participants were compared to the subjective listening effort rating in ESCU (see Fig. 5). Besides the large scatter with very different EDA values for the same ESCU, Spearman's rank correlation revealed a low but significant correlation of *r*=0.337 for the *z*-values but no significant correlation for the relative peak rate.

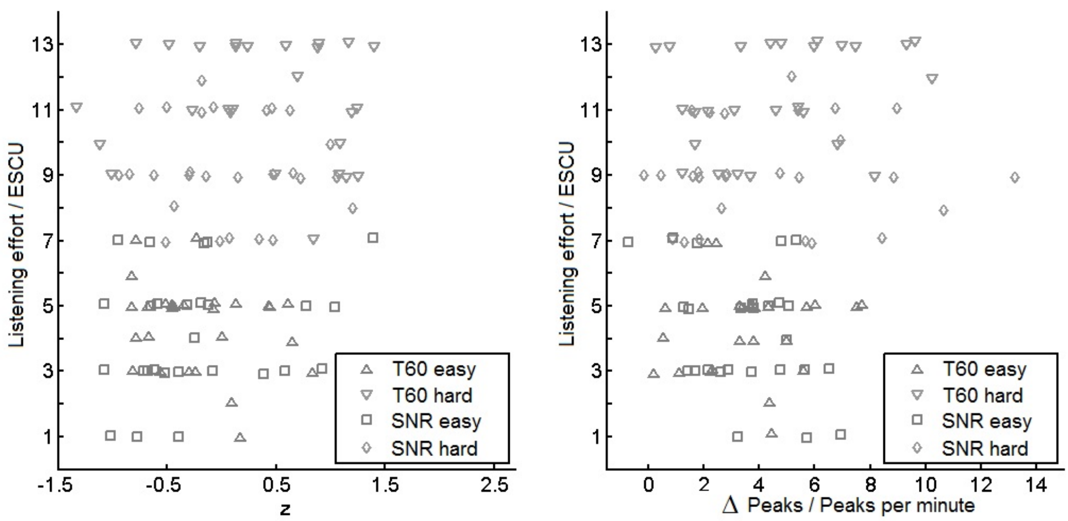


Fig. 5: Scatter plots of subjective listening effort in ESCU and EDA in *z*-values (left) and in relative peak rate (right) for all four test conditions.

DISCUSSION AND CONCLUSIONS

The subjective ratings of listening effort distinguish very well between the easy and the hard test conditions for both subject groups. Nevertheless, the differences between the test conditions are less pronounced in the older HI subjects compared to the younger NH subjects. This might be related to the older age of the HI subjects compared to the NH subjects and their respective listening experience in hard conditions (see, e.g., Larsby *et al.*, 2005). In contrast to the subjective ratings, both analysis methods of the EDA show a large scatter of the results as well as small or absent differences between the four test conditions and between the subject groups. Nevertheless, the results for the NH subjects indicate the same tendencies in the EDA data as in the subjective ratings where, for the HI subjects, differences in the EDA were observed only for some of the test conditions. Even though the test conditions were selected to manifest very different subjective listening efforts, differences in the EDA were difficult to demonstrate, especially for the group of HI subjects. One reason might be that the HI subjects perceived less stress in the hard conditions due to their experience with listening difficulties in everyday life. In addition, the lab situation without any interfering factors might cause less stress than usual listening experiences, as expressed at least by one subject. Another reason might be the older age of the HI subjects, which is frequently accompanied by skin alterations and therefore possible problems in recording the EDA. The lack of stress is also supported by the very weak relationship between subjective ratings and EDA recordings for all test conditions and participants. The EDA does not seem to be directly related to the subjective ratings, but might, in addition, be influenced by other so far unknown factors. The precise mechanism in skin conductance variations, and therefore also the applicability of the EDA in the lab, remains to be explained.

ACKNOWLEDGEMENTS

Supported by governmental funding initiative “Niedersächsisches Vorab” of the Lower Saxony Ministry for Science and Culture, research focus “Hören im Alltag Oldenburg (HALLO)”. We thank the Akustikbüro Oldenburg for the impulse responses, and Henning Schepker, Jan Rennies, Hartmut Meister, and Magdalene Ortmann for their support. English language support provided by www.stels-ol.de.

REFERENCES

Bruns, T. and Praun, N. (**2002**). *Biofeedback: Ein Handbuch für die therapeutische Praxis,* First Edition (Vandenhoeck & Ruprecht, Göttingen), Chapter 1,4.

Critchley, H.D. (**2002**). “Electrodermal Responses: What happens in the Brain,” The Neuroscientist, **8**, 132-142.

Goldstein, D.S. and Kopin, I.J. (**2007**). “Evolution of concepts of stress,” Stress, **10**, 109-120.

Gramann, K. and Schandry, R. (**2009**). *Psychophysiologie: Körperliche Indikatoren psychischen Geschehens,* Fourth Edition (Beltz PVU, Winheim).

Houtgast, T. and Steeneken, H.J.M. (**1985**). "A review of the mtf concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am., **77**, 1069-1077.

Larsby, B., Hällgren, M., Lyxell, B., and Arlinger, S. (**2005**). "Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects," Int. J. Audiol., **44**, 131-143.

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W.A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., and Spriet, A. (**2010**). "Multicenter evaluation of signal enhancement algorithms for hearing aids," J. Acoust. Soc. Am., **127**, 1491-1505.

Mackersie, C.L. and Cones, H. (**2011**). "Subjective and psychophysiological indexes of listening effort in competing-talker task," J. Am. Acad. Audiol., **22**, 113-122.

Mackersie, C.L., PacPhee, I.X., and Heldt, E.W. (**2015**). "Effects of hearing loss on hart rate variability and skin conductance measured during sentence recognition in noise," Ear Hearing, **36**, 145-154.

Rennies, J., Holube, I., and Verhey, J.L. (**2013**). "Loudness of speech and speech-like signals," Acta Acust. United Acust. **99**, 268-282.

Rennies, J., Schepker, H., Holube, I., and Kollmeier, B. (**2014**). "Listening effort and speech intelligibility in listening situations affected by noise and reverberation," J. Acoust. Soc. Am., **136**, 2642-2653.

Schandry, R. (**1989**). *Lehrbuch der Psychophysiologie: Körperliche Indikatoren psychischen Geschehens,* Second Edition (Psychologie Verlags Union, München-Weinheim).

Schepker, H., Haeder, K., Rennies, J., and Holube, I. (**2015**). "Listening effort and speech intelligibility in reverberation and noise for hearing-impaired listeners," Int. J. Audiol. (submitted).

Wagener, K., Brand, T., and Kollmeier, B. (**1999**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Optimierung des Oldenburger Satztests," Z. Audiol., **38**, 4-15.

Lateralized speech perception in normal-hearing and hearing-impaired listeners and its relationship to temporal processing

GUSZTÁV LŐCSEI[1,*], JULIE HEFTING PEDERSEN[2], SØREN LAUGESEN[2], SÉBASTIEN SANTURETTE[1], TORSTEN DAU[1], AND EWEN N. MACDONALD[1]

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

[2] *Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark*

This study investigated the role of temporal fine structure (TFS) coding in spatially complex, lateralized listening tasks. Speech reception thresholds (SRTs) were measured in young normal-hearing (NH) and two groups of elderly hearing-impaired (HI) listeners in the presence of speech-shaped noise and different interfering talker conditions. The HI subjects had either a mild or moderate hearing loss above 1.5 kHz and reduced audibility was compensated for individually in the speech tests. The target and masker streams were presented as coming from the same or from the opposite side of the head by introducing 0.7-ms interaural time differences (ITD) between the ears. To assess the robustness of TFS coding, frequency discrimination thresholds (FDTs) and interaural phase difference thresholds (IPDTs) were measured at 250 Hz. While SRTs of the NH subjects were clearly better than those of the HI listeners, group differences in binaural benefit due to spatial separation of the maskers from the target remained small. Neither the FDT nor the IPDT tasks showed a clear correlation pattern with the SRTs or with the amount of binaural benefit, respectively. The results suggest that, although HI listeners with normal hearing in the low-frequency range might have elevated SRTs, the binaural benefit they experience due to spatial separation of competing sources can remain similar to that of NH listeners.

INTRODUCTION

Normal-hearing (NH) listeners are extremely skillful in following a particular talker in the presence of multiple interfering acoustic sources. Through the use of binaural cues, interaural level differences (ILDs) and interaural time differences (ITDs), listeners can segregate sources that are spatially separated. While NH listeners can exploit spatial cues to aid robust speech identification in cocktail-party scenarios, hearing loss has been shown to negatively affect spatial perception of speech (Neher *et al.*, 2011). Furthermore, speech intelligibility performance can vary substantially across individual hearing-impaired (HI) listeners with similar audiograms. One potential explanation for this are individual differences in temporal fine structure (TFS) coding (e.g., Strelcyk and Dau, 2009; Papakonstantinou *et al.*, 2011).

*Corresponding author: guloc@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

The purpose of this study was to investigate the relationship between monaural and binaural TFS coding and speech intelligibility in lateralized conditions. To assess the robustness of low-frequency TFS coding, we measured frequency discrimination thresholds (FDTs) and interaural phase difference detection thresholds (IPDTs) for pure tones at 250 Hz. In addition, a speech intelligibility experiment was conducted where the stimuli were presented over headphones and "spatialized" with frequency-independent ITD cues only. We hypothesized that listeners who have elevated pure tone IPD detection thresholds will have limited capabilities to exploit ITD disparities between target and masker streams, and thus have a reduced spatial release from masking (SRM) once maskers are separated from the target.

METHODS

Listeners

10 young NH (21-29 yrs, mean: 23 std: 3.01) and 19 older HI (55-85 yrs, mean: 71.7, std: 7.19) listeners participated in the study. Members of the NH group had audiometric thresholds lower than 20 dB HL at octave frequencies between 125 and 8000 Hz. The HI listeners had normal hearing or a mild hearing loss below 1.5 kHz and a mild-to-moderate hearing loss at frequencies above 1.5 kHz. For each listener the difference in audiometric thresholds between the ears was at most 15 dB. The HI group was divided into two age-matched subgroups: those having pure-tone average thresholds (PTAs) less or equal to 40 dB HL above 1.5 kHz on average were classified as mildly impaired (HI_{mild}, 8 listeners) and the others were classified as moderately impaired (HI_{mod}, 11 listeners), respectively. This homogeneity of audiograms within groups was desirable in order to minimize audibility confounds at high frequencies once investigating the results of the speech intelligibility experiments.

Speech tests

SRTs were measured using target sentences uttered by a female talker from the Danish DAT corpus (Nielsen *et al.*, 2014). We used the "Dagmar" sentences as targets in the presence of the following interferers: speech shaped noise (SSN), reversed speech with 2, 4, or 8 competing male talkers from the Grid corpus (Cooke *et al.*, 2006) and forward speech with single sentences uttered by the 2 other female talkers from the DAT corpus. Target and masker stimuli were presented as coming from a lateral direction by introducing 0.7-ms ITDs between the ears for each of the streams. Two spatial configurations were used in each masker conditions: target and maskers leading on the same side (later referred to as co-located conditions) and target and maskers leading on opposite sides (separated condition). The side of the target was randomized from trial to trial. Spatial conditions with each masker type were clustered into separate blocks and the SRT tracking procedure for the different spatial conditions within these blocks were run on an interleaved manner. The notations S_1, C_2, C_4, C_8, and D_2 are used to denote the set of conditions where SSN, reversed speech of 2, 4, or 8 competing talkers, or 2 interferers from the DAT corpus are used as maskers, respectively. When referring to a specific spatial condition within each of these sets, the "co" and "sep" indicators will be used as superscripts (e.g., S_1^{co} refers to the condition with the SSN masker, where target and masker are presented as coming from the same side).

The maskers in the S_1, C_2, C_4, and C_8 conditions were spectrally shaped to have the same long-term average spectrum of the target talker. For the S_1 conditions, 50 tokens of 5 seconds were generated. The actual masker tokens in the S_1 conditions were randomly selected from these on each trial. For the C_2, C_4, and C_8 conditions, continuous streams of sentences were generated from each of the first eight male talkers from the Grid corpus. Low-energy intervals were removed and the resulting recordings were time-reversed. 50 non-overlapping tokens of 5 seconds were selected from each of these talkers. When generating masker tokens, single random tokens were drawn from the pre-generated pool of tokens for each of the first 2, 4, or 8 Grid talkers, which were then mixed. Similarly to the S_1 conditions, this was done trial-to-trial. Finally, in the D_2 conditions, randomly selected full sentences were used as maskers. In the SSN and reversed speech conditions, maskers started 1 s before the onset of the target sentence and ended with the target sentence. The D_2 maskers started at the same time as the target.

The stimuli were presented over headphones. To simulate free-field presentation, the target sentences were first set to a nominal level of 65 dB SPL "free field", mixed with the maskers, and finally amplified by adding open ear gain components (Moore *et al.*, 2008). The elevated hearing thresholds of the HI subjects were compensated for by applying frequency dependent linear gains based on their audiograms and the long-term average spectrum of the target speech (Neher *et al.*, 2011; Nielsen *et al.*, 2014). The audibility criterion was set to 15 dB at and below 3 kHz which was reduced to 4 dB at 8 kHz by logarithmic interpolation at the intermediate frequencies. SRTs corresponding to the 50% sentence correct values were tracked by adapting the masker level in 2-dB steps. SRTs were estimated based on the performance over one list in each condition. The speech tests were performed in two sessions and subjects were trained on 3 lists before each visit. We tested the S_1, C_2, and C_4 conditions during the first and the C_8 and D_2 conditions during the second visit. Within each visit, the presentation order of the conditions was balanced using a latin-square design. List numbers used for the target sentences were balanced between conditions with the same technique.

Temporal processing

To assess the robustness of monaural and binaural TFS coding, frequency discrimination thresholds (FDTs) and interaural phase discrimination thresholds (IPDTs) were measured at 250 Hz, respectively.

The FDT test was similar to that of Papakonstantinou *et al.* (2011). A 3-interval 3-alternative forced-choice (3I-3AFC) paradigm was applied in combination with a multiplicative one-up two-down tracking rule. Listeners had to indicate the target tone that had a higher frequency than the two references, which were presented at 250 Hz. The initial difference between target and reference was set to 25%, and the initial step-size to 2. The step-size was reduced by a factor of 0.75 after every other reversal. The minimum step-size was 1.125, which was used for the last 8 reversals. Thresholds were calculated as the geometrical mean of these reversal points. Overall, 5 runs were performed by each subject. The final threshold was calculated as the geometrical mean of the thresholds in the last 3 runs. All stimuli were presented monaurally at 65 dB SPL to the ear with the lower audiometric threshold at the test frequency. FDTs were not measured for two of the HI_{mild} and three of the HI_{mod} listeners.

The IPDT test was based on the TFS-LF test (Hopkins and Moore, 2010). Listeners were requested to pick the target stimulus containing an interaural phase shift of $\Delta\phi$ degrees in a 2I-2AFC task using a multiplicative 1-up 2-down tracking rule. Both target and reference stimuli consisted of four 200-ms long pure tones presented binaurally, each ramped with a 20-ms long Hann window and separated by 100-ms silent intervals. For the reference stimuli, each of the four tones had 0° interaural phase. For the target stimuli, the interaural phase of the second and fourth tone was changed to $\Delta\phi$. Initially, $\Delta\phi$ was set to 90°. The initial step-size was 3.375 and was decreased to 2.25 and 1.5 after the first and second reversals. 8 reversals were made with this final step-size. The threshold was estimated by taking the geometrical mean of these reversal points. Listeners completed 5 threshold estimation tests and the final threshold was calculated as the geometrical mean of the last 3 runs. The stimuli were presented at 30 dB SL.

RESULTS

The SRTs for the NH (white), HI_{mild} (light grey), and HI_{mod} (dark grey) groups are shown in Fig. 1. In the box plots, the thick black lines denote the medians and the boxes extend to the 25th and 75th percentiles. The thin lines extend to the most extreme data points within 1.5 interquartile ranges from the 25th and 75th percentiles. A repeated-measures ANOVA was performed on the SRTs with masker type and spatial distribution as within-subject factors and listener group as between-subject factor. The degrees of freedom were adjusted with Greenhouse-Geisser correction where the assumption of sphericity was violated.

In most of the tested conditions, NH listeners performed the best, followed by the HI_{mild} group and then by the HI_{mod} listeners. This was supported by the significant main effect of listener group [$F(2,26)=24.171$, $p<0.001$]. Differences between groups were smallest in the S_1 conditions and greatest in the C_2 conditions. NH listeners yield the lowest SRTs in the C_2 conditions, while HI listeners performed best in the S_1 conditions. Despite the inherent spectro-temporal fluctuations in the C_8 backgrounds, all groups had elevated thresholds as compared to the stationary S_1 conditions. While the NH listeners performed better as the number of reversed interferers decreased from 8 to 4 to 2, the HI listeners performed similarly in all of these conditions. Consistent with these observations, SRTs differed significantly between the various masker types [$F(3.11,80.95)=28.02$, $p<0.001$], and the interaction between masker type and listener group was also significant [$F(6.23,80.96)=5.03$, $p<0.001$].

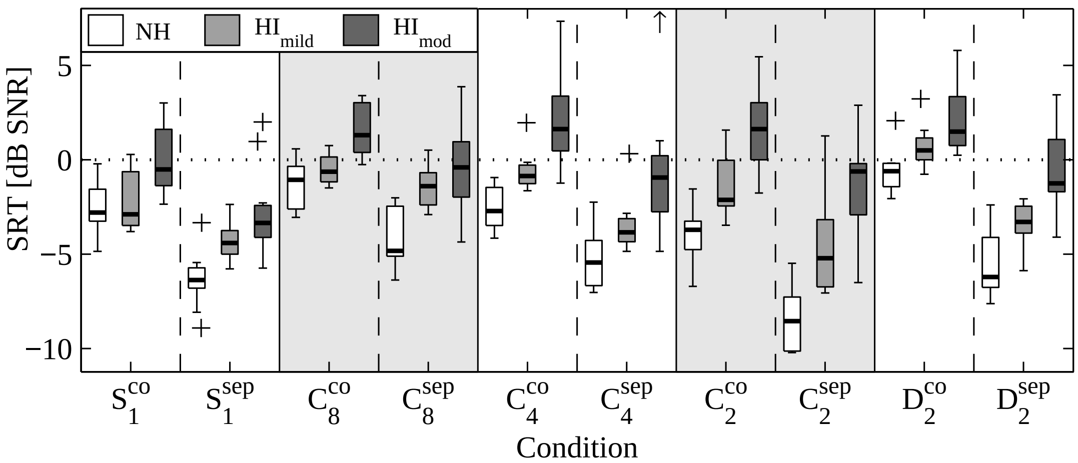


Fig. 1: SRTs across conditions (white: NH, light grey: HI_{mild}, dark grey: HI_{mod}). Shaded areas denote condition groups with the same type of background noise, while the dashed lines separate the spatial distributions within the noise groups (left: co-located, right: separated).

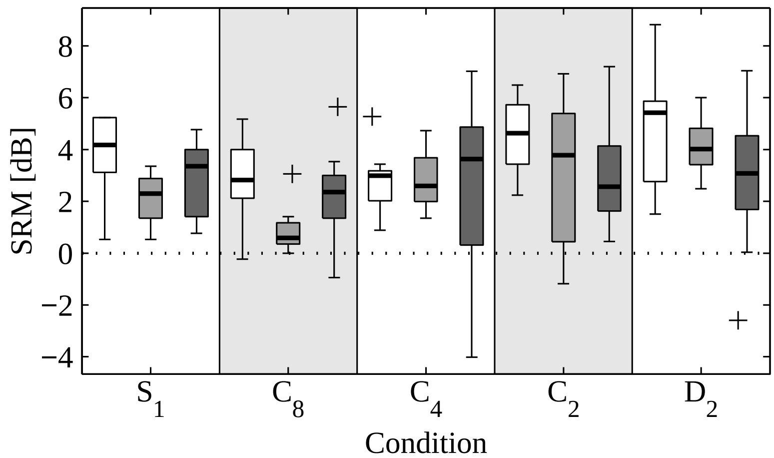


Fig. 2: SRMs across conditions (white: NH, light grey: HI_{mild}, dark grey: HI_{mod}). SRM was calculated as the difference in SRTs between the co-located and separated spatial configurations within noise condition groups.

In Fig. 2 the spatial release from masking (SRM), calculated as the difference in SRTs between spatially separated and co-located target-masker conditions, is shown for the different noise conditions. All of the listener groups benefited from spatially separating the maskers. The benefit varied between 1 and 5 dB, depending on noise condition and listener group. Except for the C_4 condition group, the NH listeners benefited the most from spatial separation. On average, the NH group yielded 3.78 dB

SRM, while the HI_{mild} and HI_{mod} groups yielded 2.66 and 2.62 dB, respectively. The main effect of spatial distribution was significant [$F(1,26)=311.41$, $p<0.001$]. On average, the greatest SRM was obtained in D_2 and the smallest in C_8. Both interactions between spatial distribution and masker type [$F(3.08,79.96)=4.57$, $p=0.005$] and between spatial distribution and listener group [$F(2,26)=4.91$, $p=0.016$] were significant.

The FDT and IPDT results are presented in Fig. 3. For both measures, the NH group performed significantly better than the HI group (HI_{mild} and HI_{mod} collapsed). The mean thresholds were 1.22 and 2.87 percent in the FDT (two-tailed t-test: $p=0.014$), and 11.47 and 19.5 degrees in the IPDT experiment (two-tailed t-test: $p=0.0186$), respectively.

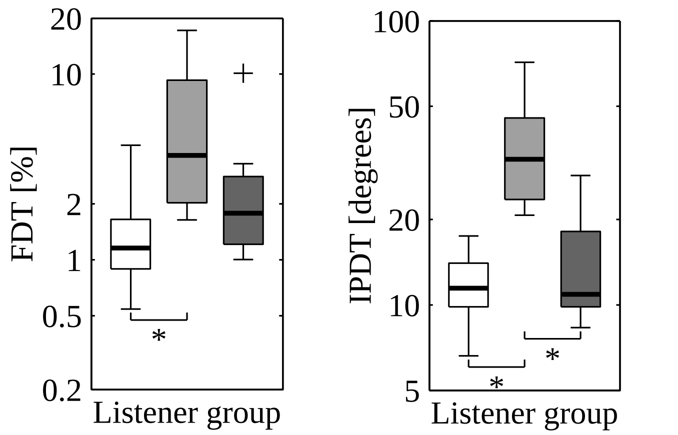


Fig. 3: FDT (left) and IPDT (right) results for NH (white), HI_{mild} (light grey) and HI_{mod} (dark grey). Asterisks denote statistically significant differences in means.

Figure 4 shows scatter plots of SRTs averaged across all conditions (SRT_{avg}) vs. PTAs at octave frequencies from 0.5 to 4 kHz (left panel), average SRTs across co-located conditions (SRT_{co}) vs. FDT (middle panel), and the average SRM benefit of all conditions (SRM_{all}) vs. IPDT (right panel). The correlations between FDT and SRT_{co} and between IPDT and SRM_{all} were not significant. The correlation between PTA and SRT_{avg} was significant when the HI_{mild} and HI_{mod} groups were pooled together ($r=0.55$, $p=0.015$). The slope of the regression line was 0.11, showing that, on average, a 9-dB increment in PTA yielded about 1-dB increment in SRT. This correlation was not significant when the HI_{mild} and HI_{mod} groups were considered separately.

DISCUSSION

The aim of the current study was to investigate the relationship between TFS processing and speech perception in lateralized conditions. We hypothesized that reduced FDTs and IPDTs would be associated with elevated SRTs in conditions where target and maskers are co-located or with reduced SRMs, respectively. Individualized linear gains were applied to all speech stimuli to reduce possible confounds due to stimulus inaudibility.

The HI listeners showed elevated SRTs as compared to the NH population, and the differences were largest in fluctuating masker condtions. No correlation was found between SRTs in the co-located conditions and FDTs. This contradicted our hypothesis and some previous results (Papakonstantinou *et al.*, 2011). Instead, SRTs were positively correlated with audiometric thresholds. It is still likely that these differences in the SRTs arose to some extent from impairment factors not directly related to reduced audibility, as these have been partly compensated for. One such reason could be the broadening of auditory filters at the higher sound pressure levels of the stimuli persented to the HI subjects (Studebaker *et al.*, 1999).

HI listeners experienced less SRM than NH, but the difference was small. Thus, HI listeners retain some benefit from large ITDs differences between target and maskers. While the HI group performed worse in the IPDT experiments, the IPDTs and SRM scores were not correlated. One reason for this could be the relatively large ITDs used to trigger different spatial positions. These time differences were clearly detectable to almost all of our HI subjects at 250 Hz. The effect of reduced binaural TFS coding on SRM might be more pronounced when the ITD differences between the target and maskers are relatively small.

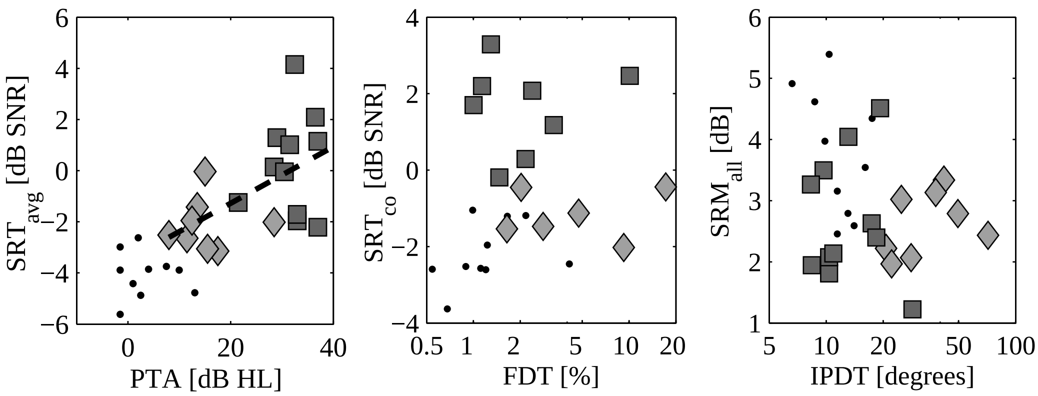


Fig. 4: Scatter plots between audiometric thresholds or TFS coding and speech reception performance (dots: NH, diamonds: HI_{mild}, squares: HI_{mod}). The dashed regression line was fitted to the data of the HI group (HI_{mild} and HI_{mod} collapsed). See text for further details.

The pattern of differences in the FDT and IPDT tests between the HI_{mild} and HI_{mod} listeners is surprising considering that these groups were age-matched and had the same hearing threshold levels at 250 Hz. Listeners in the HI_{mod} group performed similarly to those in the NH group. A significant overlap between the spread of data of NH and HI have been observed in earlier studies as well (Hopkins and Moore, 2011; Papakonstantinou *et al.*, 2011). Given the relatively small number of subjects in each group, it might be that this distribution of the data was just a casual result of partitioning the HI group into two subgroups.

CONCLUSIONS

Consistent with earlier studies (Neher *et al.*, 2011), the results of the speech experiments revealed that HI listeners experience difficulties in spatial listening tasks. The difficulties were more pronounced in fluctuating background noise than in steady-state noise. However, in contrast to earlier studies (Papkonstantinou *et al.*, 2011), between-subject differences in the HI group could not been explained by TFS coding as measured by FDTs, but by average audiometric thresholds. It is likely that the correlations between SRTs and PTAs can be at least partly attributed to factors other than audibility (such as broader auditiory filters at higher presentation levels), as the audibility of the target stimuli has been individually compensated for. The amount of SRM was smaller for HI than for NH listeners, but only in the order of 1 dB. Low-frequency IPDTs did not correlate with SRM. SRMs in an experimental paradigm applying smaller ITDs to separate target from maskers would be more limited by IPDTs at low frequencies and may thus be a more sensitive measure to investigate the effect of binaural TFS processing on spatial speech perception.

REFERENCES

Cooke, M., Barker, J., Cunningham, S., and Shao, X. (**2006**). "An audio-visual corpus for speech perception and automatic speech recognition," J. Acoust. Soc. Am., **120**, 2421-2424.

Hopkins, K. and Moore, B.C.J. (**2010**). "Development of a fast method for measuring sensitivity to temporal fine structure information at low frequencies," Int. J. Audiol., **49**, 940-946.

Hopkins, K. and Moore, B.C.J. (**2011**). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," J. Acoust. Soc. Am., **130**, 334-349.

Moore, B.C.J., Stone, M.A., Füllgrabe, C., Glasberg, B.R., and Puria, S. (**2008**). "Spectro-temporal characteristics of speech at high frequencies, and the potential for restoration of audibility to people with mild-to-moderate hearing loss," Ear Hearing, **29**, 907-922.

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.

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.

Papakonstantinou, A., Strelcyk, O., and Dau, T. (**2011**). "Relations between perceptual measures of temporal processing, auditory-evoked brainstem responses and speech intelligibility in noise," Hear. Res., **280**, 30-37.

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.

Studebaker, G.A., Scherbecoe, R.L., and McDaniel, D.M., and Gwaltney, C. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am., **105**, 2431-2444.

Predicting masking release of lateralized speech

Alexandre Chabot-Leclerc*, Ewen N. MacDonald, and Torsten Dau

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

Lőcsei et al. (2015) [Speech in Noise Workshop, Copenhagen, 46] measured speech reception thresholds (SRTs) in anechoic conditions where the target speech and the maskers were lateralized using interaural time delays. The maskers were speech-shaped noise (SSN) and reversed babble with 2, 4, or 8 talkers. For a given interferer type, the number of maskers presented on the target's side was varied, such that none, some, or all maskers were presented on the same side as the target. In general, SRTs did not vary significantly when at least one masker was presented on the same side as the target. The largest masking release (MR) was observed when all maskers were on the opposite side of the target. The data in the conditions containing only energetic masking and modulation masking could be accounted for using a binaural extension of the speech-based envelope power spectrum model [sEPSM; Jørgensen *et al.*, 2013, J. Acoust. Soc. Am. 130], which uses a short-term equalization-cancellation process to model binaural unmasking. In the conditions where informational masking (IM) was involved, the predicted SRTs were lower than the measured values because the model is blind to confusions experienced by the listeners. Additional simulations suggest that, in these conditions, it would be possible to estimate the confusions, and thus the amount of IM, based on the similarity of the target and masker representations in the envelope power domain.

INTRODUCTION

Listeners benefit from listening with two ears compared to a single ear in complex listening situations. This binaural benefit is usually explained in terms of "better-ear" (BE) and binaural unmasking (BU) concepts. The former relies on interaural level differences (ILDs) caused by the acoustical "shadow" cast by the head, which creates an advantageous signal-to-noise ratio (SNR) at the ear contra-lateral to the masker. In the latter, the interaural time differences (ITDs) give the hearing system the ability to increase the effective SNR by "cancelling" some of the masker signals (equalization-cancellation (EC) theory; Durlach, 1963).

The BE benefits are typically modeled in terms of audibility (Beutelmann *et al.*, 2010; Lavandier and Culling, 2010; Wan *et al.*, 2014), with a decision metric such as the speech intelligibility index (SII; ANSI, 1997). In other words, those models consider only energetic masking (EM), where EM is defined as masking of the

*Corresponding author: alech@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

peripheral representation of the signal. However, Stone *et al.* (2012) showed that noises that are typically considered "steady", such as speech-shaped noise (SSN), actually behave more as modulation maskers than as energetic maskers, i.e., they provide "modulation masking" (MM). Yet, EM and MM may not be sufficient to account for speech intelligibility data for some masker types, such as speech, in which case the unaccounted-for masking is labeled as "informational masking" (IM). According to Watson (2005), IM can be divided into two categories, uncertainty and similarity. Uncertainty is explained as a listener's inability to identify the target, whereas similarity prevents a listener from segregating the target and the masker. Multiple factors can reduce the similarity between target and masker, such as spatial separation and fundamental frequency (F_0) information, and thus reduce IM (Bronkhorst, 2000).

The present study investigated the contributions of MM and IM and their interactions in an ITD-only binaural condition with a variable number of maskers (Lőcsei *et al.*, 2015) using a binaural extension of the multi-resolution speech-based envelope power spectrum model (mr-sEPSM; Jørgensen *et al.*, 2013; Chabot-Leclerc *et al.*, 2015). The mr-sEPSM framework considers MM using the SNR in the envelope domain (SNR_{env}) as the decision metric and was shown to account well for intelligibility where IM was not the dominating factor, such as with SSN maskers, sinusoidally modulated maskers, or multi-talker babble. Here, the maskers under consideration were SSN and time-reversed speech maskers, the latter known to produce informational masking, although not as much as regular speech (Rhebergen *et al.*, 2005). In particular, the focus was to analyze how well the SNR_{env} metric could capture the intelligibility change as a function of the total number of maskers and the masker configuration and what could be attributed to IM.

MODEL DESCRIPTION

The structure of the proposed model is presented in Fig. 1. It consists of two monaural realizations of the mr-sEPSM (Jørgensen *et al.*, 2013) and a binaural unmasking pathway implemented as an EC process (Wan *et al.*, 2014).

The model takes as input the noisy speech and the noise-alone signals for each ear. Each signal is processed through a filterbank of 22 gammatone filters covering the frequency range from 63 Hz to 8 kHz with a third-octave spacing. The sub-band envelopes are then extracted using half-wave rectification followed by a fifth-order Butterworth low-pass filter with a cutoff frequency of 770 Hz (Breebaart *et al.*, 2001). Jitter in the time and amplitude domain is applied independently to each sub-band envelope to limit the efficacy of the EC process; all jitters are zero-mean Gaussian processes with standard deviations of $\sigma_\delta = 105$ µs for the temporal jitter and of $\sigma_\varepsilon = 0.25$ for the amplitude jitter (Durlach, 1963). In the monaural pathways, the envelopes are further processed by a modulation filterbank consisting of eight second-order band-pass filters with octave spacing between 2 and 256 Hz. A third-order low-pass filter with a 1-Hz cutoff frequency is applied in parallel to the filterbank.

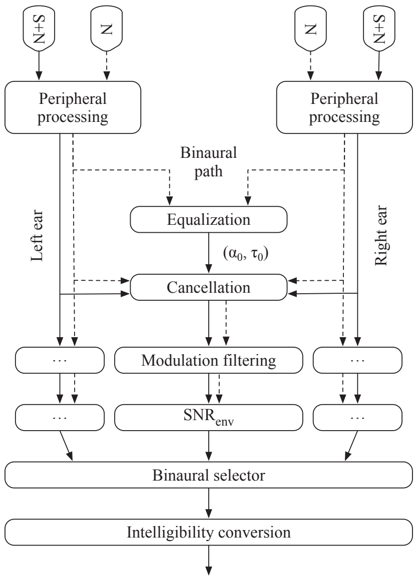


Fig. 1: Structure of the proposed model.

Only modulation filters with center frequencies below one-fourth of their respective peripheral-filter center frequency are used (Verhey *et al.*, 1999). The output of each modulation filter is then divided into non-overlapping segments of duration inversely proportional to the modulation filter's characteristic frequency, i.e., the output of the 4 Hz filter is divided into 250-ms segments. The power of each segment is calculated as its variance and the lower limit of the envelope power is set to −30 dB relative to 100% modulation. The $\mathrm{SNR}_{\mathrm{env}}$ for each segment, i, peripheral channel, p, and modulation channel, n, is calculated as:

$$\mathrm{SNR}_{\mathrm{env},i}(p,n) = \frac{P_{\mathrm{env},S+N,i}(p,n) - P_{\mathrm{env},N,i}(p,n)}{P_{\mathrm{env},N,i}(p,n)}, \qquad \text{(Eq. 1)}$$

where $P_{\mathrm{env},S+N}$ is the power of the noisy speech mixture and $P_{\mathrm{env},N}$ is the power of the noise alone.

The binaural unmasking stage is implemented as described in Wan *et al.* (2014). The jittered envelopes at the output of the peripheral filterbank are the inputs to the EC process, which is applied independently in each channel as well as in short 20-ms time frames, k. For each time-frequency frame, the equalization stage selects the optimal ITD, τ_0, and ILD, α_0, using the following equations:

$$\tau_0 = \arg\min_{\tau} \{\rho\}, \quad |\tau| < \frac{\pi}{\omega}, \text{ and} \qquad \text{(Eq. 2)}$$

$$\alpha_0 = \sqrt{\frac{E_{N,L}}{E_{N,R}}}, \qquad \text{(Eq. 3)}$$

where ρ is the normalized cross-correlation function of the left and right ears within the frame, $E_{N,L}$ and $E_{N,R}$ are the masker energies for the left and right ear, respectively, and ω is the center frequency of the channel of interest. Subsequently, the sub-band signal, B_p, is reconstructed for each channel by summing over all frames.

The unmasked outputs for the noisy speech and the noise alone are then used as inputs to the modulation filtering stage of the mr-sEPSM and processed similarly to the monaural pathways, yielding a binaurally unmasked $\mathrm{SNR_{env}}$, BU-$\mathrm{SNR_{env}}$.

A selection stage then selects the best $\mathrm{SNR_{env}}$ of the left, right and binaural pathways, yielding the complete model's output, the B-$\mathrm{SNR_{env}}$. The B-$\mathrm{SNR_{env}}$ is then averaged across time, and combined optimally across modulation and peripheral filters:

$$\text{B-SNR}_{\text{env}} = \left[\sum_{p=1}^{22} \sum_{n=1}^{9} \text{B-SNR}^2_{\text{env}}(p,n) \right]^{1/2}. \qquad \text{(Eq. 4)}$$

The final B-$\mathrm{SNR_{env}}$ is then converted to intelligibility using a Gaussian psychometric function. The left- and right-ear pathways are combined and converted similarly, yielding alternate model outputs for each ear.

More details about the mr-sEPSM framework and the EC process implementation can be found in Jørgensen *et al.* (2013) and Wan *et al.* (2014), respectively.

METHODS

In this experiment, the speech and masker signals were lateralized individually to the left or right using fixed 33-sample delays (687.5 µs) and the spatial distribution of maskers was systematically varied. The speech material was the DAT corpus (Nielsen *et al.*, 2014), sampled at 48 kHz and recorded by female speakers. The DAT corpus consists of unique meaningful Danish sentences built as a fixed carrier sentence with two interchangeable target words. The maskers were either of one stationary SSN, denoted as $\mathbf{S}_{xy}$ conditions, or 2, 4, or 8 time-reversed sentences from the GRID corpus (Cooke *et al.*, 2006), denoted as $\mathbf{C}_{xy}$ conditions, where y is the total number of maskers and x is the number of maskers on the same side as the target. Both the SSN and the GRID material were shaped to have the same long-term spectrum as the target speech material. The maskers were either all on the same side as the target (e.g., $\mathbf{C}_{44}$), half on the same side (e.g., $\mathbf{C}_{24}$), or all on the opposite side (e.g., $\mathbf{S}_{04}$). The target level was fixed at 65 dB SPL and the maskers were summed before their levels were adjusted to the desired SNR. Model predictions were calculated for 30 randomly selected sentences and for SNRs ranging from -12 to 9 dB in 3-dB steps. The predicted SRT was the average across target sentences. The mean and standard deviation of the psychometric function were fitted to minimized the square error between the "left-ear" of the model and the word-scores as a function of SNR in the collocated condition ($\mathbf{S}_{11}$), as measured by Lőcsei *et al.* (2015).

RESULTS

Figure 2 shows the speech reception thresholds (SRTs) measured by Lőcsei *et al.* (2015) (open squares), the predictions by the proposed model (B-sEPSM; filled squares), as well as the predictions by the left- and right-ear outputs of the B-sEPSM (left- and right-pointing triangles, respectively) for each masker type and configuration. In the $\mathbf{S}_{x1}$ conditions with SSN maskers, the B-sEPSM predicted SRTs lower than the data by 0.5 to about 3 dB, but captured the MR when the maskers were moved to the opposite side. In the $\mathbf{C}_{x8}$ condition, the B-sEPSM accurately captured the MR when 4 and then all 8 reversed-speech maskers were lateralized to the other side. In the $\mathbf{C}_{x4}$ condition, the B-sEPSM predicted a similarly progressive MR as in the $\mathbf{C}_{x8}$ condition, as 2 or all 4 maskers were lateralized to the other side. This is in contrast to the data, where the SRT was constant at about -2.5 dB when 4 or 2 of the maskers were on the same side as the target and then there was about 5 dB of MR once all maskers were on the other side. In the $\mathbf{C}_{x2}$ condition, the B-sEPSM predicted constant SRTs of about -10 dB, irrespective of the masker arrangement. In contrast, the data SRTs were about the same when 2 or 1 masker(s) were collocated with the target at about -4 dB — not significant differences, $p < 0.05$ (Lőcsei *et al.*, 2015) — and then decreased by 4 dB once all maskers were on the other side, similar to the $\mathbf{C}_{x4}$ condition. The SRTs predicted by the left- and right-ear models (left- and right-pointing triangles) depended only on the total number of masker and masker type, irrespective of their configuration. The SRTs were highest in the $\mathbf{C}_{x8}$ and lowest in the $\mathbf{C}_{x2}$ condition, consistent with the increased number of dips in the two-masker condition. Overall, the Pearson correlation coefficient between the B-sEPSM predictions and the data was 0.78 and the mean absolute error was 2.24 dB.

DISCUSSION

The B-sEPSM could account well for the MR due to lateralization in the conditions with the SSN masker ($\mathbf{S}_{x1}$ conditions) and also accurately predicted the SRTs and MR in the $\mathbf{C}_{x8}$ conditions. However, the model was "too good" once the number of maskers was small enough such that IM became the dominating factor, i.e., in the conditions $\mathbf{C}_{x4}$ and $\mathbf{C}_{x2}$. A possible explanation framework has been put forward by Best *et al.* (2013), where it was suggested that intelligibility has a "lower limit" (of SRT) corresponding to the EM/MM present in the condition. In this case, the model's failure can be explained by the fact that it is blind to IM, and thus predicts the lower limit of intelligibility, given EM and MM only.

It is assumed that the mr-sEPSM framework has "perfect segregation" due to its access to the noisy-speech mixture and the noise-alone signals. Therefore, if most of the IM is due to confusion caused by the similarity between the target and maskers, and not to uncertainty about the target, then the B-sEPSM is blind to those confusions (Watson, 2005). An estimate of those confusions in the model would allow it to account for some of the IM in the listener. A possible approach would be to use a model of streaming, such as Elhilali and Shamma (2008) or

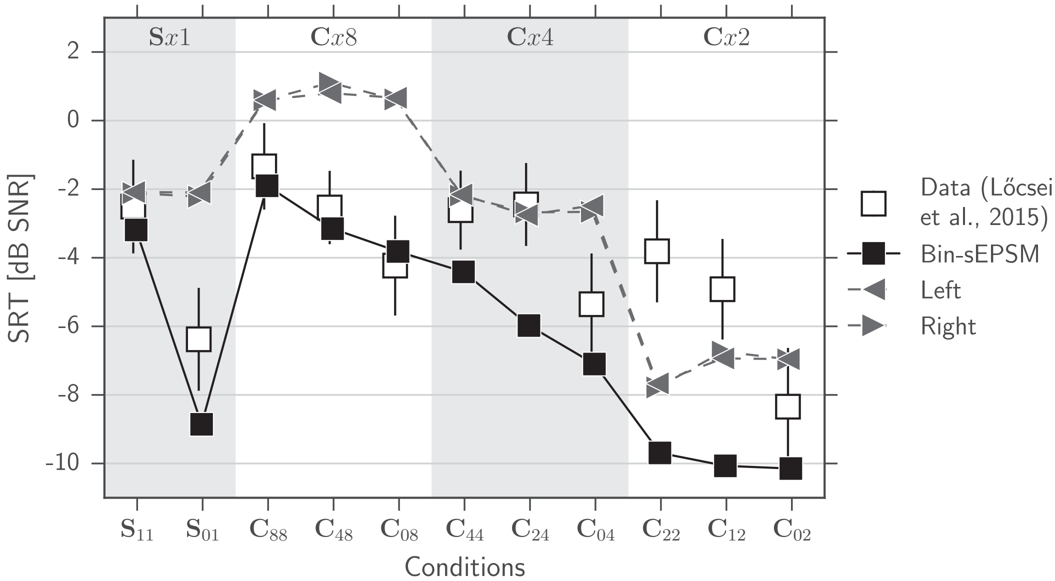


Fig. 2: Measured SRTs (open squares; Lőcsei *et al.*, 2015) and predictions by the B-sEPSM (filled squares) and the 'left-' and 'right-ear' models (triangles) for each condition. **S** conditions are with SSN maskers and **C** conditions are with reverse-babble maskers.

Christiansen *et al.* (2014), and to combine its output with the intelligibility model's output; a single-stream percept would lead to worse intelligibility than a multi-stream percept. Although that approach might prove powerful and possibly more realistic, it would greatly increase the complexity of the models, to the extent that two internal representations would be required. Figure 3 shows a potential similarity measure, calculated as a "modulation distance" between the speech estimate (i.e., $(S+N)-N$) and the noise-alone representations, as a function of the SNR and for different masker configurations. Given the three-dimensional representation of the envelope power as a function of sub-band frequency, modulation frequency, and time frames, the "modulation distance" is calculated as the Euclidean distance between the sub-band and modulation frequency representation (i.e., a 2D matrix) of the speech estimate and the noise for each time frame: The "distance" is then averaged across all time frames.

In Fig. 3, the black lines show the distance for the $\mathbf{C}_{x2}$ condition, where most IM was observed. The distance was largest in $\mathbf{C}_{02}$ condition (dashed line), whereas the distances for conditions $\mathbf{C}_{22}$ and $\mathbf{C}_{12}$ (solid and dotted lines) were almost the same. This mirrors the data, where an MR was observed once all maskers were not collocated with the target, i.e., confusions were resolved once spatial cues were available. In contrast, the distance varied much less as a function of masker location when MM was the dominating factor, such as in the SSN maskers conditions (dark gray lines, $\mathbf{S}_{x1}$) and in the eight-reversed speech masker conditions (light gray lines, $\mathbf{C}_{x8}$).

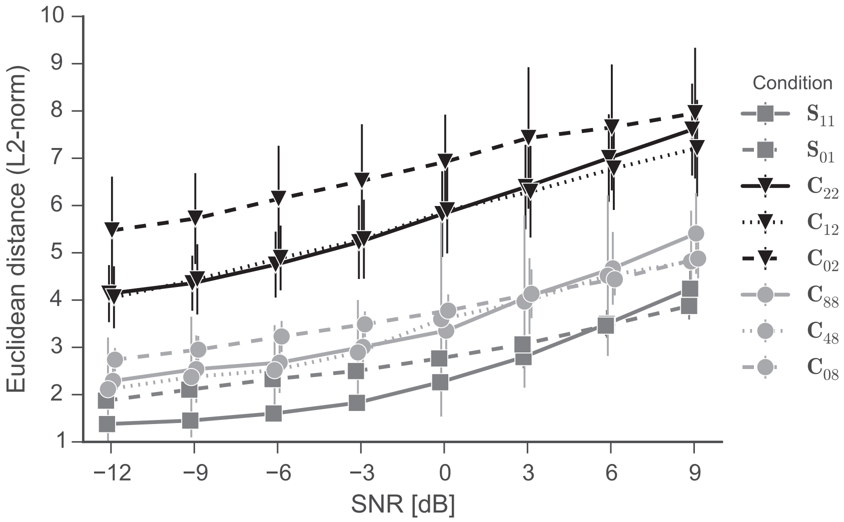


Fig. 3: Euclidean distance between the speech estimate and the noise in the envelope power domain, as a function of SNR. Each line represents a different condition.

In summary, the B-sEPSM could accurately predict SRTs when the dominating factor was modulation masking, but failed when IM became more prevalent. It seems that similarity information between the target estimate and the maskers is available in the multi-resolution envelope power representation and that it could be used to account for some of the IM. However, more work is required in order to combine this information with the binaural model predictions.

ACKNOWLEDGMENTS

This research was supported in part by the National Science and Engineering Research Council of Canada (NSERC), Phonak, and the Technical University of Denmark.

REFERENCES

ANSI (**1997**). *American National Standard Methods for Calculation of the Speech Intelligibility Index.* ANSI S3.5, American National Standards Institute, New York.

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.

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.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am., **110**, 1074-1088.

Bronkhorst, A. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acta Acust. United Ac., **86**, 117-128.

Chabot-Leclerc, A., MacDonald, E.N., and Dau, T. (**2015**). "Predicting binaural speech intelligibility using the signal-to-noise ratio in the envelope power domain," J. Acoust. Soc. Am., submitted.

Christiansen, S.K., Jepsen, M.L., and Dau, T. (**2014**). "Effects of tonotopicity, adaptation, modulation tuning, and temporal coherence in "primitive" auditory stream segregation," J. Acoust. Soc. Am., **135**, 323-333.

Cooke, M., Barker, J., Cunningham, S., and Shao, X. (**2006**). "An audio-visual corpus for speech perception and automatic speech recognition," J. Acoust. Soc. Am., **120**, 2421-2424.

Durlach, N. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am., **35**, 1206-1218.

Elhilali, M. and Shamma, S.A. (**2008**). "A cocktail party with a cortical twist: How cortical mechanisms contribute to sound segregation," J. Acoust. Soc. Am., **124**, 3751-3771.

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.

Lavandier, M. and Culling, J.F. (**2010**). "Prediction of binaural speech intelligibility against noise in rooms," J. Acoust. Soc. Am., **127**, 387-399.

Lőcsei, G., Hefting Pedersen, J., Laugesen, S., Santurette, S., Dau, T., and MacDonald, E.N. (**2015**). "Lateralized speech perception, temporal processing and cognitive function in NH and HI listeners," Poster presented at Speech in Noise Workshop (Copenhagen, Denmark).

Nielsen, J.B., Dau, T., and Neher, T. (**2014**). "A danish open-set speech corpus for competing-speech studies," J. Acoust. Soc. Am., **135**, 407-420.

Rhebergen, K.S., Versfeld, N.J., and Dreschler, W.A. (**2005**). "Release from informational masking by time reversal of native and non-native interfering speech," J. Acoust. Soc. Am., **118**, 1274-1277.

Stone, M.A., Füllgrabe, C., and Moore, B.C.J. (**2012**). "Notionally steady background noise acts primarily as a modulation masker of speech," J. Acoust. Soc. Am., **132**, 317-326.

Verhey, J.L., Dau, T., and Kollmeier, B. (**1999**). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model'," J. Acoust. Soc. Am., **106**, 2733-2745.

Wan, R., Durlach, N.I., and Colburn, H.S. (**2014**). "Application of a short-time version of the equalization-cancellation model to speech intelligibility experiments with speech maskers," J. Acoust. Soc. Am., **136**, 768-776.

Watson, C.S. (**2005**). "Some comments on informational masking," Acta Acust. United Ac., **91**, 502-512.

Perceptual equivalence of test lists in a monosyllabic speech test with different hearing thresholds

ALEXANDRA WINKLER[1,3,*], BRITTA JENSEN[1,3], THOMAS BRAND[2,3], AND INGA HOLUBE[1,3]

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

[2] *Medizinische Physik, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany*

[3] *Cluster of Excellence "Hearing4All"*

EN ISO 8253-3 (2012) describes the requirements and validation of speech material for speech audiometry. Although speech tests are typically applied to listeners with hearing impairment, the validation is conducted with listeners with normal hearing abilities. The aim of this study was to determine the effect of hearing thresholds on the validation results. Since hearing thresholds of listeners with hearing impairment show a large variability, groups of participants with normal hearing listened to the Freiburg monosyllabic speech test (Hahlbrock, 1953) preprocessed with two simulated homogenous hearing losses, as well as to the original speech material. Discrimination functions were fitted to the results and speech levels for speech recognition scores of 50% were determined. According to EN ISO 8253-3 (2012), the perceptual balance of the lists is given when the confidence interval of the speech levels is within 1 dB from the median across all lists. This criterion is not fulfilled for several test lists, which partly differed for the hearing-loss configurations. When taking the measurement accuracy of the experiment into account, consistent deviations are observed in four test lists. The results suggest that if perceptual balance is fulfilled for participants with normal hearing, this might not be valid for participants with hearing impairment. Predictions of speech recognition using the Speech Intelligibility Index could not replicate test list differences.

INTRODUCTION

The German monosyllabic speech test (Hahlbrock, 1953) is a standard test in hearing diagnostics and in the validation of hearing aid fitting. This test consists of 20 lists with 20 monosyllables. For comparison of different settings and/or hearing aids, speech material should be perceptually and phonemically balanced. EN ISO 8253-3 "specifies requirements for the composition, validation and evaluation of speech test materials" used in speech audiometry (EN ISO 8253-3, 2012, p. 1) for

*Corresponding author: alexandra.winkler@jade-hs.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark.
Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4.
The Danavox Jubilee Foundation, 2015.

listeners with normal hearing. Nevertheless, speech audiometry is usually applied to listeners with hearing impairment, who show a large variety in hearing thresholds. Therefore, simulated hearing losses were used in this study to reduce the variability and to be able to resolve test list differences.

PARTICIPANTS

The requirements for the listeners according to EN ISO 8253-3 (2012) regarding age and hearing threshold were:

- Age: 18 - 25 years
- Hearing threshold ≤ 10 dB HL between 0.25-8 kHz and ≤ 15 dB HL at maximum two frequencies

These requirements were fulfilled by all listeners. In total, 120 listeners (80 female, 40 male, median age 23 years) participated in this study. The participants were separated into three groups of 40 listeners each (NH: 31 ♀, 9 ♂, SIM A: 25 ♀, 15 ♂ and SIM B: 24 ♀, 16 ♂). Median hearing thresholds for NH, SIM A, and SIM B are shown in Fig. 1. For group SIM B, discomfort levels were measured in addition to the pure tone thresholds. The discomfort level for this listener group had to be at least 90 dB HL at 500 Hz and 1 kHz to avoid too loud levels for the processed stimuli which were presented via headphones during the test.

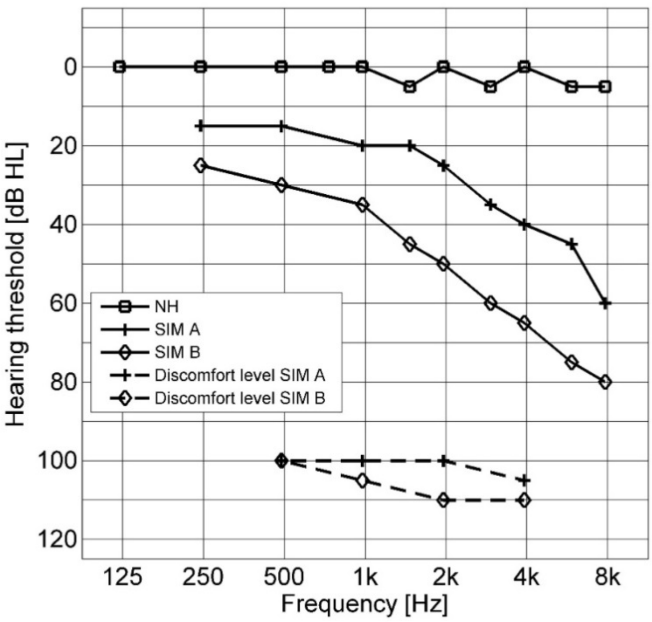


Fig. 1: Median hearing thresholds for NH (squares) and simulated hearing thresholds with uncomfortable loudness levels for SIM A (crosses) and SIM B (diamonds).

In this study the software SIM PRO (HörTech gGmbH) was used to simulate the hearing thresholds and to filter the original speech material. SIM PRO is based on the Master Hearing Aid (Grimm *et al.*, 2006). Multichannel dynamic signal processing and spectral smearing modify test signals using expansion instead of compression, to simulate different hearing losses. For the purpose of presenting sounds to a participant, the original speech material is filtered according to a pure tone audiogram and a discomfort level. The simulated hearing loss SIM A (see Fig. 1) was based on average hearing thresholds of a population aged 65 to 74 years from von Gablenz and Holube (2015). The discomfort levels for SIM A were specified according to Pascoe (1988). For SIM B (see Fig. 1), customer data acquired by German hearing aid acousticians (Nüsse *et al.*, 2014) were used. Those data included hearing thresholds as well as levels of discomfort and were selected to meet speech recognition scores between 30 and 80% for standard diagnostic test levels. All test lists for SIM A and B were processed separately, depending on the presentation level.

TEST SIGNALS

Freiburg monosyllables (recordings from 1969) according to DIN 45626-1 (1995a) and DIN 45621-1 (1995b) were presented monaurally via headphones (Sennheiser HDA200). The levels are overall sound pressure levels (SPL) measured in an ear simulator. All 20 lists were tested with all participants, five lists at four levels each:

- NH: Original speech material presented at 17.5 dB SPL, 23.5 dB SPL, 29.5 dB SPL, and 35.5 dB SPL.
- SIM A: Filtered speech material at 39.5 dB SPL, 45.5 dB SPL, 51.5 dB SPL, and 57.5 dB SPL.
- SIM B: Filtered speech material at 65 dB SPL, 80 dB SPL, 90 dB SPL, and 95 dB SPL.

SPEECH INTELLIGIBILTY INDEX (SII)

The SII (ANSI S3.5, 1997) estimates speech recognition based on the amount of speech contained in each frequency band. Hearing thresholds or different speech material can be used as input for the model. In this study, normal hearing ability was assumed and the band important function of the Northwestern University Auditory Test No. 6 (NU6-monosyllables) was chosen. The speech material for the SII prediction was the same as for the experiments.

RESULTS

Speech recognition curves

Based on the speech recognition scores for the different presentation levels, discrimination functions for NH and SIM A were fitted to the data of all 20 lists separately (Brand and Kollmeier, 2002). For SIM B, linear interpolation between data points above and below 50% speech recognition per list was used, because recognition scores were well below 100%, even for very high presentation levels.

Fig. 2 shows all 20 discrimination functions for NH, SIM A, and SIM B together with the expected data for SIM B. The expected data represent the range of the recognition scores of the customer data (Nüsse *et al.*, 2014) from whom the hearing thresholds and levels of discomfort were selected.

The fitted functions were used to estimate the level for 50% speech recognition (L_{50}). The median L_{50} was 25.2 dB for NH, 50.2 dB for SIM A, and 73.5 dB for SIM B. The range between the lowest and highest L_{50}, i.e. the easiest and the most difficult list, respectively, varied between 4.2 dB (NH), 7.8 dB (SIM A), and 16.5 dB (SIM B).

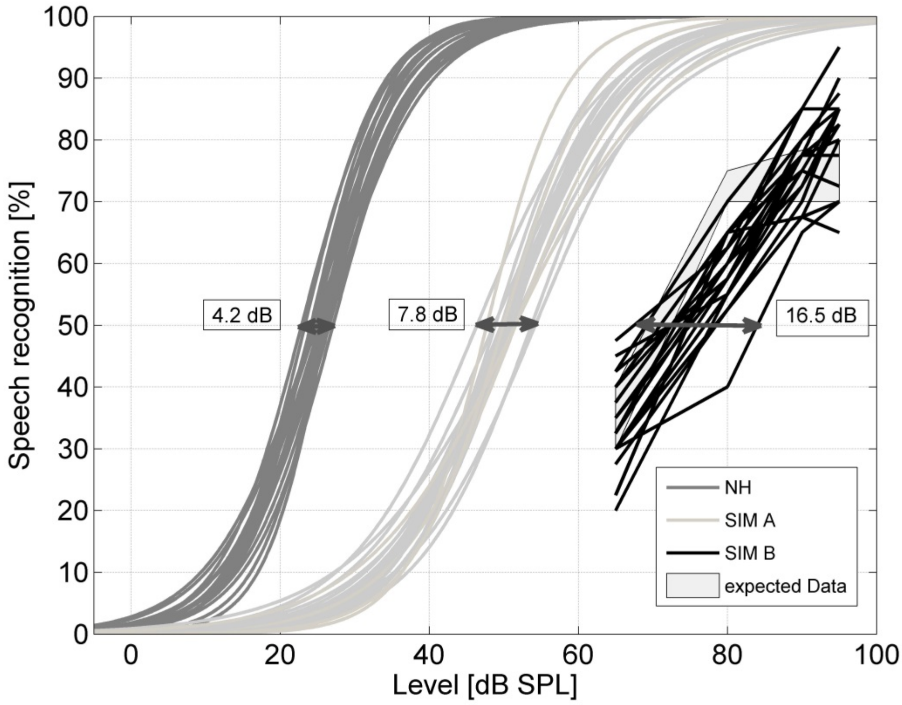


Fig. 2: Discrimination functions for all 20 lists: NH (dark grey), SIM A (gray), SIM B (black), and expected data in light gray. Ranges between the lowest and highest L_{50} are marked by arrows.

Measurement inaccuracy and perceptual balance across lists

Binomial distribution and Gaussian error propagation leads to an inaccuracy in the estimated L_{50} (σ in Fig. 3) according to equation 4 in Brand and Kollmeier (2002). For the calculation of σ for each list, speech recognition scores for one level below and one level above L_{50}, as well as the slope at L_{50} and the number of data points, were required. The number of data points was given by 20 words per list and ten listeners (per list) each. For every subject group, the highest value for σ of all 20 test lists was selected. This led to inaccuracies of from 1.4 to 7.5 dB and therefore of more than 1 dB as predefined in EN ISO 8253-3 (2012, clause 4.9). Nevertheless, several lists showed consistent deviations larger than σ for at least two subject groups (marked lists in Fig. 3).

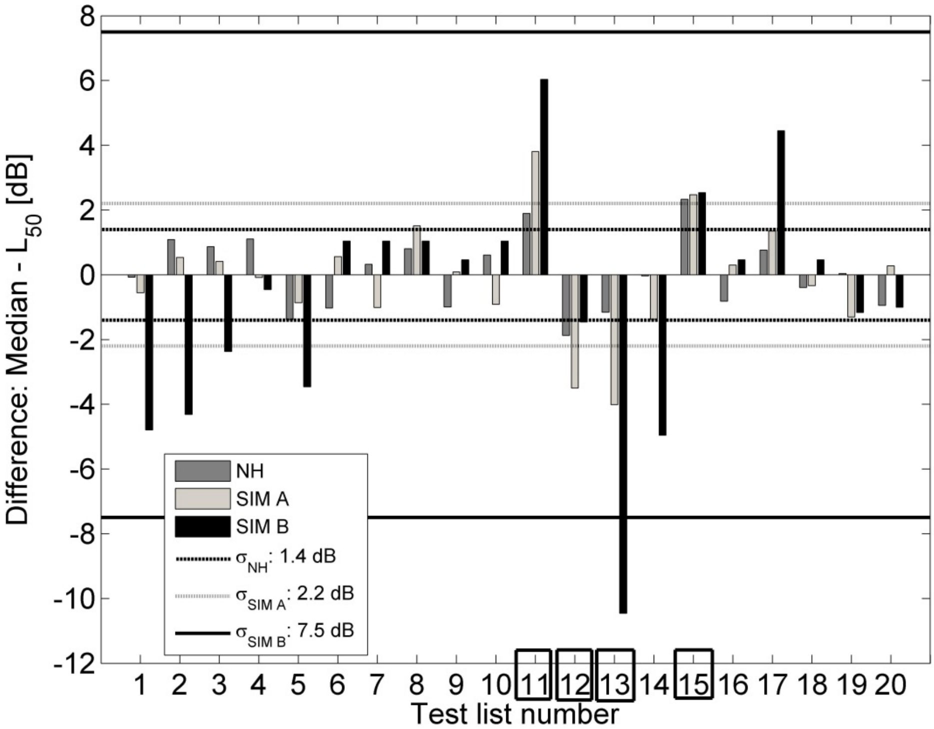


Fig. 3: Deviation of L_{50} from median L_{50} for all lists including inaccuracy (σ) of L_{50}. Marked test list numbers (11, 12, 13, and 15) are above σ for at least two groups of listeners.

Speech Intelligibility Index (SII)

The SII was applied to analyse whether the differences in L_{50} can be predicted by spectral deviations in the test lists for the three subject groups. Frequency analysis (third octave bands) of the lists revealed minor spectral differences for all thresholds and presentation levels. Fig. 4 shows an example for SIM A at 45.5 dB SPL presentation level.

The SII was calculated for each test list, subject group, and presentation level. The SII values were converted to predict speech recognition scores by using the average discrimination function of the original speech material for group NH. Then, linear interpolation was used to calculate the predicted L_{50} for every test list. These values were compared to the L_{50} estimation of the measured speech recognition scores. The correlation between estimated and predicted L_{50} are given in Table 1. The correlation coefficient for SIM A is larger than for the other groups, but none of the correlations are significant ($p > 0.05$).

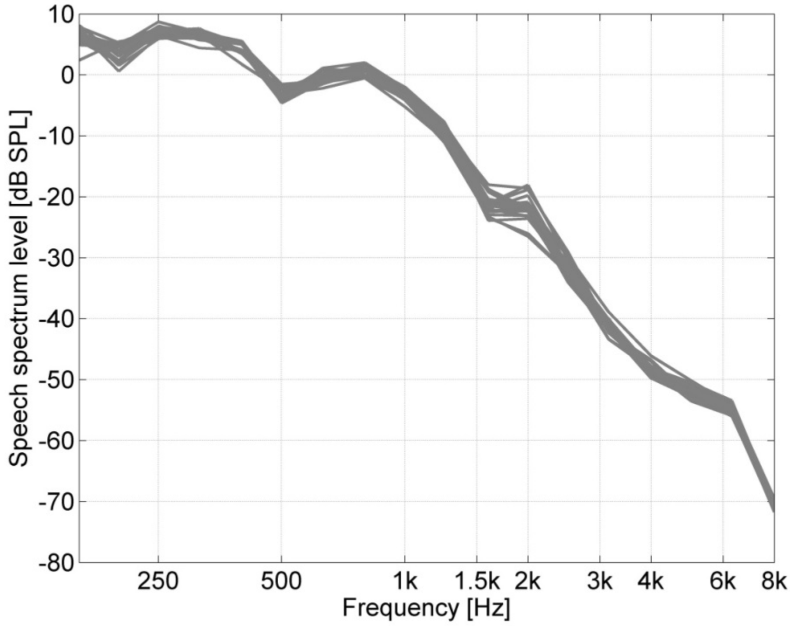


Fig. 4: Third octave spectra of 20 lists for SIM A at 45.5 dB SPL.

NH	SIM A	SIM B
$r = 0.23$ $p = 0.335$	$r = 0.42$ $p = 0.066$	$r = 0.02$ $p = 0.428$

Table 1: Pearson's correlation between estimated and predicted L_{50}.

SUMMARY AND DISCUSSION

In summary, the results of the measurements showed that the Freiburg monosyllables are not perceptually balanced for participants with normal hearing and hearing impairment. This conclusion was drawn although the hearing impairment was simulated and it is questionable whether the recognition scores are similar to those for "real" hearing impairments. Nevertheless, recognition scores for SIM B are well within the range of data from customers of hearing aid acousticians and the conclusion of perceptual imbalance is drawn not only from the results of one subject group.

The L_{50} of several test lists deviate by more than 1 dB from the median values as defined by EN ISO 8253-3 (2012). On the other hand, the shallow slopes of the discrimination functions led to a measurement inaccuracy of up to 7.5 dB. Even for NH, the measurement inaccuracy was calculated to be up to 1.4 dB, which could only be improved to 1 dB by increasing the number of participants from 40 to 80.

Despite this measurement inaccuracy, lists 11, 12, 13, and 15 still deviate noticeably for at least two groups of listeners each and should be avoided in future applications of the Freiburg monosyllabic speech test.

To further analyze the test list deviations, the SII was used as an objective measure based on the spectrum of the speech material. Unfortunately, none of the exceptional lists (11, 12, 13, and 15) differ or shows larger variation compared to the other lists of the test (s. Fig. 4). Hence, the predicted L_{50} of the different lists within one group was very similar. Therefore, there seems to be no direct relation between the measurement and the prediction, even though there was a tendency towards correlation of predicted and measured L_{50} for the mild hearing loss (SIM A).

Other approaches to explain observed test list deviations might be, for example, a possible phonemic imbalance of the test lists – even though the phoneme distribution was taken into account by Hahlbrock (1953) – or possible differences in word popularity or knowledge. These criteria will need further examination.

ACKNOWLEGDEMENTS

We thank Daniel Berg (HörTech gGmbH), Lena Haverkamp, Sascha Bilert for support and data collection. This project was funded by Ph.D. program Jade2Pro. English language support by www.stels-ol.de.

REFERENCES

Deutsches Institut für Normung **(1995a)**. "DIN 45626-1; Tonträger mit Sprache für Gehörprüfung, Teil 1: Tonträger mit Wörtern nach DIN 45621-1," Beuth Verlag, Berlin.

Deutsches Institut für Normung **(1995b)**. "DIN 45621-1; Sprache für Gehörprüfung. Teil 1: Ein- und mehrsilbige Wörter," Beuth Verlag, Berlin.

American National Standards Institute **(1997)**. "ANSI S3.5-1997; Methods for calculation of the Speech Intelligibility Index (SII)," Acoustical Society of America, New York.

Brand T. and Kollmeier B. **(2002).** "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am., **111**, 2801-2810.

European Committee for Standardization **(2012)**. "EN ISO 8253-3; Acoustics – Audiometric test methods – Part 3: Speech Audiometry," Beuth Verlag, Berlin.

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.

Hahlbrock K.H. **(1953)**. "Über Sprachaudiometrie und neue Wörterteste", ArchivOhr-usw. Heilk. u. Z.Hals- usw. Heilk., Bd. **162**, 394-431.

HörTech gGmbH: http://www.hoertech.de (accessed on 31st of July 2015).

Nüsse T., von Gablenz P., Holube I., and Bonsel H. **(2014)**. "Analyse tonaudiometrischer Daten von Kunden einer Hörgeräteakustikerkette," Hörakustik, **10**, 8-13.

Pascoe, D.P. **(1988)**. “Clinical measurements of the auditory dynamic range and their relation to formulas for hearing aid gain,” in Proc. 13th Danavox symposium: *Hearing Aid Fitting: Theoretical And Practical Views*. Ed. J. Jensen. Copenhagen, Denmark.
von Gablenz, P. and Holube, I. (**2015**). “Hearing threshold distribution and effect of screening in a population-based German sample,” Int. J. Audiol., 1-16, doi: 10.3109/14992027.2015.1084054

Investigating low-frequency compression using the Grid method

MICHAL FERECZKOWSKI*, TORSTEN DAU, AND EWEN N. MACDONALD

Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark

There is an ongoing discussion about whether the amount of cochlear compression in humans at low frequencies (below 1 kHz) is as high as that at higher frequencies. It is controversial whether the compression affects the slope of the off-frequency forward masking curves at those frequencies. Here, the Grid method with a 2-interval 1-up 3-down tracking rule was applied to estimate forward masking curves at two characteristic frequencies: 500 Hz and 4000 Hz. The resulting curves and the corresponding basilar membrane input-output (BM I/O) functions were found to be comparable to those reported in literature. Moreover, slopes of the low-level portions of the BM I/O functions estimated at 500 Hz were examined, to determine whether the 500-Hz off-frequency forward masking curves were affected by compression. Overall, the collected data showed a trend confirming the compressive behaviour. However, the analysis was complicated by unexpectedly steep portions of the collected on- and off-frequency forward masking curves.

INTRODUCTION

There is an ongoing debate concerning the characteristics of human basilar-membrane input-output (BM I/O) functions in the low frequency (<1000 Hz) range, particularly when they are obtained using forward-masking experiments. These methods rely on an assumption that the response of the BM is linear for a stimulus whose frequency is approximately an octave lower than the characteristic frequency (CF) for that position (Robles and Ruggero, 2001). Thus, BM I/O functions are characterized using two conditions in a temporal masking curve (TMC) paradigm (Nelson *et al.*, 2001). In the "on-frequency" condition, the masker frequency is the same as that of the masked signal. In the "off-frequency" condition, the masker frequency is set approximately one octave below that of the signal and the resulting threshold is taken as a linear reference for the corresponding on-frequency threshold. The thresholds obtained from the off- and on-frequency TMCs are paired by the masker-signal gap and the resulting scatterplot is assumed to approximate a BM I/O at the cochlear site corresponding to the signal's frequency.

A key assumption of the TMC method is that the rate of recovery from forward masking is independent of both the level and the frequency of the masker. Recently,

*Corresponding author: mfer@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

both aspects of this assumption have been questioned. Wojtczak and Oxenham (2009) observed that, for a high-level (e.g., above 83 dB SPL) off-frequency masker in normal-hearing (NH) listeners, the rate of recovery from forward masking was level dependent. Stainsby and Moore (2006) measured TMCs in three listeners with a flat mild-to-severe hearing loss and found that the rate of recovery from forward masking was larger for low (500 Hz) than for high (4000 Hz) center frequencies. Consequently, the BM I/O compression ratio (CR) would be significantly smaller at low than at high frequencies. In contrast, Lopez-Poveda and Alves-Pinto (2008) found evidence that the rate of recovery was not frequency dependent and concluded that the CR at low and high frequencies is similar.

Both physiological experiments and psychophysical data suggest that BM I/O is linear at very low levels (Plack and O'Hanlon, 2003; Lopez-Poveda and Alves-Pinto 2008). Therefore, it is assumed that the nonlinear gain at these levels is constant and that the slopes of on-frequency TMCs at masker levels near hearing threshold should be similar to the corresponding off-frequency TMC slopes, provided that the off-frequency TMCs reflect linear processing at BM. If the off-frequency reference was influenced by compression, the BM I/O curve derived from the off- and on-frequency curve pair would show slopes higher than 1 at the very low input levels.

Fereczkowski (2015) developed the Grid method as an alternative to other tracking methods. The most important difference between the standard methods and the Grid is that the latter varies more than one experimental parameter during a single experimental run. The main advantage of this approach is its relatively high time efficiency. As shown in Fereczkowski (2015), the method allows the experimenter to locate and track a single TMC threshold curve within 2-4 minutes, which is comparable to the time needed to estimate 2-3 single thresholds by means of the Single-Interval Up-Down (SIUD) method (Lecluyse and Meddis, 2009) and maximally one threshold when using transformed up-down paradigms.

The characteristics of the BM I/O estimates obtained here were compared to those of previous studies. Moreover, it was hypothesised that if the off-frequency TMC is affected by compression, the BM I/O functions obtained from pairing the low-level linear on-frequency and the off-frequency curves would show an expansive characteristic in the low-level region.

METHOD

Listeners, stimuli, and procedure

Individual ears were tested from 8 clinically normal-hearing (audiometric thresholds <20 dB HL) listeners (7 males and 1 female with a mean age of 27.8 years). All listeners provided written informed-consent and the procedure was approved by the National Research Ethics Committee of Denmark.

Masking curves were measured at two signal frequencies: 500 and 4000 Hz. In the off-frequency condition, the masker frequency was set an octave below that of the signal. The masker duration was 200 ms. The signal duration was 16 ms and 24 ms (raised cosine gating, no steady state) when the signal frequencies were 4000 and 500 Hz, respectively. Onset and offset ramps of the masker tone were the same as those for the corresponding signal. The signal level was 7 dB sensation level (SL).

The maximum masker level allowed in the procedure was 85 dB SPL for the on-frequency condition and 95 dB SPL for the off-frequency condition. The minimum level allowed was 10 dB below the individual's probe threshold. Finally, the set of all possible levels was created between these limits, with 2 dB resolution. Possible durations of the masker-signal gap (measured between zero-amplitude points) belonged to the following set: 1, 2, 3, 4, 6, 8, 12, 16, 24, 32, 48, 64, 96, 128, 160, 192, and up to 352 ms in 32-ms steps. All stimuli were generated on a PC running Matlab and a 24 bit RME soundcard. The presentation was monaural via Sennheiser HDA200 headphones in a double walled booth.

Since the maximum masker level in the on-frequency condition was set to 85 dB SPL, the BM I/O thresholds were approximated by two-section fits (i.e., two straight lines that intersected at a knee point, KP). The mean and standard deviation of the fitted parameters were estimated using the bootstrapping method. In the variant used here, a single fit to the complete data set was provided, consisting of N points, and the fitted value of the parameter under investigation was used as the estimate of the mean. Subsequently, fits were performed to all N possible N-1 element sets in order to estimate the standard deviation of the mean.

The experimental procedure consisted of three steps. First, the absolute thresholds for the signals were measured. Subsequently, the listeners were trained, for at least two hours, in the forward masking task. In the data-collection phase, three repetitions of each of the four TMC conditions were run. In each run the threshold curve was sampled once (from lowest to highest levels). The tracking rule used was 3-up 1-down, 2-alternative forced-choice (AFC) and feedback was provided to the listener after each response.

RESULTS

Using linear interpolation, TMC thresholds were estimated for each experimental run and then averaged. The left panel of Fig. 1 presents mean TMCs collected for all listeners (NH1-NH8). Each row represents data collected for a single listener and each column corresponds to a different combination of TMCs. The left column presents TMC thresholds obtained for a 4-kHz signal. The right column presents TMC thresholds obtained for a 500-Hz signal. The squares represent the on-frequency TMCs and the triangles represent the off-frequency thresholds. The filled symbols represent thresholds for which the masker-signal gap was no greater than 10 ms. The distinction has been highlighted in both panels because, in some cases, the TMC slopes increase markedly below the 10-ms gap. The mean on and off-frequency TMCs, collected for a single frequency, were paired by the masker-signal gap to estimate individual BM I/O thresholds. The right panel of Fig. 1 presents mean BM I/O thresholds, along with the corresponding two-section fits. For clarity, only the fits performed on the complete sets are shown. The leftmost columns present BM I/Os estimated from the data shown in the corresponding columns of the left panel. The rightmost column represents BM I/Os derived by pairing on-frequency TMC thresholds for a 500-Hz signal with the off-frequency TMC thresholds for a 4-kHz signal. Each row presents a single listener's data.

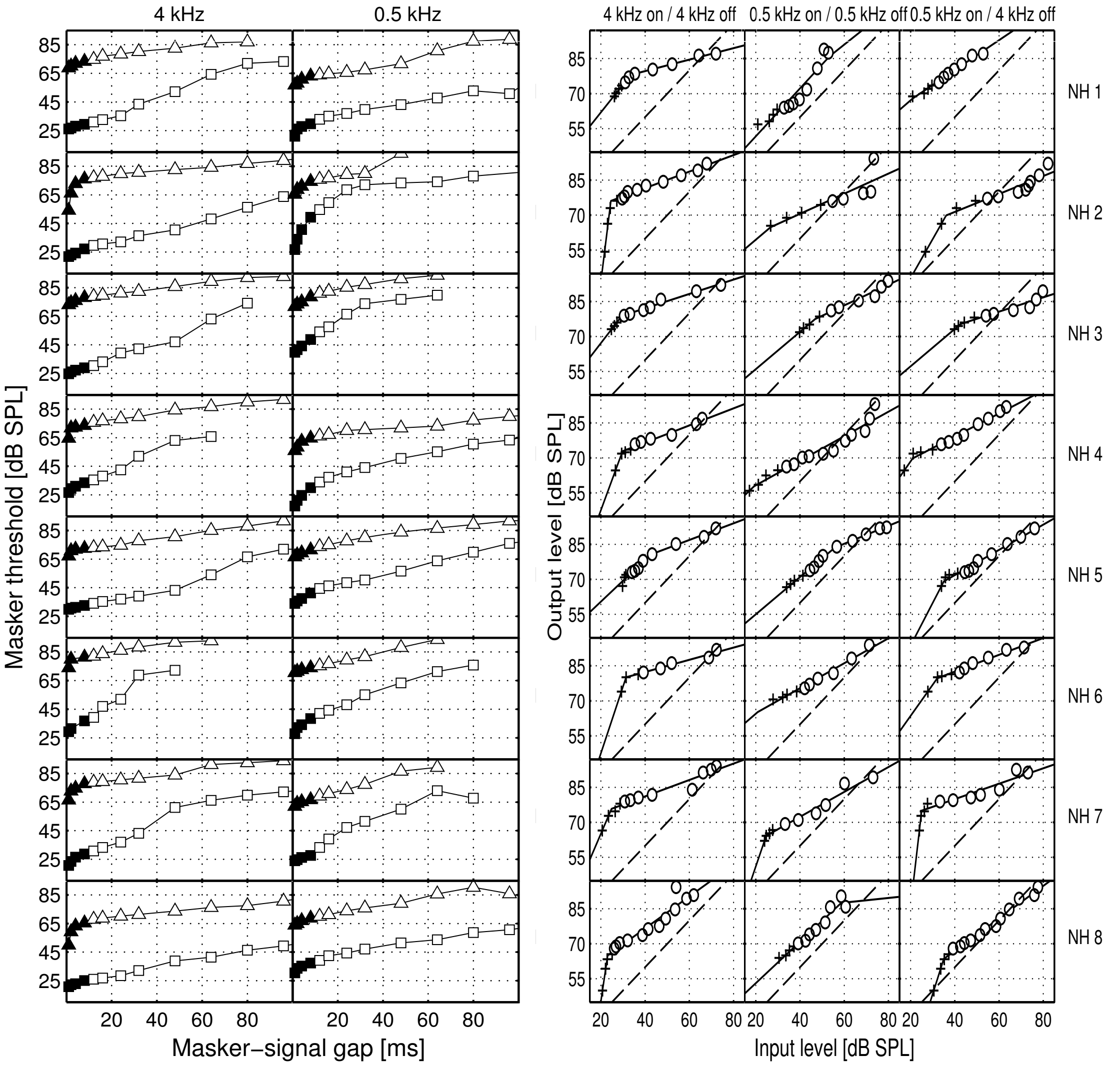


Fig. 1: TMCs (left panel) and estimated BM I/O functions (right panel) for each individual listener. In the left panel, the data for a 4-kHz and 0.5-kHz signal are shown in the left and right column, respectively. The squares and triangles represent the on- and off-frequency thresholds, respectively. In the right panel, the two leftmost columns present BM I/O functions from the two columns in the left panel. The rightmost column presents the BM I/O estimate for a 500-Hz tone using the 4000-Hz off-frequency curve as a reference. In both panels, the filled symbols or crosses indicate results obtained for masker-signal gaps shorter than or equal to 10 ms.

Each two-section fit was defined by four parameters: a linear slope at low levels, a compressive slope at medium levels, and the two coordinates of the KP. A constrained minimization routine of the fitted mean square error to the collected data-points was run. The constraints were chosen based on assumptions regarding human BM I/O functions. The slope of the "linear" part of the I/O function was assumed to be greater than 0.8 and greater than the compressive slope. Thus, the compressive slope was assumed to be between 0.1 and 0.8. This corresponds to the CR ranging between 1.25 and 10. Thus, the lower bound of the CR was assumed to be below the values typically found for NH (2.5-6, e.g., Lopez-Poveda *et al.*, 2003). This was allowed in case the behaviour of the off-frequency TMC was compressive. Finally, the input level at KP was expected to be lower than 60 dB SPL (Lopez-Poveda and Johannesen, 2012). If a fitted KP (and thus one of the two fitted slopes) lay outside the level range of the corresponding data points, such values of KP and slope were discarded.

Table 1 presents the details of the two-section fits from the right panel of Fig. 1. The last row of Table 1 presents the averages computed from individual data.

4 kHz signal			**0.5 kHz signal**			**0.5 kHz on / 4 kHz off**		
KP **[dB SPL]**	**CR**	**Lin. slope**	**KP** **[dB SPL]**	**CR**	**Lin. slope**	**KP** **[dB SPL]**	**CR**	**Lin. slope**
34 /34	4.1/4.1	1.2/1.3	- /52	1.3/1.6	- /1.4	- / 47	1.6/6.8	- /0.8
25 /32	2.8/3.1	7.6/1.0	- / -	2.3/1.6*	- / -	36/ -	2.6/2.0	1.7/ -
30 /31	3.2/3.3	1.2/0.9	47 / -	2.2/2.2	0.8/ -	43/ -	3.3/2.9	0.8/ -
30 / -	2.7/3.0	2.6/ -	20 / -	1.9/1.8	0.9/ -	21/ -	2.0/1.8	1.3/ -
43 /42	2.8/2.8	0.9/1.0	56 /53	2.6/2.6	0.8/1.1	36/51	1.9/1.8	1.4/0.8
31 /39	3.8/3.8	2.9/0.8	- / -	1.9/1.6	- / -	30/47	2.8/3.3	2.5/0.8
25 / -	2.8/2.9	2.2/ -	25 / -	1.9/1.8	2.8/-	25/ -	3.0/2.8	8.3/ -
23 / -	1.5/1.5	6.4/ -	59 /58	10*/10*	0.9/1	35/ -	1.4/1.4	3/ -
30 /36	***3.0/3.1***	***3.1/ 1.0***	***41 /55***	***2.0/1.8***	***1.2/1.2***	***33/48***	***2.3/2.9***	***2.5/0.8***

Table 1: Individual parameters from BM I/O fits plotted in Fig. 1. Additional fits were performed to the BM I/O data points represented by circles alone. The values of the corresponding parameters are shown after a slash (/). A hyphen is used when the fitted parameter value was discarded, as described in the Method. Unreliable CR estimates (i.e., with standard deviations greater than 1) are marked with a star. The last row presents the averages of the estimates in the corresponding columns.

Applying the Grid method to discrimination tasks

Stainsby and Moore (2006) performed their forward masking experiments using a 3-interval forced choice procedure with a 3-up 1-down tracking method and their average threshold acquisition time was 10 minutes. Here, a 2-interval 3-up 1-down method was tested and the average threshold acquisition time was about 1 minute. Taking the difference in number of intervals into consideration, the average threshold acquisition time for the Grid method can be estimated to be about 1.5 minutes, which suggests that the Grid method can be to 6-7 times faster than the reference method. However, this comparison of the time-efficiency between the methods is not complete, because it does not take possible differences in the accuracy of both methods into account. In Fereczkowski (2015), the Grid method was shown to offer a similar accuracy as the SIUD method. Further experiments and/or Monte Carlo simulations are needed for a more direct comparison of the accuracy of the transformed up-down and Grid methods.

DISCUSSION

KP and CR estimates

When the 4-kHz off-frequency TMC data were used as the linear reference in BM I/O estimations, the KP estimates could be found in 15 out of 16 tested cases. The average estimated KP levels were 33 and 30 dB SPL for 0.5 and 4 kHz, respectively. This difference was not statistically significant.

The individual CR estimates reported by Lopez-Poveda *et al.* (2003) varied between 2.5 and 6 and most of the estimates fell between 3 and 5. Further, it was found that the CR estimated for a CF of 500 Hz was similar to that estimated at 4 kHz. Moreover, the off-frequency TMCs collected for the 500-Hz signal were found to be steeper than those collected for the 4-kHz signal. Thus, the CRs estimated based on the 500-Hz reference would be lower than those estimated using the 4-kHz off-frequency TMC as the reference. All these findings were replicated in the present study. When using the 4-kHz reference, CR estimates for CFs of 0.5 and 4 kHz were not significantly different ($p = 0.076$). However, the CRs estimated at 0.5 and 4 kHz obtained with the 0.5 and 4-kHz linear references, respectively, were significantly different ($p = 0.013$) with the CR estimate at 4 kHz being, on average, greater by a factor of 1.7. However, the average CR values estimated here are lower than those reported in Lopez-Poveda et al. (2003). An explanation for the discrepancy could be that here, in some cases, the masker-signal gap ranges tested in the on-frequency condition were greater than the corresponding ranges tested in the off-frequency condition. Since the collected off-frequency TMCs were not extrapolated, the dynamic range of the corresponding BM I/O functions that were tested was limited. This can be seen in the case of the BM I/O functions estimated for listener NH8 and the two BM I/O functions of listener NH1, where the 500-Hz on-frequency TMC was used. For these three cases, the values of CR estimated were the lowest of all of the CR estimates. Note that the estimate of CR = 10 was omitted as it was considered unreliable due to the large uncertainty of the estimate. However, excluding listener NH8 and the two BM I/O curves of listener NH1 where the 500-Hz on-frequency TMC was used, did not affect the conclusions.

Behaviour of the BM I/O curves at low levels

The main motivation for choosing a very low probe level (7 dB SL) was to test very low masker levels and thus enable testing very low BM I/O levels at 500 Hz. It was expected that, if the 500-Hz off-frequency TMC characteristic was influenced by compression, the corresponding BM I/O function would show an expansive characteristic at the low-level input range. The average of the fitted low-level slopes of the collected 500 Hz BM I/Os was 1.2, which supports this hypothesis. However, there are reasons to question the reliability of this result.

First, the corresponding average slope measured at 4 kHz was 3.1. This is much higher than any of the values reported in the literature for high CFs (e.g., 1.46 at 8 kHz in Lopez-Poveda *et al.*, 2003). Second, examination of the TMCs collected at 4 kHz revealed very steep (up to 10 dB/ms) portions of the TMCs for very low masker-signal gap values (1-10 ms, listeners 2, 4, 6-8 at 4 kHz and listeners 1, 2, and 4 at 500 Hz). The reason for this behaviour is not clear. It is unlikely that this is an effect of the test procedure, since the effect was not found in all cases. In some cases it was observed in the off-frequency functions but not in the corresponding on-frequency functions (the cases at 4 kHz). In some cases, it was observed in both TMCs (listeners 2 and 4 tested at 500 Hz). It is unlikely that the effect was due to insufficient training as listeners 6 and 8 had more than 10 hours of experience in forward masking tasks prior to conducting the present experiment.

In order to further investigate this effect, linear extrapolation was used to find the expected masker level of the collected on-frequency TMCs at 0 ms masker-signal gap. The extrapolated value was compared to the signal level. It was found that the mean difference between the compared values was 1.8 ± 5.1 dB, similar to the expected masker threshold in the simultaneous masking task. Thus, the observed effect might be due to a difference in the acoustic cues used by listeners for very short vs. longer masker-signal gaps.

The analysis of the collected TMC thresholds was repeated but without the thresholds collected for gaps ≤ 10 ms. The CR estimates from this second analysis did not differ significantly (paired *t*-test returned $p = 0.45$). This was not surprising since the compressive region of the BM I/O generally corresponds to the thresholds obtained for the masker-signal gaps above 10 ms. However, while 20 low-level slope estimates were obtained from the original analysis, only 11 were obtained in the re-analysis. In the case of TMCs obtained for the 4-kHz signal (on- and off-frequency), 5 low-level slope estimates were obtained and their average was 1.0. In the case of TMCs obtained for the 0.5-kHz signal (on- and off-frequency), 3 low-level slope estimates were obtained and their average was 1.2, with the minimum of 1.0. This supports the initial hypothesis that the off-frequency TMC collected for the 500-Hz signal is affected by compression, which is in line with the conclusions of

Lopez-Poveda and Alves-Pinto (2008) regarding the nature of compression at low CFs.

SUMMARY

In this study, TMCs were obtained with a 2-interval 3-up 1-down Grid procedure and the corresponding BM I/O functions were estimated for 8 NH listeners at two frequencies: 500 and 4000 Hz. The KP and CR estimates derived from the BM I/O estimates were found to be comparable to those from the literature. The time-efficiency was estimated to be 6-7 times higher than that of the reference AFC method.

Some of the obtained TMCs exhibited steep portions for low masker-signal gaps, which was not expected and inconsistent with the data shown in other studies. This behaviour may be due to listeners using different cues when performing the task with the small masker-signal gaps. The collected data showed a trend confirming the hypothesis that the off-frequency TMCs at low CFs may be subject to cochlear compression.

REFERENCES

Fereczkowski (**2015**). *Time-Efficient Behavioural Estimates of Cochlear Compression.* Ph. D. thesis, Technical University of Denmark.

Lecluyse, W. and Meddis, R. (**2009**). "A simple single-interval adaptive procedure for estimating thresholds in normal and impaired listeners," J. Acoust. Soc. Am., **126**, 2570-2579.

Lopez-Poveda, E.A., Plack, C.J., and Meddis, R. (**2003**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am., **113**, 951-960.

Lopez-Poveda, E.A. and Alves-Pinto, A. (**2008**). "A variant temporal-masking-curve method for inferring peripheral auditory compression," J. Acoust. Soc. Am., **123**, 1544-1554.

Lopez-Poveda, E.A. and Johannesen, P.T. (**2012**). "Behavioral estimates of the contribution of inner and outer hair cell dysfunction to individualized audiometric loss," J. Assoc. Res. Otolaryngol., **13**, 485-504.

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.

Plack, C.J. and O'Hanlon, C.G. (**2003**). "Forward masking additivity and auditory compression at low and high frequencies," J. Assoc. Res. Otolaryngol., **4**, 405-415.

Robles, L. and Ruggero, M.A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev., **81**, 1305-1352.

Stainsby, T.H. and Moore, B.C. (**2006**). "Temporal masking curves for hearing-impaired listeners," Hear. Res., **218**, 98-111.

Wojtczak, M. and Oxenham, A.J. (**2009**). „Pitfalls in behavioral estimates of basilar-membrane compression in humans," J. Acoust. Soc. Am., **125**, 270-281.

Simultaneous measurement of auditory-steady-state responses and otoacoustic emissions to estimate peripheral compression

RAUL H. SANCHEZ AND BASTIAN EPP*

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

Assessment of the compressive nonlinearity in the hearing system provides useful information about the inner ear. Auditory-steady state responses (ASSR) have recently been used to estimate the state of the compressive nonlinearity in the peripheral auditory system. Since it is commonly assumed that outer hair cells in the inner ear play an important role in the compressive nonlinearity, it is desirable to selectively obtain information about the inner ear. In the current study, the signal in the ear canal present during ASSR measurements is utilized to extract sinusoidally-amplitude modulated otoacoustic emissions (SAMOAEs). It is hypothesized that the stimulus used to evoke ASSRs will cause acoustic energy to be reflected back from the inner ear into the ear canal, where it can be picked up as an otoacoustic emission (OAE) and provide information about cochlear processing. Results indicate that SAMOAEs can be extracted while measuring ASSRs using sinusoidally-amplitude modulated tones. However, comparison of simulations using a transmission model and the data show that the SAMOAE measured above 50 dB SPL are strongly influenced by the system distortion. A robust extraction and evaluation of SAMOAE in connection with ASSR may be possible by a proposed method to minimize the distortion. The ability to evaluate SAMOAE over a large input level range during ASSR measurement will provide information about the state of the peripheral auditory system without the need of additional measurement time.

INTRODUCTION

The healthy auditory system exhibits a nonlinear behavior related to the frequency selectivity and the sensitivity to soft sounds. In psychoacoustical experiments, it is commonly assumed that outer hair cells are the main contributor to the compressive nonlinearity. The growth of neural responses suggests however, that peripheral compression also occurs at retro-cochlear stages (Cooper and Yates, 1994). Since psychoacoustical experiments such as growth of masking (Plack and Oxenham, 1998) and temporal masking curves (Nelson and Schroder, 2004) allow the evaluation of the system as a whole, they should be interpreted as the total compression of the system rather than exclusively of the inner ear. In addition, comparison across measures is

*Corresponding author: bepp@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

difficult since different stimuli (pure tones, band-limited noises, etc.) are used in the different techniques.

Physiological non invasive measurements using sinusoidally amplitude modulated (SAM) tones revealed that the basilar membrane grows compressively as a function of input level for these stimuli (Rhode and Recio, 2001). If peripheral compression is, at least partially, due to cochlear compression, the amplitude of auditory steady state responses (ASSR) measured using SAM tones as a function of level, can be assumed to reflect the compressive growth of the cochlear nonlinearity. Recently, Encina Llamas *et al.* (2014) showed compressive input-output functions by measuring ASSR evoked by SAM tones as a function of stimulus level. In the same study, distortion product otoacoustic emissions (DPOAE) level growth functions were measured for the same listeners, showing smaller compression than the ASSR data. However, the nonlinear nature of generation of the DPOAE complicates a direct comparison of the results, especially on an individual basis.

In order to avoid this difficulty, evaluation of OAEs at the stimulus frequencies might help to facilitate the interpretation. Such stimulus frequency otoacoustic emissions (SFOAE) have been measured as a function of the presentation level (Schairer *et al.*, 2003). Their data also show a compressive growth as a function of stimulus level. The ability to extract information about cochlear compression from SAM tones rather than from pure tones will enable the simultaneous measurement of SAMOAE and ASSR, and hence provide two sources of information about auditory processing in the inner ear without the need of additional measurement time.

In order to estimate OAE, the stimulus sound pressure P_0^x at the ear canal needs to be estimated. In SFOAEs, this is commonly done using either a suppression or compression paradigm (Kalluri and Shera, 2007). In the suppression paradigm, the OAE is extracted by comparison of the ear canal sound pressure in the presence and the absence of a suppressor tone, aimed to suppress the basilar membrane vibration at the cochlear partition corresponding to the stimulus frequency. In the compression paradigm, the OAE is extracted by scaling and subtraction of two intervals at different stimulus levels, assuming compressive growth of the OAE and linear scaling of the stimulus pressure in the ear canal. Since SAM tones are similar to pure tones in terms of bandwidth, this technique might also be applicable to SAM tkones. Nevertheless, ASSR recordings require long steady state intervals in order to capture the envelope-following responses.

The current study presents data on SAMOAEs measured following a method that allows simultaneous ASSR and OAE recordings and presents an approach to improve the signal-to-noise ratio of the SAMOAEs by reduction of transducer artefacts.

METHOD

Otoacoustic emissions evoked by SAM tones were measured at four different carrier frequencies. The modulation frequencies were chosen to match the stimuli used in

Encina Llamas *et al.* (2014). To verify the applicability of the compression paradigm for SAM tones, SFOAE at a frequency of 2 kHz were measured. To estimate the influence of system distortion, the measurements were repeated using an ear simulator (coupler B&K type 4157).

Measurement setup

Stimuli were generated by a custom software written in MATLAB, using a 24-bit soundcard (RME Fireface 800) with sample rate 48 kHz. After pre-amplification (HB7) the stimuli were transmitted to an Etymotic ER-10B+ probe. The recording signal was obtained by the probe microphone with +20 dB amplification, and band-pass filtered using an analogue bandpass filter between 0.3 and 6 kHz. Calibration was performed using the ear simulator mentioned above for each frequency and stored in the software.

Subjects

Five subjects with normal hearing thresholds (age: 24-31 years) were recruited for the experiment. Subjects were seated in an armchair in a double-wall isolated booth. Subjects were allowed to sleep or read. The time of the whole protocol was limited to three hours including breaks between conditions. All experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-3-2013-004).

Stimulus conditions SAMOAE

Stimuli were presented in separated pairs with a level difference of 6 dB. For each set of conditions stimulus levels, the lower level was set from 10 to 70 dB SPL in steps of 10 dB. In the second sequence, the level increment of 6 dB was obtained by playing the same stimulus phase-matched to both channels of the probe (Fig. 1). Recordings were made for four different center frequencies: $f_c = 1002$ Hz modulated by $f_m = 87$ Hz, $f_c = 2005$ Hz modulated by $f_m = 93$ Hz, $f_c = 4011$ Hz modulated by $f_m = 98$ Hz, and $f_c = 498$ Hz modulated by $f_m = 81$ Hz. The modulation depth was $m = 0.85$ for all conditions.

Stimulus conditions SFOAE

A three-interval paradigm using suppression and compression was included in the protocol (see Kalluri and Shera, 2007). The three intervals consisted of a 7-second sequence with a 0.25-s ramp-in and out at the beginning and end of each interval. The first interval contained the stimulus, the second interval the suppressor, with a fixed level of $L_s = 65$ dB and a ratio between the frequencies of the suppressor and the probe $f_s/f_p = 0.88$. The third interval contained the stimulus with a level increment of 6 dB by sending the signals to both channels.

In order to reduce the influence of the small differences between the transducers, a compensation method was included. The transfer function of both channels (H_1

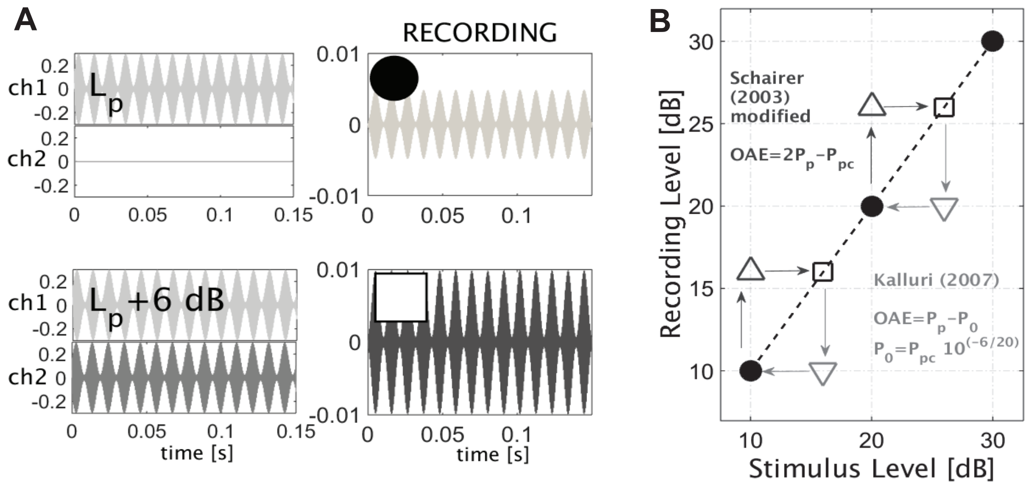


Fig. 1: Compression paradigm to extract SAMOAEs (adapted from Schairer *et al.*, 2003). SAMOAE recording consisted of steady state measurements: P_p(circle) by 1 transducer and P_{pc}(square). A) SAMOAE stimuli and and recording signals. In the first long interval, P_p was played to channel one. The second interval contained the same signal in both channels P_{pc}. In the recoding, the stimulus level L_p had a difference of 6 dB between the two interval due to acoustic constructive interference. B) Measurements were performed at L_p and $L_p + 6$ dB in 10-dB steps. Then by using either up or down scaling, the complex difference between the two measurements provides the SAMOAE. Two recorded intervals are needed for each SAMOAE point.

and H_2) and the transfer function between both channels (H_{12}) were recorded in the coupler by using random white noise at each of the studied conditions. An algorithm for correcting the frequency and phase differences between the two channels was implemented as follows:

$$\mathrm{OAE} = \mathrm{P_p} + \mathrm{P_p H_{12}} - \mathrm{P_{pc}}, \qquad \text{(Eq. 1)}$$

where P_p is the ear canal sound pressure measured at the probe and P_{pc} is the ear canal sound pressure in the compression interval recording.

Analysis

Measurements were divided in 1-second epochs. The first and last epoch were discarded. Epochs were also treated by a custom artifact rejection algorithm that removed the epochs with clear artifacts. After time averaging, the OAEs were extracted by using the suppression and compression method. The level of the three spectral components (carrier and sidebands) in dB SPL was obtained from the frequency domain signal with a resolution of 1 bin/Hz.

RESULTS

Average results of the 5 subjects are shown in Fig. 2. The data are clearly separated from the noise level for all conditions. At levels above 50 dB, OAE levels were similar to the coupler residuals.

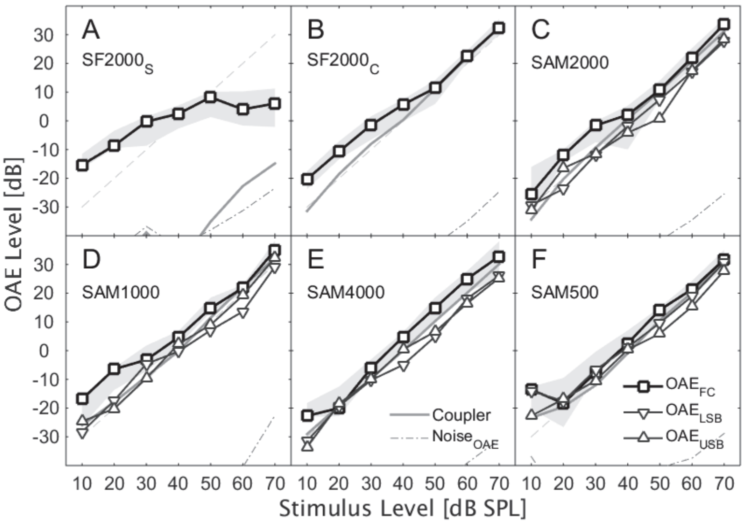


Fig. 2: Median results. SFOAEs measured by suppression (A) and compression (B). Panels C-F show SAMOAEs for centre frequencies of 2005 Hz, 1002 Hz, 4011 Hz, and 498 Hz. The thick grey line indicates coupler residual. The coupler residual was found to grow linearly in all compression conditions. The shadowed area shows the inter-quartile interval.

Figure 3 shows the results of subject APJ after applying the channel difference compensation. Coupler residuals appear closer to the noise than to the measured OAE. However, a considerable influence of the transducer distortion remains above 55 dB SPL.

DISCUSSION

At levels below 50 dB, the measured OAEs could be clearly separated from the transducer once the channel difference is compensated. This indicates, that OAEs evoked by SAM tones can be extracted using paradigms developed for SFOAEs. At levels above 50 dB, the transducer distortion seemed to dominate the OAEs. The distortion was likely due to either the use of acoustic summation in the compression interval and small differences between the transducers or the intermodulation distortion of each loudspeaker, violating the linearity assumption.

In order to investigate the contribution of the distortions to the OAEs at high stimulus levels, simulations were performed with a non-linear transmission model (Epp *et al.*,

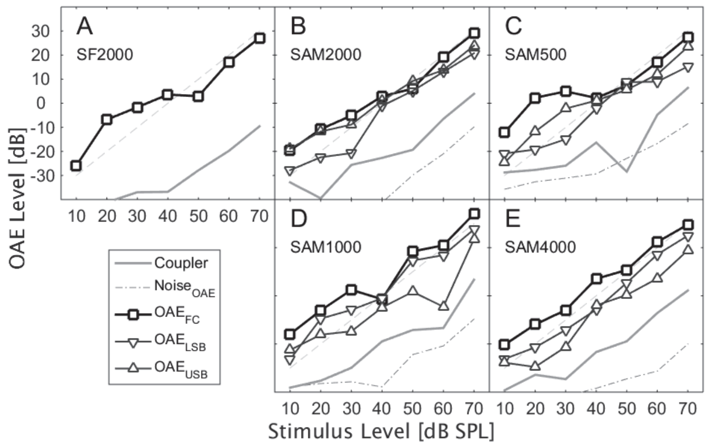


Fig. 3: Results for subject APJ, SFOAEs measured by compression (A). Panels B-E show SAM-OAEs for centre frequencies of 2005 Hz, 498 Hz, 1002 Hz, and 4011 Hz. The method involves a compensation of the difference between transducers (H_{12}).

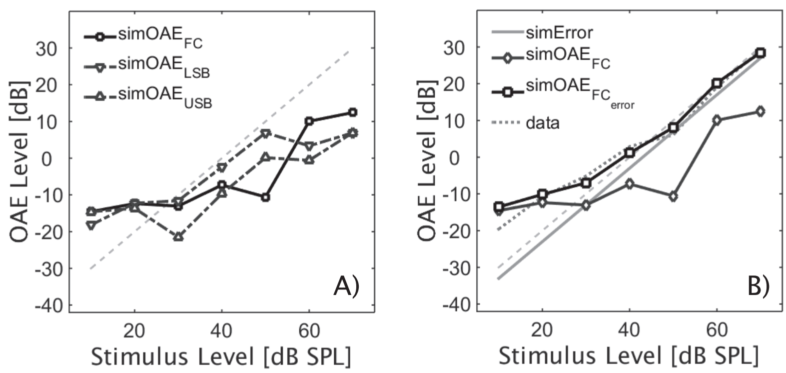


Fig. 4: A) Simulation results for SAMOAE at 2 kHz. B) Simulation results including the coupler residual as an error source.

2010) capable of generating SFOAEs. OAEs were simulated using the same procedure as in the experiment for the condition SAM at 2 kHz. To compare experimental and simulated data, a linearly growing transducer distortion was assumed and added to the simulated OAEs. Simulation results (Fig. 4) suggest that the distortion of the transducer not only affects the results above 55 dB SPL but also leads to an obscure result at lower levels.

One way to reduce the influence of coupler distortion is to make only use of one transducer. In a pilot experiment the compression stimuli was delivered into the ear by the same channel as the probe but a level ($L_{com} = L_p + 10$ dB). As a result, the OAEs were clearly separated from the coupler residuals below 30 dB SPL. At higher levels the nonlinearlity of the transducer still dominated the response (Fig.5A).

Another method to reduce the influence of coupler distortion might be the application of a 3-interval 2-evoked (2E OAE) OAE paradigm used in Schairer *et al.* (2003) where the influence of the transducer's distortion for SFOAEs was only found significant above 60 dB. The 2E OAE method involves two measurements of the P_p, one with each of the transducers, and the P_{pc} by using both transducers at the same time. If the same probe is to be used, the generation the stimuli in the ASSR measurement may be modified in order to involve a sequence of these three measurements (Fig.5B).

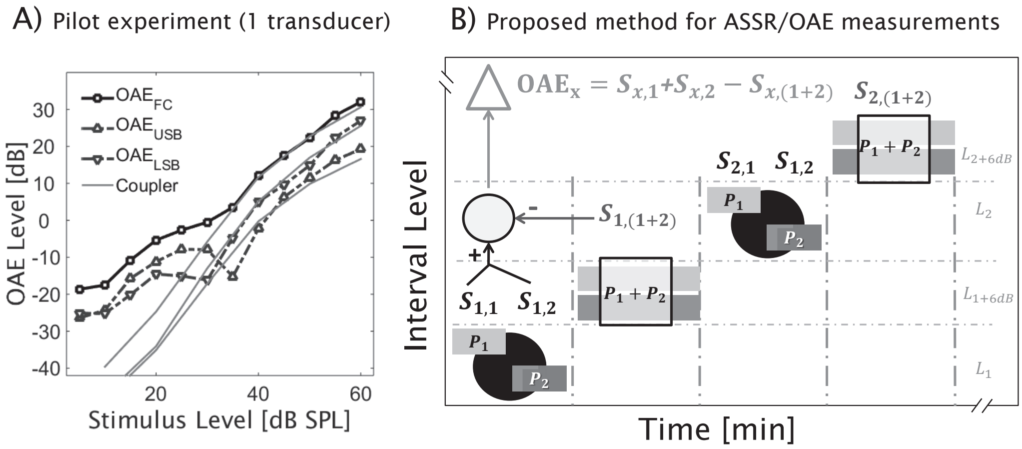


Fig. 5: A) SAMOAE for the 2-kHz condition measured by using only one transducer and $L_{com} = L_p + 10$ dB. B) Proposed solution: the three intervals needed for the OAE measurement by using the 2E OAE (Schairer *et al.*, 2003) are included in the ASSR procedure. P_x denotes the probe and $S_{y,x}$ the recording where *x* is the channel and *y* the measurement point.

CONCLUSION

Extraction of OAE using SAM tones is possible in consecutive steady state intervals. However, due to the transducers' distortion, results were obscured at levels above 50 dB SPL. A proposed alternative method may minimize this problem. If the influence of the transducer distortion on the measured OAEs can be reduced, the simultaneous measurement of ASSR and SAMOAE might provide a more detailed insight into the mechanisms contributing to peripherial compression.

ACKNOWLEDGEMENTS

We thank Gerard Encina Llamas, François Guérit, and Andreu Paredes Gallardo for their helpful suggestions, discussions, and help with the measurement-system.

REFERENCES

Cooper, N.P. and Yates, G.K. (**1994**). "Nonlinear input-output functions derived from the responses of guinea-pig cochlear nerve fibres: variations with characteristic frequency." Hear. Res., 78, 221-234.

Encina Llamas, G., Epp, B., and Dau, T. (**2014**). "Comparison of peripheral compression estimates during auditory steady state responses (ASSR) and distortion-product otoacoustic emissions (DPOAE)," 37th Annual MidWinter Meeting of the Association for Research in Otolaryngology, San Diego, CA, USA.

Epp, B., Verhey, J.L., and Mauermann, M. (**2010**). "Extraction of OAEs during multi-frequency ASSR recordings with the goal to estimate peripheral compression," J. Acoust. Soc. Am., **128**, 1870-1883.

Kalluri, R. and Shera, C. (**2007**). "Comparing stimulus-frequency otoacoustic emissions measured by compression, suppression, and spectral smoothing," J. Acoust. Soc. Am., **122**, 3562-3575.

Nelson, D.A. and Schroder, A.C. (**2004**). "Peripheral compression as a function of stimulus level and frequency region in normal-hearing listeners," J. Acoust. Soc. Am., **115**, 2221-2233.

Plack, C.J. and Oxenham, A.J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am., **103**, 1598-1608.

Rhode, W. and Recio, A. (**2001**). "Basilar-membrane response to multicomponent stimuli in chinchilla," J. Acoust. Soc. Am., **110**, 981-999.

Schairer, K., Fitzpatrick, D., and Keefe, D. (**2003**). "Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears," J. Acoust. Soc. Am., **114**, 944-966.

Contralateral masking for monaural speech intelligibility measurements with hearing aids in free-field speech conditions

GERTJAN DINGEMANSE[1,*] AND BAS FRANCK[2]

[1] *Erasmus University Medical Center, Department of Otorhinolaryngology, Rotterdam, The Netherlands*

[2] *Radboud University Medical Center, Department of Otorhinolaryngology, Nijmegen, The Netherlands*

Patients with asymmetrical hearing loss or unilateral hearing loss often suffer from bad hearing at the poor side, from localisation problems, and from poor speech understanding in noise. In many cases speech audiometry in free field can be an effective tool to decide whether speech understanding is equivalent for both aided ears, making binaural interaction possible, but only if the speech intelligibility is measured for each ear separately. However, it is difficult to evaluate the effectiveness of each (aided) ear individually. This is due to the fact that sound generated in free field can reach both ears, i.e., also the non-test ear. The sound can reach the non-test ear in three ways: directly from the loudspeaker, indirectly by transcranial transmission via the test ear (cross-hearing), or via the skull. In many clinics the non-test ear is "masked" by a foam plug and/or earmuffs. This method helps to minimise the effect of hearing direct sound at the non-test ear. However, transcranial transmission cannot be ruled out by this method. We suggest a new method of contralateral masking, while stimulating in free field. Theoretical considerations are outlined to determine the masking levels necessary to mask sufficiently, and to avoid too much masking (over-masking). For most asymmetric hearing losses a simple rule can be used.

INTRODUCTION

We often use speech audiometry in free field condition to measure the effect of amplification of a hearing aid on speech intelligibility. Speech is presented by a loudspeaker, received and amplified by the hearing aid, and presented to the ear. This way we investigate speech intelligibility with hearing aids.

In many cases it is useful to measure the contribution of each hearing aid separately. When wearing two hearing aids we like to know whether both hearing aid settings produce equal speech intelligibility. By measuring the effect separately for each hearing aid and for both hearing aids together, it is possible to compare the hearing aid settings and measure the binaural advantage. Asymmetry between hearing loss in

*Corresponding author: g.dingemanse@erasmusmc.nl

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

both ears is common. Also in that case it is good to know whether comparable intelligibility can be realised for both ears.

However, measuring aided speech intelligibility for each ear is not simple, because the non-test ear can also contribute to the result. There are three ways in which the speech from the loudspeaker can reach the non-test ear:

1. Speech from the loudspeaker reaches the non-test ear directly by air.
2. Speech from the loudspeaker is amplified by the hearing aid in the test ear and is transcranially transmitted to the cochlea of the non-test ear.
3. Speech from the loudspeaker is conducted by the skull to the non-test ear.

The main question is how to make sure that speech coming from the loudspeaker is not heard by the non-test ear.

In the following parts we will discuss what happens when the test ear is not plugged or masked. After that we will discuss how to avoid hearing in the non-test ear. Then we will discuss problems with masking, i.e., over-masking. Finally, the clinical relevance and rules will be discussed.

MEASURING MONAURALLY WITHOUT PLUGGING THE NON-TEST EAR

As mentioned in the introduction, speech from the loudspeaker reaches also the non-test ear. Whether speech is intelligible also in the non-test ear depends on the size of the hearing loss in the non-test ear.

Using the speech audiogram (Fig. 1) we can observe when speech is intelligible in the non-test ear at normal speech levels. From the example of Fig. 1 we observe from the middle panel that speech at a level of 65 dB SPL is partly intelligible for the left ear. It must be kept in mind that the speech audiogram in Fig. 1 applies to monaurally presented speech by TDH-39 headphones. For speech presented by a loudspeaker in a free-field condition the normal curve will approximately be 3 dB better (ANSI S3.6). However for a level of 3 dB above the speech threshold the speech score is still very low. Therefore, we do not consider this shift. So, from Fig. 1 we observe that measurement of speech intelligibility with a hearing aid on the left is possible at 65 dB SPL because there is no speech intelligibility in the right ear at this speech level. On the other hand, the effect of a hearing aid on the right side on speech intelligibility in free field cannot be reliable measured, because the free-field speech is also intelligible in the left ear. Blocking or masking of the left ear is needed for measurement of the effect of the hearing aid on the right side.

ATTENUATION USING AN EAR PLUG (OR EAR MUFF)

If the non-test ear is plugged with an earplug that attenuates with an amount of D dB, then the speech level in the non-test ear is S−D dB. The widely used foam plugs have an attenuation of at least 10 dB in the lower speech frequencies and even more for higher frequencies. This is the assumed protection value, APV, which is calculated (per frequency) as the mean attenuation minus the standard deviation (3M

Occupational Health Group, 2009; 2010; Berger, 1984). When covering the ear by an extra earmuff, the APV increases to 25-30 dB (Berger, 1984). In the example of Fig. 1, it means that the use of an earplug in the left ear is not sufficient to make free-field speech unintelligible for a speech level of 75 dB SPL in all individuals. But the combination of an ear plug and an ear muff would be a suitable solution.

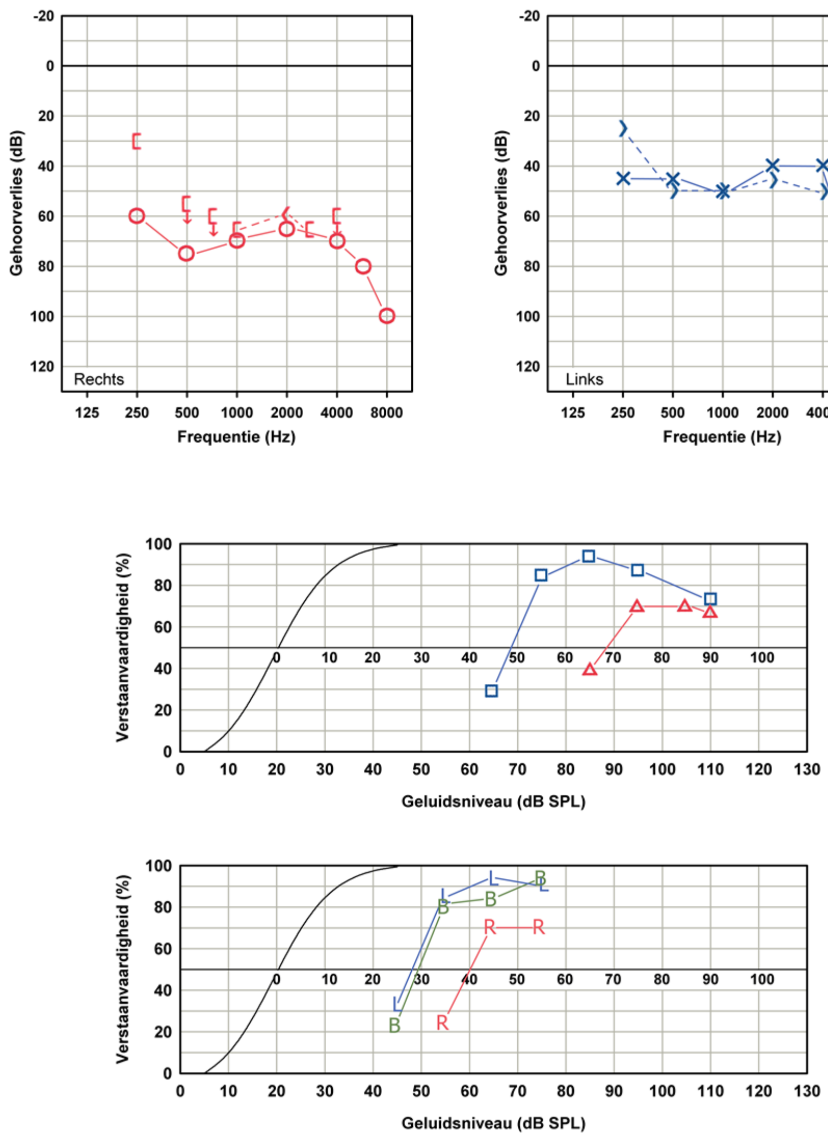


Fig. 1: Upper panel: tone audiogram. Middle panel: Speech audiogram including normal psychometric curve. The monaurally measured phoneme score (% correct) of Dutch CVC words is plotted as a function of speech level in dB SPL. Bottom panel: Free-field speech intelligibility measurements with hearing aids. Masking of the contralateral ear with an insert phone at a level of N = S – 10 was applied for the ‘R’ curve.

In cases with a large conductive hearing loss in the non-test ear the use of ear plug and/or muff is not sufficient, because the skull conducts free-field sound to the cochlea. The minimum attenuation of this signal path is 45 dB, according to Zwislocki (1957) and Berger *et al.* (2003). Thus, the speech may be above the nearly normal bone conduction thresholds in the non-test ear.

THE USE OF MASKING NOISE

From the foregoing it appears that speech in the range of 50-80 dB SPL might be intelligible in the non-test ear, despite using a plug or muff. Therefore masking is required in the non-test ear in such cases.

For presentation of masking noise, insert phones like Etymotics ER-3A and 5A are very suitable. These insert phones have foam plugs that have relatively high attenuation. This means that less masking noise is needed. According to the data sheet of insert phone ER-3A an attenuation of at least 20 dB is possible in practically all ears at any frequency in the range 125-4000 Hz. But the effective attenuation depends on insertion depth (Clark and Rosser, 1988). For more shallow insertion the attenuation is less than 20 dB. To minimize the risk of insufficient damping an attenuation value of 10 dB will be used and a careful insertion is highly recommended.

The speech level in the non-test ear is S–D when using insert phones.

Therefore, speech will just be masked in the non-test ear when the noise level N is:

$$N = S - D \text{ or } N = S - 10 \qquad \text{(Eq. 1)}$$

Notice that we assume that the speech noise used for masking is calibrated such that, for a speech-to-noise ratio of 0 dB on the audiometer dials, speech is not intelligible.

For ear problems with increased ear volume (radical cavity, tympanic membrane tube, and tympanic membrane perforation) the actual noise level of the insert phone may be less (Voss *et al.*, 2000). For these cases, we choose N = S.

AVOIDING HEARING IN THE NON-TEST EAR BY CROSS-HEARING

As already mentioned in the introduction, there is a second way how speech from the loudspeaker can reach the non-test ear. Speech is amplified by the hearing aid in the test ear and is transcranially transmitted to the cochlea of the non-test ear. The question now is: Is the masking noise level N = S – D in the non-test ear large enough to prevent hearing the transcranially transmitted sound?

Speech with level S is amplified by the hearing aid with a certain gain. The amplified speech is attenuated with an intra-aural attenuation of the amplified sound (IA_{HA}, Interaural Attenuation hearing aid) before it reaches the non-test ear. IA_{HA} depends on the tightness of the fit of the earmould and is also frequency-dependent. For tightly fitted ear moulds and deeply placed in-the-canal hearing aids IA_{HA} can be as low as 45 dB, but typically the interaural attenuation is at least 60 dB (Fagelson *et al.*, 2003; Valente *et al.*, 1995; Gudmundsen, 1997; Munro and Contractor, 2010).

The speech level in the non-test ear due to transcranial transmission is:

$$S + \text{Gain} - IA_{HA} \qquad \text{(Eq. 2)}$$

Speech at this level is only intelligible, at least in part, if it is above the bone conduction levels of the non-test ear.

From this formula, it is clear that the risk that speech might be intelligible in the non-test ear is greatest if the gain of the hearing aid is large and the bone conduction thresholds in the non-test ear are low.

The amount of noise that is needed for proper masking of this transcranially transmitted speech is:

$$N = S + \text{Gain} - IA_{HA} + \text{ABG\_non-test ear} \qquad \text{(Eq.3)}$$

where ABG_non-test ear is the air-bone gap in the non-test ear. A volume correction might be applied for increased ear volumes (0 or 10 dB).

Now we have two masking rules: Eq. 1 for calculation of the masking level that prevents speech intelligibility in the non-test ear when speech reaches the non-test ear directly, and Eq. 3 that is used to calculate the masking level necessary to prevent speech intelligibility in the non-test ear that reaches the ear transcranially.

The highest masking level from the two calculations has to be used.

Equation 3 prescribes more masking noise than Eq. 1 if:

$$\text{Gain} + \text{ABG\_non-test ear} > IA_{HA} - D \qquad \text{(Eq. 4)}$$

With IA_{HA} = 60 dB and D = 10 dB:

$$\text{Gain} + \text{ABG\_non-test ear} > 50 \qquad \text{(Eq. 5)}$$

Thus, in most cases the masking level from Eq. 1 is sufficient. Only in special cases like the application of a power hearing aid or a large air-bone gap in the non-test ear, the noise level needs to be calculated from Eq. 3.

IS OVER-MASKING POSSIBLE?

It is important to verify that masking noise in the non-test ear can be overheard in the test ear. This will mask the speech in the test ear. The risk of over-masking is most prominent when an air-bone gap is present in the test ear. This is due to the relatively favourable bone conduction thresholds in the test ear.

In order to have a near 100% speech score the masking noise in RMS value should be 10 dB lower (Fig. 2). Thus, to avoid over-masking there should be a signal-to-noise ratio (SNR) of 10 dB at the cochlea of the test ear. This corresponds to an SNR on the audiometer dial settings of 30 dB, due to calibration of the noise as a fully masking noise at dial settings that are equal for speech and noise. These values were derived from the psychometric curve of speech in noise with Dutch CVC words (Fig. 2). Other speech materials possibly need other values.

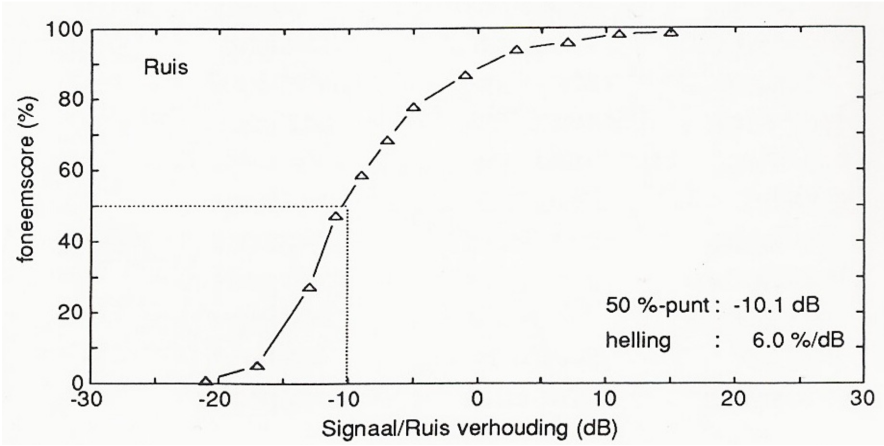


Fig. 2: Normal curve for Dutch CVC words in speech-shaped noise. The phoneme score (% correct) is plotted as a function of the signal to noise ratio (SNR in dB). For an SNR of −10 dB a percentage of 50% of the words are reported correctly.

For ears with a sensorineural hearing loss the SNR must be even better than for normal hearing. We assume that there will be over-masking when:

$$S - N \leq 40 \quad \text{(Eq. 6)}$$

In the test ear the effective speech level is:

$$S + \text{Gain} - \text{ABG\_test ear} \quad \text{(Eq. 7)}$$

The effective noise level is equal to the noise level in the non-test ear minus the interaural attenuation of the insert phone (IA_{IP}). This inter-aural attenuation is at least 55 dB (Munro and Contractor, 2010; Munro and Agnew, 1999; Sklare and Denenberg, 1987).

If Eq. 1 is used to determine masking noise level N, then there is a risk of over-masking when:

$$(S + \text{Gain} - \text{ABG\_test ear}) - (S - 10 + \text{volume correction} - IA_{IP}) \leq 40 \quad \text{(Eq. 8)}$$

Hence, there is a risk of masking when:

$$\text{ABG\_test ear} - \text{Gain} + \text{volume correction} \geq 25 \quad \text{(Eq. 9)}$$

A hearing aid will compensate (largely) for the air-bone gap in the test ear, so with hearing aids, over-masking is very unlikely.

If Eq. 3 is used to determine the masking noise level N, then there is a risk on over-masking when:

$$(S + \text{Gain} - \text{ABG\_test ear}) - (S + \text{Gain} - IA_{HA} + \text{ABG\_non-test ear} + \text{volume correction} - IA_{IP}) \leq 40 \quad \text{(Eq. 10)}$$

Test ear	Non-test ear	Masking
Sensorineural	Normal	$N = S - 10$ Under-masking and over-masking not possible
Sensorineural	Sensorineural	$N = S - 10$ Under-masking and over-masking not possible
Sensorineural	Conductive or Mixed	$N = S - 10$ + volume correction (0 or 10 dB) Risk of under-masking when: Gain + ABG_non-test ear ≥ 50 dB. Then use: N = S + Gain – 60 + ABG_non-test ear + volume correction (0 or 10 dB). Over-masking is very unlikely
Conductive or Mixed	Normal or Sensorineural	$N = S - 10$ Risk of under-masking when: Gain ≥ 50 dB. Then use: N = S + Gain – 60 + volume correction (0 or 10 dB). Over-masking is very unlikely
Conductive or Mixed	Conductive or Mixed	$N = S - 10$ + volume correction (0 or 10 dB) Risk of under-masking when: Gain + ABG_non-test ear ≥ 50 dB. Then use: N = S + Gain – 60 + ABG_non-test ear + volume correction (0 or 10 dB). Risk of over-masking with this second rule when: ABG_test ear + ABG_non-test ear + volume correction ≥ 75 dB Transcranial hearing cannot be avoided in cases with a large air-bone gap in both ears!

Table 1: Masking rules for all possible combinations of hearing losses for the case that a hearing aid is on the test ear and an insert phone is in the non-test ear for masking. Also rules are given that indicate when there is a risk of under-masking and over-masking. The alternative rules are given that apply in these cases.

Hence:

$$\text{ABG\_test ear} + \text{ABG\_non-test ear} + \text{volume correction} \geq 75 \qquad \text{(Eq. 11)}$$

This shows that if there is an air-bone gap for only one ear, over-masking is very unlikely. Only when a large air-bone gap is present for both ears, over-masking is possible.

CLINICAL IMPLICATIONS

For clinical practice it is important to use a simple rule that is applicable for most cases. Therefore, the measuring procedure is simplified: always use masking with insert phones, even though this is not strictly necessary in all cases.

The basic masking rule is:

$$N = S - 10 + \text{volume correction (0 or 10 dB)} \qquad \text{(Eq. 12)}$$

For the attenuation of the insert foam tip a conservative value of 10 dB is chosen, to guarantee that the rule is valid for different fitting in various ears.

In the case of conductive or mixed hearing losses, in certain cases there is a risk of under-masking or over-masking. This should be taken into account when interpreting the measurements. Table 1 gives masking rules for all combinations of hearing losses. In this table also rules are given when there is risk of under-masking or over-masking. When so, alternative rules can be found. Again, these rules are based on conservative estimations of interaural attenuation values from the literature in order to make the rules safe for various ears and patients.

REFERENCES

3M Occupational Health Group. (**2009**). *Hearing Protection Catalog 2009*.

3M Occupational Health Group. (**2010**). *Technical Datasheet E-A-R™ Classic™ Roll-Down Earplugs*, www.3m.com.

ANSI S3.6. (**2004**). *American National Standard Specifications for Audiometers.*

Berger, E.H. (**1984**). "Attenuation of earplugs worn in combination with earmuffs," Occup. Health Saf., **16**, 72-73.

Berger EH, Kieper RW, and Gauger D. (**2003**). "Hearing protection: Surpassing the limits to attenuation imposed by the bone-conduction pathways," J. Acoust. Soc. Am., **114**, 1955-1967.

Clark, J.L. and Rosser, R.J. (**1988**). "Three studies comparing performance of the ER-3A tubephone with the TDH-50P earphone," Ear Hearing, **9**, 268-274.

ER-3A Product Sheet. Etymotic Research, Inc., www.etymotic.com

Fagelson, M.A., Noe, C., Murnane, O., and Blevins, J. (**2003**). "Predicted gain and functional gain with transcranial routing of signal completely-in-the-canal hearing aids," Am. J. Audiol., **12**, 137-145.

Gudmundsen, G.I. (**1997**). "Physical options," In: Tobin, H. (Ed.). *Practical Hearing Aid Selection and Fitting*, Washington, D.C.: Department of Veterans Affairs, pp. 1-16.

Munro, K. and Agnew, N. (**1999**). "A comparison of inter-aural attenuation with the Etymotic ER-3A insert earphone and the Telephonics TDH-39 supra-aural earphone," Br. J. Audiol., **33**, 259-262.

Munro, K.J. and Contractor, A. (**2010**). "Interaural attenuation with insert earphones," Int. J. Audiol., **49**, 799-801.

Sklare, D.A. and Denenberg, L.J. (**1987**). "Interaural attenuation for Tube Phone (R) insert earphones," Ear Hearing, **8**, 298-300.

Valente, M., Potts L.G., Valente, M., and Goebel, J. (**1995**). "Wireless CROS versus transcranial CROS for unilateral hearing loss," Am. J. Audiol., **4**, 52-59.

Voss, S.E., Rosowski. J.J., Merchant, S.N., Thornton, A.R., Shera, C.A., and Peake, W.T. (**2000**). "Middle ear pathology can affect the ear-canal sound pressure generated by audiologic earphones," Ear Hearing, **21**, 265-274.

Zwislocki J. (**1957**). "In search of the bone-conduction threshold in a free sound field," J. Acoust. Soc. Am., **29**, 795-804.

Audio-visual speech stimuli for the study of lip-reading and multi-sensory integration abilities in hearing-impaired individuals

MAREN STROPAHL[1,*] AND STEFAN DEBENER[1,2]

[1] *Department of Psychology, Carl von Ossietzky University, Oldenburg, Germany*

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

Audio-visual integration of speech is frequently investigated with the McGurk effect. Incongruent presentation of auditory and visual syllables may result in the perception of a third syllable, reflecting fusion of visual and auditory information. However, perception of the McGurk effect depends strongly on the stimulus material used, making comparisons across groups and studies difficult. To overcome this limitation we developed a large set of audio-visual speech material, consisting of eight different speakers (4 females and 4 males) and 12 syllable combinations. The quality of the material was evaluated with 24 young and normal-hearing subjects. The McGurk effect was studied in eight adult cochlear implant (CI) users and compared to 24 normal-hearing individuals using a probabilistic model. The comparison confirmed previous reports of stronger audio-visual integration in CI users. The audio-visual material developed in this study will be made freely available.

INTRODUCTION

In daily life situations the integration of information from multiple senses is necessary to interact with the environment (Driver and Noesselt, 2008). In real-life communication most of the speech signal is encoded by the auditory input. Nevertheless, it has been shown that visual information such as lip movements can improve speech intelligibility especially in noisy situations (Sumby and Pollack, 1954). Audio-visual integration therefore plays a major role for communication and auditory restoration. Cochlear implants (CIs) are biomedical devices that allow individuals with a profound sensorineural hearing loss to regain parts of their hearing ability. Despite the electrical input, CI users are able to show improved speech recognition shortly after implantation (Sandmann *et al.*, 2014). Nevertheless, speech understanding in noisy situations remains difficult for the majority of CI users (Fu *et al.*, 1998). The deficit of CI users in their auditory processing may also be reflected in a different use of visual speech cues compared to normal hearing (NH) controls. There is evidence that CI users are better in lip reading and in integrating audio-visual stimuli (Rouger *et al.*, 2007; Stropahl *et al.*, 2015).

*Corresponding author: maren.stropahl@uni-oldenburg.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

One way to investigate audio-visual integration is the McGurk effect which became a popular tool over the past decades (MacDonald and McGurk, 1978; McGurk and MacDonald, 1976). If individuals are presented with incongruent audio-visual syllables such as an auditory “Ba” and a visual “Ga” they may perceive neither the auditory nor the visual component but a third, different syllable (e.g., “Da”). This perception of a fused auditory and visual component is called the McGurk effect. Behavioral studies with CI users showed a bias towards the visual component of incongruent audio-visual McGurk stimuli and an altered audio-visual integration (Rouger *et al.*, 2008; Tremblay *et al.*, 2009). For by far most studies investigating the McGurk effect, research groups recorded their own stimulus material which comprises typically only one male or female speaker and very few syllables (MacDonald and McGurk, 1978; McGurk and MacDonald, 1976; Rouger *et al.*, 2008; van Wassenhove *et al.*, 2005). Basu Mallick *et al.* (2015) recently reported that the perception of the McGurk illusion strongly depends on the stimuli used. They recorded eight different McGurk stimuli from four female and four male speakers and compared the amount of fusion percepts for a large sample of 165 participants. A high variability of the amount of fusion of individuals was clearly evident across the different stimuli. Furthermore most of the participants (77%) almost always or almost never perceived the illusion, so the distribution deviates from normality (Basu Mallick *et al.*, 2015).

To account for stimulus differences and to correctly identify individual differences the noisy encoding of disparity (NED) model was proposed (Magnotti and Beauchamp, 2015). The NED model classifies each stimulus in its estimated likelihood that the auditory and the visual component evoke the McGurk effect (stimulus disparity). Furthermore the model estimates two individual parameters: the sensory noise of encoding the audio and the visual component and the individual disparity threshold which is the prior probability of an individual to encode the audio-visual incongruent stimulus as a fused percept. The individual disparity threshold is a fixed value along the stimulus disparity. Both individual parameters are assumed to be consistent across stimuli (Magnotti and Beauchamp, 2015). The two individual parameters of the model allow researchers to compare groups in their audio-visual integration independent of the presented stimulus. Using this approach, we developed a large battery of audio-visual stimuli and applied the NED model. This enabled us to investigate audio-visual integration in hearing and hearing-impaired individuals and describe group effects independent from stimulus effects. Specifically, a subgroup of eight adult, experienced CI users was compared to a control group (N = 24).

METHODS

Stimuli

To test the McGurk illusion, a set of audio-visual stimuli was recorded. Eight syllables were selected. The selection was based on the second study of MacDonald and McGurk (1978). The syllables were spoken from eight trained speakers (four females) with education in singing or theater playing, ensuring high professionalism in narrating the material. A Canon HF100 HD (CAM) high definition camera with a resolution of 1920 × 1080 (MPEG4 H.264, 25fps) was used, as well as the 26TK microphone (G.R.A.S.). Audio and video materials were synchronized offline and processed to optimized stimulus quality. The audio-visual videos obtained all start with a still image of the speaker (last frame before movement onset), followed by the spoken syllable, giving a total duration of approx. 2s for each clip. In total twelve combinations of audio-visual incongruent stimuli were used to test the McGurk illusion.

Data acquisition

To evaluate the recorded stimulus set, a control group of 24 NH students (15 females; mean age 26 ± 5.9 years) was tested. The participants did not report any neuropsychological abnormalities, had normal hearing thresholds and normal or corrected-to-normal vision. A second group consisted of eight CI users (four females) that were all post-lingually deafened and had received their implant at least one year before testing. All CI users were unilaterally implanted and seven used an additional hearing aid on the non-implanted ear which was activated during testing. The mean age of the CI group was 47 ± 24.5 years. The CI users showed a variety of hearing loss etiologies. Five CI users had presumable hereditary causes of hearing loss which was in three cases further impelled by loudness damage, two cases might have undergone a probable oxygen loss at birth, and one CI user suffered from a Gusher syndrome. The study was conducted in accordance with the local ethical committee guidelines of the University of Oldenburg and in agreement with the declaration of Helsinki. Participants gave written informed consent before the experiment. Participants were seated in a sound-shielded booth 1.5 m away in front of a screen. Audio signals were presented binaurally in a free-field setting. Three different conditions were tested in randomized order; the participants had to respond in a four-alternative forced-choice to either auditory only or visual only syllables or the percept for incongruent audio-visual syllables. Participants were instructed to select one of the four syllables presented on the screen after each trial. In the audio-visual condition the participants were instructed to indicate what they perceived aurally. Each stimulus was presented five times for each of the eight speakers, giving a total number of 800 trials (120 audio only (A_{only}), 200 visual only (V_{only}), 480 A-V incongruent (McGurk)).

Data analysis

The correct phoneme identification frequency was calculated for each condition and compared between groups. To test group differences, the Mann-Whitney-U-Test (MWU-Test) was applied. This non-parametric test is suitable for not normally distributed data and unequal group sizes. To further analyze the results and to account for group differences, the NED model by Magnotti and Beauchamp (2015) was applied. The probabilistic model allows separating individual and stimulus differences. The NED uses the individual fusion proportion for each stimulus which was defined as neither the auditory component nor the visual component but a percept of a new combination of the auditory and the visual component (originally defined as an illusion by McGurk and MacDonald (1976)). Three parameters are estimated based on the behavioral fusion data: (1) The audio-visual disparity for each stimulus estimating the differences between the auditory and the visual component and therefore the likelihood of the two components to be fused to the McGurk illusion; (2) The individual sensory noise describing the noise while processing the visual and auditory component of the audio-visual stimulus. The sensory noise is assumed to be constant for a person across stimuli; (3) The disparity threshold as the prior probability of each individual to integrate auditory and visual features (resulting in a fusion percept). The individual disparity threshold is independent of the stimulus disparity. The NED model considers stimulus differences and therefore allows comparing multi-sensory integration across individuals and across groups. The model fitting was done in R based on source code provided by Magnotti and Beauchamp (2015).

RESULTS

Correct phoneme identification

The group average result for correct phoneme identification is shown in Fig. 1. For the NH controls, the correct identification in the A_{only} condition was overall very high, with a mean of M_{NH} = 97.1%. NH individuals easily identified the audio stimuli which confirms the good quality of the audio material. The CI users on the other hand showed a significant reduction in correctly identified phonemes (M_{CI} = 68.7%, U = -4.09, p < .001). The V_{only} condition revealed a clearly diminished correct identification rate for both groups. As can be seen in Fig. 1, the groups did not differ in their ability to discriminate the V_{only} phonemes (M_{NH} = 31.3%, M_{CI} = 31.69%, U = -.22, p = .848). When evaluating the results of the AV incongruent (McGurk) condition, the correct answer would be the audio stimulus. A significant group difference could be observed for the McGurk condition. The NH controls correctly identified the audio stimuli despite the incongruent visual input with M_{NH} = 46.48%. In contrast the CI group showed a lower number of correctly identified phonemes (M_{CI} = 6.43 %), U = -3.53, p < .001. To further explore the difference in the McGurk condition, the response types of both groups were split up to determine if the participants either perceived the correct audio stimulus, the visual stimulus or a fused percept (see Fig. 2). The NH controls reported for the incorrect trials in M_{NH} = 9.7% the visual component and reported in M_{NH} = 43.83% of the trials a fused percept. The CI users reported the visual component in M_{CI} = 24.94% of the trials and a fused percept in M_{CI} = 68.62%. The CI users hence showed an overall stronger reliance on the visual component and a higher proportion of fusing the auditory and the visual component. Comparing the amount of fusion between groups and independent from the stimulus material is important. Fused percepts were therefore further analyzed with the NED model.

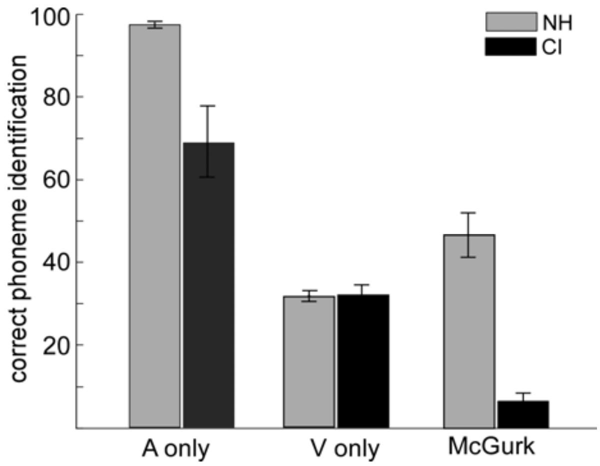


Fig. 1: Correct phoneme identification (with standard error of mean) of the NH controls (grey) and the CI users (black) for the three conditions of audio only (A only), visual only (V only) and the incongruent audio-visual combination (McGurk). CI users showed a significant deficit in understanding the correct phoneme in the A only and the McGurk condition. The visual only condition did not reveal a group difference.

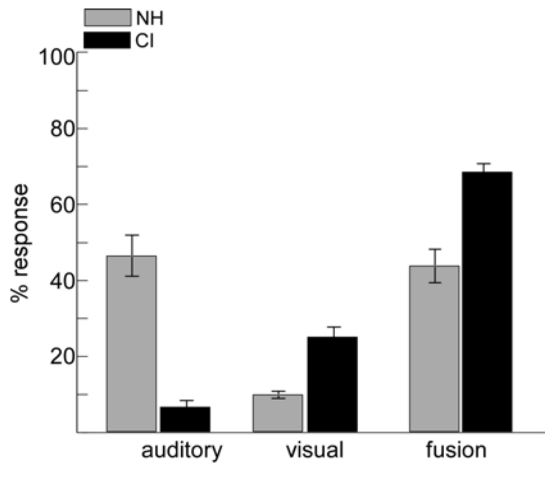


Fig. 2: Response types (with standard error of mean) for the incongruent audio-visual condition (McGurk) separated for the two groups (NH grey, CI black). The correct answer was the auditory component. For wrong answers, NH controls reported more often a fused percept and barely the visual component, whereas CI users were more focused on the visual component and showed a higher amount of fusion percept compared to NH controls.

Group comparison based on the NED-model

The comparison between the NH control group and the CI group was based on the NED model, which accounts for stimulus differences. The estimated parameters are based on the amount of fusion for each individual for each stimulus. The individual parameters, which are stimulus independent, are the sensory noise and the disparity threshold. Both parameters were estimated for each individual and the mean of the groups was compared. The MWU-Test revealed a significant group difference in the sensory noise of encoding the auditory and the visual component (U = -4.18, p < .001) as well as in the individual prior probability to perceive the McGurk illusion (U = -3.57, p < .001). The group difference is shown in Fig 3.

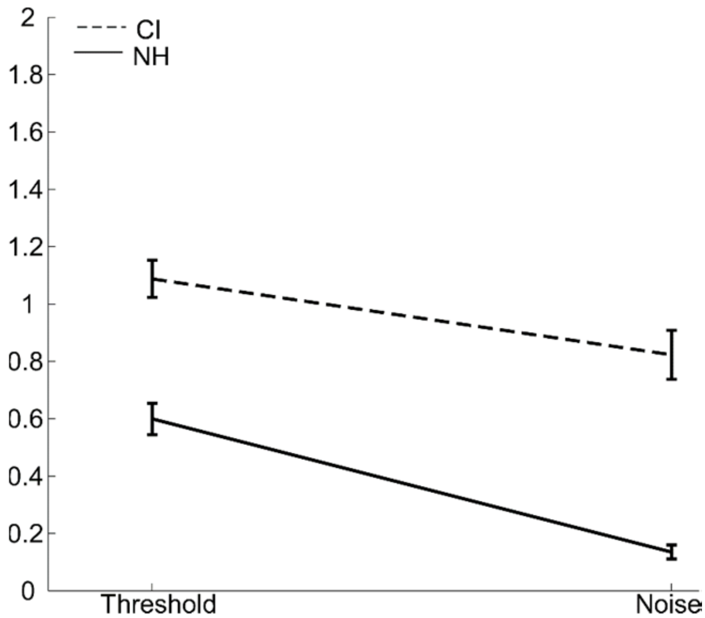


Fig. 3: Group comparison of the NED-Model parameters sensory noise (Noise) and individual disparity threshold (Threshold) plotted with standard error of mean. CI users (dashed lines) showed a significantly higher noise as well as a higher disparity threshold compared to the NH controls (solid line), which reflects differences in audio-visual speech integration.

DISCUSSION

The present study evaluated the McGurk effect tested with newly developed audio-visual stimuli. A group of NH controls and a small subgroup of CI users were compared. CI users showed a deficit in identifying the correct syllable in the A_{only} condition and showed an altered response behavior in the incongruent conditions compared to NH controls. The NED model revealed further group differences in the sensory noise of encoding the auditory and the visual component as well as in the individual probability of perceiving the McGurk illusion which further indicates differences in audio-visual integration between hearing-impaired and hearing individuals. Importantly, these measures aim to reveal a more stimulus-independent characterization of audio-visual integration.

The CI users showed a deficit in the auditory only condition which might be due to the degraded input of the CI (Fishman *et al.*, 1997). Moreover, syllables provide sparse linguistic information compared to meaningful words, hence they may be more problematic to identify for the CI users (Rouger *et al.*, 2008). Interestingly, the visual only condition revealed no group differences between the groups although previous studies suggested superior lip reading abilities even after many years of CI use (Rouger *et al.*, 2007; Stropahl *et al.*, 2015). However, also for the visual only condition the stimuli were meaningless syllables providing only little linguistic information. The better lip reading abilities might therefore result from a strong integration of lexical, semantic, and syntactic information usually provided by the audio-visual stimulus for example in daily-life communication (Rouger *et al.*, 2008). The ability of CI users to identify the correct phoneme (based on the auditory percept) in the audio-visual incongruent conditions was significantly reduced compared to NH controls. By splitting up the responses of the incongruent condition it could be shown that the CI users relied more often on the visual component in the case of ambiguous auditory input, which is in line with other studies (Rouger *et al.*, 2008; Tremblay *et al.*, 2009). In contrast the NH controls relied more often on the auditory component of the incongruent stimulus. The fact that CI users reported more fusion percepts indicates an altered, possibly stronger pattern of audio-visual integration. This interpretation is supported by the NED analysis, which also showed a significant difference in audio-visual integration of the CI users. In a study by Tremblay *et al.* (2009) the CI users did not show an overall higher fusion in the incongruent conditions, whereas descriptively the better CI users showed higher fusion proportions. Nevertheless, the amount of fusion highly depends on the stimulus material used (Basu Mallick *et al.*, 2015) which makes group comparisons within one study and across individuals and studies rather difficult if the amount of fusion is considered without taking into account stimulus effects.

We plan to make the stimulus material freely available in the near future. This will allow others to select McGurk stimuli most appropriate for specific research questions. Furthermore, an extended study investigating audio-visual integration of CI users is under way. Identifying the neural correlates of the stronger McGurk illusion in CI users may help to guide hearing restoration rehabilitation efforts.

REFERENCES

Basu Mallick, D., Magnotti, J.F., and Beauchamp, M.S. (**2015**). “Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type,” Psychon. B. Rev., doi:10.3758/s13423-015-0817-4.

Driver, J. and Noesselt, T. (**2008**). “Multisensory interplay reveals crossmodal influences on “sensory-specific”brain regions, neural responses, and judgments,” Neuron, **57**, 11-23.

Fishman, K.E., Shannon, R.V., and Slattery, W.H. (**1997**). “Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,” J. Speech Lang. Hear. Res., **40**, 1201-1215.

Fu, Q.-J., Shannon, R.V., and Wang, X. (**1998**). “Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,” J. Acoust. Soc. Am., **104**, 3586-3596.

MacDonald, J. and McGurk, H. (**1978**). “Visual influences on speech perception processes,” Percept. Psychophys., **24**, 253-257.

Magnotti, J.F. and Beauchamp, M.S. (**2015**). “The noisy encoding of disparity model of the McGurk effect,” Psychon. B. Rev., 22, 701-709. [Source code: http://openwetware.org/wiki/Beauchamp:NED]

McGurk, H. and MacDonald, J. (**1976**). “Hearing lips and seeing voices,” Nature, **264**, 746-748.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., and Barone, P. (**2007**). “Evidence that cochlear-implanted deaf patients are better multisensory integrators,” Proc. Nat. Acad. Sci. USA, **104**, 7295-7300.

Rouger, J., Fraysse, B., Deguine, O., and Barone, P. (**2008**). “McGurk effects in cochlear-implanted deaf subjects,” Brain Res., **1188**, 87-99.

Sandmann, P., Plotz, K., Hauthal, N., de Vos, M., Schönfeld, R., and Debener, S. (**2014**). “Rapid bilateral improvement in auditory cortex activity in postlingually deafened adults following cochlear implantation,” Clin. Neurophysiol.. **126**, 594-607.

Stropahl, M., Plotz, K., Schönfeld, R., Lenarz, T., Sandmann, P., Yovel, G., De Vos, M., and Debener, S. (**2015**). “Cross-modal reorganization in cochlear implant users: Auditory cortex contributes to visual face processing,” NeuroImage, **121**, 159-170.

Sumby, W. H. and Pollack, I. (**1954**). “Visual contribution to speech intelligibility in noise,” J. Acoust. Soc.Am., **26**, 212-215.

Tremblay, C., Champoux, F., Lepore, F., and Théoret, H. (**2009**). “Audiovisual fusion and cochlear implant proficiency,” Restor. Neurol. Neuros., **28**, 283-291.

van Wassenhove, V., Grant, K.W., and Poeppel, D. (**2005**). “Visual speech speeds up the neural processing of auditory speech,” Proc. Nat. Acad. Sci. USA, **102**, 1181-1186.

A smartphone-based, privacy-aware recording system for the assessment of everyday listening situations

SVEN KISSNER*, INGA HOLUBE, AND JOERG BITZER

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

When trying to quantify hearing difficulties in every-day listening situations, mostly questionnaires are used to collect and evaluate subjective impressions. Obtaining objective data outside a laboratory is relatively difficult, given the required equipment and its proper handling as well as privacy concerns emerging from long-term audio recordings in a non-regulated and populated environment. Therefore, a smartphone-based system was developed that allows long-term ecological momentary assessment. Microphones are placed close to the ears to obtain signal characteristics, e.g., interaural level differences, similar to those perceived by a listener. Currently, root-mean-square, averaged spectra and the zero crossing rate are calculated. Additional features can be implemented and the flexibility of the smartphone itself allows for additional functionality, e.g., subjective ratings on predefined scales. A simple user interface ensures that the system can be properly handled by non-tech-savvy users. As only the extracted features but not the audio-data itself are stored, screening and approval of the recorded data by the test subject is not necessary. Furthermore, additional standard features, e.g., the spectral centroid, can be computed offline, utilizing the recorded features.

INTRODUCTION

Capturing an acoustical environment regarding its physical characteristics as well as how it is perceived by a subject can be a valuable tool. It allows for improving existing hearing systems, the refinement of fitting procedures and their evaluation, as well as studies on individual experiences and behavior in given situations. In practice, however, situations are often assessed using questionnaires retroactively or under laboratory conditions. Without objective measurements, it is difficult to establish a proper relation between a situation and its perception. Delayed feedback can lead to biased and vague results and a controlled environment does not necessarily reflect real-life conditions or evoke similar reactions. To circumvent those issues, data has to be captured directly in the respective situations (Ecological Momentary Assessment; Shiffman *et al.*, 2008). If a study does aim to capture objective data in-situ, the handling of more or less user-friendly technical equipment can frustrate subjects. To ensure privacy, screening of recorded data and/or a declaration of consent from all parties involved is required, the former being time consuming while the latter often is impractical in public spaces.

*Corresponding author: sven.kissner@jade-hs.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

We developed a system that overcomes the problems described above to a certain extent. The primary goal was a mobile recording system that is easy to use, even for non-tech-savvy users. It should be relatively inexpensive to build multiple devices and make them easily replaceable if needed, not relying on specific components. Programming should be flexible and functionality easily extendable if required. Last but not least, the system is engineered with a subject's privacy in mind.

HOW IT WORKS

The main goal is a system which is able to retrieve certain acoustic parameters from everyday listening situations without storing the audio data itself. Following is a description of the hard- and software (also see Fig. 1) as well as an overview of the chosen features.

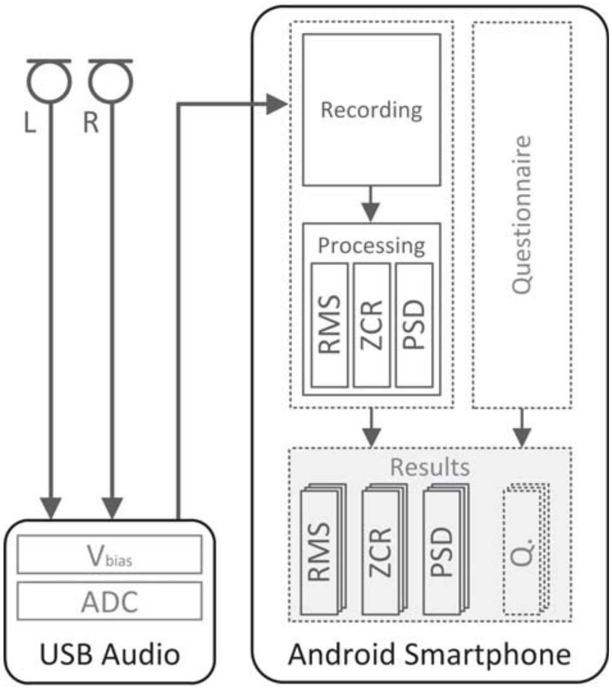


Fig. 1: Schematic of the recording system outlining its hardware components as well as the basic operational sequence.

Hardware

The system consists of three main hardware components. An Android-based smartphone, a USB audio interface, and hearing aid dummies. The Android platform was chosen due to its openness and flexibility. There is a multitude of smartphones and tablets available, fitting almost every conceivable requirement. Unfortunately, most Android-based consumer hand-held devices do not support stereo-input using the standard 3.5-mm audio jack. To circumvent this, we decided to use an external

USB audio interface. This limits the number of eligible devices as the Android-device in question has to support USB-OTG, i.e., can act as a host to USB devices. On the other end, the USB audio device must be class compliant and therefore not require proprietary drivers. Additionally, the interface has to support stereo input, while many audio interfaces with the required small form factor only offer one input channel.

We selected the Moto G by Motorola, an affordable, mid-range smartphone as well as thumb-sized USB audio interfaces of type USB-MA by Andrea Communications.[1] The interface was fitted with an micro-USB port to allow for a direct connection to the smartphone without the need for an additional adapter. The behind-the-ears hearing aid dummies each house a microphone of type EK-23024 by Knowles Electronics. The audio interface supplies a bias voltage of 2.2 V, so no additional power source is required. The dummies are connected to the audio interface using a 3.5-mm stereo audio jack.

Software

The software used for data acquisition and processing was developed in Java using the Android SDK. Currently there is no sufficiently sensible and robust way to access external USB audio interfaces using the API defined by Google. To circumvent the time-consuming implementation of a driver, a third-party product based on libusb was purchased (Dr-Jordan-Design).

Upon opening the Android app, the user is presented with a simple and clean user interface featuring a large button to start or stop data analysis. The button as well as text indicate the current status. The core of the app is a background service. The user interface connects to an existing service or starts a new instance if none is running. Due to the way Android handles lifecycles of an app or activity, a service enables processes to run continuously, even if another app (i.e., camera or questionnaire) becomes active. The service manages audio acquisition as well as processing which both run parallel in their respective threads. Raw audio data is recorded continuously, with a sampling rate of 16 kHz and cached in chunks of 60 s. Each completed chunk is reported to the service which in turn starts a new thread processing the cached data, i.e., sequentially calculating the implemented features and writing the data to the device's storage. Each processed chuck is again reported to the service which deletes the cached audio chunk and, if available, starts processing the next. This is repeated until analysis is stopped and the cached audio data is processed and discarded.

Due to the influence of low-frequency noise the signal is high-pass filtered ($f_0 = 100$ Hz, 2nd-order Butterworth) before processing.

Features of interest

For the selection of acoustical parameters we focused on Kates (2008), who discusses various features for acoustic classification in hearing aids. We also considered *computational complexity* of a certain feature. The system should not back up on cached data, i.e., processing of a single chunk should be finished before the next is completely cached. This also affects power consumption and therefore the maximum duration of continuous recording sessions before the device has to be recharged. As the calculated features are stored on the device, the available *storage space* must be taken into acount as well. A test subject should be able to use the system autonomously for a given amount of time, i.e., four days, eight hours each day, without the need need to daily retrieve the data to free up space. The phone used in the current system provides about 5.5 GB of usable flash-storage. The features currently implemented generate about 130 MB of data per hour, allowing for roughly 40 hours of data. This, of course, can be mitigated by current smartphones offering more internal memory or support for external memory. Another aspect in feature selection is the ability to *derive additional features offline* to save both processing power and storage space.

Therefore, three features being calculated on the recording device. The broadband power (root-mean-square; RMS), the zero crossing rate of the signal and its derivative (ZCR / ΔZCR; Kates, 2008), as well as spectral information in the form of power spectral density (PSD) of left and right channels as well as their cross power spectral density (CPSD; Welch, 1967). Table 1 shows the parameters used for feature extraction.

	RMS	ZCR/ΔZCR	PSD/CPSD
Blocksize in ms	25	25	25/125*
Overlap in ms	12.5	12.5	12.5
nFFT in samples	•	•	512

Table 1: Processing parameters for the selected features. (* the smoothed block is equivalent to 125 ms, see Sec. “Privacy”)

PERFORMANCE

To determine the performance of the system, we measured transfer function, noise level, and total harmonic distortions in an anechoic chamber. To eliminate the speaker’s transfer function (NTi TalkBox and Fostex 6301B) as well as to obtain reference measurements in silence, a G.R.A.S 40AF free-field microphone was used. The results are shown in Fig. 2. Noise floor and frequency response are smoothed in equivalent-rectangular-bandwidth for display.

The frequency responses are reasonably flat within 1 dB up to 3 kHz, rising slightly beyond. The noise of the various sensors also behaves similarly, up to 45 dB SPL at low frequencies (unfiltered), falling below 35 dB SPL beyond 1 kHz. Additionally, the noise is shown after high-pass filtering ($f_0 = 100$ Hz, 2nd order Butterworth) as applied before feature extraction, as well as with A-weighting.

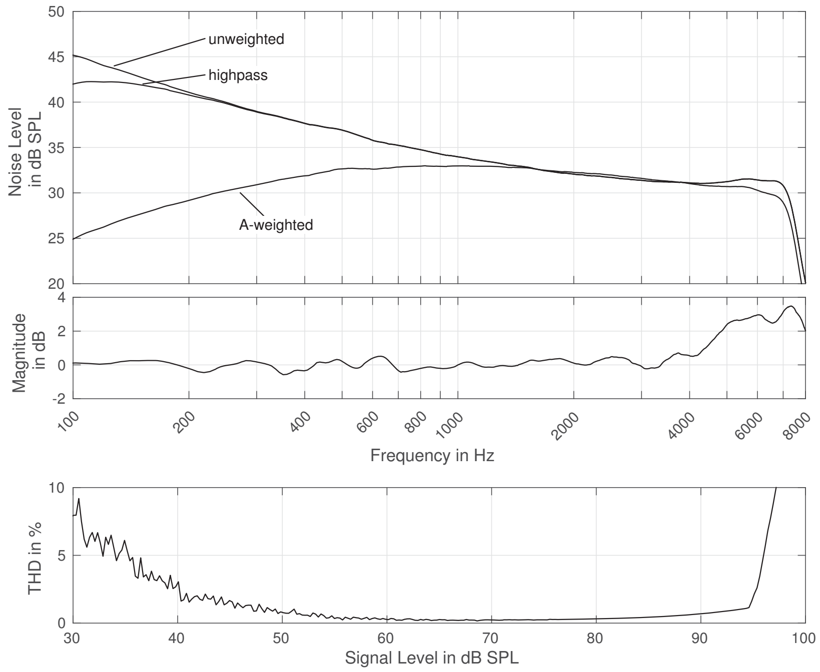


Fig. 2: *Top:* Transfer function. *Middle:* Noise level as a function of frequency for the unweighted, highpass filtered, and A-weighted noise floor, mean over all microphones. *Bottom:* Total harmonic distortion as a function of sound pressure level for a 1-kHz sine. Mean over all measurement systems (10 microhones)

The total harmonic distortions, measured using a 1-kHz amplitude swept sine and calculated from the first four harmonics relative to the fundamental frequency, are dominated by the noise for low frequencies. Beyond a signal level of 40 dB SPL, the signal emerges from the noise until clipping starts abruptly at around 95 dB SPL. The 40AF shows a similar behavior for low levels while it exhibits an extended dynamic with a shallow increase of THDs towards 100 dB SPL. Considering noise levels and THD, the system offers a usable dynamic range of 45 to 55 dB.

Figure 3 shows the selected features, RMS, ZCR and ΔZCR, as well as PSD, as calculated by the system for different situations. Depicted is only one channel. Seconds 0 to 20 show the results for an office with normal background noise, running computers, typing, etc. Seconds 20 to 40 show the results for the same office with two people conversing. Seconds 40 to 60 show the results walking besides a road with steady traffic. The PSD shows increased levels for low frequencies in the quiet office, corresponding with low-frequency microphone noise. Specific events are clearly visible and correspond well over all features, be it keystrokes and speech in the office or loud vehicles passing and a steady background noise for the traffic situation.

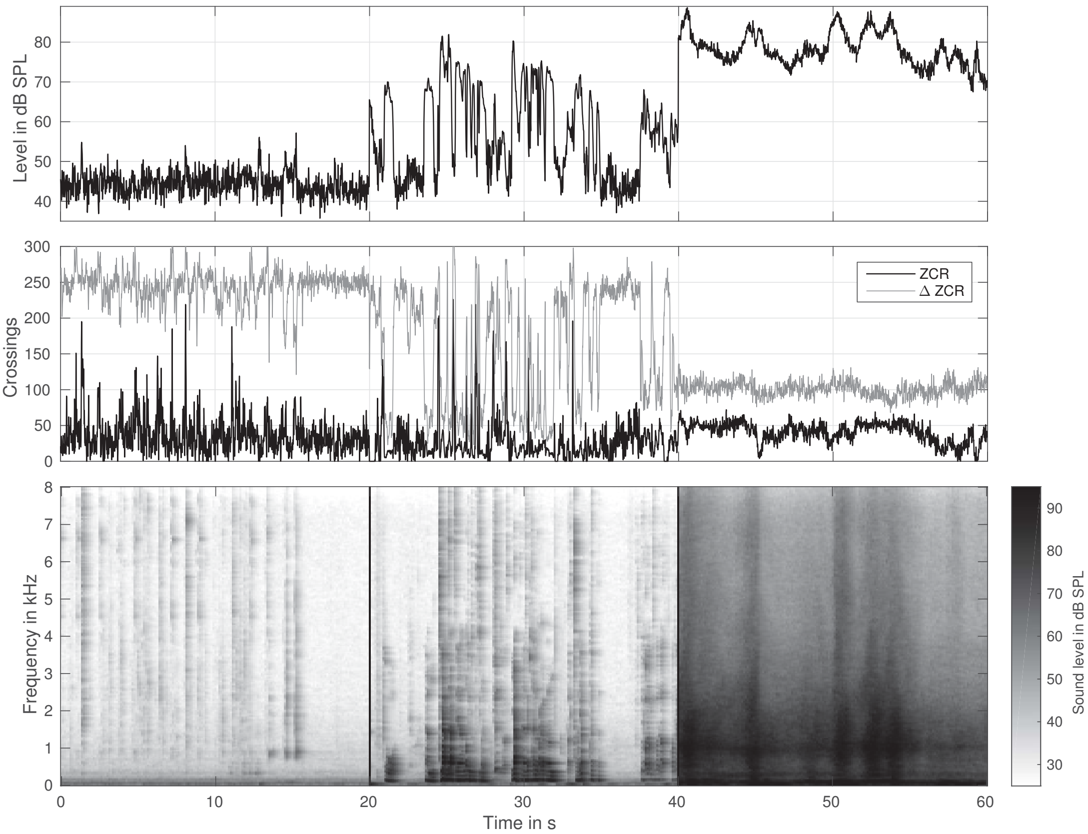


Fig. 3: Examplary RMS (top), ZCR and ΔZCR (middle), as well as PSD (bottom) for one channel as calculated from different situations. *Seconds 0-20:* The author's office with normal background noise, typing, etc. *Seconds 20-40:* A conversation between the author and a collegaue in the same office. *Seconds 40-60:* Walking besides a road with steady traffic.

PRIVACY

While we do not store audio data with respect to a subject's and third-parties' privacy, taking a closer look at the extracted features shows that while broadband RMS and zero crossing rate contain no privacy sensitive information, PSDs are more revealing. With little effort, we are able to reconstruct the audio signal from the stored PSDs to an extent where speech is intelligible and semantic information laid open. To circumvent this, we decided to apply additional smoothing to the PSDs to the point where reconstruction yields no sensitive information.

To determine the required time constant, a listening test was conducted using the Göttingen sentence test (GÖSA; Kollmeier and Wesselkamp, 1997). The speech material was presented to each listener via headphones (HDA200), driven by an amplifier (HB7, Tucker Davis Technologies) and a additional stereo headphone amplifier (MicoAmp HA400, Behringer) adjustable by the test subject. The test was controlled using the Oldenburg Measurement Application (OMA, Hörtech gGmbH). All speech material was presented at a base level of 70 dB SPL without additional backgound noise. For each test condition, the subject was presented one test list and instructed to adjust the volume for best intelligibility using the HA400. After an appropriate level was found, a first list with the unprocessed GÖSA sentences was presented followed by the processed sentences, in randomized order. The sentence lists were also selected randomly. Ten normal-hearing listeners (three female, seven male, age 20-27 years) participated in the test. The respective audiograms showed thresholds of 10 dB HL or below from 125 Hz to 4 kHz and 20 dB HL or below up to 8 kHz.

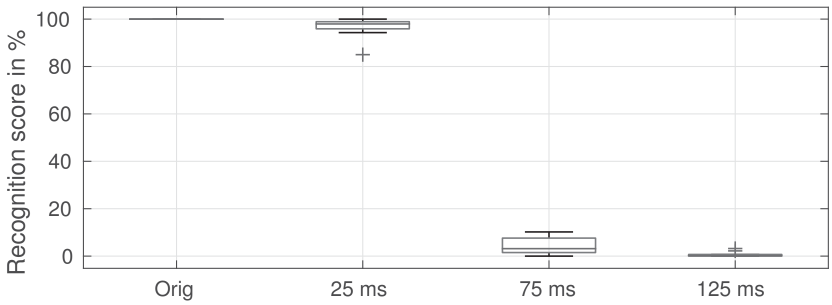


Fig. 4: Recognition score for original and processed sentences ($\tau = 25, 75,$ and 125 ms). The boxes show median (boxed line), lower and upper quartile (respective boundary of the box), lowest and highest values within 1.5 times the quartile range relative to the quartiles (whiskers), and outliers (+).

Figure 4 shows the correctly recognized words in percent for each test condition. As expected, the unprocessed sentences were fully recognized. While the median drops slighlty to 97.2% for $\tau = 25$ ms, two listenes still reach a score of 100%. For 75 and 125 ms, the score drops to 1.6% and 0.6% respectively. For the latter, five test listeners could not repeat one single word correctly. While $\tau = 75$ ms also appears to be sufficiently unintelligible, we choose $\tau = 125$ ms for additional headroom in case of uncommon circumstances like exceptionally slow speech.

CONCLUSIONS

This paper describes a well behaved system for long-time analysis of everyday listening situations. It delivers objective acoustic parameters at high resolution while maintaining the privacy of the subject and third parties. The software is easily extendable to capture additional features or provide enhanced functionality. An implementation of a basic online scene-analysis might be used to perform certain actions, e.g., trigger a questionnaire when a the acoustic environment changes, or the user might be prompted to take a picture using the smartphone's camera to capture the scene, of course in accordance with specific privacy regulations.

While the hard- and software are very easy to use, initial field tests showed that elderly people with little or no experience with handheld computers and/or touch-devices sometimes have difficulties operating the system. While there is acoustic and tactile feedback if the audio interface is not plugged in or analysis not started, there is still room to improve the handling as well as instruction of test subjects.

ENDNOTES

[1] The authors are in no way affiliated with the companies mentioned here or have any special interest in promoting a certain product. References are for documentation purposes only.

ACKNOWLEDGEMENTS

Many thanks to Anne Schlüter and Christoph Eike for conducting the listening test and to Audiofon Hearing Systems for supplying the hearing aid dummies.

Supported by governmental funding initiative "Niedersächsisches Vorab" of the Lower Saxony Ministry for Science and Culture, research focus "Hören im Alltag Oldenburg (HALLO)".

REFERENCES

Kates, J.M. (**2008**). *Digital Hearing Aids*. ISBN: 978-1-59756-317-8. (Plural Publishing).

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**, 2412-1421.

Shiffman, S., Stone, A.A., and Hufford, M.J. (**2008**). "Ecological momentary assessment," Ann. Rev. Clin. Psychol., **4**, 70-73.

Welch, P.D. (**1997**). "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," IEEE Trans. Audio Electroac., **15**, 70-73.

A new tool for subjective assessment of hearing aid performance: Analyses of Interpersonal Communication

RICHARD PALUCH[1, 2], MATTHIAS LATZEL[3,*], AND MARKUS MEIS[1]

[1] *Hörzentrum Oldenburg, Oldenburg, Germany*

[2] *University of Oldenburg, Working Unit "Sociological Theory", Oldenburg, Germany*

[3] *Phonak AG, Stäfa, Switzerland*

The performance of two different adaptive beamformer approaches in environments close to reality were investigated. They were subjectively evaluated via questionnaires and focus group discussions. Additionally, a new tool was tested, to assess how well video analyses with external rating of subjects' communication behavior, related to the grounded theory approach, generate new measures to describe the communication behavior using the different hearing aid algorithms. With this methodology, the results show different behavior of the participants between the algorithms in loud environments only. The new assessment tool was found to be a valuable method for obtaining a deeper insight into subjects' behavior and a new promising outcome tool for audiology.

INTRODUCTION

Directional microphone systems in hearing devices improve the speech intelligibility in complex listening situations. This has been confirmed in various studies in defined situations in the laboratory (e.g., Ricketts and Mueller, 1999; Ricketts and Henry, 2002; Bentler, 2005; Picou *et al.*, 2014). However, the question remains as to how relevant these results are for real life. Common measuring tools (e.g., questionnaires) used during clinical field trails are not sensitive enough and produce results with high variability, depending on the prudence of the subjects while filling in the questionnaire and on the situations occurring during the field trial. Research systems for evaluation in real life (e.g., Hasan *et al.*, 2013) are able to verify the situation by collecting physical data of the environment. They turned out to be a step forward but there are still the subjects' uncertainties which cannot be avoided/controlled by such systems.

Another approach is the simulation of real talking and listening situations in a laboratory and the use of head trackers, to get an objective measure of the influence of the systems by monitoring head movements (Cohen *et al.*, 2014). To overcome this uncertainty, it is necessary to use methodologies which do not make use of the subject's ratings of the test systems itself, but instead, provide measures demonstrating the effect of the test systems on the subjects' behavior objectively. This would then lead to conclusions regarding the performance of the test systems.

*Corresponding author: matthias.latzel@phonak.com

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

Two studies comparing the same directional microphone systems show contradictive results in the laboratory and home trial (Appleton-Huber and König, 2014; Latzel, 2015). In the lab, the system with the binaural beamformer showed favorable results compared to the system with the monaural beamformer, particularly in the areas of objective and subjective speech understanding. In contrast, in the home trial, where subjective results were obtained via questionnaires, the results were the opposite, especially for the situation “speech in (loud) noise”.

To improve interpretation of the data, a new methodology was transferred for the use with hearing aids in order to compensate for the disadvantages of a field test and the limited clinical relevance of lab measurements. This methodology has been used previously with a stronger focus on ethnographic field observation to analyse if and how social robots are experienced as social actors (Lindemann and Matsuzaki, 2014). Therefore a meaningful combination of the advantages of home and laboratory trials was set up.

Our study had two main objectives:

1. To investigate how two different adaptive beamformer approaches perform in environments close to reality when they are subjectively evaluated with questionnaires and focus group discussions.
2. To assess how well video analyses with external rating of the subjects’ communication behavior related to the grounded theory approach, generate new measures to describe the hearing performance of different hearing aid algorithms.

METHOD

A subgroup of the subjects from the beamformer study described in Latzel (2015) was invited to a moderated group discussion session. All participants were present at the same session. The subjects consisted of five experienced and two inexperienced hearing aid users. Six of them were male and three were female. All subjects had a moderate to severe hearing loss: better ear (4HFA), 43.8 dB HL (SD: 6.5 dB); worse ear (4HFA), 49.0 dB HL (SD: 6.4 dB). The mean age was 76.0 years (range 56-78 years). During the former beamformer studies, subjects had perceived differences of at least two scale points between the different beamformer approaches in daily life.

During the testing, subjects wore Phonak Audeo V90 312 hearing aids which were fitted according to the Phonak Adaptive Digital fitting formula (Latzel, 2013). The hearing aids were set with two programs:

- Program 1: Adaptive Monaural Beamformer: adaptive UltraZoom (aUZ)
- Program 2: Adaptive Binaural Beamformer: adaptive StereoZoom (aSZ)

Two difficult listening situations were simulated with the use of CAS (Communication Acoustic Simulator) at the Hörzentrum Oldenburg.

The first one was a laboratory scenario (S1) simulating a coffee house with an average sound level of 55 dB (LA_{eq}).

The second one was a loud laboratory scenario (S2) simulating a supermarket with an average sound level of 67 dB (LA_{eq}).

The first outcome measure was named Analyses of Interpersonal Communication in Realistic Acoustical Experimental Settings (AICRAS©) and consisted of a questionnaire and focus group discussion. The participants were encouraged to discuss the following topics, assuming a general interest of all participants, so that all would be both active (talking) and passive (listening) participants at the discussions:

- Topic 1: “Important hearing situations” (in the quieter lab scenario)
- Topic 2: “Experiences with hearing aids” (in the quieter lab scenario)
- Topic 3: “Communication in noise” (in the louder lab scenario)
- Topic 4: “Needs for future hearing aids” (in the louder lab scenario)

Subjects were firstly asked to fill out a questionnaire individually, rating different dimensions of hearing aid performance on a scale of 1-7 or −4 (too soft) to +4 (too loud). During discussion topics 1 and 3, they tested aUZ and during topic 2 and 4, they tested aSZ. They were not allowed to change the program to receive absolute ratings. Following this, subjects filled in one questionnaire as a group, where each of them judged aUZ in comparison to aSZ with regards to several different hearing aid performance dimensions (loudness of speech, speech intelligibility, listening effort, sound, loudness of the environment, and overall satisfaction) for both quieter and louder scenarios. Hearing aid performance dimensions and scales, from −5 (aUZ is better) to +5(aSZ is better), were shown or a board and participants were asked to give their ratings by placing stickers on the board.

In addition a second outcome measure was used named Video-based Analyses of Interpersonal Communication in Realistic Acoustical Experimental Settings (VIB-AICRAS©): An external rater watched a video of the focus group of subjects and rated their communication behavior based on the grounded theory approach by Glaser and Strauss (1967).

An example of the coding process for the grounded theory approach based on the video of the study can be seen in Fig. 1.

Recording	Phenomena	Indicators	Concepts
	Two persons are situated side by side. The left one leans his head slightly to the side and the right one moves his lips. Both of them are wearing glasses.	The left person listen closely to what the other person had said. The right person leans forward, to be better understood.	A verbal interaction.
			Face-to-face.
			loud environment.

Fig. 1: Example of the grounded theory approach for a certain section of the video from the study.

The external rater judged the communication behavior of the subjects according to two theoretical aspects indicated in Glaser and Strauss (1967) and Strauss (1987) respectively. The first aspect was 'forms of interaction', where the raters judged the amount of symbolic gestures (e.g., waving hands, "blocking behaviour") being used versus the amount of verbal communication. The second aspect was 'interdependence', where the rater judged the amount of face-to-face communication compared to group interaction. Both aspects were judged for aSZ and aUZ in both S1 and S2 lab scenarios. The rater identified, in total, 286 scenes and the analyses of the two hours of video material took approximately two weeks.

RESULTS

All box plots which follow show minimum, maximum, median, 25^{th} and 75^{th} quartiles. Results of the AICRAS© outcome measure can be seen in Figs. 2-4. As a general remark, all users reported noticing clear differences between aUZ and aSZ especially in the louder lab scenario (S2). Inexperienced users preferred aSZ in S2, due to more of the loud background sound being suppressed.

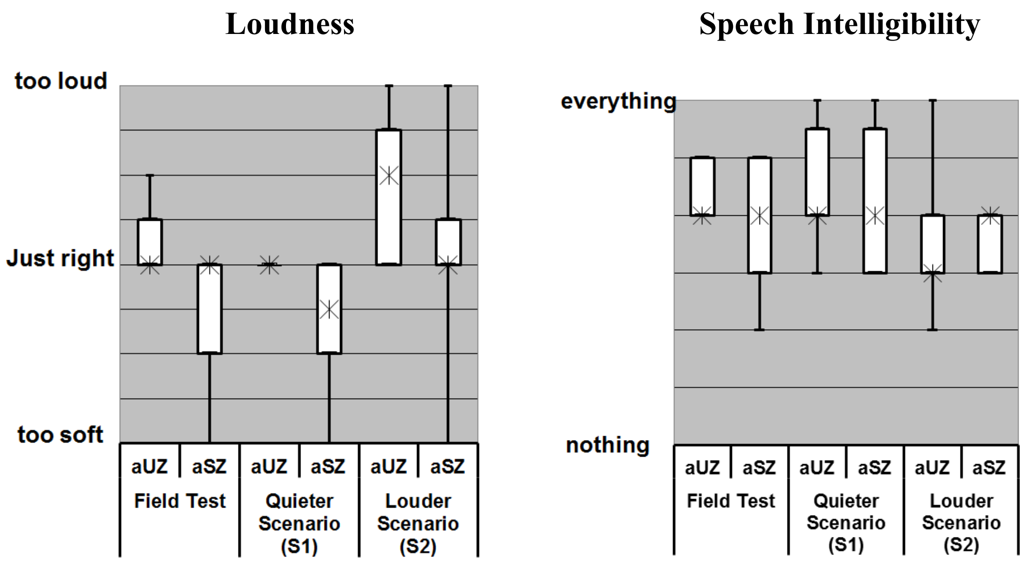


Fig. 2: Results of questionnaires completed individually. Comparison of results from the home trial with results in quieter and louder lab scenarios. Loudness was measured on a 9-point scale (−4 to +4). Speech Intelligibility was measured on a 7-point scale.

Figure 2 shows the results of the questionnaires which the subjects filled in individually. The home trial results were obtained in a prior study (Latzel, 2015). Loudness was perceived as "too loud" with aUZ and "adequate" with aSZ in S2. aSZ was perceived as slightly "too soft" in the quieter scenario (S1). Speech intelligibility was rated better with aSZ in contrast to aUZ in S2 but the speech intelligibility was rated lower in S2 than in the home trial.

Interestingly, participants rated the algorithms the same as they had in the home trial, when tested in the quieter lab situation, S1. This suggests that, during the home trial, the participants were mainly only in quieter situations because they deliberately avoided louder ones. They had been instructed to test the hearing aids also in louder situations. Nevertheless, they apparently did not. This would explain why the laboratory and home trial results from Appleton-Huber and König (2014) and Latzel (2015) were contradictory.

Figures 3 and 4 show the results of the questionnaires which subjects filled out as a group. There was a preference for aUZ in the quieter scenario for all dimensions with a general shift towards a preference of aSZ for the louder lab scenario. That is a second indication that the system is doing what it is intended to do (aUZ in softer noise environments, aSZ in louder environments) and that participants may have avoided louder situations during the home trial, so that that the advantages of aSZ could not be perceived.

Remark: The low rating of speech intelligibility in the louder lab scenario for aSZ may be due to the more “frontal” communication with the tester during the group assessment. The tester was standing quite far away and was therefore out of the radius so that the directional microphone was no longer effective anymore.

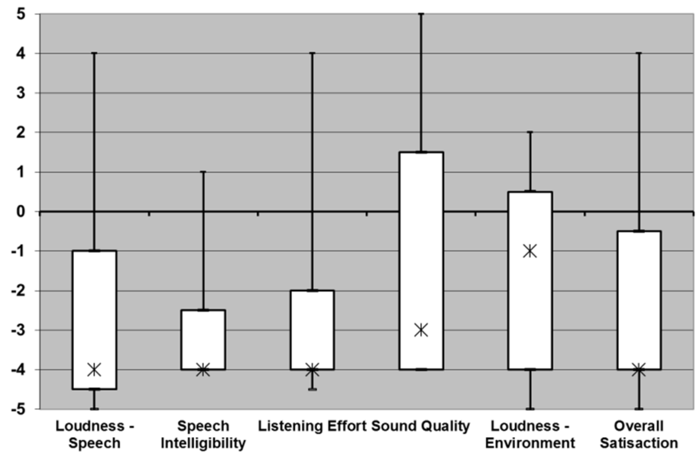


Fig. 3: Results of the questionnaire completed as a group for the quieter scenario (S1). Subjects rated preference of aUZ or aSZ for each dimension using a scale of -5 to +5.

Results of the VIB-AICRAS© outcome measure can be seen in Figs. 5 and 6.

In the quieter scenario, the behavior of the participants was very similar for both algorithms. This indicates that in quieter situations, the performance difference between the two algorithms is too small to make a difference in the behavior of the participants.

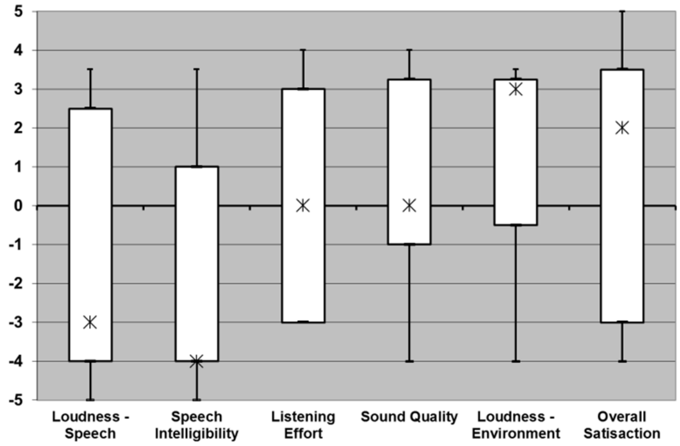


Fig. 4: Results of the questionnaire completed as a group for the louder scenario (S2). Subjects rated preference of aUZ or aSZ for each dimension using a scale from −5 to +5.

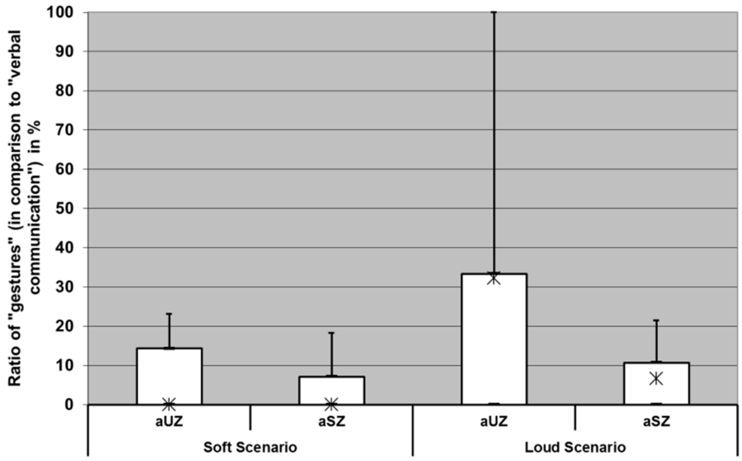


Fig. 5: VIB-AICRAS© 'Forms of interaction'. Ratio of symbolic gestures (compared to verbal communication) for aSZ and aUZ in both lab scenarios.

The external rater noticed a higher ratio of non-verbal communication (ratio of symbolic gestures to spoken words) for aUZ (mean = 28.4%) in comparison to aSZ (7.1%) in the louder lab scenario (p = 0.11, Wilcoxon, see Fig. 5), indicating more difficulty communicating in this situation. However, analyses also showed side-effects to using aSZ: The subject had to turn himself significantly more often towards his neighbor, in order to understand better. (p = 0.02, Wilcoxon, statistically significant).

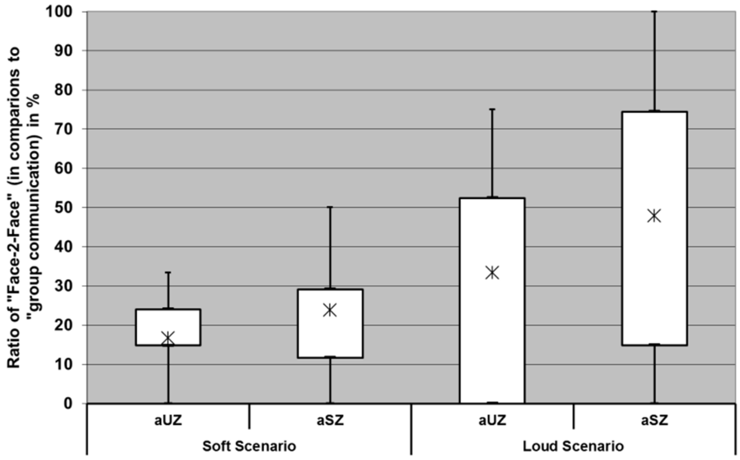


Fig. 6: VIB-AICRAS© 'Interdependence'. Median of face-to-face communication ratio (compared to group interaction) for aSZ and aUZ in both lab scenarios.

The external rater also noticed that the ratio of face-to-face communication in comparison to group interactions increases with increasing noise level. There was a higher ratio of face-to-face communication with aSZ (mean = 46.6%) than with aUZ (mean = 30.4%) with $p = 0.17$ (Wilcoxon, see Fig. 6), which leads to the side-effect described above.

Consequently, in the louder scenario, the difference between the algorithms is apparently larger and therefore can be seen in differences in participant behaviour. This indicates that the use of a narrower beamformer results in less group communication and more communication with the person sitting opposite.

CONCLUSIONS

Based on the questionnaire data, a slight overall preference in loud situations for aSZ was observed. This preference was based on subjects perceiving the environmental sound as smoother.

The home trial results for the dimension "situation with loud noise" is more highly correlated with the results of the quieter than of the louder lab scenarios. Subjects did not experience (were avoiding) loud situations during the home trial which explains the contradictive results.

In quieter situations there is preference for aUZ in all dimensions, whereas aSZ was preferred more in louder situations. This was observed especially for inexperienced hearing aid users.

The results lead to the conclusion that focusing only on maximum speech intelligibility by a narrower beamformer is not always favorable. It depends on the situation and the subjects' individual experiences/preferences and this should be something to consider during the hearing aid fitting procedure.

In general, the questionnaire tool AICRAS© and, especially the video tool VIB-AICRAS©, can be seen as valuable tools to obtain new outcome measurements in audiology.

REFERENCES

Appleton-Huber, J. and König, G. (**2014**). "Improvement in speech intelligibility and subjective benefit with binaural beamformer technology," Hearing Review, **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.

Cohen, J., Walden, T., and Brungart, D. (**2014**). "Effective communication in real-world environments with hearing aids," Poster at International Hearing Aid Conference (IHCON), Lake Tahoe, CA, August 2014.

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

Hasan, S.S., Lai, F., Chipara, O., and Wu, Y.H. (**2013**). "AudioSense: Enabling real-time evaluation of hearing aid technology in-situ," Proc. IEEE Int. Symp Comput. Based Med. Syst., 167-172.

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. and Matsuzaki, H. (**2014**): "Constructing the robot's position in time and space – the spatio-temporal preconditions of artificial social agency," Sci. Technol. Innov. Studies, **10**, 85-106.

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.

Ricketts, T. and Mueller, G. (**1999**). "Making sense of directional microphone hearing aids," Am. J. Audiol., **8**, 117-127.

Ricketts, T.A. and Henry, P. (**2002**). "Evaluation of an adaptive, directional-microphone hearing aid," Int. J. Audiol., **41**, 100-111.

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

Evaluating outcome of auditory training

ANNA LETH KRISTIANSEN[1] AND CARSTEN DAUGAARD[2,*]

[1] *Widex A/S, Lynge, Denmark*

[2] *DELTA Technical-Audiological Laboratory, Odense, Denmark*

Various articles suggest that better speech understanding can be obtained by auditory training. Auditory training typically consists of training speech perception in varying background noise levels or with degraded speech signals. Recently a Danish material for auditory training has been developed. This material consists of music training examples as well as speech training exercises. The rationale behind adding music training is variation in exercises as well as calling attention to details in auditory perception that could be valuable in speech perception as well as in hearing aid fitting. The results presented in this poster originate from examination of the benefits this material can provide on speech perception. Results from the investigation show an average benefit of auditory training, but with a large interpersonal variation, suggesting that a preselection of the individuals better suited for auditory training is needed. A battery of cognitive tests has been applied pre- and post-training, results from these tests are presented and discussed, in order to determine if there is correlation between cognition in general, improvement in cognition by auditory training, and obtaining better speech understanding by auditory training.

HISTORY OF AUDITORY TRAINING

Auditory training links naturally to hearing rehabilitation. The attention to the field grew in the USA around World War II, where better diagnostic capabilities and means of rehabilitation of hearing casualties from military service was severely needed. Skills such as lip-reading and “listening practice” would accompany the prescription of hearing aids to minimize the perceived handicap of the hearing loss. As hearing aids were improved during the eighties the auditory training as a unique part of the rehabilitation disappeared. In the late nineties, however, auditory training in the USA had a revival based on computer controlled learning programs and new scientific results.

The basic concept, which makes the training of hearing possible, is the auditory plasticity; reorganizing neural connections in the brain on the basis of input – and behavioural changes (Musiek, 2002). The argument is that a ski-sloping hearing loss, for example, deprives the stimulation of sound at high frequencies, thus causing the neurons to reorganize based on a bass dominated input. Restoring the treble by

*Corresponding author: cd@delta.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

means of a hearing aid will not find the right path in the brain until the connections regarding treble input are restored. Training might improve the speed of these changes.

COGNITION AND HEARING

Sweetow and Henderson Sabes (2004) have introduced a hierarchical communicat-ion model illustrating the build-up of acoustical communication from access to sound up to deriving meaningful information through the communication. The model shown below is a slightly modified version of the original (Sweetow and Henderson Sabes, 2004).

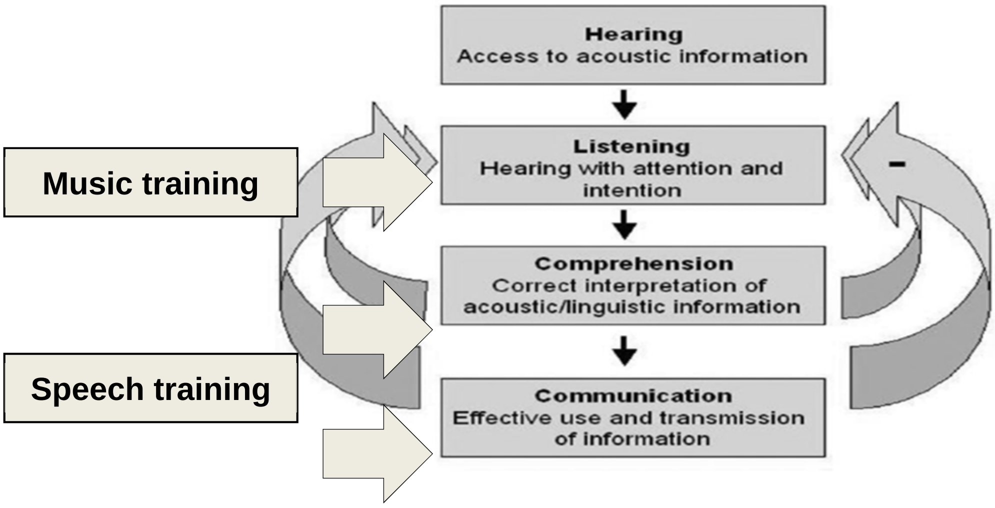


Fig. 1: Hierarchical model of communication form Sweetow and Henderson Sabes (2004), modified with indications of proposed entrance levels for speech and music training (Kristensen, 2013).

Taking this hierarchy into account, it is fair to propose that auditory training with speech signals aims at promoting the understanding at the higher levels in the hierarchy, while music training could be introduced as a way to sharpen the attention of details in the sound signal, as well as a break from speech perception tasks.

It could also be argued that the music training helps to establish connections between listening cues and words describing them, thus enhancing the ability to describe the performance of the hearing aid. This could help in the process of the best possible adjustment of the hearing aid.

In 2007 the Auditory Cognitive Neuroscience Society was founded, acknowledging the need for more multidisciplinary research in cognition and hearing. Cognition represents the mental processes and skills of requiring knowledge. The most prominent cognitive functions are: memory (including working memory), attention, executive functions (self-regulating functions), language functions and floating memory (the genetic preconditions for learning) (Banich, 2004). In acoustical communication these functions enhance our ability to extract the meaning of an acoustical signal in complex listening situations, and thus play an important role in speech perception, as indicated in Fig. 1. Despite the research in the area, so far a clear identification of the cognitive functions most relevant for listening in complex situation has not yet been revealed (Arlinger *et al.*, 2009).

TRAINING MATERIAL

The training material used for this project is based upon a Danish training material designed with speech perception tasks and music listening tasks. The speech part of the material is based upon the Danish DAT speech material. The user task is to identify the two last words of a sentence in the noise of one or two competing sentences. In the user screen it is possible to vary the signal to noise ratio from −10 to +10 dB. A few speech tests with variable speech speed are also found in the material. In the music tasks the user is presented for original as well as degraded music. The user must range the music pieces as more or less degraded (distortion, vibration, and tone) (Kristensen, 2013). In the original training material, the idea of the music listening tasks was to introduce the user to expressions describing sound. In this project the music tasks are only used for variation.

The training material is presented in PowerPoint, which eases the access to systems it can run on but limits the user interaction considerably. Based upon the feedback from the participants in this project it can be concluded to be problematic that only very limited feedback can be given to the user and that it is impossible to adaptively adjust the difficulty for the user.

TEST SET-UP

The current project has investigated if auditory training of hearing aid users has any effect on speech intelligibility in noise, cognitive abilities, communication skills, and degree of hearing handicap in hearing aid users.

Furthermore, it is investigated whether some people benefit more from the auditory training than others do and if so, which factors and personal characteristics can be used to identify those individuals most likely to benefit from auditory training.

To evaluate the effect of the auditory training program, a quantitative experimental study was performed. A participant group of 15 hearing aid users aged 55-81 years was selected to train with the program for two months. Their hearing loss had an average PTA of 55 dB HL varying from 5 dB HL to 95 dB HL. Their discrimination score was on average 74%, varying from 100% to 32%. The inclusion criteria were somewhat loose, as the focus was to recruit as many participants as possible willing to do the training for two months. By coincidence all participants wore different hearing aids. The hearing aids worn by the participants were coincidentally all different newer products from the leading European manufacturers.

Prior to and after the period of training the participant group was presented with a test battery to assess the benefit of the exercises. The test battery consists of both objective and subjective tests in areas where improvement due to the auditory training could be expected. Only off-task tests were selected to reveal a more general effect of the training rather than a learning effect.

Cognitive test	Test modality	Measured cognitive ability
Visual forward digital span test	Visual	Working memory
Jaeggi-Bushkuehl dual n-back task	Audiovisual	Working memory and floating intelligence
Fast counting test	Visual	Visual perception
Go/no-go auditory reaction time test	Auditory	Auditory attention and processing efficiency
Eriksen flanker test	Visual	Information processing and selective attention

Table 1: Overview of the selected cognitive tests, their modality and which cognitive ability they measure.

The test battery consisted of two speech in noise tests – Dantale II and Just Follow Conversation (JFC) – five cognitive tests, and two subjective tests – the NSH question-naire and the Hearing Handicap Inventory for the Elderly (HHIE).

The choice of cognitive tests for this project was not straight forward. First of all the tests had to be in Danish, it should not require skilled personnel (psychologist or equivalent) to perform them, and they should be available at a reasonable price. The cognitive tests in this project were selected from a website (www.cognitivefun.com) which contains a large collection of cognitive tests, testing different functions of cognition. Here it was possible to choose five different tests each focusing on different abilities. The chosen cognitive tests, their test modalities, and the cognitive skills tested in each test, are shown in Table 1.

RESULTS

The results from the pre- and post- speech tests have been summarized in Figs. 2 to 5, showing the difference score for each participant. Bars above the horizontal line indicate improvement from the training, bars below the line indicates that the participant did worse in the post-test. The bar isolated at the right is the average.

The graph for Dantale II (Fig. 2) shows an improvement for roughly half of the participants, and a small set-back or no improvement for the other half. In the JFC case (Fig. 3), all improved or did at least as good in the post-test as in the pre-test. The improvement is significant for JFC (paired *t*-test, p=0.001) but not for Dantale II (paired *t*-test, p=0.051). A fair correlation between the improvement in the two tests for the participants is seen. This indicates improved speech perception for some of the participants. However the test-retest variation of the Dantale II test might be too high to track the small improvements.

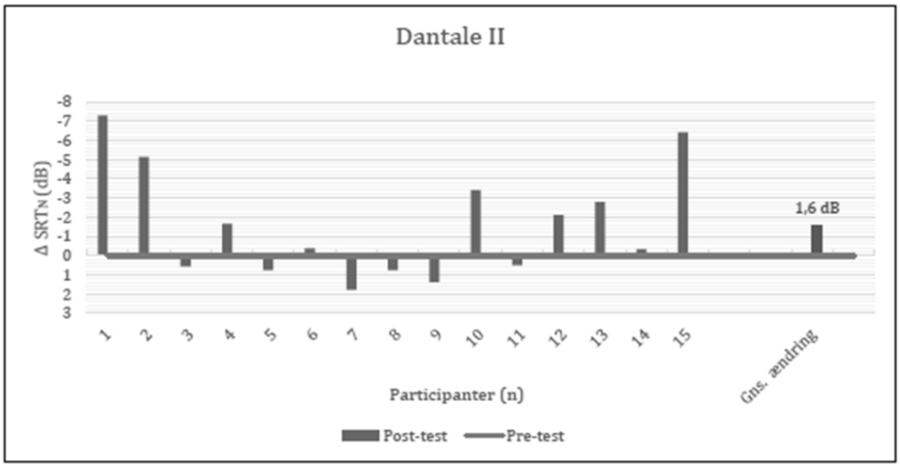


Fig. 2: Improvement per participant and on average ("Gns. ændring", isolated at the right), measured by the Dantale II (Hagermann) speech test.

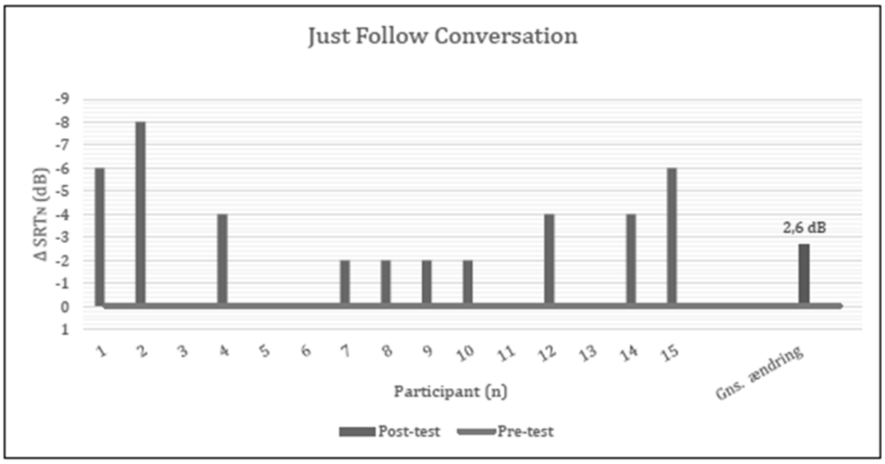


Fig. 3: Improvement per participant and on average ("Gns. ændring", isolated at the right), measured by the JFC speech test.

The graphs for the difference between the pre- and post- answers from the two questionnaires are shown in Fig. 4. In the NSH questionnaire a majority of participants indicate very little effect from the training. No significant improvement could be found. (paired *t*-test, p=0.255). The HHIE shows a significant average improvement (paired *t*-test, p=0.019). Further analysis reveals that improvement primarily origins from situational rather than emotional questions of the HHIE.

Graphs representing the cognitive tests are shown in Fig. 5. Both tests show steady or improved performance for the majority of participants.

Looking at the general results from the test battery, an improvement due to auditory training seems to be plausible. The improvement is most clearly visible in the JFC speech test, in the situational questions of the HHIE, and in some of the cognitive tests. From the tests it is also clear that some participants seem to benefit more from the training than others, with different participants showing improvement in different tests. Thus it is difficult to find a clear pattern of which participants in general improved in the tasks trained in the test battery. A hint to which factors and personal characteristics can be used to identify those individuals can be derived from a correlation analysis (Pearson's correlation coefficient). Table 2 presents a matrix of the Pearson corrrelation of the improvement of variables with demographic factors and pre-test scores. It indicates that tone loss and years with hearing loss correlate with improvement in the go/no go auditory test. As the go/no-go auditory test is testing auditory attention and processing efficacy, it is fair to speculate that people with larger and longer lasting hearing losses will benefit from the training because it sharpens their auditory attention.

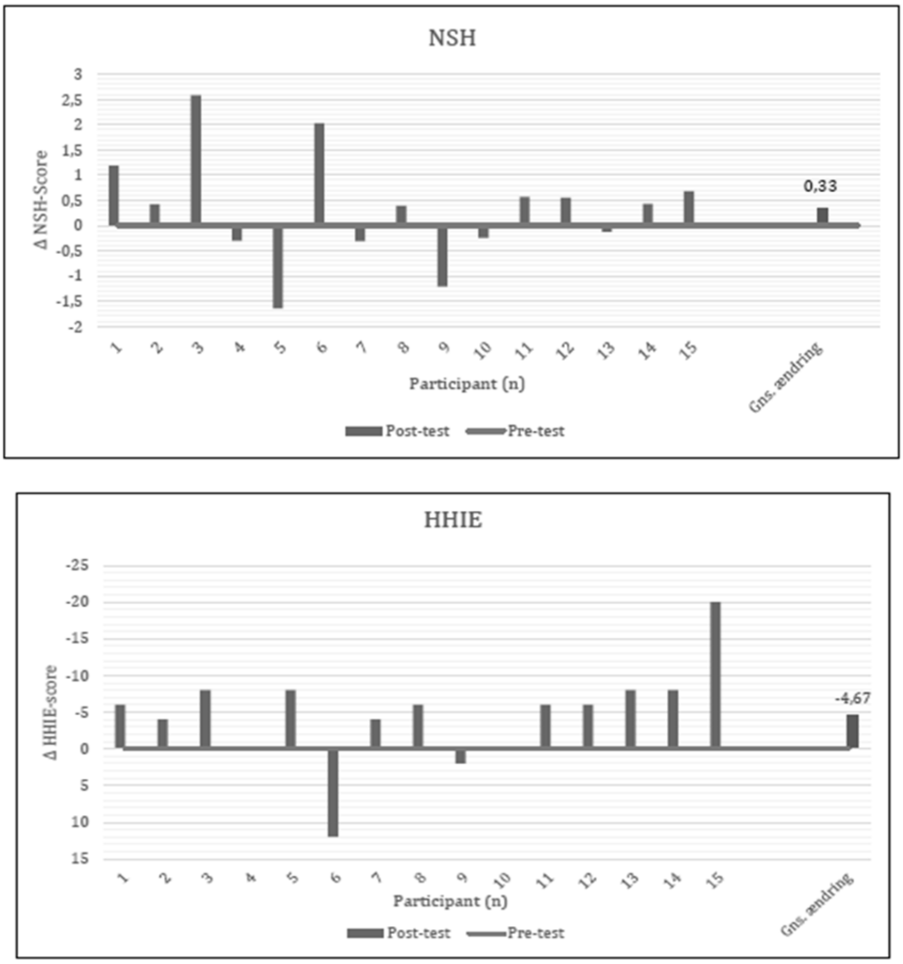
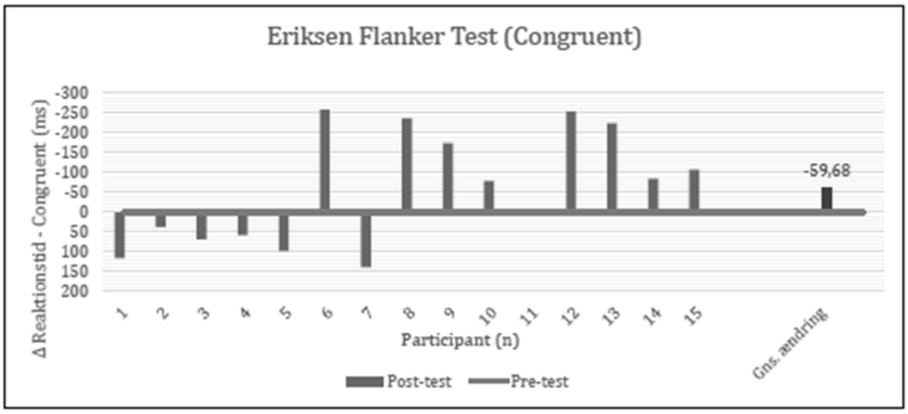


Fig. 4: Improvement per participant and on average ("Gns. ændring", isolated at the right), measured by the NSH (top panel) and the HHIE (bottom panel) questionnaires.



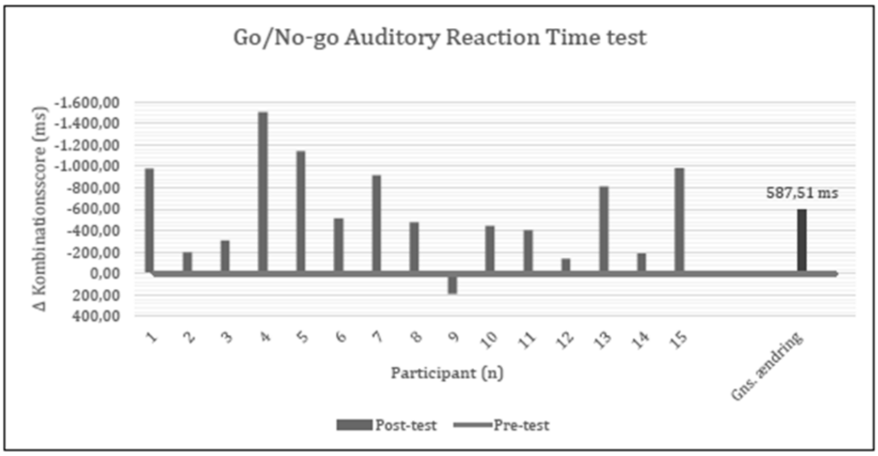


Fig. 5: Improvement per participant and on average ("Gns. ændring", isolated at the right), measured by two of the cognitive tests. These figures were selected for this article as the correlation analysis indicates that improvement in these cognitive tests correlates with severity of hearing loss and Dantale II score.

The table also shows that discrimination score correlates with the congruent part of the Eriksen flanker test. The cognitive ability tested in this test is information processing and selective attention, which again indicates that auditory training enhances attention and processing speed, and the more pronounced the hearing loss the more benefit.

It is also interesting to note that age and hours of training do not seem to influence the measured benefit of the training. If the latter is true it calls for a much more adaptive approach to the training than the current Danish training material at present can offer.

CONCLUSION

From the feedback from the participants it is clear that the training material should have a more adaptive difficulty level and should provide more feedback. The use of music tasks is a good variation of the training. From the questionnaires answered it seems that the training has only limited influence on the participants' perceived improvement from the training.

The results from this project indicate that auditory training can improve cognitive skills related to speech understanding and performance in speech tests. However, the benefit of the training varies considerably among the participants. Correlation analysis hints that more severe, longer lasting hearing losses undergo the biggest improvement in auditory attention and information processing ability from the training.

		Improvement in dependent variables (from pre to post test)									
		Dantale II (SRT_N)	JFC (SRT_N)	Go/No-go Auditory test	Fast Counting	Eriksen Flanker test (Congurent)	Eriksen Flanker test (Incongurent)	Jaeggi-Buschkuel	Visual digit span	NSH Questio-narie	HHIE
Demographics and variables pretest score	Age	0,31	0,09	-0,18	-0,06	0,26	0,18	0,12	-0,14	0,35	0,18
	Tone-loss	-0,17	-0,13	0,55*	-0,19	-0,33	-0,23	0,01	0,08	0,16	-0,1
	Discrimination Score	0,08	-0,03	0,5	0,2	0,56*	0,41	-0,18	0,09	0,08	0,17
	Years with hearing loss	0,04	0,04	0,73***	-0,1	-0,32	-0,31	-0,39	-0,18	0,08	0,09
	Years with hearing aids	-0,04	-0,2	0,29	0,06	-0,12	0,09	-0,03	0,25	0,48	-0,28
	Hours of auditory training	0,38	0,41	0,5	0,11	-0,22	-0,15	-0,46	0,14	0,22	0,1
	Dantale II (SRT_N)	-0,04	0,05	0,55*	-0,37	-0,55*	-0,42	-0,13	-0,04	0,13	-0,31
	JFC (SRT_N)	-0,29	0,07	0,42	-0,36	-0,49	-0,3	0,08	-0,35	-0,16	-0,32
	Go/No-go Auditory test	0,16	-0,07	0,9	-0,01	-0,31	-0,52	-0,21	-0,07	-0,32	0,28
	Fast Counting	0,01	0,01	0,49	-0,45	-0,46	0.59*	0,16	0,16	-0,28	-0,16
	Eriksen Flanker test (Congurent)	-0,2	-0,06	-0,32	0,13	0,76***	0,61**	0,11	0,21	-0,11	0,14
	Eriksen Flanker test (Incongurent)	-0,35	-0,16	-0,44	0,09	0.78***	0,78***	0,06	0,19	0,01	-0,09
	Jaeggi-Buschkuel	0,18	-0,25	-0,13	0,21	-0,01	0	-0,3	0,52*	0,53*	0,03
	Visual digit span	-0,21	-0,23	-0,09	-0,16	-0,64**	-0,37	0,27	-0,47	0,01	-0,07
	NSH Questionarie	-0,13	-0,29	-0,22	0,25	0,19	0,07	0,21	0,27	-0,49	-0,06
	HHIE	0,05	0,39	0,32	-0,07	-0,44	-0,35	0,1	-0,11	-0,35	0,36

Table 2: Pearson's correlation matrix between demographic data and pre-test score of dependent variables with improvements of dependent variables. Circles show significant correlations between variables. Correlations between pre-test score and improvement for the same variable are not highlighted.

REFERENCES

Arlinger, S., Lunner, T., Lyxell, B., and Pichora-Fuller, M.K. **(2009)**. "Background and basic processes – The emergence of cognitive hearing science". Scand. J. Psychol., **50**, 372-384.

Banich, M.T. **(2004)**. *Cognitive Neuroscience and Neuropsychology*, 2nd edition. Boston: Houghton Mifflin Company.

Kristensen, S.T. **(2013)**. *Høretræning – Auditory Training*. Master's thesis, Odense: University of Southern Denmark.

Musiek, F.E., Shinn, J., and Hare, C. **(2002)**. "Plasticity, auditory training, and auditory processing disorders," Semin. Hear., **23**, 263-275.

Sweetow, R.W. and Henderson Sabes, J. **(2004)**. "The case for LACE: Listening and auditory communication enhancement training," Hearing Journal, **57**, 32-40.

Sensitivity to angular and radial source movements in anechoic and echoic single- and multi-source scenarios for listeners with normal and impaired hearing

MICHA LUNDBECK[1,2,*], GISO GRIMM[1,2], VOLKER HOHMANN[1,2], SØREN LAUGESEN[3], AND TOBIAS NEHER[1]

[1] *Medizinische Physik and Cluster of Excellence Hearing4all, Oldenburg University, Oldenburg, Germany*

[2] *HörTech gGmbH, Oldenburg, Germany*

[3] *Eriksholm Research Centre, Snekkersten, Denmark*

So far, very little is known about the perception of spatially dynamic sounds, especially under more complex acoustic conditions. Therefore, this study investigated the influence of reverberation and the number of concurrent sources on movement perception of listeners with normal and impaired hearing. Virtual listening environments were simulated with the help of a higher-order Ambisonics-based system that allows rendering complex scenarios with high physical accuracy. Natural environmental sounds were used as the stimuli. Both radial (near-far) and angular (left-right) movement perception were considered. The complexity of the scenarios was varied by adding stationary sound sources as well as reverberation. As expected, hearing-impaired listeners were less sensitive to source movements than normal-hearing listeners, but only for the more complex acoustic conditions. Furthermore, adding sound sources generally resulted in reduced sensitivity to both angular and radial source movements. Reverberation influenced only radial movement detection, for which elevated thresholds were observed. Altogether, these results illustrate the basic utility of the developed test setup for studying factors related to spatial awareness perception.

INTRODUCTION

Sensorineural hearing loss can lead to a multitude of hearing deficits, particularly under more complex listening conditions. For example, hearing-impaired (HI) listeners are known to experience great difficulty with listening in multi-source conditions and with judging distance and movement, and these problems appear to be related to their experience of handicap (Gatehouse and Noble, 2004).

Even though a number of studies have addressed distance and movement perception in normal-hearing (NH) listeners (e.g., Perrott and Saberi, 1990; Chandler and Grantham, 1992) the same is not true for HI listeners. Also, the studies that have

*Corresponding author: micha.lundbeck@uni-oldenburg.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4.

been conducted so far have generally focused on simple situations: anechoic single-source scenarios. An exception to this is a recent study of Brungart *et al.* (2014) who presented multiple environmental sounds in various pseudo-dynamic arrangements to their listeners by adding to or removing sounds from the auditory scene. They found a decrease in performance with increased task complexity and the number of sound sources presented.

In the current study, we piloted a novel test setup that we developed for studying factors related to the perception of spatial dynamics. To that end, we used a toolbox that allowed us to simulate virtual acoustic environments with high physical accuracy. Our focus was on the perception of moving sounds, i.e., sounds that varied in terms of their angular (left-right) or radial (near-far) position. In particular, we investigated the influence of the number of concurrent sound sources as well as reverberation on the ability of young normal hearing (NH) and elderly hearing impaired (HI) listeners to detect changes in angular (left-right) or radial (near-far) source position. Our hypotheses were as follows:

1. Young NH listeners will generally outperform elderly HI listeners in terms of their thresholds for source movement detection
2. An increased number of sources will result in higher thresholds for source movement detection
3. Reverberation will generally also affect source movement detection

METHODS

Participants

The participants were eight young NH listeners (2 male, 6 female) aged 23-29 yrs (mean: 25.8 yrs) and 10 paid elderly HI listeners (6 male, 4 female) aged 64-79 yrs (mean: 74.5 yrs). Five of them were experienced hearing aid users with 2-6 yrs of experience. The NH listeners had normal audiometric thresholds (≤ 25 dB HL) from 0.125 to 8 kHz. The HI listeners had symmetric moderate-to-severe sensorineural hearing losses, as depicted in Fig. 1.

Setup

The “Toolbox for Acoustic Scene Creation and Rendering” (TASCAR; Grimm *et al.*, 2015) was used to simulate the virtual environments. TASCAR allows rendering complex scenarios with high physical accuracy, including moving sound sources. The acoustic environment was based on an entrance hall (approx. 10.5 m × 6 m × 2.8 m with solid walls, glass and wooden floor). The head of the virtual listener was placed 1 m away from the middle of the shorter wall facing along the longer side. The target source was located 1 m away from, and directly in front of, the listener (height of 1.5 m). A schematic top-down view of the room is shown in Fig. 2. A change in complexity of the scenario was achieved by adding two or four stationary sound sources at a distance of 1 m each and azimuths of ±30° and ±60° relative to the frontal direction. The room could be changed from an anechoic to an echoic (T_{60} = 0.8 s) environment.

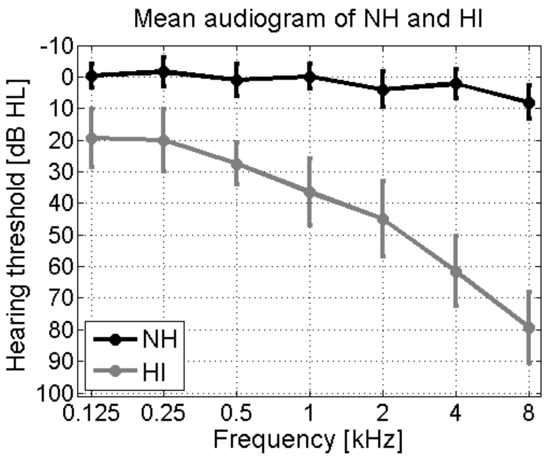


Fig. 1: Average hearing thresholds for the NH (black) and the HI (grey) group. Error bars denote standard deviations.

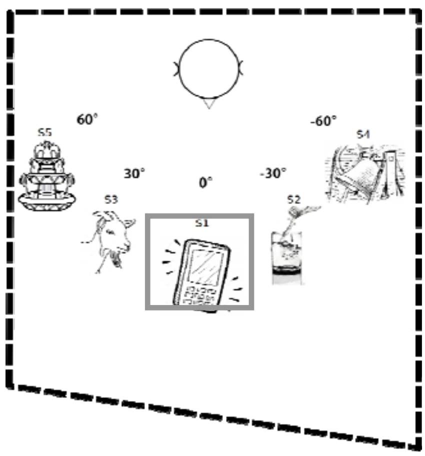


Fig. 2: Schematic top-down view of the simulated room, showing the virtual listener and the five sound sources at 0° (S1), ±30° (S2, S3), and ±60° (S4, S5). S1 (telephone sound) was moving either in the left-right or near-far direction (see text for details).

Stimuli

For reasons of comparability with the literature (Chandler and Grantham, 1992), we made reference measurements using a one-octave band of noise centered at 3 kHz as the stimulus. In this case, the velocity of the source movement was fixed at 20°/s for angular and 7 m/s for radial movements. The tracking variable in the adaptive procedure (see below) was the stimulus duration. The measured values of stimulus duration were then multiplied with the velocity to obtain the Minimum Audible Movement Angle (MAMA) or the Minimum Audible Movement Distance (MAMD) (e.g., 0.45 s × 20°/s = 9° of arc).

In addition, we made measurements with up to five different environmental sounds (similar to Brungart *et al.*, 2014). A ringing phone served as the target sound in all measurements (see Fig. 2). The other sound sources (soda pouring, goats, church bells, and a fountain) were fixed in location. Each sound was presented at an overall level of 65 dB SPL (nominal). For the measurements with the environmental sounds the stimulus duration was fixed at 3 s. To vary the extent of source movement the velocity was varied in the adaptive procedure.

Procedure

Initially, the hearing thresholds from 0.125 Hz to 8 kHz of all participants were determined. They were then seated in a soundproof booth in front of a screen where they could use a graphical user interface to provide their answers. Their task was to indicate whether or not they perceived the target sound source to move. For this, a single-interval 2-alternative-forced-choice (AFC) paradigm with an adaptive 1-up 2-down rule (Levitt, 1971) was implemented in the software framework "psylab" (Hansen, 2006). In half of the trials, the target sound source was simulated to move. In the angular movement conditions, the direction of movement (towards the left or right) was randomized, whereas in the radial movement conditions a withdrawing movement was always simulated. Playback was via a 24-bit Edirol UA-25 soundcard, a headphone preamplifier (Tucker-Davis HB-7), and a pair of Sennheiser HDA 200 headphones. For the HI listeners linear amplification was provided via the Master Hearing Aid research platform (MHA; Grimm *et al.*, 2006) according to the NAL-RP fitting rule to ensure adequate audibility.

Initially, a training run was completed for every new condition, i.e. before the reference measurements and whenever the movement direction (left-right to near-far or vice versa) was changed. Each participant completed two blocks of measurements divided into angular and radial movement measurements with a preceding reference condition. As apparent from Table 1, 12 environmental scenarios were tested (in randomized order). After two to three weeks a set of retest measurements was performed to assess test-retest reliability. The whole experiment took about four hours.

Spatial movement dimension	**Number of sound sources**	**Degree of reverberation**
Left-right (MAMA) vs. Near-far (MAMD)	1 source (moving or not) vs. 3 sources (1 moving or not) vs. 5 sources (1 moving or not)	Anechoic vs. Echoic

Table 1: Experimental variables chosen for the simulation of the different environmental scenarios. A total of 12 scenarios were tested.

Data analysis

In accordance with Chandler and Grantham (1992), a criterion was set to accept or exclude thresholds estimated by the adaptive procedure. In their study, thresholds were only accepted if the standard deviation of the tracking variable at the reversal points did not exceed one-third of the corresponding threshold value. Due to the fact that we observed large tracking excursions for some of our participants, we raised the criterion value to one-half of the threshold. As a result, a total number of 11 thresholds had to be excluded (out of 484 estimated thresholds).

Because of the relatively small sample size and the non-normal distribution of some datasets we performed non-parametric tests. To test for group differences we performed Mann-Whitney *U*-tests for independent samples. To test for the influence of the number of sound sources we performed Friedman’s ANOVAs, while for testing the influence of reverberation within each group we performed Wilcoxon tests for dependent samples.

Two of the HI participants had great difficulties to hear out the target sound source in the multi-source scenarios. For these conditions, they therefore had to be excluded from the data analysis.

RESULTS

Reference measurements

In Fig. 3 the reference measurements for the two movement dimensions are depicted. The left panel shows the MAMA thresholds for both groups in comparison to a reference data point taken from Chandler and Grantham (1992). The right panel shows the MAMD thresholds for the two groups. The difference between the NH and HI thresholds was not significant for either reference condition (MAMA: $p = 0.54$; MAMD: $p = 0.17$).

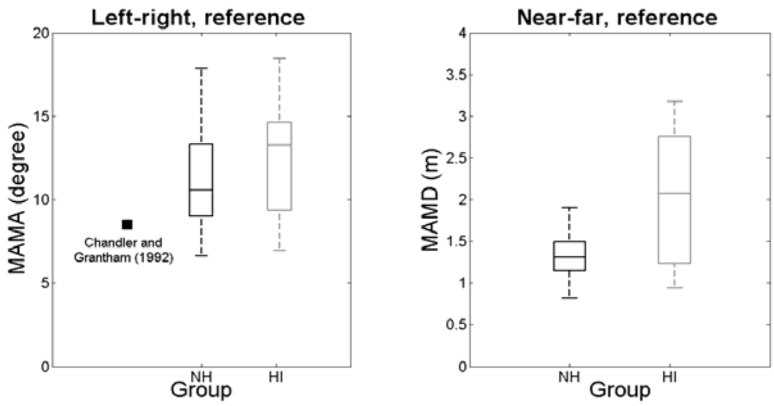


Fig. 3: Boxplots of reference measurements. Left: MAMA thresholds (fixed velocity of 20°/s) for NH and HI listeners. The black square shows a reference threshold value from Chandler and Grantham (1992). Right: MAMD thresholds for NH and HI listeners.

Compared to the literature value, the median MAMA threshold measured with our setup was slightly elevated. However, the reference value falls clearly within the range of our dataset. For the MAMD measurements, no corresponding literature data are available.

MAMA measurements with environmental sounds

Fig. 4 shows the thresholds for the angular movement detection task.

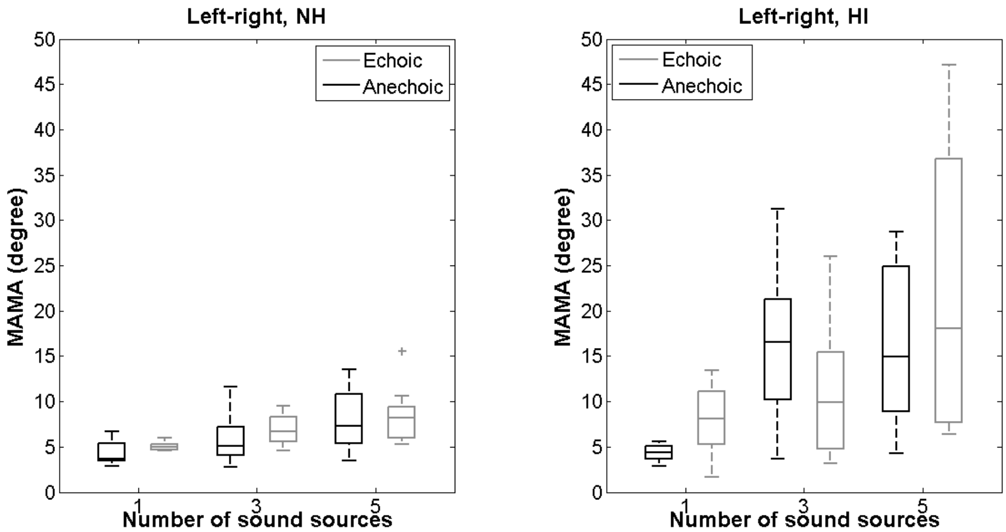


Fig. 4: MAMA thresholds for environmental sounds. NH data are depicted in the left and HI data in the right panel. Shown is the MAMA for different numbers of sound sources and degree of reverberation (black: without reverb; grey: with reverb).

A Mann-Whitney *U*-Test (2-tailed) performed on the data pooled across all conditions revealed a significant difference between the two groups ($U = -2.9$, $p = 0.004$). Furthermore, a significant change in threshold was found when the number of sound sources was increased (pooled across the two reverberant conditions). This was true for both groups and all conditions (all $p < 0.05$) except for the comparison of three and five sound sources within the HI group ($p = 0.3$). However, no influence of the degree of reverberation was found (NH: $p = 0.22$; HI: $p = 0.9$). A possible explanation for this could be that listeners may quickly 'learn' room reverberation patterns, enabling them to suppress spatial cues of signal components that have been corrupted by reflections (cf. Shinn-Cunningham, 2000).

MAMD measurements with environmental sounds

The results for the near-far movement detection task are depicted in Fig. 5. The data were analyzed in the same manner as the MAMA thresholds. Again, a significant difference between the two groups was found ($U = -2.6$, $p = 0.01$). Furthermore, unlike in the MAMA results, reverberation had a significant influence on both groups (both $p < 0.001$). Also, a general influence of the number of sound sources was found. Only the comparison of the 3- and 1-source scenario for the NH group and the 3- and 5-source scenario for the HI group was non-significant (all other $p < 0.05$). Interestingly, the thresholds for the 1-source scenarios were similar for the two groups. Sensitivity worsened for the multi-source scenarios, especially so for the HI group. Participants reported that they depended on level changes of the target stimulus and that it was difficult to imagine the withdrawing movement in the virtual environment. The combination of additional masker sounds and reverberation led to a more diffuse sound field that lowered the direct-to-reverberant sound ratio, an important cue for distance perception in rooms (Bronkhorst and Houtgast, 1999; Zahorik, 2002). Hence, the detection of level changes presumably became more difficult.

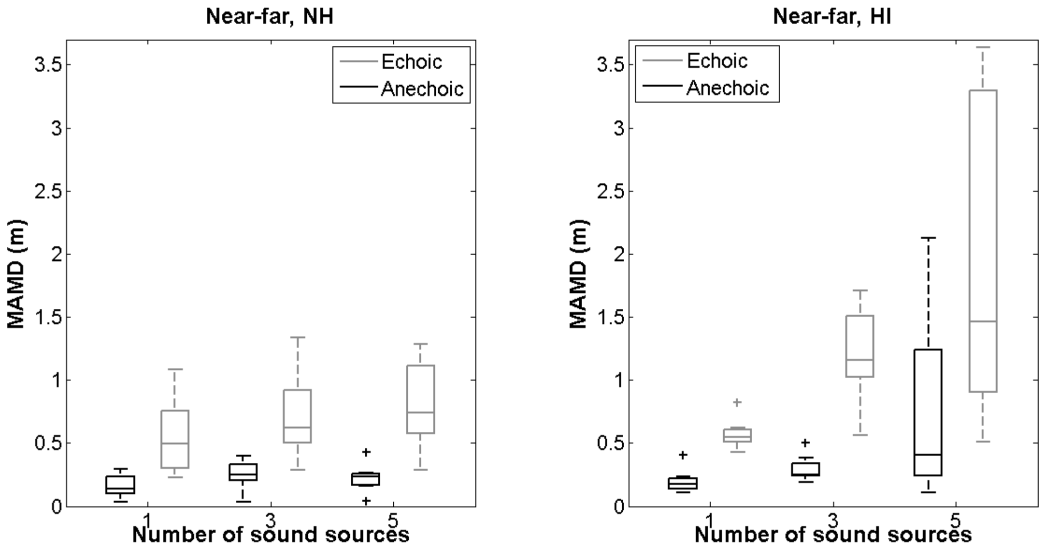


Fig. 5: MAMD thresholds for environmental sounds. NH data are depicted in the left and HI data in the right panel. Shown is the MAMD for different numbers of sound sources and degree of reverberation (black: without reverb; grey: with reverb).

SUMMARY

This study investigated the influence of the number of sound sources and reverberation on source movement perception in listeners with normal and impaired hearing. Comparison to some literature data showed that our (TASCAR-based) setup can be used for the assessment of spatial dynamics, as the thresholds we obtained were of comparable magnitude to those from the literature measured with a free-field setup. Results for the environmental sounds generally showed the expected differences between NH and HI listeners insomuch as the NH listeners were more sensitive to angular and radial source movements in the multi-source conditions. Furthermore, an increase in the number of concurrent sound sources generally resulted in higher thresholds (except for the NH listeners in the near-far conditions). Finally, the expected change in thresholds under reverberant conditions was found for the near-far conditions, but not for the left-right conditions. Altogether, this study shows promise regarding the assessment of movement perception in complex listening scenarios with the developed test setup.

ACKNOWLEDGEMENTS

This research was funded by the Oticon Foundation and the DFG Cluster of Excellence EXC 1077/1 “Hearing4all”.

REFERENCES

Bronkhorst, A.W. and Houtgast, T. (**1999**). “Auditory distance perception in rooms,” Nature, **397**, 517-520.

Brungart, D.S., Cohen, J., Cord, M., Zion, D., and Kalluri, S. (**2014**). “Assessment of auditory spatial awareness in complex listening environments,” J. Acoust. Soc. Am., **136**, 1808-1820.

Chandler, D.W. and Grantham, D.W. (**1992**). “Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity,” J. Acoust. Soc. Am., **91**, 1624-1636.

Gatehouse, S. and Noble, W. (**2004**). “The speech, spatial and qualities of hearing scale (SSQ),” Int. J. Audiol., **43**, 85-99.

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., Ewert, S., and Hohmann, V. (**2015**). “Evaluation of spatial audio reproduction schemes for application in hearing aid research,” Acta Acust. United Ac., **101**, 842-854.

Hansen, M. (**2006**). “Lehre und Ausbildung in Psychoakustik mit psylab: Freie Software für psychoakustische Experimente,” Fortschritte der Akustik, **32**, 591.

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

Perrott, D.R. and Saberi, K. (**1990**). „Minimum audible angle thresholds for sources varying in both elevation and azimuth,” J. Acoust. Soc. Am., **87**, 1728-1731.

Shinn-Cunningham, B. (**2000**). “Learning reverberation: Considerations for spatial auditory displays,” Proc. International Conference on Auditory Display, Georgia Institute of Technology.

Zahorik, P. (**2002**). “Assessing auditory distance perception using virtual acoustics,” J. Acoust. Soc. Am., **111**, 1832-1846.

Effect of harmonic rank on the streaming of complex tones

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, England*

The effect of the rank of the harmonics on sequential stream segregation of complex tones was investigated for normal-hearing participants with no musical training. It was hypothesized that stream segregation would be greater for tones with high pitch salience, as assessed by fundamental frequency (f_0) difference limens. Pitch salience is highest for tones containing some low (resolved) harmonics, but is also fairly high for tones containing harmonics of intermediate rank. The tones were bandpass filtered between 2 and 4 kHz and harmonic rank was varied by changing the f_0. There was a significant trend for less stream segregation with increasing harmonic rank. The amount of stream segregation was inversely correlated with the f_0 difference limens, consistent with the hypothesis.

INTRODUCTION

Fundamental frequency (f_0) discrimination, which provides a measure of pitch salience, is better for complex tones with low harmonics (low harmonic rank) than for tones with only high harmonics. This has often been interpreted in terms of spectral resolvability, i.e., f_0 difference limens (f_0DLs) are smaller for complex tones that contain resolved harmonic components (Bernstein and Oxenham, 2006a;b). However, Bernstein and Oxenham (2003) found that f_0DLs were similar for complex tones presented dichotically, with even harmonics presented to one ear and odd harmonics to the other, and for complex tones with all harmonics presented to both ears, even though dichotic presentation should lead to greater resolvability of the harmonics. They argued that harmonic rank and not resolvability is the key factor governing the magnitude of f_0DLs.

The stream segregation of sequences of sounds is facilitated by perceived differences between the sounds, such as differences in frequency, spectrum and f_0 (Moore and Gockel, 2002). This, combined with the fact that f_0 discrimination is better for tones that contain low harmonics, leads to the hypothesis that stream segregation of complex tones would also be affected by harmonic rank. However, this is not consistent with the results of Vliegen and Oxenham (1999), who found that subjective judgements of stream segregation were similar for pure tones, complex tones with low harmonics, and complex tones with only high unresolved harmonics. The present study investigated subjective stream segregation for pure

*Corresponding author: samkma@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

tones and complex tones with variable harmonic rank and compared the results to f_0DLs measured for the same stimuli.

EXPERIMENT 1: SUBJECTIVE STREAM SEGREGATION

Method

The stimuli were sequences of ABA triplets where the f_0s of the A and B tones were varied (see Fig. 1). Each tone had a duration of 90 ms, including 10-ms raised-cosine onset and offset ramps, with gaps of 20 ms within each triplet and 110 ms between triplets. Each sequence lasted approximately 8 s and contained 19 triplets.

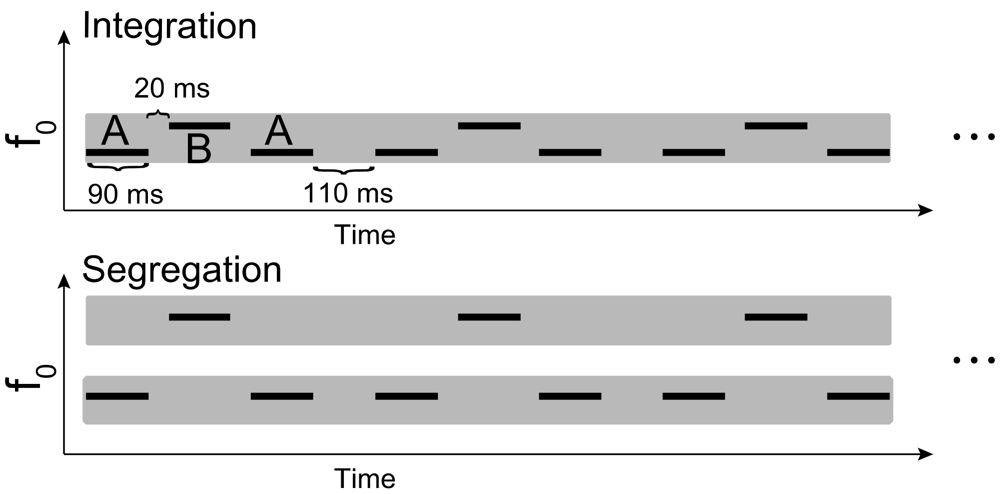


Fig. 1: Schematic illustration of stream segregation and the stimuli used to investigate it.

Both pure tones and complex tones were used. The complex tones were bandpass filtered between 2000 and 4000 Hz. The filter had a spectral slope of 30 dB per octave for frequencies 100 Hz from the edges of the flat bandpass region and 50 dB per octave for frequencies farther away from the passband edges. The harmonic rank of the complex tones was varied by changing f_0. One pure-tone condition with an A-tone frequency of 2000 Hz and five complex tone conditions with A-tone f_0s of 80, 100, 150, 250, and 500 Hz were tested. The B-tone f_0 was always higher than that of the A-tones. Six B-tone f_0s were used with each A-tone f_0, resulting in 36 conditions. The tones had an overall sound pressure level (SPL) of 80 dB and a threshold equalising noise (TEN) with a level of 55 dB SPL/ERN_n was used to mask combination tones and to limit the audibility of stimulus components falling in the filter skirts. An uncorrelated TEN with a level of 25 dB SPL/ERN_n was presented to the other ear.

Nine normal-hearing participants (four female) with audiometric thresholds ≤ 20 dB hearing level (HL) were tested. The participants were between 21 and 28 years of age and had no musical training. The participants were instructed to try to hear out one stream as separate from the other and to indicate via a keyboard key press when their perception changed between one and two streams.

The stimuli were sampled at 44100 Hz and played via a Fireface UXC sound card (RME, Haimhausen, Germany) through Sennheiser HD650 headphones (Wedemark, Germany) in a sound-attenuating booth. The tone stimuli were presented monaurally to the ear that had the lowest mean audiometric threshold across 2, 3, and 4 kHz. Each condition was tested three times in each of 12 blocks. Each participant was trained for at least one two-hour session. The amount of training needed for each participant and the number of sessions was determined based on the mean of the standard deviations calculated for the arcsine-transformed proportions across all trials within a block for each condition. If the value of this measure was greater than 0.2 in one of the three blocks tested within a session, that session was repeated. However, the results were included here if a value greater than 0.2 occurred for only a single block when it was preceded by at least four blocks that had values less than 0.2.

Results

The individual results are shown in Fig. 2; the percentage of time that a given stimulus was perceived as segregated is plotted as a function of the A-tone f_0. Symbols indicate the f_0 difference, Δf_0. The results show large variability across participants. However, there are some general tendencies. The tendency to hear stream segregation usually increased with increasing f_0 difference between the A and B tones. Furthermore, pure tones and complex tones with a high f_0 (more low harmonics) were generally more likely to be perceived as segregated than complex tones with a low f_0.

The effect of f_0 was tested using a one-way within-subjects ANOVA based on a measure of the overall score across all differences in f_0 between the A and B tones after arcsine transformation of the percent scores (in proportions). This measure is called the “normalised segregation score” and its value varies from 0 (no segregation) to 1 (complete segregation). There was a significant effect of f_0 [$F(1,5) = 38.4$, $p < 0.001$]. Bonferoni-corrected pairwise comparisons showed that there were significant differences between the scores for the conditions with f_0s of 500 Hz and 150 Hz [$p = 0.004$], 2000 Hz and 150 Hz [$p = 0.004$], 2000 Hz and 100 Hz [$p < 0.001$], 2000 and 80 Hz [$p < 0.001$], 500 Hz and 100 Hz [$p = 0.002$], 500 Hz and 80 Hz [$p = 0.001$], 250 Hz and 150 Hz [$p = 0.022$], 250 Hz and 100 Hz [$p = 0.0052$], 250 Hz and 80 Hz [$p = 0.0038$], and 150 Hz and 80 Hz [$p = 0.0039$].

EXPERIMENT 2: F_0DLS

Method

Each trial contained three successive tones, two with a base f_0 (or frequency) and one with a higher f_0 or frequency. Each tone had a duration of 500 ms and each was temporally centred in a 700-ms TEN. The tones were separated by a 400-ms gap

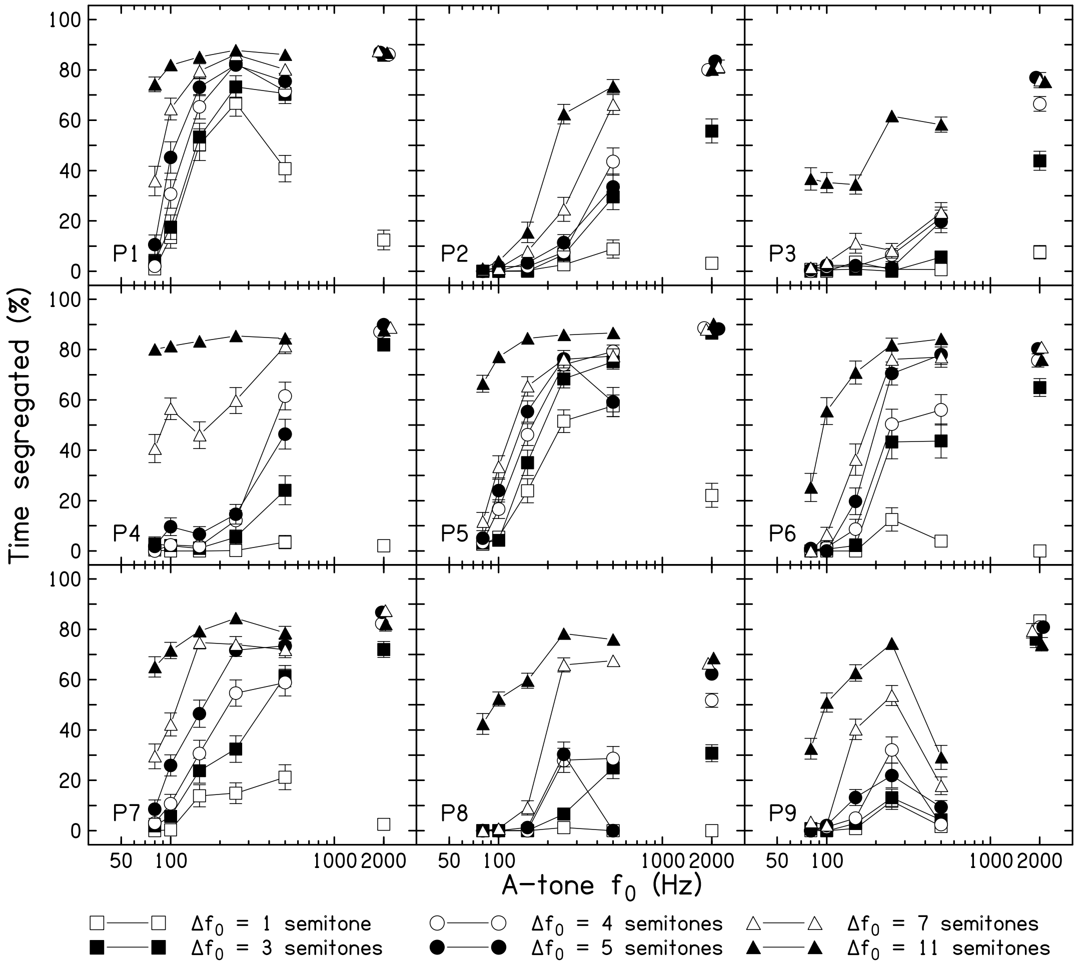


Fig. 2: Percentage of time that participants indicated that they perceived two streams. The mean and standard errors are shown for each participant. Δf_0 is the difference in f_0 between the A and B tones.

The spectrum and level of the tones and the TEN were the same as for experiment 1. The base f_0s of the complex tones were 80, 100, 150, 250, and 500 Hz and the frequency of the pure tone was 2000 Hz, the same as for the A-tones in experiment 1. The participant was instructed to identify the tone with the higher f_0 or higher frequency. A three-interval three-alternative forced-choice weighted up-down adaptive procedure was used to track the 70% correct point on the psychometric function (Kaernbach, 1991). The f_0DL was estimated as the geometric mean of the f_0 differences at the last six reversal points.

As in the study of Bernstein and Oxenham (2006a), the base f_0 was roved over the range ± 5% across trials (uniform distribution) to encourage the participants to listen to the current stimulus instead of comparing the stimulus to the memory of previous stimuli. Also, the levels of the tones were roved within each trial (uniform distribution with a range of ± 2.5 dB) to reduce any loudness cues.

Data are presented for five of the nine participants from experiment 1 since f_0DL measurements have currently not been completed for the remainder. Each condition was tested five times. The participants were trained for one two-hour session and runs were repeated if the standard deviation across the reversal points used to estimate the DL was greater than 0.25.

Results

The f_0DLs are shown in Fig. 3. The f_0DLs varied across participants and were markedly larger for participant nine than for the other participants. Thresholds tended to be lowest (best performance) for the pure tone (2000 Hz) and for the highest f_0 for the complex tones. The f_0DLs were roughly constant for f_0s up to 150 Hz, but decreased somewhat for the f_0 of 250 Hz, even though the lowest harmonic in the passband for the 250-Hz f_0 was the 8th, and this would have been barely, if at all, resolved. This is consistent with the idea that harmonic rank rather than resolvability governs f_0 discrimination (Bernstein and Oxenham, 2003).

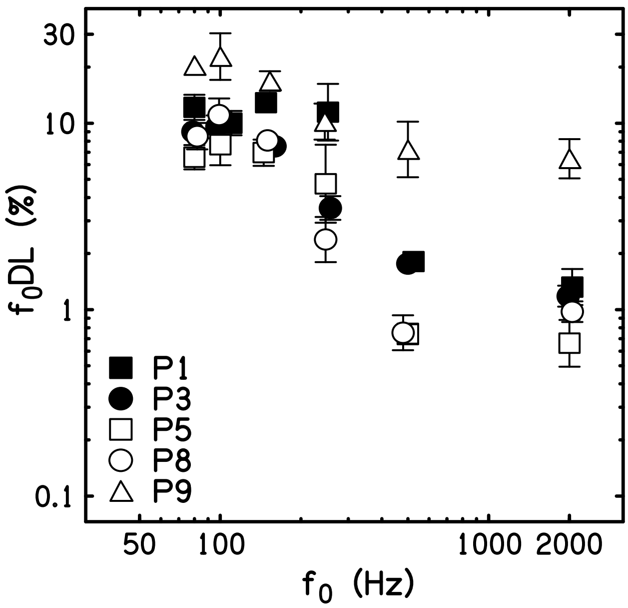


Fig. 3: f_0 discrimination thresholds. Means and standard errors are shown for each participant.

DISCUSSION

The results showed that stream segregation can occur for complex tones with only high harmonics, consistent with results from earlier studies (Vliegen *et al.*, 1999; Vliegen and Oxenham, 1999). However, Vliegen and Oxenham (1999) reported that subjective stream segregation, as measured in our experiment 1, was similar for pure tones, complex tones with low (resolved) harmonics and complex tones with only high (unresolved) harmonics. This contrasts with the results from the present study. Their measurements were made without a noise masker, and so may have been influenced by combination tones, especially for the complex tones with only high harmonics. They also measured stream segregation in the presence and absence of a background noise that would mask combination tones for two conditions, pure tones and complex tones with only high harmonics. Stream segregation seemed to be greater for the pure tones than for the complex tones (as observed in the present study), but this was not discussed in their paper. Furthermore, both the previous and the present study showed large variability across participants, which may also help explain the difference across studies.

Figure 4 shows the normalised segregation score (experiment 1) plotted against the log-transformed f_0DL (experiment 2), for each participant and each f_0 (open symbols). The figure also shows the means across participants (filled squares).

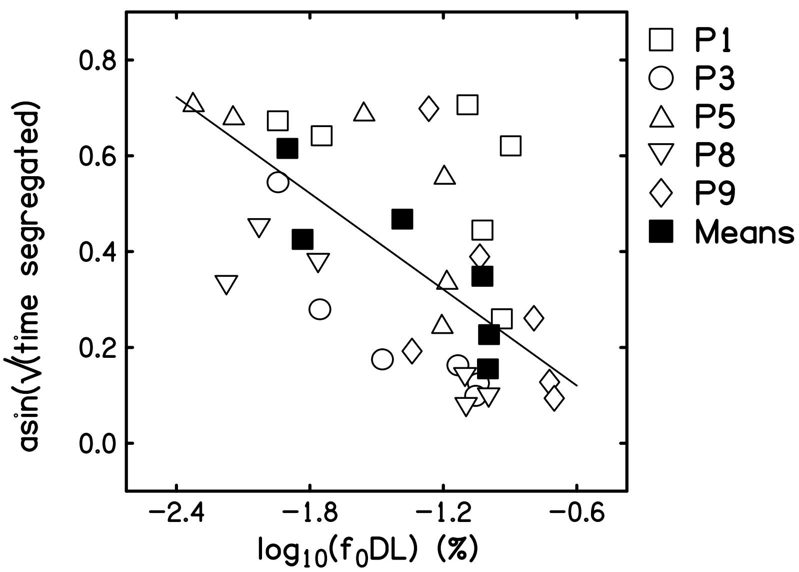


Fig. 4: Arcsine-transformed normalised stream segregation estimate as a function of the log-transformed f_0DL for each participant and the mean across participants (solid squares). The line is a least-squares fit to the mean data.

For the mean across participants, there was a significant Pearson correlation between the normalized segregation score and the f_0DLs [$r = -0.84$, $p = 0.034$], indicating that stream segregation is more likely for tones with small f_0DLs. For the individual participants, the correlations were $r = -0.50$ [$p = 0.31$], $r = -0.89$ [$p = 0.017$], $r = -0.76$ [$p = 0.08$], $r = -0.92$ [$p = 0.01$], and $r = -0.60$ [$p = 0.21$] for P1, P3, P5, P8, and P9, respectively. All correlations were negative, confirming that small f_0DLs, indicating high pitch salience, are associated with greater stream segregation.

CONCLUSIONS

There was a significant effect of harmonic rank on the tendency for a sequence of complex tones to be heard as two streams. More stream segregation occurred for complex tones with resolved harmonics than for complex tones with unresolved harmonics. Also, stream segregation was greater for complex tones that led to low f_0DLs, suggesting that good f_0 discrimination is associated with greater streaming.

ACKNOWLEDGEMENTS

This research was supported by the Oticon Centre of Excellence for Hearing and Speech Sciences (CHeSS). BCJM was supported by the Rosetrees Trust.

REFERENCES

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

Bernstein, J.G. and Oxenham, A.J. (**2006a**). "The relationship between frequency selectivity and pitch discrimination: effects of stimulus level," J. Acoust. Soc. Am., **120**, 3916-3928.

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.

Kaernbach, C. (**1991**). "Simple adaptive testing with the weighted up-down method," Percept. Psychophys., **49**, 227-229.

Moore, B.C.J. and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta Acust., **88**, 320-333.

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.

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

An objective measurement of TMTF by using envelope of cABR

TAKASHI MORIMOTO[1,*], GAKU YABUSHITA[1], YOH-ICH FUJISAKA[1], TAKESHI NAKAICHI[1], YASUHIDE OKAMOTO[2,3], AYAKO KANNO[2,3], SHO KANZAKI[3], AND KAORU OGAWA[3]

[1] *Department of Hearing Aid Algorithms Development, RION Co., Ltd., Tokyo, Japan*

[2] *Department of Otorhinolaryngology, Inagi Municipal Hospital, Tokyo, Japan*

[3] *Department of Otorhinolaryngology, School of Medicine, Keio University, Tokyo, Japan*

The temporal modulation transfer function (TMTF) has been proposed as a means for estimating temporal resolution. There are several problems that need to be overcome in the measurement of TMTF. For example, the threshold may be misjudged due to a lack of concentration by the measurement person as many judgment efforts are required, and the measurement task may be misunderstood due to limited language recognition ability. An appropriate objective measurement method is needed to avoid interference in the measurement process. We focused on cABR (auditory brainstem response to complex sounds) for objective measurement of TMTF because cABR faithfully represents several temporal acoustical features, including the envelope component of complex sound stimuli. The results for the temporal characteristic of cABR using SAM noise as complex sound stimuli can be used as an objective measurement of TMTF because the degree of cABR fluctuation was found to be related to the modulation depth of the stimuli and might be related to the modulation detection thresholds derived from TMTF.

INTRODUCTION

Some hearing-impaired people have reduced temporal resolution. Zeng *et al.* (1999) said that it is difficult for listeners with reduced temporal resolution to understand speech because speech recognition depends on the detection of temporal cues. The temporal modulation transfer function (TMTF) was proposed as a means for measuring the ability to detect temporal resolution (Viemeister, 1979). The TMTF means the threshold for detecting the amplitude modulation depth as a function of modulation frequency. Bacon and Viemeister (1985) reported that the modulation detection thresholds for hearing-impaired listeners were higher than those for normal-hearing listeners. Use of the TMTF in clinical diagnosis could make it possible to describe the auditory characteristics of hearing impaired patients more precisely, and this information might be useful in the fitting of hearing aids and be applicable to new

*Corresponding author: t-morimoto@rion.co.jp

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

hearing instrument algorithms.

However, application of the TMTF to clinical diagnosis is difficult because its measurement requires seven modulation detection thresholds. Its measurement thus usually requires many subjective trials that take about 40 minutes in total. A simplified TMTF measurement method (S-TMTF), which takes about 10 minutes, was proposed (Morimoto *et al.*, 2013) to reduce the measurement time. Using S-TMTF reduces the total measurement time by about 75 %.

Even though the time has been greatly reduced, much effort is still required to make the many judgments required to measure the modulation detection thresholds. The effort required to make these judgments may affect the measurement person's concentration, causing him or her to make some misjudgments. In addition, it is difficult to measure the TMTF for infants and foreign nationals due to their inability to comprehend the task accurately. This means that an appropriate objective measurement, such as auditory steady-state response or auditory brainstem response, which are established as pure tone audiometry for objective measurement, is needed to avoid interference in the measurement process.

To establish an objective measurement of TMTF, we focused on cABR (auditory brainstem response to complex sounds) as it faithfully represents several acoustical features, including the envelope component fluctuation of complex sound stimuli (Aiken and Picton, 2008). In this paper, two hypotheses are presented and validated through the measurement of cABR. This validation means that it may be possible to estimate the threshold of temporal resolution by using cABR.

Hypothesis 1 : The degree of fluctuation in cABR varies with the modulation depth of the stimuli.

Hypothesis 2 : The fluctuation in cABR disappears as the modulation depth of the stimuli approaches the threshold for the measurement subject.

TEMPORAL MODULATION TRANSFER FUNCTION

Generally, a sinusoidal amplitude modulated broadband noise (SAM noise) is used in the measurement of TMTF to estimate the modulation detection thresholds for each modulation frequency. Figure 1 shows an actual experimental data of TMTF for normal-hearing and hearing-impaired listeners. The modulation detection threshold is often measured using modulation frequencies of 8, 16, 32, 64, 128, 256, and 512 Hz (Eddins, 1993; Shen and Richards, 2013). These thresholds are expressed in decibels as $20\log_{10}(m)$, where m is the modulation depth parameter. If m equals 1.0, the signal is 100%. If m is 0.5 or 0.1, the modulation depth is expressed as -6 dB (50%) or -20 dB (10%), respectively. The modulation detection thresholds are almost constant from a modulation frequency of 8 Hz to about 50 Hz. Above 50 Hz, the thresholds increase at a rate of about 3 dB per octave of modulation frequency (Bacon and Viemeister, 1985).

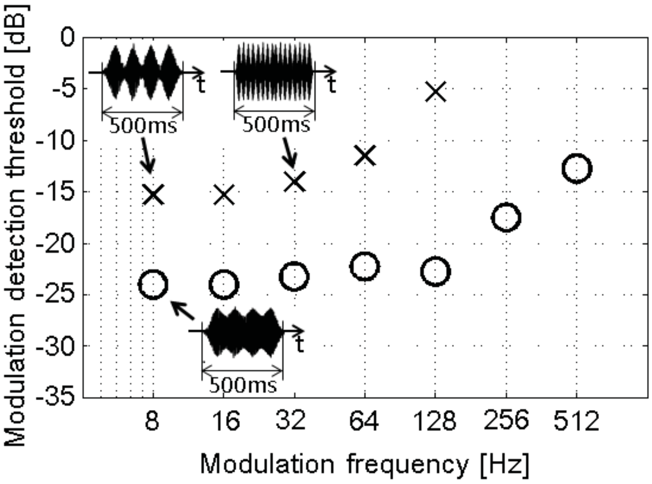


Fig. 1: Example of TMTF measurement data. Circles represent results for normal-hearing listeners, and x-marks represent results for hearing-impaired listeners. Waveforms show an image of sinusoidal amplitude modulated noise for each graph space.

The modulation detection thresholds for hearing-impaired listeners increase more than that of normal-hearing ones, as shown in Fig. 1. This difference reflects the degradation in their ability to recognize temporal resolution (Zeng *et al.*, 1999, 2005). Therefore, it is difficult for hearing-impaired listeners to measure the modulation detection threshold at high modulation frequencies. In contrast, it is easy for them to detect the fluctuation from 8 to 50 Hz because the modulation detection thresholds in this range remain low, as shown in Fig. 1. The modulation detection threshold should thus be measured at 8 Hz because it can be measured even if the subject is hearing impaired.

EXPERIMENT

SAM noises in our cABR experiments were also used as stimuli in the measurement of TMTF in order to validate our two hypotheses. The measured cABR was compared with the characteristics of the stimuli.

Subjects

Seven normal-hearing subjects participated. They ranged in age from 24 to 31 years. The stimuli were presented to their right ear. The subjects had hearing thresholds better than 15 dB HL in the tested ear at all audiometric frequencies from 125 to 8000 Hz.

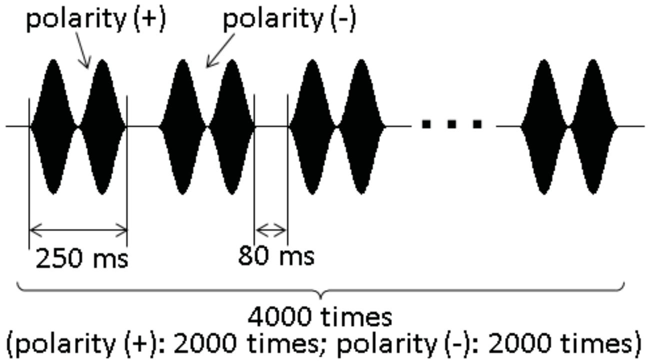


Fig. 2: Example of stimulus waveform with 0-dB modulation depth.

Stimuli and equipment

The stimuli were three SAM noises. The carriers were maximum length sequence (Mseq) noises. The modulation frequency was 8 Hz because it is easy to detect fluctuations at 8 Hz, as mentioned above. The noise duration was 250 ms, including 2.0-ms rise/fall cosine ramps, and the inter stimulus interval (ISI) was 80 ms. The stimuli consisted of alternating condensation and rarefaction polarities. The modulation indices of the stimuli were 0, -5, and -10 dB. A stimulus waveform example with a 0-dB modulation depth is shown in Fig. 2.

The stimuli were generated and presented via MATLAB and delivered through a 16-bit digital-to-analog converter (OCTA-CAPTURE, Roland) and headphone amplifier (A20, Beyerdynamic). The stimuli were presented to the test ear through an insert earphone (ER-3A, Etymotic Research) at an intensity of 80 dB SPL.

Recording and data analysis of cABR

Continuous electroencephalographic (EEG) signals were acquired with a data acquisition system (MP150, Biopac Systems) from Cz-to-ipsilateral earlobe with the forehead as the ground and digitized at 20,000 Hz. All recordings were made with electrodes (Cz and earlobe: EL258S; forehead: EL258; Biopac Systems) (impedance < 5 kΩ). A bandpass filter (from 70 to 2000 Hz) was applied to the recordings to isolate the brain-stem response frequencies. The EEG signals were then divided into 330-ms epochs (40-ms pre-stimulus onset to 290-ms post-stimulus onset). An artifact criterion of ± 20 μV was applied to reject epochs that contained myogenic artifacts.

The stimuli were presented in alternating polarities, allowing for the creation of responses comprised of both the added and the subtracted of the two polarities. When the added response was created, the envelope component of the response was enhanced; conversely, when the subtracted response was created, the temporal fine structure component was enhanced (Aiken and Picton, 2008). To investigate the fluctuations in the cABR, we used added responses with two polarities. We calculated

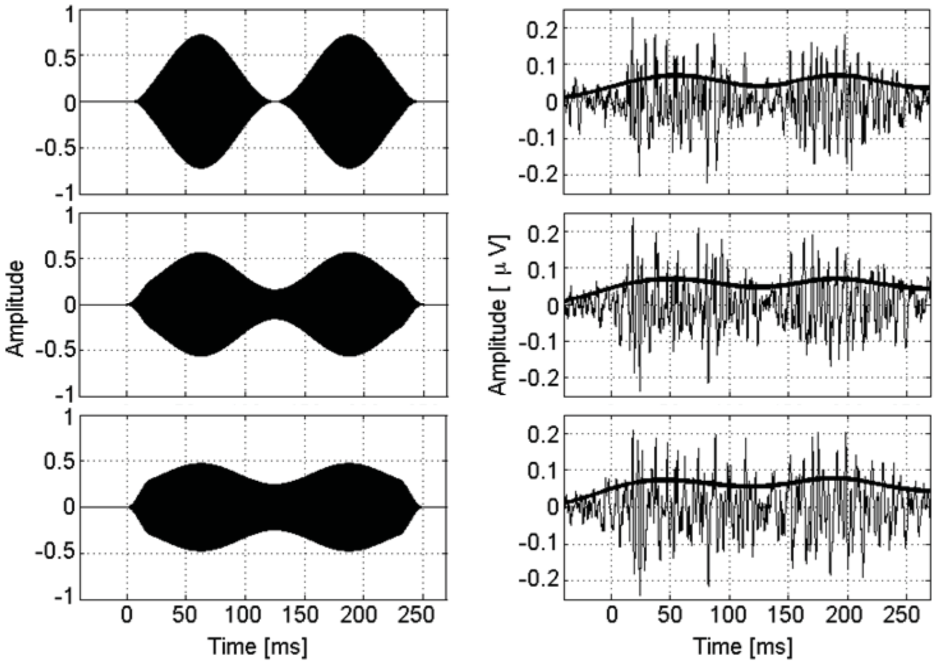


Fig. 3: Example of cABR measurements. Upper, middle, and lower panels on left show waveforms of SAM noise for modulation depths of 0, −5, and −10 dB, respectively. Those on right show added responses for modulation depths of 0, −5, and −10 dB, respectively. In the right panels, thin lines represent added responses and heavy lines represent the envelope of added responses.

the envelope component of the added response by using the Hilbert transform and a lowpass filter with a cutoff frequency of 10 Hz. The difference between the minimal (from 62.5- to 187.5-ms post-stimulus onset) and maximal (from 125- to 250-ms post-stimulus onset) of envelope component of the added response was used as the degree of fluctuation.

RESULTS

Figure 3 shows an example of our cABR measurements. The figures on the left show the waveforms for each stimulus (modulation depths of 0, −5, −10 dB), and the ones on the right shown the added responses from the cABR. The degree of fluctuation in the added responses attenuated as the modulation depth decreased. The transition in the degree of fluctuation for each subject, shown in Fig. 4, was confirmed in order to investigate the attenuation tendency. The heavy solid line represents the average degree of fluctuation, and the error bars represent the standard deviation for each modulation depth. The average of degree of fluctuation exhibited a similar attenuation tendency. In addition, the degree of fluctuation converged to a certain value at a modulation depth of approximately −10 dB.

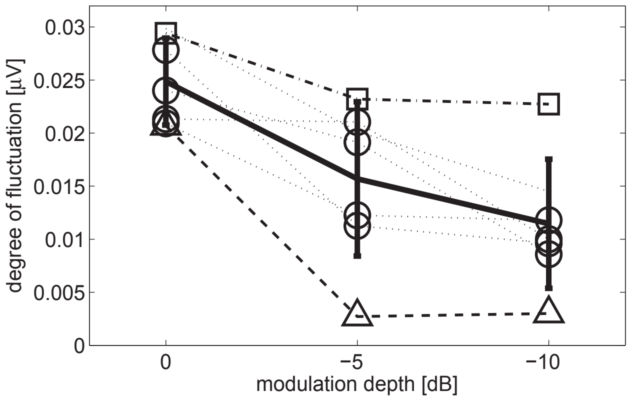


Fig. 4: Transition in degree of fluctuation as function of modulation depth for each subject and average. Dotted lines, dash-dot line, and dashed line show results for each subject. Squares and dash-dot line represent results for the subject with the largest fluctuation, and triangles and dashed line represent results for the subject with the smallest fluctuation at modulation depths of −5 and −10 dB. Heavy solid line represents average of degree of fluctuation, and error bars represent standard deviations for each modulation depth.

DISCUSSION

The degree of fluctuation was attenuated as the modulation depth decreased for most of the subjects. In addition, the degree of fluctuation converged to a certain value at a modulation depth of approximately -10 dB. These results support hypothesis 1, i.e., the envelope component of the added response of cABRs from two polarity stimuli varies depending on the fluctuation of the stimuli.

Considering hypothesis 2, if the fluctuation in cABR disappears as the modulation depth of the stimuli approaches a modulation detection threshold, the threshold can be estimated from the cABR. Therefore, the modulation detection thresholds derived

Subject	Modulation detection threshold [dB]	Marker used in Fig. 4
A	−27.5	□
B	−26.5	○
C	−26.3	○
D	−25.8	○
E	−25.3	○
F	−25.2	○
G	−23.8	△

Table 1: Modulation detection thresholds at modulation frequency of 8 Hz.

from the TMTF measurement at a modulation frequency of 8 Hz were measured for comparison with the degree of fluctuation in cABR. The measured modulation detection thresholds are shown in Table 1.

The subject with the largest fluctuation had the highest threshold while the one with the smallest fluctuation had the lowest threshold. This indicates that the degree of fluctuation in cABR is related to the threshold derived from the TMTF measurement. It might therefore be possible to estimate the modulation detection threshold for each subject from the convergent value, although there was some difference between the threshold and the convergent value derived from our cABR measurements.

REFERENCES

Aiken, S.J. and Picton, T.W., (**2008**). "Envelope and spectral frequency-following responses to vowel sounds," Hear. Res., **245**, 35-47.

Bacon, S.P. and Viemeister, N.F. (**1985**). "Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners," Audiology, **24**, 117-134.

Eddins, D.A. (**1993**). "Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region," J. Acoust. Soc. Am., **93**, 470-479.

Morimoto, T., Nakaichi, T., Harada, K., Okamoto, Y., Kanno, A., Kanzaki, S., and Ogawa, K., (**2013**). "A simplified measurement method of TMTF for hearing impaired listeners," In Proc. ISAAR 2013: *Auditory Plasticity - Listening with the Brain*. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Eds. T. Dau, S. Santurette, J.C. Dalsgaard, L. Tranebjerg, T. Andersen, and T. Poulsen. ISBN 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Shen, Y. and Richards, V.M. (**2013**). "Temporal modulation transfer function for efficient assessment of auditory temporal resolution," J. Acoust. Soc. Am., **133**, 1031-1041.

Viemeister, N.F., (**1979**). "Temporal modulation transfer function based upon modulation thresholds," J. Acoust. Soc. Am., **66**, 1364-1380.

Zeng, F.G., Oba, S., Grade, S., Sininger, Y., and Starr, A. (**1999**). "Temporal and speech processing deficits in auditory neuropathy," Neuroreport, **10**, 3429-3435.

Zeng, F.G., Kong, Y.Y., Michalewski, H.J., and Starr, A. (**2005**). "Perceptual consequences of disrupted auditory nerve activity," J. Neurophysiol., **93**, 3050-3063.

Discrimination scores used in medico-legal assessment of hearing disabilities

ELLEN RABEN PEDERSEN[1,*], PETER MØLLER JUHL[1], RANDI WETKE[2,3], AND TURE ANDERSEN[2,3]

[1] *The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, Denmark*

[2] *Department of Audiology, Odense University Hospital, Odense, Denmark*

[3] *Institute of Clinical Research, University of Southern Denmark, Odense, Denmark*

Objective: Examination of Danish data for medico-legal compensations regarding hearing disabilities. The purposes are: 1) to investigate whether discrimination scores (DSs) relate to patients' subjective experience of their hearing and communication ability, 2) to compare DSs from different discrimination tests (auditory/audio-visual perception and without/with noise), and 3) to discuss the handicap scaling used for compensation purposes in Denmark. *Design:* Data for 466 patients from a 15 year period (1999-2014) were analysed. From the data set 50 patients were omitted due to suspicion of exaggerated hearing disabilities. *Results:* The DSs were found to relate well to the patients' subjective experience of their speech perception ability. As expected the least challenging test condition (highest DSs) was the audio-visual test without an interfering noise signal, whereas the most challenging condition (lowest DSs) was the auditory test with noise. The hearing and communication handicap degrees were found to agree, whereas the measured handicap degree tended to be higher than the self-assessed handicap degree. *Conclusions:* The DSs can be used to assess patients' hearing and communication abilities. In order to get better agreements between the measured and self-assessed handicap degrees it may be considered to revise the handicap scaling.

INTRODUCTION

Even though noise-induced hearing loss is a significant work related injury in many industrialized countries, there is no standard way of assessing a person's hearing disabilities regarding medico-legal compensation purposes across countries. In Denmark an ENT doctor has to fill in a special medical examination form. The form is filled in for all kinds of medico-legal assessments of hearing disabilities regardless of whether the hearing disability is work related, due to an accident, or a treatment

*Corresponding author: erpe@mmmi.sdu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

injury. Thus, the form is used by both the Danish National Board of Industrial Injuries and Danish private insurance companies.

In order to assess the patient's hearing disability, and thereby the amount of financial compensation to be paid, the form uses the scaling of the hearing handicap (HH) and the communication handicap (CH) proposed by Salomon and Parving (1985). The HH degree is based on the patient's ability to comprehend speech auditorily without the help from visual cues, whereas the assessment of the CH degree is based on the patient's audio-visual speech comprehension. Both the HH and the CH are determined 'self-assessed' by an interview and 'measured' by results from discrimination tests. For both types of handicaps an overall degree is stated as the mean of the self-as-sessed and the measured degree.

This study analyses data from a large number of patient forms collected over a 15-year period (1999-2014). The research purposes of the study are:

1) to investigate whether discrimination scores (DSs) relate to patients' subjective experience of their hearing and communication ability.
2) to compare DSs from different discrimination tests (auditory/audio-visual perception, and without/with noise).
3) to discuss the handicap scaling used for compensation purposes in Denmark.

METHODS

Patients

The medical examinations for 466 patients form the basis of this study. The patients were referred to medico-legal examinations due to hearing disabilities mainly caused by work related noise exposure. For a minor part of the patients the hearing difficulties were due to an accident or a treatment injury. From the data set 50 patients were omitted due to suspicion of exaggerated hearing disabilities. Thus, the analyses include data for 416 patients (376 men and 40 women, aged 10-80 years with an average age of 54 years).

Interview

In the medical examination form the HH and the CH degrees are determined 'self-assessed' by an interview containing three questions:

QI Are you able to understand speech one-to-one in a quiet environment?

QII Are you able to understand speech one-to-one despite background noise, speech, music or other everyday noises?

QIII Are you able to follow a group conversation at home?

For each patient the three questions were posed twice, first regarding auditory perception and then regarding audio-visual perception. For patients having hearing aids the questions were posed two additional times. The answers regarding hearing aid use are in this study used for the handicap scaling only (i.e., regarding research purpose 3).

The patient answered each question with ‘Yes’ or ‘No’. If the patient answered Yes with a reservation, the answer was recorded as ‘(Yes)’. Thus, a (Yes)-answer refers to the patients being able to perceive speech but only under certain circumstances, e.g., depending on the character of the voice, the noise type or the placement of the talker. The doctor assesses whether a (Yes)-answer is interpreted as Yes or No.

Discrimination tests

In the medical examination form the HH and the CH degrees are determined ‘measured’ by results from discrimination tests. The discrimination tests were performed using the Dantale word lists and Dannoise (Elberling *et al*., 1989). They were performed for two listening conditions:

- Without interfering noise (in analogy with QI), speech level = 65 dB SPL
- With interfering noise (in analogy with QII), signal-to-noise ratio = 0 dB (both the speech signal and the noise signal were 65 dB SPL)

For each patient the two discrimination tests were performed twice, first regarding auditory perception and then regarding audio-visual perception. For patients having hearing aids the two discrimination tests were performed two additional times. The discrimination tests performed using hearing aids are in this study used for the handicap scaling only (i.e., regarding research purpose 3). The result of each test is stated as the discrimination score (DS), i.e., the percentage of correctly answered words.

Handicap scaling

The degrees of the HH and the CH were assessed based on the patient’s answers to the questions in the interview as well as on the results from the discrimination tests. Thus, for each person four handicap degrees were determined: HH self-assessed, CH self-assessed, HH measured, and CH measured. The handicap degree classification was: 0 = no handicap, 1 = slight handicap, 2 = mild to medium handicap, 3 = considerable handicap, 4 = severe handicap, and 5 = total handicap.

Table 1 shows how each of the four handicaps were assessed. The answers to the three questions are in the columns marked QI, QII, and QIII, whereas the discrimination scores marked DSI and DSII are for the conditions without and with an interfering noise, respectively. The abbreviation A is for auditory perception (HH), whereas AV is for audio-visual perception (CH). The column “HA use” refers to whether the questions were answered regarding hearing aid/the discrimination tests were performed with hearing aid. Note that setting the handicap degree using the table is not always unambiguous.

RESULTS

Fig. 1 shows for all three questions that the percentages of Yes-answers are larger for the audio-visual than for auditory perception. It also shows that the percentage of Yes-answers is largest for question I representing good listening conditions and smallest for question III representing poor listening conditions.

Degree	HA used	Self-assessed						Measured			
		HH			CH			HH		CH	
		QI A	QII A	QIII A	QI AV	QII AV	QIII AV	DSI A	DSII A	DSI AV	DSII AV
0	No	Yes	Yes	No	Yes	Yes	Yes	Normal	> 80%	Normal	> 80%
1	No	Yes	No	No	Yes	Yes	No	Normal	< 60%	Normal	> 60%
2	Yes	Yes	Yes	No	Yes	Yes	Yes	90-95%	> 60%	Normal	> 90%
3	Yes	Yes	No	No	Yes	Yes	No	> 60%	< 60%	> 90%	> 60%
4	Yes	No	No	No	Yes	No	No	< 60%	→ 0%	> 60%	< 60%
5	Yes	No	No	No	No	No	No	0%	0%	< 60%	0%

Table 1: Hearing handicap and communication handicap scaling; both self-assessed and measured. See the text for details. The table is a merged reproduction of Tables I, V, and VI in Salomon and Parving (1985).

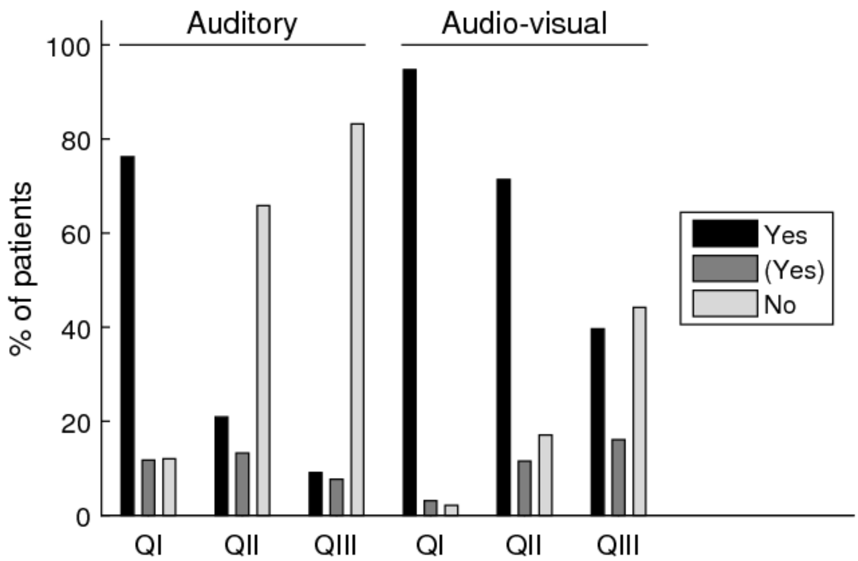


Fig. 1: Percentages of Yes-, (Yes)-, and No-answers to questions I, II, and III. Each question was posed regarding auditory and audio-visual perception, respectively.

Fig. 2 shows the DSs obtained for patients who have answered Yes, (Yes), and No to question I and II, respectively. DSs across the different answers are also shown. The DSs for each of the four conditions are selected as to reflect the listening situation of the question, e.g., for question I auditory the DSs are measured auditory without an interfering noise signal. For all four conditions the medians of DSs are found to be statistically significantly different for all three answers at the five percent level. Additionally, the medians across the different answers (market with squares) are found to be statistically significantly different for all four conditions.

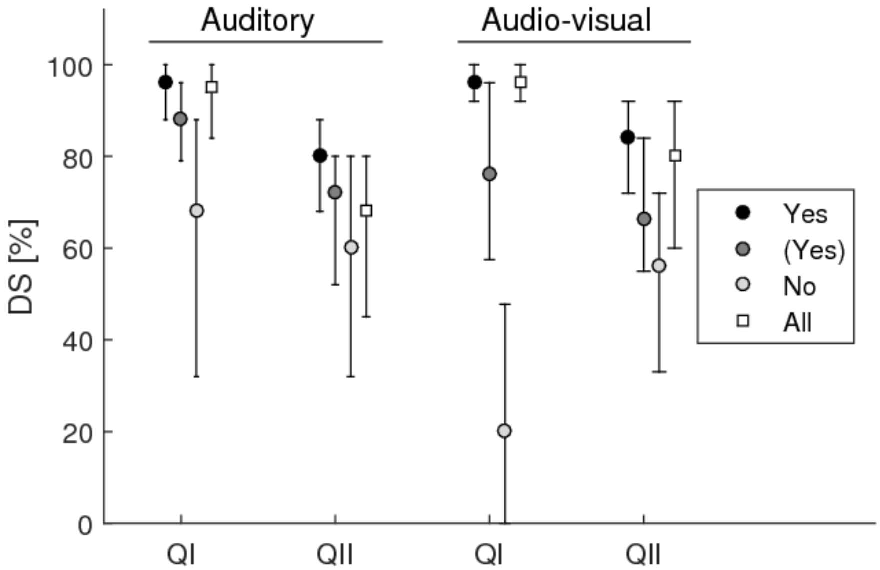


Fig. 2: Medians of the DSs for patients grouped as to their answers to question I and II. Data for the DS across the different answers are marked 'All'. The lower and upper ends of the error bars represent the 25th and 75th percentile, respectively.

Connections between DSs across the different test conditions are shown in Fig. 3. For each subfigure the most challenging test condition (of the two in concern) is on the x-axis, whereas the least challenging test condition is on the y-axis. For the persons obtaining low DSs in the most challenging condition there are relative large individual differences in the DS enhancement as the listening condition improves; especially for the test conditions in Fig. 3, a) and b). Note that, as the DS scale is censored, it is not possible to score below 0% or above 100%, termed the floor effect and the ceiling effect, respectively. Thus, persons obtaining DSs of 100% in the most challenging test condition cannot get higher scores in the least challenging test condition.

Fig. 4 shows the connection between the HH and CH degree as well as the connection between the self-assessed and the measured handicap degree. The handicap degrees were assessed using the scaling reported in Table 1. As seen a large number of the patients are assigned handicap degrees of 0 or 1. For both the self-assessed and the measured handicaps most patients obtain HH and CH degrees which are identical or differ by one degree of handicap from one another, see Fig. 4, a) and b). This agrees with the finding in Salomon and Parving (1985). The agreement between the HH and CH degrees indicates that the handicap scaling compensates for the fact that speech comprehension is easier audio-visual than auditory. For some of the patients the measured handicap degree is higher and even up to four degrees higher than the self-assessed handicap degree, see Fig. 4, c) and d).

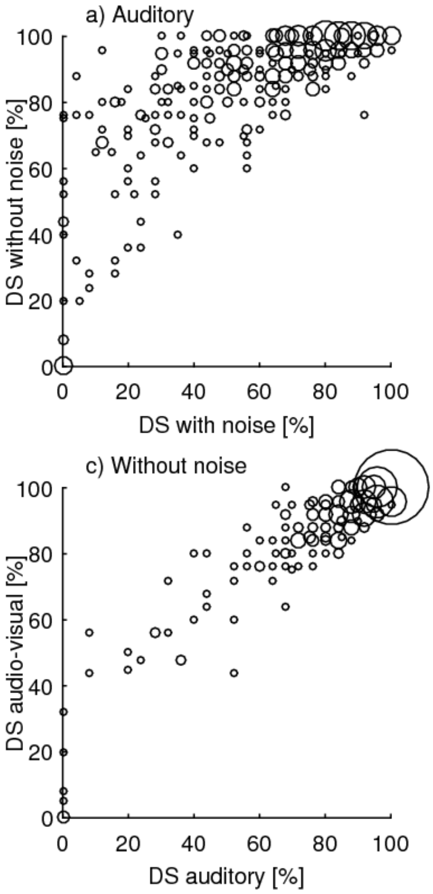


Fig. 3: Combinations of the DSs for the different discrimination test setups. The connection between the DSs with and without noise is in a) for the auditory perception and in b) for the audio-visual perception. The connection between the DSs auditorily and audio-visually is in c) for the test setup without noise and in d) for the test setup with noise. The bigger the dot (defined by the area), the more patients have obtained the same DS in the two tests in concern.

DISCUSSION

For all four test setups the patients who had answered Yes obtained the highest DSs (Fig. 2). The small variations in the DSs for the Yes-answers to question I for DS without noise can be explained by the ceiling effect, i.e., scores cannot go higher than 100%. However, lower scores can be achieved by changing the test setup, e.g., by lowering the level at which the words are played. For the discrimination tests performed with noise the DSs can be lowered by either lowering the SNR or by changing the interfering noise signal to one which is more difficult to distinguish from the speech signal.

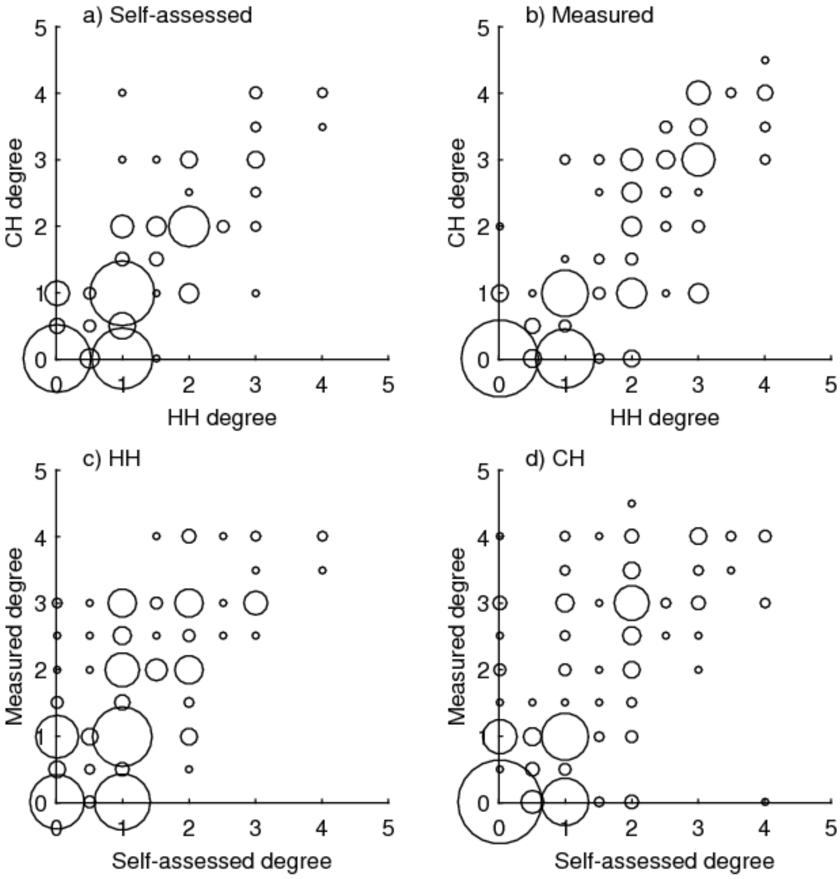


Fig. 4: Connections of the different handicap degrees (ordinal scaled values). The connection between the hearing handicap (HH) and the communication handicap (CH) is in a) self-assessed based on the patient's answer to the three questions and in b) assessed based on the DSs. The connection between the self-assessed and the measured handicap degrees is in c) for the HH, i.e., related to auditory perception and in d) for the CH, i.e., related to audio-visual perception. The bigger the dots (defined by the area), the more patients have obtained the same handicap degree.

If the handicap scaling is to be revised this should be adjusted, so that the measured and self-assessed handicap degrees are more consistent, i.e., either the self-assessed scale should be changed so a high self-assessed degree is easier to obtain or the measured scale should be changed so a high measured degree is harder to obtain. Since the HH and CH degrees correlate well, the adjustment should be made so that the proportion between the HH and CH degrees is kept fixed, for instance by applying the same adjustment to the measured degrees of both the HH and CH.

Furthermore, if the handicap scaling is to be altered, it should be framed unambiguously so that determination of the handicap degrees are uninfluenced by the experimenter's subjective evaluation, i.e., not as today where the measured handicap degree can fall outside a degree or into two degrees.

CONCLUSIONS

Data for the medical examination form filled in over a 15-year period were analysed. The data set includes data for 466 patients, from which 50 were omitted due to suspicion of having exaggerated their hearing disabilities. Analysing the data for the remaining 416 patients gave the following answers to the three research purposes listed in the introduction:

1) The DSs relate well to the patients' subjective experience of their speech perception ability. This was found for all four investigated test conditions.
2) The patients obtained higher DSs when the discrimination tests were performed without noise than with noise, and slightly higher when performed audio-visually than auditorily.
3) In order to get better agreements between the measured and self-assessed handicap degrees it may be considered to revise the scaling for either the HH or the CH. The handicap scaling should be framed unambiguously.

ACKNOWLEDGEMENTS

The authors thank Hilde Christin Wangen, Department of Audiology, Aarhus University Hospital for carrying out the majority of the audiological tests. The authors also thank Torben Poulsen, Centre for Applied Hearing Research, Technical University of Denmark for valuable discussions. This study was supported by the Oticon Foundation, grant number 12-1361.

REFERENCES

Elberling, C., Ludvigsen, C., and Lyregaard, P.E. (**1989**). "DANTALE: a new Danish speech material," Scand. Audiol., **18**, 169-175.

Salomon, G. and Parving, A. (**1985**). "Hearing disability and communication handicap for compensation purposes based on self-assessment and audiometric testing," Audiology, **24**, 135-145.

List of authors