SPEECH IN NOISE WORKSHOP

5-6 January 2017

Oldenburg, DE

Abstracts

The 9th Speech in Noise Workshop is organised by selfless volunteers from the University of Oldenburg, Germany:

- *Tim Jürgens*
- *Anna Warzybok*
- *Rebecca Carroll*
- *Melanie Zokoll*
- *Melanie Tipping*
- *Jana Müller*

Coordination: Etienne Gaudrain, Lyon Neuroscience Research Center, CNRS — Université Lyon 1, France.

Contact information: info@spin2017.de

The Speech in Noise Workshop is generously supported by:

Programme

Thursday, January 5

08:30 – 09:00	Registration and coffee
09:00 – 09:15	Welcome / Introduction
09:15 – 09:45	“Towards individual aided performance predictions for the Matrix Sentence Test” — Marc René Schädler — *University of Oldenburg, DE* ... *5*
09:45 – 10:15	“Semantic processing during speech-in-speech perception and links with executive functions” — Fanny Meunier — *CNRS, FR* ... *6*
10:15 – 10:45	Coffee break, (Foyer, hang up posters for both days)
10:45 – 11:15	“Assessment of cognitive load during listening and findings related to hearing-aid use” — Jana Besser — *Sonova, CH* ... *7*
11:15 – 11:45	“Spatial release from informational and energetic masking in bimodal and bilateral cochlear implant users” — Rene H. Gifford — *Vanderbilt University, US* ... *7*
11:45 – 12:15	**Recipient of the Colin Cherry Award 2016** “Assessment of spectral ripple discrimination in cochlear implant users: the untold story” — Alejandro Lopez Valdez — *Trinity College Dublin, IE* ... *8*
12:15 – 13:45	Lunch (Foyer)
13:45 – 14:45	**Keynote lecture** “A versatile view on the aging auditory system for auditory processing and speech perception” — Astrid van Wieringen — *ExpORL, KU Leuven, BE* ... *9*
14:45 – 15:00	Overview of session 1 posters
15:00 – 15:30	Coffee break
15:30 – 17:30	Poster session 1 ... *15*
19:00 – ...	Dinner at Beisserei (Poststrasse 1, Oldenburg)

Friday, January 6

08:30 – 09:00	Coffee
09:00 – 09:30	"The role of high-level processes for oscillatory phase entrainment to speech sounds" — Benedikt Zoefel — *Medical Research Council, Cambridge, GB* *10*
09:30 – 10:00	"Binaural intelligibility prediction for noisy and non-linearly processed speech" — Asger Andersen — *Oticon, DK* *11*
10:00 – 10:15	Overview of session 2 posters
10:15 – 10:30	Coffee break (Foyer)
10:30 – 12:30	Poster session 2 .... *15*
12:30 – 13:45	Lunch (Foyer)
13:45 – 14:15	"The impact of noise on speech processing over extended periods of time: A developmental perspective" — Johanna Barry — *MRC Institute of Hearing Research, University of Nottingham, GB* *11*
14:15 – 14:45	"Modeling speech intelligibility based on envelopes derived from auditory spike trains" — Christoph Scheidinger — *Danish Technical University, Copenhagen, DK* *13*
14:45 – 15:15	Coffee break (take down all posters)
15:15 – 15:45	"Silent speech interfaces" — Tanja Schultz — *University of Bremen, DE* *13*
15:45 – 16:15	"Context shapes neural speech coding: Effects of visual speech, prior auditory speech and listener's goals on cortical and subcortical speech processing" — Inga Schepers — *University of Oldenburg, DE* *14*
16:15 – 16:45	**Announcement of the Colin Cherry Award 2017 recipient, Discussion**

Thursday 5 January, 09:15—09:45

Towards individual aided performance predictions for the Matrix Sentence Test

M. R. Schädler, A. Warzybok, D. Hülsmeier, M. Buhl, S. Hochmuth, M. Zokoll-van der Laan, B. Kollmeier
CvO Universität Oldenburg, Germany

With world-wide increasing life expectancy and aging societies, (age-related) hearing disorders, their impact on interpersonal communication, and possible treatments are becoming a concern to many citizens. Consequently, there is an increasing demand for high-performance hearing algorithms that measurably improve the everyday communication experience of their users. The development of hearing aid algorithms and their individual fitting still strongly depend on costly empirical data from speech recognition tests, mainly due to the lack of suitable and accurate models of the speech perception of listeners with impaired hearing. In this context, a universal model of human speech recognition that (a) accounts for the most relevant individual factors of (impaired) hearing and their interaction with signal processing algorithms, (b) is able to predict the outcome of speech intelligibility measures in sufficiently realistic environments, and (c) does so without the need of any empirical data or knowledge that is not available to human listeners, is highly desirable. Unfortunately, such a model does not exist.

However, towards this goal, the simulation framework for auditory discrimination experiments (FADE) was proposed, which accurately predicted the outcome of the matrix sentence test in different languages in stationary and fluctuating noise conditions, and basic psycho-acoustic experiments for listeners with normal hearing. Because FADE simulates the whole speech recognition process, using an automatic speech recognition system, it does require neither empirical data nor knowledge that is not available to human listeners to predict the outcome of the matrix sentence test. Accepting the matrix sentence test in fluctuating noise conditions as sufficiently realistic, the remaining missing property of FADE is the adequate integration of signal processing deficiencies that are related to hearing impairment in order to correctly model their interaction with different signal pre-processing algorithms.

In this contribution the recent progress in extending FADE with signal processing deficiencies to model impaired hearing as well as the evaluation of predictions for individual benefits in SRT for listeners with impaired hearing using a range of binaural pre-processing algorithms, will be presented. While the benefit for the group of normal hearing listeners and the group of hearing impaired listeners can be predicted, the results indicate, that the audiogram is not sufficient to explain the individual speech recognition abilities of listeners with impaired hearing.

Semantic processing during speech-in-speech perception and links with executive functions

F. Meunier
CNRS, Nice, France

In this talk I will present a study in which we investigated the links between speech-in-speech perception capacities and four executive function components: response suppression, inhibitory control, switching and working memory. We used a cross-modal semantic priming paradigm using a written target word and a spoken prime word. Each prime word was implemented in only one of two overlapping pronounced sentences (cocktail party situation). Participants had to perform a lexical decision task on visual target words and simultaneously listen to only one of two pronounced sentences. The onset of the visual target presentation corresponded to the offset of the pronounced prime word. The attention of the participant was our first manipulated variable (noted attention): The prime was in the pronounced sentence listened to by the participant (attended condition) or in the ignored pronounced sentence (ignored condition). Our second manipulated variable (noted relatedness) was the semantic link between the written target word and the spoken prime word: The prime and target were semantically related or unrelated. In addition, we asked to participants to perform both an executive function task (the anti-saccade task of Bialystok et al., 2006, allowing the simultaneous measurement of switching cost, inhibitory-control cost and response-suppression cost) and working memory span task (extracted from WAIS IV). A correlation analysis was performed between the executive and priming measurements. Our results showed a significant interaction effect between attention and the relatedness variable. We observed a significant priming effect in the attended but not in the ignored condition. Only this last priming condition (ignored) was significantly correlated with the three executive measurements of the anti-saccade task. However, no correlation between priming effects and working memory scores was found. These results are in line, first, with the role of attention for semantic priming effect and, second, with the implication of executive functions in speech-in-noise understanding capacities. These findings allow to specify further the nature of the links between speech in noise comprehension and executive functions.

Assessment of cognitive load during listening and findings related to hearing-aid use

J. Besser
Sonova, Stäfa, Switzerland

The talk will give a short introduction into the concepts of listening effort and cognitive load during listening, which have received increasing attention in cognitive hearing research over the past decade. Different types of assessment methods have been applied to measure cognitive load during listening, i.e., self-report, behavioral, and physiological measures. The different types of assessments will be discussed regarding their individual (dis)advantages and some methodological considerations will be given. Furthermore, insights from some previous experimental studies will be presented with a focus on results related to amplification and signal processing. Potential areas of applying measures of cognitive load during listening in combination with hearing aids will be exemplified.

Spatial release from informational and energetic masking in bimodal and bilateral cochlear implant users

R. Gifford
Vanderbilt University, School of Medicine, Nashville, TN, USA

K. D'Onofrio
Vanderbilt University, Nashville, TN, USA

Spatially separating the target and masker(s) increases speech understanding in noise. While bilateral cochlear implant (CI) users may receive some benefit from spatial release from masking (SRM), the degree of benefit is lower and more variable than normal-hearing listeners. Poor transmission of temporal fine structure with current CI technology limits transmission of interaural time differences (ITDs)—a likely reason for reduced SRM. For bimodal CI users, access to fine structure from the non-CI ear may facilitate speech understanding in noise via glimpsing and/or segregation. Because real-world listening involves informational masking with various spatial separations between talkers and distracters, we aimed to investigate the interaction between two types of maskers (informational and energetic) and SRM in a group of bimodal and bilateral CI users with symmetric noise configurations. Our hypotheses were (1) CI recipients would exhibit less SRM than shown in the literature with asymmetric noise configurations, and (2) bimodal listeners would exhibit greater release from informational masking than bilateral CI users.

Speech understanding was assessed in the presence of speech (informational) and signal-correlated noise (energetic) maskers for 22 adult CI users (8 bilateral, 14 bimodal). A single female or male talker presented at 60 dBA originated from 0 degrees and the distracters (two different male talkers or signal-correlated noise) were positioned at either 0 degrees, 45 & 315 degrees, or 90 & 270 degrees. The signal-to-noise ratio (SNR) was determined individually to yield approximately 50% correct with the best CI alone. Results revealed that neither bimodal nor bilateral CI users exhibited SRM, and bimodal listeners exhibited greater release from informational masking than bilateral CI users, particularly for male talkers. Previous studies investigated SRM using asymmetric noise configurations allowing for better ear listening to dominate. In symmetric noise conditions, listeners rely on binaural cues, which are not generally available to CI users. For this reason, CI users with acoustic hearing preservation may fare better in symmetric noise configurations if they are able to use ITD cues.

Bimodal and bilateral CI users exhibit similar release from informational masking for a female target talker. When the target and masker were both male, bimodal listeners exhibited greater release from informational masking than bilateral CI users. Bimodal listeners may have been able to combine segregation and/or glimpsing—via fine structure in the non-CI ear—thereby increasing susceptibility to informational masking with the speech maskers.

Thursday 5 January, 11:45—12:15

Assessment of spectral ripple discrimination in cochlear implant users: the untold story

A. Lopez Valdes
Eriksholm Research Centre, Denmark | Trinity College Dublin, Ireland

M. Mc Laughlin
Trinity College Dublin, Dublin, Ireland

C. Simoes-Franklin, L. Flood, J. Smith, P. Walshe, L. Viani
Beaumont Hospital, Dublin, Ireland

R. B. Reilly
Trinity College Dublin, Dublin, Ireland

Introduction — In Ireland, following the introduction of the new born hearing screening programme in 2014, deaf children can have access to sound via bilateral CIs from as young as one year of age. Therefore, there is an increasing need for objective non-linguistic methods to asses CI performance. Previously, we developed a novel method to acquire cortical auditory evoked potentials (CAEPs) from CI users, significantly minimizing the effect of the electrical artefact. This opened the possibility to implement non-linguistic spectra ripple discrimination (SRD) assessment in an objective manner by contrasting the CAEP response to inverse ripple stimuli. Here, we evaluate newly implanted CI users, longitudinally, and demonstrate the clinical applicability of objective non-linguistic assessment of SRD as well as the relative time course of the development of SRD abilities, and speech perception. This provides great potential for objective screening and customisation of CI devices across ages of the population, which to date has not been possible.

Materials and Methods — Nine adult CI users attended seven research sessions, during their first year of rehabilitation. Behavioural SRD thresholds were measured using a two-alternative forced-choice paradigm. Speech perception in quiet and talker-babble noise (10dB SNR and 5dB SNR) was measured using AzBio sentences. Neural SRD was measured in response to spectrally rippled broadband noise at 0.25, 0.5, 1 and 2 ripples per octave. Stimuli were presented unilaterally, at most comfortable level, via the auxiliary input of the speech processor.

Results — Repeated measures ANOVA indicates that there is a significant time effect in the evolution of SRD ($F(7,56)=6.65$, p-value<0.01) and speech perception in quiet and in noise at 10dB SNR ($F(7,16.21)=4.45$, p-value<0.01; $F(6,55.82)=3.16$, p-value<0.01). Post-hoc tests using the Bonferroni correction revealed that there is a statistically significant increase in SRD abilities two months after implantation onwards, see Figure 1A. Pearson's correlation analysis revealed that SRD at one week after switch on correlates with speech perception in quiet ($r=0.878$, p-value=0.002) and in noise at 10dB SNR ($r=0.707$, p-value=0.05) up to nine months, see Figure 1B. Preliminary analysis of CEP responses suggests changes in morphology and latencies over time.

Discussion — This study demonstrates that longitudinal assessment of SRD can be used in the clinic to assess CI performance and predict speech perception abilities after one week from implant switch-on. An objective non-linguistic assessment like the one presented here may have a great impact on the assessment of CI performance in young children where behavioural testing is unreliable.

Thursday 5 January, 13:45—14:45

A versatile view on the aging auditory system for auditory processing and speech perception

A. van Wieringen
ExpORL, KU Leuven, Belgium

In daily life most speech communication information has to be captured in non-optimal acoustical environments and in the presence of interfering sound sources. In the auditory system, the complex speech signal needs to be analyzed acoustically and neurally, separated from the noise, and mapped to phonemes and words. There is a broad consensus in the literature that speech perception problems in the aging population originate from a combination of hearing impairment, cognitive decline, and central auditory processing deficiencies.

In the key talk, data on neural temporal processing, speech perception, and binaural processing are presented in order to disentangle effects of age-related peripheral hearing deficits and central processing changes. Auditory sensitivities from the peripheral (brainstem) to the central level (cortex) are investigated across age using a combination of objective neurophysiological and behavioral performance measures. Both objective and behavioral data are obtained from different adult normal hearing and hearing impaired age-cohorts, taking into account age-matching where appropriate. Understanding these neural correlates will lead to improved strategies to enhance central auditory plasticity in seniors (through training programs) and will support the development and application of better auditory prostheses (digital hearing aids and cochlear implants).

Funding bodies : Research Council of KU Leuven through project OT/12/98 and FWO-aspirant grant to Tine Goossens (grant number 11Z8815N)

Friday 6 January, 09:00—09:30

The role of high-level processes for oscillatory phase entrainment to speech sounds

B. Zoefel

MRC Cognition and Brain Sciences Unit, Cambridge, UK

Neural oscillations adjust their phase to rhythmic stimulation, a phenomenon called phase entrainment. This mechanism seems to be of particular importance for the processing of speech: Assumed to underlie speech comprehension, phase entrainment is omnipresent in current theories of speech processing. Nevertheless, speech is a complex stimulus and both low- and high-level processes might contribute to phase entrainment as it is commonly reported in the literature. Our aim was to disentangle these processes and provide a detailed characterization of the neural mechanisms underlying phase entrainment to speech. For this purpose, we constructed speech/noise stimuli without systematic fluctuations in sound amplitude or spectral content (here termed “low-level” features), while keeping both fluctuations in high-level features (including phonetic information) and intelligibility. In human psychophysical and electroencephalographic (EEG) data as well as primate intracranial recordings, we were able to show that phase entrainment can be observed in response to speech sounds in which systematic fluctuations in low-level features have been removed. This “high-level” entrainment shows specific characteristics and seems to reflect a particularly efficient mechanism of speech processing which is conserved across species. Finally, the relation between phase entrainment and speech comprehension remains debated. Based on the data presented here and elsewhere, we discuss possible reasons (and solutions) for this controversy and propose how brain stimulation techniques can help to clarify the role of oscillatory phase entrainment for the comprehension of speech sounds.

Friday 6 January, 09:30—10:00

Binaural intelligibility prediction for noisy and non-linearly processed speech

A. Andersen
Oticon A/S | Aalborg University, Denmark

Speech intelligibility prediction is becoming an increasingly popular tool within the speech processing community, as an alternative to time consuming and costly listening experiments. The short-time objective intelligibility (STOI) measure has enjoyed particular popularity, due to its simplicity and high performance in a range of key scenarios. However, the STOI measure lacks the ability to predict binaural advantage (i.e. the advantage of listening with two ears in conditions with spatial separation between target and masker), which is important in many applications. We therefore propose a binaural version of the STOI measure, based on extending the STOI measure with a modified version of the equalization cancellation (EC) model. The binaural STOI measure is shown to retain many of the favorable properties of the STOI measure in diotic conditions. On top of this, the measure can predict the advantage gained from spatial separation between a talker and a point source masker in an anechoic environment. Lastly, as an example of an application, we show how the measure can predict the outcome of a listening experiment, comparing the intelligibility of speech processed by different hearing aids. In this case, the binaural STOI measure is able to predict the relative performance of both normal hearing and hearing impaired listeners quite accurately.

Friday 6 January, 13:45—14:15

The impact of noise on speech processing over extended periods of time: A developmental perspective

J. Barry
MRC Institute of Hearing Research, University of Nottingham, UK

The majority of a school-aged child's listening and learning happens in classrooms where signal-to-noise ratios have been estimated to range from around +5 to -7 dB (Crandell et al. 2000). Though typically-developing children rarely complain about not being able to hear or understand what is going on (Klatte et al. 2010), there is considerable evidence to suggest they have more difficulty than adults in perceiving and understanding speech in such conditions. Here, we use an ecologically relevant task to investigate how children's processing efficiency varies as a function of time when listening in noise.

Four different groups of children (n = 20 x 4 groups) were required to detect mispronunciations and nonsense word targets when listening to a story presented in noise. The mispronunciations were either closely related to the immediately preceding context (i.e. predictable) or were not closely-related (i.e. unpredictable). Three different noise conditions were assessed: four-speaker babble (either +6 dB or +2dB) and speech-shaped modulated noise (SMN) +2 dB. Listening efficiency in these conditions was compared with performance on the same task presented without added noise.

We hypothesised an important role for language in supporting listening efficiency (perception (target detection errors) x processing rate (reaction time)). We therefore predicted more reliable and faster detection of predictable than unpredictable mispronunciations in all conditions. We further predicted that nonsense words would be detected more reliably and faster than either the predictable or unpredictable targets, since they would be less susceptible to noise-masking effects.

The greatest number of missed targets was observed in the +2 dB babble. Benefit for context was only observable in reaction times and only in the babble, not SMN, conditions. Predictable targets were detected more quickly than unpredictable or nonsense targets, though the benefit of context faded over time. By contrast, nonsense words were detected more quickly in the SMN and this effect was maintained throughout the task. Finally, most detection errors were observed for nonsense words regardless of noise masker, suggesting a tendency for children to accept these targets as real unknown words rather than as nonsense.

Overall, the findings suggest listening efficiency in children reflects a dynamic interrelationship between noise-related masking, processing load but also, factors associated with the developing language system. These findings will be discussed in the context of current research aimed at understanding of how noise impacts children's ability to listen and learn in their everyday listening environments.

References

Crandell, C. C., Smaldino, J. J. (2000). Classroom acoustics for children with normal hearing and with hearing impairment. Language, Speech, and Hearing Services in Schools, 31, 362-370.

Klatte, M., Lachmann, T., Meis, M. (2010). Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. Noise and Health, 12, 270-282.

Modeling speech intelligibility based on envelopes derived from auditory spike trains

C. Scheidiger, J. Zaar, T. Dau
DTU, Denmark

Speech intelligibility models aim to predict the human ability to understand speech in adverse listening conditions. The presented work combines the back-end processing of the multi-resolution speech-based envelope power spectrum model (mr-sEPSM; Jørgensen et al., 2013) with an auditory nerve (AN) model (Zilany and Bruce, 2014). The presented work calculated signal-noise-ratios in the envelope domain (SNRenv) for normal-hearing listeners based on different envelope representations derived from the AN model: (i) instantaneous firing rates (ii) peristimulus time histogram (PSTH) of auditory nerve spike trains, and (iii) SUMCOR neural metrics (Heinz and Swaminathan, 2009). The SNRenv patterns showed good agreements compared to the SNRenv patterns calculated from the acoustic (i.e. Hilbert) envelope (Heinz, 2016). Furthermore, speech intelligibility for normal-hearing listeners based on envelopes derived from PSTHs for CLUE sentences (CLUE; Nielsen and Dau, 2009) was predicted accurately in speech shaped noise (SSN), sinusoidally amplitude modulated noise (SAM) and speech-like noise (ISTS; Holube et al., 2010). Effects of hearing loss resulted in poorer speech intelligibility predictions. The work provides a foundation for quantitatively modeling individual effects of inner and outer hair cell loss on speech intelligibility.

Silent speech interfaces

T. Schultz
University of Bremen, Germany

Speech is a complex process emitting a wide range of biosignals, including, but not limited to, acoustics. The interpretation of biosignals - stemming from the articulators, the articulator muscle activities, the neural pathways, or the brain itself - can be useful to circumvent limitations of conventional speech processing systems, particularly when communication takes place in noisy environments or in silence.

In my talk I will present ongoing research at the Cognitive Systems Lab (CSL), where we explore Biosignals like muscle and brain activity to facilitate human-machine as well as machine-mediated human communication. Several applications will be described such as Silent Speech Interfaces that rely on articulatory muscle movement captured by electromyography to recognize and synthesize silently produced speech, Brain-to-text interfaces that use brain activity captured by electrocorticography to recognize speech and brain computer interfaces based on near infrared spectroscopy. We hope that our research will lead to a better understanding of the neural basics of speech production and will result in alternative interfaces that offer ways to communicate silently without disturbing bystanders or compromising privacy, and may be used for rehabilitation of speech-disabled people in the future.

Friday 6 January, 15:45—16:15

Context shapes neural speech coding: Effects of visual speech, prior auditory speech and listener's goals on cortical and subcortical speech processing

I. M. Schepers

Applied Neurocognitive Psychology Lab, Department of Psychology and Cluster of Excellence "Hearing4All" and Research Center Neurosensory Science, Oldenburg University, Germany

Natural speech perception occurs in a complex context of co-occurring visual speech information, prior auditory speech information and the goals of the listener. I will present results from several experiments, which investigated the effect of different contextual factors on neural speech coding. First, results from an electrocorticography (ECoG) study on audiovisual sentence processing will be presented. All subjects showed electrodes that entrained either to the auditory sentence envelope or the visual lip movements. Several electrodes over posterior superior temporal gyrus (pSTG) showed entrainment to both modalities and enhanced high-gamma band responses to audiovisual compared to auditory-only speech. These results provide evidence that posterior superior temporal gyrus is important for multisensory integration of audiovisual speech. Next, results from a functional magnetic resonance imaging (fMRI) study will be presented, which show that a distributed network is activated when prior auditory speech information is integrated to understand degraded sublexical speech and that successful speech comprehension correlates with greater activation in early auditory cortex. Finally, I will highlight the important role of subcortical structures such as the thalamic centromedian-parafascicular complex (CM-Pf) in goal-oriented behavior selection, including the selection of a speaker-stream. Electrodes in the CM-Pf exhibited transient neural responses in the alpha/beta range to speech and non-speech cues that required a goal-oriented behavior. Taken together, these findings show which cortical and subcortical regions play a role in the integration of different contextual information during speech perception.

Posters

*Posters of the session 1, Thursday 5 January, 14:45, are marked in **green**.*
*Posters of the session 2, Friday 6 January, 10:30, are marked in **magenta**.*

13 Noise-adaptive near-end listening enhancement for normal-hearing and hearing-impaired listeners

J. Rennies
Fraunhofer IDMT, Oldenburg, Germany

H. Schepker, D. Huelsmeier, J. Drefs, S. Doclo
University of Oldenburg, Department for Medical Physics and Acoustics, Germany

Speech played back in noisy environments is often impaired by environmental noise, resulting in poor speech intelligibility, e.g., for public-address announcements at train stations or mobile calls in noisy rooms. In such conditions listeners with impaired hearing face considerable challenges extracting the speech information. In such listening conditions it is usually not possible to reduce the noise present at the listener's location. One solution is to pre-process the target speech signal to enhance speech intelligibility depending on the current noise, which is referred to as near-end listening enhancement (NELE). This contribution presents the results of a series of listening tests to evaluate the NELE algorithm AdaptDRC. The AdaptDRC algorithm consists of several processing stages including dynamic range compression and frequency shaping, which are adaptively controlled by short-term estimates of the speech intelligibility index. If the estimated intelligibility is good, then no pre-processing is applied. If the estimated speech intelligibility index decreases, then the pre-processing is gradually increased. In its standard implementation the algorithm works under an equal-rms-power constraint, i.e., the rms level of each block of the processed speech is the same as the rms level of the unprocessed block. An additional implementation (AdaptDRCplus) further allows for an adaptively controlled increase in rms level by up to 6.5 dB while keeping the peak level of each block constant. Speech intelligibility was measured using the Oldenburg sentence test with unprocessed stimuli and stimuli processed by the AdaptDRC and AdaptDRCplus algorithm in different types of background noise and different SNRs. Both normal-hearing and unaided hearing-impaired listeners participated in the experiments. The results indicated that large benefits can be obtained even when the speech level was kept constant, but this benefit strongly depended on background noise type and SNR. The benefit due to the AdaptDRC algorithm was generally smaller for hearing-impaired than for normal-hearing listeners, but interindividual differences were quite large. All subjects benefited from the adaptive rms level increase indicating that NELE algorithms constraint to keeping the output rms level constant may not be sufficient for hearing-impaired listeners in some conditions.

14 Effects of hearing-aid compression on amplitude modulation processing and speech recognition

A. Wiinberg, B. Epp
Centre for Applied Hearing Research, Technical University of Denmark

M. L. Jepsen
Widex A/S

T. Dau
Centre for Applied Hearing Research, Technical University of Denmark

Despite recent advances in hearing-aid technology, hearing-aid users continue to find it difficult to follow conversations in noisy and reverberant environments. It is hypothesized that some of these difficulties originate from a reduced ability to process the linguistic information conveyed in the temporal envelope waveform. This can be a consequence of damages to the auditory system, or a consequence of the signal processing in the hearing-aid. To investigate this, the relationships between temporal envelope sensitivity, speech recognition and fast-acting wide-dynamic range compression (WDRC, used to account for loudness recruitment) were evaluated in normal-hearing (NH) and hearing-impaired listeners (HI) with a mild to moderately-severe cochlear hearing loss.

Two measures of temporal envelope sensitivity were considered: (i) temporal modulation transfer functions (TMTFs) with tonal carriers where amplitude modulation detection thresholds were obtained as a function of modulation frequency and (ii) "Supra-threshold" modulation-depth discrimination (MDD) thresholds, where the just-noticeable increase in modulation depth from a (supra-threshold) standard modulation depth were measured as a function of the modulation frequency. The frequency of the carrier was either 1 kHz or 5 kHz, and the modulation frequency was between 8 to 256 Hz. To estimate the impact of WDRC, TMTFs and MDD thresholds were obtained with and without WDRC. Speech recognition in both stationary and fluctuating noise was obtained for the same listeners with and without multi-band WDRC using a Danish version of the hearing in noise test (HINT).

Lower modulation detection thresholds were obtained for the HI listeners compared to the NH listeners at low modulation frequencies for the 5 kHz carrier, possibly as a consequence of reduced auditory peripheral compression. Conversely, elevated MDD thresholds were obtained for the HI listeners compared to the NH listeners. The latter result indicates that hearing-impairment can be associated with a reduced fidelity of supra-threshold temporal envelope coding. No relationship between temporal envelope sensitivity and the benefit of fast-acting WDRC in speech recognition was found. However, a strong relationship between temporal modulation discrimination and speech recognition was observed for the NH listeners. In contrast, no corresponding relationship was observed for the HI listeners. This indicates that, even though the fidelity of supra-threshold temporal envelope coding is reduced, other limitations in the damaged auditory system, such as reduced frequency resolution, seem to be more detrimental and thereby limiting factors in relation to speech recognition.

15 Evaluating fast-acting compression in basic psychoacoustic and speech tasks

B. Kowalewski, M. Fereczkowski, E. MacDonald
Hearing Systems group, Department of Electrical Engineering, Technical University of Denmark,

O. Strelcyk
Sonova U.S. Corporate Services

T. Dau
Hearing Systems group, Department of Electrical Engineering, Technical University of Denmark,

Multi-channel wide dynamic-range compression (WDRC) is involved in the signal processing schemes of most modern hearing aids. Commonly used nonlinear-gain prescriptions (e.g., DSL v5.0, NAL) do not make specific recommendations for other parameters of compression amplification, such as time constants or the number of processing channels.

Currently, there is no consensus regarding optimal values of these parameters. Studies using speech recognition as an outcome measure have shown mixed results, partly due to the differences in testing conditions (e.g. overall input levels and signal-to-noise ratios), differences in individual cognitive abilities, and differences in available cues for speech recognition. The latter likely depend on the degree and configuration of the hearing loss.

Here, a model of auditory processing (Jepsen et al. 2011) was used to compare linear and various compression schemes in terms of individual listener's performance in basic psychoacoustic tasks. Specifically, time constants and the number of channels were chosen to provide the best possible improvement of the simulated aided performance of hearing-impaired listeners in tasks assessing spectral and temporal resolution. The simulation outcomes indicated that compression with a sufficiently large number of channels and short time constants can provide a benefit over linear amplification. These results were verified in aided psychoacoustic experiments.

Further, speech recognition performance of the same group of hearing-impaired listeners with compressive and linear amplification was investigated. Speech reception thresholds (SRTs) were measured using two different speech corpora and maskers. Dantale II (Wagener et al., 2003) sentences were presented with ICRA7 fluctuating noise and resulting SRTs were expected to be negative. Using the DAT corpus (Nielsen et al., 2014), the target talker was presented with two competing talkers and positive SRTs were expected in this condition. It was hypothesized that compressive processing may be more beneficial at negative signal-to-noise ratios (SNRs), hence the relative benefit of compressive over linear processing should be larger for the Dantale corpus.

Even though no significant interactions were observed, a trend was seen in the data suggesting that the relative benefits of compression and linear amplification might depend on the SNR. Studying the effect in more subjects and across a broader range of SNRs might help resolve these issues.

References

Jepsen M.L.; Dau T.; J Acoust Soc Am 129(1) 2011 pp. 262-281

Nielsen J.B.; Dau T.; Neher T.; J Acoust Soc Am 135(1) 2014 pp. 407-420

Wagener K.; Josvassen J.L.; Ardenkjær R.; Int J Audiol 42 2003 pp. 10-17

16 Effect of low-frequency gain on speech intelligibility and sound quality in a competing voices situation

J. Lochner
Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark

S. Santurette
Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark | Department of Otorhinolaryngology, Head & Neck Surgery, and Audiology, Rigshospitalet, Denmark

E. MacDonald
Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark

L. Bramsløw
Eriksholm Research Centre, Oticon A/S, Denmark

In hearing aids, low-frequency (LF) gain in the octave bands centered at 125, 250, and 500 Hz is often reduced either because of open fittings or to minimize upward spread of masking. However, pitch cues extracted from LF resolved harmonics are considered important for speech intelligibility (SI) in competing voices situations. Furthermore, the sound quality is typically described as more natural with higher LF gain. Here, the effects of modifying LF gain on SI and sound quality for a female target voice masked by either a same-sex or different-sex masker voice were assessed.

Ten young normal-hearing (NH) listeners and 11 older hearing-impaired (HI) listeners with sensorineural and moderate LF hearing loss participated. Two conditions with an increase in LF gain of either 4 dB or 8 dB relative to the prescribed-gain condition were tested. In the prescribed-gain condition, the HI listeners were fitted individually on both ears using a linear CAMEQ rationale, while no gain was applied for the NH listeners. A self-scored closed-set Danish matrix-type SI test (Dantale II) was used. The gain was applied directly to the stimuli and the speech stimuli were presented via headphones. The percentage of correctly identified words was evaluated at four fixed signal-to-noise ratios (SNRs). All NH listeners were measured at the same absolute SNR while all HI listeners were measured at constant SNRs relative to their individual speech reception thresholds.

For NH listeners, no significant effect of increased LF gain on SI was found. At 0 dB SNR in the prescribed-gain condition, the HI listeners scored significantly lower with the same-sex versus different-sex masker, whereas the scores were similar for the two maskers at SNRs different from 0 dB. The HI listeners did not benefit either from higher LF gains in terms of SI. However, an interaction between LF gain and masker type (same sex versus different sex) was found. This is because a 4-dB increase in LF gain was beneficial in the different-sex masker condition but detrimental in the same-sex masker condition.

A subjective quality rating of the different LF gain conditions was also performed by the same HI listeners using the speech and masker material from the SI test at 0 dB SNR. The quality ratings were significantly higher with increased LF gain compared to the prescribed-gain condition.

Overall, the results suggest that a higher LF gain does not affect SI but may be beneficial in terms of sound quality.

17 Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users

T. Goehring
University of Southampton, UK

F. Bolner
Cochlear Technology Centre, Mechelen, Belgium

J. J. M. Monaghan
University of Southampton, UK

B. Van Dijk
Cochlear Technology Centre, Mechelen, Belgium

A. Zarowski
European Institute for ORL-HNS, Antwerp, Belgium

S. Bleeck
University of Southampton, UK

Speech understanding in noisy environments is still one of the major challenges for cochlear implant (CI) users in everyday life. In this study, we propose a speech enhancement algorithm based on neural networks (NNSE) for improving speech intelligibility in noise for CI users. The algorithm decomposes the noisy speech signal into time-frequency units, extracts a set of auditory-inspired features and feeds them to the neural network to produce an estimation of which frequency channels contain more perceptually important information (higher signal-to-noise ratio, SNR). This estimate is used to attenuate noise-dominated and retain speech-dominated CI channels for electrical stimulation, as in traditional n-of-m CI coding strategies. The proposed algorithm was evaluated by measuring the speech-in-noise performance of 14 CI users in three different types of background noise. Two distinct NNSE algorithms were compared in this experiment: a speaker-dependent algorithm, that was trained on the target speaker used for testing, and a speaker-independent algorithm, that was trained on different speakers. Significant improvements in speech intelligibility in stationary and fluctuating noises were found over the unprocessed condition for both speaker-dependent and speaker-independent algorithms, with the first algorithm providing bigger improvements. Results indicate that the proposed algorithm has the potential to improve speech intelligibility in noise for CI users and proves to generalise to a range of acoustic conditions, whilst meeting the requirements of low computational complexity and processing delay in CI devices.

18 Congenital unilateral hearing impairment - Bone Anchored Hearing Implants in complex listening environments

E. Muzzi, R. Marchi, C. Falzone, E. Orzan
Audiology and Otorhinolaryngology, Institute for Maternal and Child Health "Burlo Garofolo", Trieste, Italy

Congenital unilateral hearing impairment (UHI) affects auditory performance in complex listening environments. UHI leads to auditory pathways reorganization and aural preference syndrome, as binaural plasticity occurs mostly in the first years of life. Children with UHI may have increased rates of grade failure, need for additional educational assistance, behavioural issues in the classroom and speech and language delays. However, there is a significant gap of knowledge on how to evaluate and rehabilitate children suffering from UHI.

Our objective was to assess improvements in speech understanding in complex listening environments with bone anchored hearing implants (BAHI) for congenital UHI.

Study design was a non-randomized, open, prospective case series. Setting was a tertiary referral center. Between November 2014 and November 2016, 16 children (mean age 10+/-4 years) were tested with BAHI for congenital UHI. Nine children were affected by Single Sided Deafness (SSD) and 7 suffered from Moderate Conductive Hearing Impairment (MCHI). Intervention:Transcutaneous or percutaneous BAHI. A quick adaptive procedure was used to test head shadow effect and interaural attenuation, loudness summation, binaural squelch on azimuth (90°) and elevation (30°), and spatial release from masking. Data logging informations were analyzed. AFHAB and SSQ hearing-related quality of life questionnaires were administered.

All children showed poorer-than-normal speech in noise performance without BAHI. Mean improvement of intelligibility with BAHI was 19.2% at loudness summation, 20.0% at squelch on azimuth, 15.0% at squelch on elevation with SNR+10. Mean improvement of Maximum Speech Reception Threshold (Max-SRT) with BAHI was 2.0 dB at binaural loudness summation, 3.0 dB at squelch on azimuth, 5.3 dB at squelch on elevation. On average, children obtained a speech release from masking of 1.3 dB on azimuth and of 4.8 dB on elevation. Data logging, AFHAB and SSQ results disclose an advantage with BAHI in real life conditions. Pseudo-binaural hearing enhances speech in noise performance in children. The gain is measurable on the whole test battery and it is more evident on elevation. The effectiveness of BAHI in complex listening environments is supported by data logging and quality of life measures.

19 Speech performance versus speech perception: Comparison of two different speech-in-noise tests using different microphone settings in cochlea implants

A. Büchner
Medical University of Hannover, Germany

T. Bräcker
Med-El Research Department Hannover, Germany

M. Schwebs, T. Lenarz
Medical University of Hannover, Germany

Objectives — The Oldenburg Sentence test (OLSA) is a speech-in-noise test primarily to measure the speech understanding threshold at which 50% of speech can be understood (SRT50). Speech understanding with both hearing aids (HA) and cochlea implants (CI), respectively, often reaches negative signal-to-noise ratios (SNR) using the OLSA. Unfortunately the OLSA is not very useful any more for negative SNRs to show speech benefits of new algorithms in fields of hearing research, e.g. noise reduction. Therefor different test methods and test modifications have been implemented in the past to push the SNR of the OLSA towards a positive range.

Methods — Two different speech-in-noise tests were used for this study: on the one hand the OLSA to evaluate speech performance, on the other hand a just following conversation test, the Just Understanding Speech Test (JUST) to evaluate speech perception. For both speech tests the results of 20 postlingual deafened cochlea implant (CI) users are shown. All of them used the audio processor named SONNET of MED-EL. They had to achieve a minimum score of 15% in the HSM sentence test in noise at their last clinical appointment to be able to participate in the study. Measurements of the OLSA and the JUST have been conducted in a sound treated room using a multi-loudspeaker array (0°, ±70°, ±135°, 180°). Three different microphone settings (natural beamformer, adaptive beamformer and omnidirectional microphone) were used for the tests.

Results — SNR results of the JUST (SRTjust) show a clear displacement to the positive direction compared to the measured SRT50 of the OLSA for all microphone settings. Furthermore the results show significant benefits for both the natural and the adaptive beamformer, respectively, compared to the omnidirectional setting.

Conclusion — *First:* Althoug the JUST is a speech test based on speech perception it seems to be an adequate method to evaluate speech intelligibility in noise in addition to the well-established OLSA. *Second:* Based on the fact of beamforming effects in general, improved speech understanding in noise will be expected for CI users using the new audio processor.

20 Bimodal speech intelligibility: Comparison of actual patients and simulations

B. Williges, S. Kliesch
Medizinische Physik and Cluster of Excellence Hearing4all, Universität Oldenburg, D-26111 Oldenburg, Germany

L. Jung
Implant Centrum Freiburg, Universitätsklinik für Hals-, Nasen- und Ohrenheilkunde, Freiburg (Breisgau), Germany

L. I. Geven
Universitätsklinik Hals-Nasen-Ohrenheilkunde am Evangelischen Krankenhaus, Oldenburg, Germany

T. Wesarg
Implant Centrum Freiburg, Universitätsklinik für Hals-, Nasen- und Ohrenheilkunde, Freiburg (Breisgau), Germany

T. Jürgens
Medizinische Physik and Cluster of Excellence Hearing4all, Universität Oldenburg, D-26111 Oldenburg, Germany

A large group of cochlear implant (CI) listeners uses a hearing aid on the ear opposite to their CI. On average, these so called bimodal CI listeners show improved speech intelligibility, sound localization and music enjoyment (Ching et al, 2007) when they use both devices. However, the same study also shows that not all bimodal CI listeners benefit from use of their two devices. The benefit from having two devices might be confounded by the large variability within the CI side (which is most often the better performing ear) or might be related to non-availability of binaural processing. Thus in this study, we compared speech intelligibility scores from actual bimodal listeners with scores from normal hearing (NH) listeners, who listened to a simulation of bimodal hearing (aided hearing impairment and CI simulations). The psychometric function of speech intelligibility was assesed using the German OLSA matrix sentence test with stationary noise either co-located or placed at -90° or 90° from the frontal speech. A combination of an adaptive procedure to measure 50% speech reception thresholds and and a non-adaptive procedure to measure speech recognition scores at a constant SNR was used. CI simulations were noise-like vocoders, which incorporate details of the signal processing in CI speech processors (Williges, 2015). For aided hearing impairment simulation, the Master Hearing Aid (Grimm et al, 2009) was used in combination with a hearing impairment simulation, which accounts for different degrees of audiometric thresholds, loudness recruitment and poorer representation of the speech envelope.

First comparisons between simulations and measurements in actual bimodal listeners show that the bimodal benefit depends on the signal-to-noise ratio and the spatial arrangements of noise and speech. The defined availability of (realistic) cues present within the simulations allowed a systematic analysis in which situations bimodal benefit is present or absent, with minimal confoundings due to inter-subject variability.

The proposed bimodal simulations could help in identifying speech intelligibility tasks, where the bimodal benefit is present, and even be extended towards models of speech intelligibility.

Supported by DFG.

21 Perception of French nasal vowels in vocoded speech: Information Transfer Rate analyses

O. Crouzet
LLING - Université de Nantes / CNRS, France

The production and perception of nasality is highly impacted in hearing-impaired listeners. Many studies have shown that, though wearing cochlear implants may partially improve production, such listeners still exhibit specific difficulties in the production of such segments (Fletcher et al, 1999; Baudonck et al., 2015). Perception of nasals is also subject to specific perturbations in CI listeners (Bouton et al. 2012; Borel, 2015). These observations hold for both consonants and vowels.

Nevertheless, the reasons why specific issues may occur with nasal sounds are still far from being understood. In order to investigate this issue further, two experiments were designed using vocoded speech with normal-hearing french listeners. Vocoded speech signals were generated in order to provide between 1 and 8 noise channels with 3 different ranges of amplitude modulation frequencies (slow MFs only, slow and mid-MFs, all-MFs). Preliminary results were presented at SpIN 2016 and can be summarized as follows. In a 20-AFC experiment targeting the identification of french consonants in a VCV context, oral and nasal consonants exhibited very similar behaviors: higher than random performance was reached with 4 channels when only low modulation frequencies were available. For oral consonants, even a single channel was sufficient when all modulation frequencies were present. For nasal consonants, 4 channels were necessary in this condition. Overall, though the perception of nasal consonants was less optimal than the perception of oral ones, response performance reached higher than random significance for both oral and nasal consonants when using 4 channels or more.

On the contrary, the perception of steady-state vowels showed strong discrepancies: though oral vowels exhibited tendencies that were similar to consonants, correct response rates for nasal vowels never reached significance.

In order to investigate this issue further, the study of "Information Transfer Rates" (Miller & Nicely, 1955; Christiansen & Greenberg, 2005) has been conducted in order to analyse the structure of confusion matrices. This let's us evaluate how much phonological features are degraded in different conditions. Concerning the [nasal] feature for example, up to 11.2% of the maximum IT rate are preserved in consonants but only 1.83% in vowels. As a comparison, [voicing] reaches 33.1% of the maximum IT rate in consonants and [close] reaches 10.53% in vowels. Though nasality exhibits particularly low levels of transmission rates in both types of segments, it seems that vowel nasality is more impacted that consonant nasality with respect to other features.

22 Better-ear glimpsing with symmetrically-placed interferers in Bilateral cochlear implant users

H. Hu
Universitat Oldenburg, Germany

M. Dietz
Western University,Canada

B. Williges, S. Ewert
Universitat Oldenburg, Germany

In symmetrically placed maskers, a long-term "better-ear" as a consequence of the head shadow effect does not exist. Nevertheless "better-ear glimpses" in time-frequency segments exist. By simulated better-ear glimpsing using an ideal monaural better-ear mask (IMBM), Brungart and Iyer (2012) showed that normal hearing (NH) listeners perform as well as in the natural binaural condition which also provides interaural time differences. Their results suggest that an optimal glimpsing strategy is efficient to explain NH binaural performance. In contrast, Glyde et al. (2013) showed that such a glimpsing strategy cannot fully account for binaural performance in conditions with high informational masking. Similar, a recent study by Best et al. (2015) investigated the same IMBM based glimpsing in bilateral sensorineural hearing impairment (HI) and suggest that better-ear glimpsing alone is not sufficient to explain all spatial release from masking (SRM). It can be expected that bilateral cochlear implant (BiCI) users might be able to utilize a glimpsing strategy.

This study compared SRM in both NH (with and without noise vocoding to simulate available cues in CI users) and BiCI listeners. Speech reception thresholds (SRTs) under spatially separated (±60°) or with an unnatural infinite ILD condition, generated with Head-related transfer functions, were obtained with and without IMBM processing for three masker types: stationary masker, single talker, and nonsense speech. SRTs were obtained using headphone presentation in NH listeners or direct stimulation in BiCI listeners. The question was to which extent a glimpsing strategy can be used by BiCI users and how potential deficits in binaural processing could be aided by IMBM processing.

Results indicate that NH with vocoder and BiCI listeners show a strongly reduced binaural benefit in the ±60° condition relative to the co-located condition when compared to NH. However, both groups greatly benefit from IMBM processing (as part of the CI stimulation strategy). Given that BiCI users benefit from IMBM processing, future binaural processing strategies mimicking glimpsing can be expected to aid speech intelligibility in BiCI users. Additionally, a processing strategy that increases the channel separation between maskers (like in the infinite ILD condition) might help BiCI listeners.

23 Channel interaction determines the best fitting strategy for cochlear implant users with ipsilateral residual acoustic hearing in a computer model

L. Zamaninezhad
Carl von Ossietzky University of Oldenburg, Germany

T. Jürgens
Carl von Ossietzky university of Oldenburg, Germany

Cochlear implant (CI) users, who have ipsilateral low-frequency acoustic hearing show better speech-in-noise performance on average than cochlear implant users without acoustic hearing. This has been shown by clinical, vocoder and model-based studies. It is, however, still a matter of debate, which frequency range of the ambient sound should be transferred via the CI if acoustic listening is present. Candidates for such a fitting strategy include transferring the full frequency range or a restricted one, defined by either audiometric data or psychoacoustic tests.

In this study, we aim at investigating the interplay of electrically stimulated frequency range with a less expected factor, namely channel interaction, using a model-based approach. We explain how the choice of frequency range influences channel interaction and consequently the speech-in-noise performance. We suggest a suitable fitting approach (in terms of crossover frequency) for different listeners. Moreover, we illuminate the role of CI-processed-speech modulation information in different auditory channels on speech perception of electric listeners and link this to the above-mentioned fitting approaches.

An auditory model predicting speech reception thresholds (SRTs) in noise for CI users with residual acoustic hearing (Zamaninezhad et al., accepted) was employed in this study. The auditory model can be considered as a hypothetical listener which allows systematic variation of different physiologically plausible parameters such as channel interaction, number of employed electrodes or residual acoustic hearing. Two different fitting approaches were simulated, one with electric stimulation of full frequency bandwidth and one with a restricted frequency range. The width of the spatial spread of electric field (and thus the channel interaction) was systematically increased to smear the spectral representation of speech and SRTs were predicted and compared across fitting strategies.

The results show that for listeners with highly interacted channels, a restricted electrically stimulated frequency range can improve speech perception in noise. For listeners with limited channel interaction, the extent of electrically stimulated frequency range is less influential on speech perception. Moreover, the results show the importance of CI-processed-speech modulation in auditory channels above 700 Hz: The model predicts that discarding low frequency CI-processed-speech information may be beneficial for some electro-acoustic listeners.

References

Zamaninezhad. L., Hohmann. V., Büchner. A., Schädler. M. R., Jürgens. T., 2016. A physiologically-inspired model reproducing the speech intelligibility benefit in cochlear implant listeners with residual acoustic hearing. Hear. Res. Accepted.

24 SPIRAL – a vocoder with a spiral ganglion for cochlear implant simulation

J. A. Grange, **J. F. Culling**
Cardiff University, UK

Vocoders for cochlear implant simulation use a filterbank to simulate the filter channels of the processor. Temporal envelopes extracted from these channels are then directed to the positions in the cochlear of a simulated electrode array by modulating narrowband carrier signals. Traditionally, there is one carrier signal for each processor channel/electrode. The SPIRAL vocoder breaks with this tradition by using many carrier signals. Each may be considered to represent a sub-population of ganglion cells. The modulation of each carrier is a combination of all the channel envelopes mixed according to an exponentially decaying spread of current along the electrode array, using formulae from Oxenham and Kreft (2014). Carrier signals can be noises or tones, but noise introduces intrinsic modulations that are independent of the signal envelope. Consequently, we have found that tonal carriers provide lower SRTs even with up to 160 tones. With the electrodes, in this sense, decoupled from the ganglion cells, the two can be independently manipulated. The positions of the electrodes can be changed without changing the set of carriers used. Dead regions can be simulated by omitting a range of carrier frequencies. Any degree of current spread can be simulated. The vocoder has been configured so that the user can specify the electrode number and spacing, as well as the depth of insertion, and different mappings of frequencies to electrodes can be explored.

References

Oxenham, A. J., & Kreft, H. A. (2014). Speech Perception in Tones and Noise via Cochlear Implants Reveals Influence of Spectral Resolution on Temporal Processing. Trends in Hearing, 18:1–14.

25 Relation between listening effort and speech intelligibility in noise

M. Krueger
Hörzentrum Oldenburg, Institute of Hearing Technology and Audiology, Jade University of Applied Sciences Oldenburg and Cluster of Excellence Hearing4All, Germany

M. Schulte
Hörzentrum Oldenburg and Cluster of Excellence "Hearing4all", Germany

T. Brand
Medizinische Physik, Universität Oldenburg and Cluster of Excellence "Hearing4all", Germany

K. C. Wagener, M. Meis
Hörzentrum Oldenburg and Cluster of Excellence "Hearing4all", Germany

I. Holube
Institute of Hearing Technology and Audiology, Jade University of Applied Sciences Oldenburg and Cluster of Excellence Hearing4All, Germany

In today´s communication society understanding of fellow humans is particularly import in noisy environments. Especially hearing impaired and elderly people report that they need additional effort in these situations in order to understand speech. Listening effort (LE) can be determined quickly and flexibly using subjective methods like categorical scaling or questionnaires. However, scaling results might be affected, for instance, by the range of signal-to-noise ratios (SNR) of the stimuli. Therefore, an adaptive scaling method for subjective LE ratings was developed. In this method, the signal-to-noise ratio (SNR) is varied adaptively to adjust the SNR range to the individual ratings using a 14-step scale from "effortless" to "extreme effort", including the additional category "only noise". The new scaling method (ACALES: Adaptive CAtegorical Listening Effort Scaling) is based on the adaptive categorical loudness scaling (ACALOS) procedure. ACALES was evaluated with young subjects with normal hearing and elderly subjects with impaired hearing (with and without hearing aids) by presenting sentences of the Oldenburg sentence test in four different background noises. As a comparison, speech intelligibility (SI) measurements were performed using the same conditions. The results show that SI increases with increasing SNR whereas LE is decreasing. The LE ratings as well as the SI scores are dependent on the type of masking noise (modulated versus continuous noise). Furthermore, the results showed that the procedure adjusted properly to different listening conditions and that ACALES detected a benefit in LE due to hearing aid

provision and is able to evaluate differences, e.g. due to noise reduction algorithm, in a range of SNRs where speech intelligibility scores are already saturated at 100%. The method is easy to use, the measurement time is similar to common speech intelligibility tests, and it is suitable for laboratory studies and hearing aid adjustments.

26 Subjective self-assessment of bimodal fusion

N. A. Janßen
Hearing Systems Group, Technical University of Denmark

L. Bramsløw
Eriksholm Research Centre, Denmark

S. K. Riis
Oticon Medical, Denmark

J. Marozeau
Hearing Systems Group, Technical University of Denmark

In recent years, implantation criteria for cochlear implants have changed and we see more and more so-called bimodal patients wearing a cochlear implant (CI) on one side and a hearing aid (HA) on the other side. Bimodal users thus listen with two devices that deliver sound in a very different manner due to their working principles. The HA and CI devices are often manufactured by different companies, use different processing strategies and are typically fitted by different professionals. Today, no universally accepted bimodal fitting scheme is available and many patients experience two sound percepts that are not matched in pitch and loudness.

From this the question arises, whether bimodal users fuse sounds from both ears, like normal hearing persons do. Are they perceiving just one sound object or do they perceive a more complex acoustic scene with two simultaneous sound objects?

In order to assess this question, 9 bimodal users were interviewed based on a self-assessment questionnaire. This questionnaire was divided into two parts: First, their auditory experiences such as speech perception in noise, distance perception and listening effort were assessed using the SSQ5 questionnaire (see Mertens 2013). Then, they were asked to rate the difference in perception for sounds of a single sound source across their two ears and additionally rate a possible difference in pitch and loudness, as well as preference for localization and sound quality. The latter was rated for six everyday situations, each in the context of having a conversation: Quiet, a busy restaurant, nature with wind, a busy street at rush-hour, listening to music and shopping in a supermarket.

The everyday situations do not yield significantly different results and are therefore considered together. Altogether, the answers indicate that a 4 out of 9 participants report to perceive a single fused sound object despite significant differences across ears in pitch and loudness. While this does not seem to affect perceived speech perception in noise or localization abilities, bimodal users that report a low difference in perception also report less cognitive effort.

More data will be collected in order to confirm the bimodal distribution of the two groups perceiving either fusion or no fusion.

27 Investigating the relationship between SpiN recognition, AM depth detection and evoked potentials in CI users

S. Waechter, A. Lopez Valdes
Trinity College Dublin, Ireland

C. Simoes-Franklin, L. Viani
Beaumont Hospital, Dublin, Ireland

R. Reilly
Trinity College Dublin, Ireland

Objectives – Extensive research studies have explored human temporal auditory processing abilities and their links to speech-in-noise (SpiN) perception. The proposed study design aims to investigate the influence of temporal processing abilities on speech-in-noise recognition scores. The temporal feature of interest is amplitude modulation (AM) depth detection, for which a potential objective measure will be explored based on electroencephalography (EEG) data. An additional loudness balanced condition within the EEG paradigm will account for inherent loudness differences between modulated and unmodulated stimuli to ensure these have no significant influence on the obtained EEG measures. At the workshop preliminary pilot data will be presented.

Hypotheses — Based on previous work in normal-hearing (NH) participants by our group, we hypothesize significant correlations between SpiN thresholds and the AM depth detection thresholds, as well as between behavioural and neural thresholds of AM depth detection in both cohorts. We hypothesize no significant impact of loudness balancing on the EEG measures.

Methods — Cochlear implant (CI) users and NH participants will be recruited for this study. The test battery will consist of four paradigms: (1) an adaptive threshold procedure to determine the behavioural AM depth detection threshold, (2) an adaptive speech-reception threshold (SRT) test to determine the signal-to-noise ratio at which 50% of keywords are identified correctly and (3) a neurophysiological mismatch negativity paradigm to determine a neural threshold of AM depth. (4) Another paradigm determines the rms-level for which the modulated and unmodulated sounds reach the same overall loudness subjectively for each participant. The loudness matched stimuli are presented in the EEG task as a fourth block for the 100% AM depth condition.

The chosen AM rate is 8Hz, based on the importance of slow-envelope fluctuations on speech-recognition, which is imposed on a broadband carrier of speech-shaped noise. The AM depths for the mismatch paradigm are set to 50%, 75% and 100%. All stimuli will be presented at 60dBA (calibrated for the unmodulated noise), for CI users via an OtocubeTM (http://otocube.com) and for NH participants monolaterally via headphones. The SRTs are determined with an adaptive procedure as in standard literature with ten-talker babble background noise. The signal-to-noise ratio (SNR) is adjusted depending on the number of correctly identified keywords, converging on the SNR which provides 50% correct identification.

28 The spread of excitation in cochlear implants is not the only limiter of the number of effective electrodes. Implications for the channel-interleaving sound coding strategy.

J. A. Grange, J. F. Culling
Cardiff University, UK

The spread of excitation (SOE) from cochlear implant (CI) electrodes along the spiral ganglion is commonly believed to cause the plateauing of CI users' speech-in-noise intelligibility as a function of the number of activated electrodes, from a knee point at around 8 effective electrodes[1]. Were this hypothesis correct, increasing SOE should move the knee-point to a lower number of effective electrodes. Simulations making use of the same number of analysis filters and reconstruction carriers (conventional vocoding) and incorporating SOE[2], failed to exhibit the expected knee-point shift. Our new SPIRAL vocoder uses a large, fixed number of carrier tones to better represent a continuous spiral ganglion. When varying only the number of analysis filters (i.e. activated electrodes), SPIRAL revealed the expected knee-point shift with either sentences or digit triplets. Speech reception thresholds (SRTs) rose significantly as SOE increased. However, without SOE, a knee-point remained, suggesting that SOE is not the only limiter of the number of effective electrodes.

Interleaving odd and even channels between two CIs originally yielded mixed outcomes[3]. After careful pitch matching, interleaving was later showed to improve bilateral CI users' spectral resolution[4]. However, our simulations of interleaving led to a much-lower-than-expected SRT improvement (≤ 1.5 dB) at current decay rates below 32 dB/octave (where SOE significantly elevated thresholds).

Our findings suggest that (1) the plateauing of speech intelligibility seen beyond 8 activated electrodes may be more fundamentally due to the redundancy in speech information provided by neighbouring channels, rather than just due to current spread; (2) current spread worsens the effect by spectrally smearing the spectro-temporal modulations that carry speech information, leading to inter-channel modulation masking and significant threshold elevation; (3) although channel interleaving can improve spectral resolution, the improvement does not transfer to speech intelligibility as well as was previously expected, possibly because (4) modulation masking due to current spread is more effective between channels that are not immediate neighbours and/or (5) significant modulation masking still occurs centrally (or contra-laterally) and to a level that defeats the object of traditional channel interleaving. Clearly, a more in-depth understanding of the mechanisms underlying inter-channel and contralateral modulation masking, is required to devise more effective sound-coding strategies.

1. Friesen et al. (2001) JASA, 110(2), p.1150.

2. Oxenham & Kreft (2014) Trends Hear., 18, p.1–14.

3. Tyler et al. (2010) J. Am. Acad. Audiol., 21(1), p.52–65.

4. Aronoff et al. (2015) Ear Hear, 37(2), p.e85-e90.

29 Multisensory integration in hearing aid users and non-users with a mild hearing loss

M. Tahden, A. Gieseler, H. Colonius
Psychology Department and Cluster of Excellence "Hearing4all", University of Oldenburg, Germany

Research has shown that hearing thresholds and age alone cannot comprehensively explain the substantial variability in speech understanding in noise, prompting the search for other factors to be involved. Several studies have focused on the contribution of cognitive abilities, clearly confirming a link. Yet, there remains unexplained variance suggesting that not all components relevant for explaining individual differences in speech-in-noise performance have been identified so far.

Current research indicates that the provisioning with a hearing aid could likewise contribute to these differences. In this regard, it is not only of interest how hearing-impaired individuals differ in speech understanding, but also whether they differ in cognitive abilities as a function of hearing aid use. Another relevant factor contributing to differences in speech-in-noise understanding could be found in multisensory integration, as recent studies have highlighted effects of age, hearing loss and cochlear implant provisioning on multisensory integration capacity.

As differences between hearing-impaired individuals with and without hearing aid provisioning have been widely neglected so far in this context, the comparison of individuals with a mild hearing loss who differ in hearing aid use is of specific interest. To address these issues, this study compared 40 hearing aid users and 40 non-users with a mild hearing loss (PTA between 26 and 40 dB HL), matched for age and gender, regarding speech understanding in noise, cognitive abilities and audiovisual integration. Speech understanding was measured with the Goettingen Sentence Test. From that, hearing aid benefit was derived by comparing unaided with aided measurements (Master Hearing Aid). Cognitive testing comprised a dementia screening test, a test of working memory and of verbal intelligence. To assess audiovisual integration, the sound-induced flash illusion (SIFI) was applied, where one or two flashes were paired with either a congruent or incongruent number of beeps. The strength of audiovisual integration was measured by the frequency of with which one flash accompanied by two beeps (1F2B) is perceived as two flashes, the so-called fission illusion.

Results revealed no significant differences between hearing aid users and non-users in any of the cognitive tests. In comparison, they differed significantly in the 1F2B condition of the SIFI such that hearing aid users perceived the fission illusion more often and thus, displayed enhanced audiovisual integration compared to non-users. Furthermore, correlations between the hearing aid benefit and cognitive tests as well as the audiovisual integration differed as a function of hearing aid supply.

30 The effect of background noise on the gestures, gaze and speech of hearing-impaired interlocutors

L. V. Hadley, W. O. Brimijoin, **W. M. Whitmer**
MRC/CSO Institute of Hearing Research - Scottish Section, Glasgow, UK

Holding a conversation in noise is a remarkably complex task, in which multiple individuals must coordinate speaking and listening turns toward a shared communicative goal. To be able to best bolster communication for hearing-impaired individuals, it is crucial to know the interplay between individuals' gestures, gaze and speech in real-world scenarios, data that can be compared with multi-modal strategies for optimal communication (e.g., to see if head movements and speech levels adjust appropriately for noise levels). We therefore measured the behaviour of 16 hearing impaired dyads and 11 triads holding conversations in background noise. Participants held semi-structured conversations while head movement, eye movement and speech were recorded. All groupings were mixed gender, with participants matched on speech-in-noise perception, hearing asymmetry, and age. Dyadic interactions involved three conversations in speech-shaped noise, while triadic interactions involved two conversations in speech-shaped noise, and two conversations in eight-speaker babble. In each conversation, noise level was varied between 54-78 dB in 15-25 s segments. Conversation topics (and in triads, noise types) were counterbalanced.

Results showed an effect of noise level on head movement behaviours. Across both experiments — dyad and triad conversations — participants leant closest together for the loudest noise level, and moved gradually apart for each subsequent noise level. Participants also focused their head more directly toward their partner at louder noise levels. Furthermore, there was an effect of noise type in triadic conversation, with participants leaning closer in the babble conditions. The acoustic changes due to such movements, however, were small (<1 dB) relative to changes in speaking level (1-2 dB), neither of which accommodated the noise-level changes (6 dB steps). Despite expressed difficulties in noisier conditions, there were few requests for clarification during conversation. Participants' eye position (with respect to their partner) and speech level will also be discussed. The data as a whole demonstrate clearly identifiable strategies in conversation to ameliorate hearing difficulty that could be exploited in current hearing rehabilitation and future multi-modal hearing prostheses.

Work supported by the Medical Research Council (grant number U135097131) and the Chief Scientist Office of the Scottish Government.

31 The processing of which-questions in noise-vocoded speech

A. Schouwenaars, E. Ruigendijk
University of Oldenburg, Germany

Whereas children with cochlear implants (CIs) score like their age matched peers on vocabulary, their understanding and use of grammatical aspects of spoken language lacks behind (Nikolopoulos, Dyar, Achbold & O'Donoghue, 2004; Friedmann & Szterman, 2011). In German, grammatical aspects such as case and verb-agreement are crucial for correct interpretation of object questions (in which the object precedes the subject). In previous research, we found that children with cochlear implants, when interpreting object questions make less use of case (e.g. Welchen Esel fängt der Tiger 'Which donkey is the tiger catching?') and/or verb-agreement cues (e.g. Welche Maus fangen die Frösche? 'Which mouse are the frogs catching?') than children with normal hearing (Schouwenaars et al., in prep). One explanation for these interpretation problems is the input quality (degraded speech input by the CI), whereas another explanation is the lack of input during a so-called sensitive period. The speech input by CIs can be simulated by noise-vocoded speech and its effect can be tested on normal hearing populations (e.g. Scott 2005). A first step and the aim of this study is to investigate whether adults with normal hearing show similar problems when interpreting which-questions in noise-vocoded speech as the children with CIs.

We tested 30 adults with noise-vocoded stimuli (simulated CI listeners) and 30 adults with the original stimuli (control group listeners). The noise-vocoded stimuli were generated by a software-implemented vocoder designed to simulate a 22 channel implant type (Langner and Jürgens, 2016). A picture selection task with eye-tracking was carried out to test comprehension of subject-, object- and passive questions. Two additional screening tasks tested comprehension of verb-agreement in canonical sentences and auditory discrimination of case (e.g. der vs den).

The simulated CI listeners (only those who performed well on the screening tasks) comprehended object questions worse than the control group listeners (GLMER: $p<0.001$; simulated CI listeners: 83%; control group listeners: 97%). The comprehension of subject- and passive questions was not decreased by the simulated CI input. At the moment, the gaze data is being analyzed to gain more insight into the online processing of which-questions by simulated CI input.

The offline results of the simulated CI listeners show a similar pattern compared to children with actual CIs, in that only the performance on object questions was decreased to a similar extent. This suggests that the degraded speech input of the CIs plays a big role in the explanation of children with CIs' problems with object questions.

32 Listening in noise: what can ability to memorise tones tell us about hearing impaired listeners?

M. Serman
Sivantos GmbH, Erlangen, Germany

D. Kunke
Max Planck Institute, Leipzig, Germany

R. L. Fischer, R. Hannemann, K. Kallisch
Sivantos GmbH, Erlangen, Germany

J. Obleser
Dept. of Psychology, Lübeck, Germany

Currently, hearing aids are adjusted on the basis of pure tone audiometry. In other words, we are trying to infer an individual's communication difficulties in every day situations, based on their ability to hear a few barely audible static pure tones. In reality, the nature of living organisms is such that reactions to dynamic changes in sensory inputs are often more important than reactions to static inputs. We are living, breathing change detectors. Tracking changes in the environment requires attention and memory, both of which enable the comparison of past and present events. This is particularly true in noisy environments.

We have been investigating the relationship between auditory attention, memory and every day listening skills in normal hearing (n = 139, mean age = 64.0 years) and hearing impaired (n = 199, mean age = 66.6 years) listeners. In developing a task that relates attention and short term memory, we combined a recognition memory for pitch task (Deutsch, 1972) and a dynamic attending task (Jones et al., 2002). The resulting memory for tones task consists of auditory rhythmic sequences of low frequency tones (a same–different judgment of a standard and a target tone, separated by a series of six intervening distracting tones). Subjective listening skills were assessed using the short form of Speech, Spatial and Qualities of Hearing Scale (Gatehouse and Noble, 2004; Noble et al. 2013), and questions assessing an individual's distractibility in noisy environments.

We have found significant correlations between the outcomes of the tone-memorisation task and self-assessed communication difficulties in noise, as well as auditory distractibility, in both normal hearing and hearing impaired listeners. Future research will further assess the ability of this task to provide a link between listening skills, individual performance and preference for different hearing aid settings.

33 Exploring the role of working memory for speech perception in energetic and informational masking: a dual-task study

A. Dryden
MRC Institute of Hearing Research, University of Nottingham, UK

H. Allen
University of Nottingham, UK

A. Heinrich
MRC Institute of Hearing Research, Nottingham, UK

Working Memory, a limited-capacity process, which supports simultaneous storage and manipulation of information, is thought to be important for any task which unfolds in time. Speech perception is one example, particularly when the speech is presented in a background noise. The nature of the background noise is often described in terms of energetic or informational masking: energetic masking occurs due to a physically overlap between target and masker signals while informational masking is mainly due to distractions and intrusions of masker information. We hypothesize that working memory is differently engaged depending on masking characteristics. Particularly, energetic masking may necessitate the engagement of the manipulation aspect of working memory in order to restore speech fragments into an intelligible signal. In contrast, informational masking may necessitate the engagement of both storage (of an intelligible but distracting masker) and manipulation (involved in the restauration of speech fragments).

We investigated these hypotheses with a dual-task paradigm. The primary task was a Speech-in-Noise task where low predictability sentences were presented in two types of noise, speech-modulated noise (energetic) and 3-talker babble (informational). The secondary task was one of four secondary working memory tasks, which required for successful completion either storage processes alone or storage and manipulation processes, and which engaged either the verbal or non-verbal domain. The different working memory aspects were operationalised by the following tasks: digit span forwards - verbal storage; digit span backwards - verbal storage and manipulation, Corsi span forwards - non-verbal storage; Corsi span backwards - non-verbal storage and manipulation.

In concordance with our hypotheses we predicted that reverse span measures would cause disruption of speech intelligibility in both types of masking whereas forward span measures would disrupt intelligibility disproportionally to informational versus energetic masking. Moreover, only verbal storage measures would affect intelligibility in information masking due to the modality specificity of the storage component.

34 Lexical frequency effects in noise-induced robust misperceptions

M. Cooke
University of the Basque Country, Spain

R. Marxer
University of Sheffield, UK

M. L. Garcia Lecumberri
University of the Basque Country, Spain

J. Barker
University of Sheffield, UK

While many studies have investigated low-level masking effects of noise on speech, the influence of prior expectations on the speech percept is less clear. One such higher-level factor is lexical frequency. For speech produced in quiet conditions, the current consensus is that misperceptions reported by listeners have a similar average word frequency as the target word [1]. However, a recent study [2] showed that misperceptions of words presented in six-talker babble were of higher frequency than target words.

The current investigation extends the study of lexical frequency effects in noise to several masker types and languages. We analyse consistent word-level misperceptions (defined as no fewer than six participants reporting the same outcome) made in the presence of stationary noise, temporally-modulated noise and four-talker babble, using both English [3] and Spanish [4] corpora of consistent confusions. Lexical frequencies in Zipfs (log10 occurrences/billion word forms) of the target and confused word were examined for some 5150 consistent misperceptions.

Across all maskers and for both languages, misperceptions had higher lexical frequency than target words, supporting the findings of [2]. However, the difference in frequency was much smaller for misperceptions elicited in babble than for speech-shaped or temporally-modulated noise. Lexical frequency effects also varied for misperceptions involving a single phoneme insertion, deletion, or substitution; deletions led to somewhat more frequent reported words than substitutions, while confusions involving insertions led to words that were as frequent as the targets.

We speculate that in situations where the principal effect of masking is to reduce the audibility of information-bearing speech elements, listeners have more freedom to hypothesise words that match the residual parts, and that word frequency statistics influence the final outcome. However, when listeners can also recruit information from the masker, the resulting misperception may be constrained by matching the word hypothesis to the acoustic evidence, leaving lexical frequency with less of a role.

[1] Tang, K (2015). Naturalistic Speech Misperceptions. PhD Thesis, UCL.

[2] Felty, R.A., Buchwald, A., Gruenenfelder, T.M., and Pisoni, D.B. (2013). Misperceptions of spoken words: Data from a random sample of American English words. JASA 134: 572-585.

[3] Marxer, R., Barker, J., Cooke, M., and Garcia Lecumberri, M.L. (in press). A corpus of noise-induced word misperceptions for English. JASA-EL.

[4] Toth, M.A., Garcia Lecumberri, M.L., Tang, Y., and Cooke, M. (2015). A corpus of noise-induced word misperceptions for Spanish. JASA 137:EL184–EL189.

35 Rhythm in plain and Lombard speech

H. R. Bosker
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

M. Cooke
Ikerbasque (Basque Science Foundation), Bilbao, Spain

Speech is an inherently rhythmic signal. Even though speech is by no means strictly periodic, energy patterns in speech are constrained by the physiological dynamics of the lips, jaw, and tongue. As such, energy fluctuations in speech typically occur within the 2-20 Hz range.

Recent research suggests that this rhythmicity in the speech signal plays a central role in comprehension, facilitating the processing of the signal. For instance, when the slow amplitude modulations present in typical speech are destroyed or filtered out, intelligibility drops considerably [1]. Neural mechanisms involving endogenous oscillations phase-locking to the energy fluctuations in speech have been suggested to account for these findings [2].

Given the beneficial effects of rhythmicity in comprehension, this study investigated whether speakers actually make use of increased rhythmicity in their speech (i.e., more regular alternations between high and low amplitude intervals) to improve their intelligibility in acoustically challenging listening conditions (e.g., background noise). Rhythmicity was operationalized by analyzing the modulation spectrum of speech, which represents the spectral content of the signal's amplitude envelope.

Four different corpora were analyzed (varying sample sizes; varying numbers of talkers), each including plain speech (sentences produced in quiet) and matched Lombard speech (same sentences produced in noise). Each sentence was first normalized in amplitude by RMS scaling, thus avoiding intensity confounds. The envelope of the normalized signal was then submitted to a Fast Fourier Transform (FFT), resulting in the modulation spectrum of that sentence. Comparing the average modulation spectra of plain and Lombard speech revealed greater power in Lombard speech in the delta band (1-3 Hz) across all four corpora. Comparison with previous analyses of the speech rate in plain and Lombard speech revealed that this power difference in the delta band could not be attributed to overall slower speaking rates in Lombard speech.

These findings suggest that speakers produce more rhythmic speech, particularly in the 1-3 Hz range, when talking in noise (vs. in quiet). Results are discussed in terms of the functional role of rhythmicity in dialogue and potential underlying neurocognitive mechanisms (e.g., neural oscillatory dynamics).

References

[1] Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Frontiers in Psychology, 3, 238.

[2] Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehensionzr by facilitating perceptual parsing. NeuroImage, 85, 761-768.

36 Effects of age and multiple talking faces on the visual speech advantage in noise

J. M. Beadle, C. Davis, J. Kim
MARCS Institute, Western Sydney University, Sydney, Australia

Understanding speech in noise is challenging, especially for older adults. Seeing a talker's face helps speech recognition in noise compared to an auditory only baseline (i.e., a visual speech advantage); however, previous work on the visual speech advantage is limited in that it has only presented listeners with a single talking face. In a cocktail party environment, more than one talking face is likely in an individual's visual field. This study aimed to evaluate how multiple sources of visual speech effect speech perception in noise for younger and older adults.

Younger and older adults completed the exact same version of the experiment (i.e., volume and SNR were consistent across groups). For each trial, participants listened to a sentence spoken in speech shaped noise (SNR=-3dB). Concurrently, 1, 2, 4, or 6 talking faces were presented. One talking face always matched the auditory signal; the other face(s) did not. Twenty percent of trials were auditory only, that is, the visual stimulus consisted of a static picture of a face or faces. Participants typed what they heard and were scored on key words correct.

For younger adults, all visual speech conditions provided a visual speech advantage. The advantage was greatest for the single face condition and declined as more talking faces were added. There was no difference between the 4 and 6 talking face conditions.

Preliminary results from a group of older adults also revealed a significant visual speech advantage for the single face condition. When any additional talking faces were added, however, older adults' scores were no better than in the auditory only condition.

These results suggest that processing a talker's visual speech requires attention. In the conditions with multiple talking faces, participants likely had to locate the matching talking face to gain a visual speech advantage. Older adults seem to be particularly distracted by multiple sources of visual speech. Susceptibility to distraction from visual stimuli could contribute to difficulties understanding speech in cocktail party listening environments.

37 The discrimination of voice cues in simulations of bimodal electro-acoustic cochlear-implant hearing

D. Başkent, A. Luckmann, J. Ceha
University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology / Head and Neck Surgery, Groningen, Netherlands

E. Gaudrain
Lyon Neuroscience Research Center, CNRS, Lyon, France

T. Tamati
University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology / Head and Neck Surgery, Groningen, Netherlands

Normal-hearing (NH) listeners enhance speech perception by taking advantage of talker's voice cues to separate and selectively attend to speech streams from multiple talkers (cocktail-party listening). The two voice cues, fundamental frequency (F0) and vocal-tract length (VTL), are particularly effective for voice discrimination. Due to the limitations of signal transmission through cochlear implants (CIs), most CI users show poor use of these cues. However, CI users with residual hearing in the non-implanted ear have been shown to benefit from additional speech information conveyed via acoustic hearing (usually with a hearing aid), even if the residual hearing is limited to low frequencies. Since some voice cues are also present in this frequency range, bimodal electric-acoustic hearing in CI users may also result in better discrimination of voice cues. We also expect that bimodal hearing may be particularly beneficial for discrimination of F0 voice cue, because low-frequency speech provides more salient cues for F0 than VTL.

In the current study, we investigated the potential benefits of bimodal hearing in the perception of F0 and VTL voice cues using acoustic simulations of CIs. The just noticeable differences (JNDs) for F0 and VTL were measured in an adaptive three-alternative forced-choice voice discrimination task using triplets of CV syllables. The task was to identify the odd triplet that differed from the standard triplets in F0 or VTL. The bimodal hearing was simulated by presenting low-pass filtered speech (LPF; cutoff frequencies 150 Hz and 300 Hz) in one ear to simulate residual hearing, and noise-band vocoded speech (Voc; 4, 8, and 16 spectral channels) in the other ear to simulate electric hearing. An unprocessed condition, LPF-only conditions, and Voc-only conditions were also included for comparison.

Results showed that F0 JNDs in the bimodal conditions were significantly smaller than in the Voc-only conditions (for all number of channels). There were no significant differences between the 150 Hz and 300 Hz LPF conditions. For VTL, no benefit for bimodal hearing was found. Thus, low-frequency information from the LPF speech improved F0 discrimination in CI-simulated speech. This suggests that low-frequency acoustic information provides a more salient F0 cue than what is conveyed by the noise-band vocoded part, and listeners can potentially exploit it to more accurately discriminate voices. The findings are consistent with previous studies showing a benefit for bimodal hearing in tasks involving a strong pitch component, such as in music or speech-in-noise perception.

38 Common Sound Scenarios – A context-driven categorization of everyday sound environments for application in hearing-device research

F. Wolters, **K. Smeds**, E. Schmidt, E. K. Christensen, C. Norup
ORCA Europe, Widex A/S

Evaluation of hearing-device signal-processing features are performed for research and development purposes, but also in clinical settings. Most people agree that the benefit experienced in the hearing-device user's daily life is most important, but laboratory tests are popular since they can be performed uniformly for all participants in a study using sensitive outcome measures. A number of research groups are currently focusing on developing laboratory tests that have the potential of indicating real-life benefit. In order to succeed with this work, there is a need for more information about the listening situations people encounter. The description has to include more than a description of the acoustical parameters that define the environment.

The purpose of the current study was to investigate the acoustic environments and listening situations people encounter and to provide a structured framework of common sound scenarios that can be used for instance when designing realistic laboratory tests.

A literature search was conducted. Focus was on studies including informants who reported or recorded information about acoustic environments or listening situations in field trials. Nine relevant references were found. In combination with data collected at our laboratory, 187 examples of acoustic environments or listening situations were found. These examples were categorized using a context-based approach. A structured framework of common sound scenarios (CoSS) was developed. Three intention categories, "Speech communication", "Focused listening", and "Non-specific" were divided into seven task categories. For each task category, two common sound scenarios were selected and described, creating in total 14 common sound scenarios. For each common sound scenario, information about occurrence, difficulty to hear, and importance to hear in the scenario was added based on the available data.

The framework has the potential to be used when developing hearing-device signal-processing features, in the evaluation of such features in realistic laboratory tests, and for demonstration of feature effects to hearing-device wearers. The presentation/poster will describe the CoSS framework, how it was developed, how it can be used, and its weaknesses. Need for further research in the area will also be outlined.

39 Influence of background noise level on speech perception

M. A. Zokoll
Medical Physics and Cluster of Excellence „Hearing4all“, Carl von Ossietzky Universität Oldenburg, Germany

N. Wardenga, H. Maier
Department of Otolaryngology and Cluster of Excellence „Hearing4all“, Hannover Medical School, Hannover, Germany

T. Brand, B. Kollmeier
Medical Physics and Cluster of Excellence „Hearing4all“, Carl von Ossietzky Universität Oldenburg, Germany

Speech audiometry in quiet assesses hearing ability, while speech audiometry in noise it is thought to assess possible supra-threshold distortion caused by the hearing impairment. In Germany, the Oldenburg sentence test (OLSA, e.g., Wagener et al., 1999) is commonly used to determine the listener's speech reception threshold (SRT) in noise, i.e., the signal-to-noise ratio (SNR) yielding 50% speech intelligibility. It uses an adaptive procedure with the OLSA noise (Olnoise) level fixed at 65 dB SPL by default and varying speech level. In an elaborate study, Wardenga and colleagues (Wardenga et al., 2015) tested 177 subjects with different hearing impairment in this standard setting. On the basis of their results they concluded that the OLSA can be applied to subjects with a wide range of hearing losses. However, with the standard noise level of 65 dB SPL, the SRT is determined by listening in noise only for listeners with PTAs below about 47 dB HL. Above 47 dB HL it is determined by listening in quiet. For routine clinical practice however, it is important to know the upper limit of the validity of the SRT in noise for any noise level applied. The current study systematically investigates the effect of different background noise level on the transition point between testing in noise and testing in quiet. 10 listeners with normal hearing and 30 listeners with PTAs (0.5, 1, 2, 4 kHz) between approximately 30 and 50 dB HL were tested in monaural headphone measurements to obtain SRTs for the stationary Olnoise, as well as for the fluctuating ICRA5_250 noise at noise level between 55 and 85 dB SPL. The results of the study suggested that the transition point between testing in noise and testing in quiet shifts linearly with the noise level for the Olnoise. Conform to Plomp (1978), differences in SRTs between listeners with different grades of hearing impairment decreased with increasing background noise level. For ICRA5_250, however, no transition point can be found and differences in SRTs between listeners of different hearing impairment were preserved across all noise levels tested, suggesting an additional diagnostic value of speech audiometry in fluctuating noise.

40 Speech intelligibility in noise in a virtual restaurant: the difficulties faced by those with hearing loss

B. Bardsley, J. F. Culling
Cardiff University, Wales, UK

To provide a more realistic speech-in-noise test, binaural room impulses responses (BRIRs) were recorded in a restaurant with interferers at varying table positions and the target sat directly opposite. IEEE sentences from a male speaker were used as the target material. Four male and four female continuous speech recordings were used to make interferers. These recordings were used to create 4 types of interferer; unmodulated speech-shaped noise, modulated speech-shaped noise, speech and reversed speech. The number of interferers was also varied, 1 male, 1 female, 2 (mix), 4 (mix) and 8 (mix) interferers were convolved with the BRIRs from different tables positions. The overall level was adjusted to compensate for increased energy from extra interferers. This approach provides acoustic features relevant to every day listening, such as reverberation and directivity of interferer, and provides detail on performance in varying speech-to-noise ratios in a realistic environment. Deviations from complete realism such as the inclusion of reversed speech also allow an analysis of different perceptual mechanisms.

In a similar study, Culling (2016) has shown that young, normal hearing listeners have Speech Reception Thresholds (SRTs) that progressively deteriorate with increasing numbers of interferers for modulated speech-shaped noise, speech and reversed speech. However, this effect was not seen with an increasing number of unmodulated speech-shaped noises. Typically the speech and reversed speech resulted in the lowest SRTs which may reflect the benefit of modulation and harmonicity. Evidence favouring informational masking was demonstrated but in a limited condition; masking using two reversed-speech interferers resulted in lower SRTs than for the similar number of speech interferers.

This study recruited older participants with sensori-neural hearing loss and age matched controls with normal hearing thresholds. The results suggest that in realistic reverberant environments, interferer number has a degrading influence on performance in the hearing impaired group, much like that seen in those with normal hearing thresholds. It appears they can take advantage of the modulations and harmonicity in the interferer. The role of audibility is also investigated.

References

Culling, J.F. (2016). Speech intelligibility in virtual restaurants. Journal of the Acoustical Society of America, 140, 2418-2426.

Supported by the BSA research grant and Action on Hearing Loss flexi grant.

41 Unaided and aided speech recognition performance evaluation across languages with hearing impaired patients using multilingual matrix sentence tests

M. A. Zokoll, A. Warzybok
Universität Oldenburg, Medizinische Physik, and Cluster of Excellence Hearing4all, Oldenburg, Germany

M. Buschermöhle
HörTech gGmbH and Cluster of Excellence Hearing4all, Oldenburg, Germany

K. C. Wagener
Hörzentrum Oldenburg and Cluster of Excellence Hearing4all, Oldenburg, Germany

B. Kollmeier
Universität Oldenburg, Medizinische Physik, and Cluster of Excellence Hearing4all, Oldenburg, Germany

Comparing speech recognition performance in patients with and without a certain hearing device across languages and countries is an important goal in order to achieve globally accepted audiological standards and models in diagnostics and auditory rehabilitation. The current contribution reviews the attempts to reach this goal with the development and the clinical validation of the matrix sentence recognition tests in numerous languages as well as its application in rehabilitative audiology. Matrix sentence tests use syntactically fixed, semantically unpredictable sentences (e.g. “Peter kept two green toys”) that allow for repeated measurements with the same speech material, e.g. for hearing instrument fitting and evaluation (see Kollmeier et al., IJA 2015 for a review). Their closed-set construction with well-described optimization principles (Akeroyd et al., IJA 2015) allows for a high comparability across languages, the possibility to objectively predict the individual performance with an Automatic-Speech-Recognition (ASR) technique (see Kollmeier et al., IHCON 2016), and the option to test a listener in a language the audiometrist does not understand. This property is especially useful in multilingual societies where the fitting and optimization process of the individual´s hearing device should not be dependent on the level of the patient´s language competence for the respective dominant language.

Currently, matrix sentence tests are available in at least 14 languages. They have been developed in a very compatible way and are characterized by similar psychometric properties. Therefore, they are well suited for hearing device evaluation studies or in general for accurate speech audiometry in noise. In order to yield reliable results, initial training with two training lists is recommended. After that, matrix sentence test measurements show a high test-retest reliability and a superior efficiency (in terms of reduction in variability in Speech Recognition Threshold as a function of measurement time) in comparison to other sentence recognition tests. As an example for a clinical cross-validation with other available clinical tests employing listeners with diverse hearing abilities, a multicenter study is rewieved that compared the American English Matrix Sentence Test with the QuickSIN and HINT. The data demonstrate a high sensitivity and specificity as well as a superior test-retest reliability of the matrix sentence test. The comparison of validation data across different languages including English, German, Russian, and Turkish proves the high comparability of the matrix sentence test across languages. The same holds for matrix sentence test applications for aided patient performance measurements that will also be presented and discussed.

42 Isolating the informational component of speech-on-speech masking

A. Calcus
University College London, UK

T. Schoof
Northwestern University, USA

S. Rosen
University College London, UK

B. Shinn-Cunningham
Boston University, USA

P. Souza
Northwestern University, USA

Speech perception in the presence of a competing talker involves both energetic (EM) and informational masking (IM). This study aimed to isolate the informational component of speech-on-speech masking.

EM can be eliminated by presenting the target and masker to opposite ears (i.e. dichotically). However, this also dramatically reduces the effects of IM by providing listeners with lateralization cues. Previous research using tonal sequences has shown that IM can be isolated by presenting the target and masker dichotically while rapidly switching the two streams across the ears. The question remains whether this technique can also be used for speech materials.

Speech reception thresholds (SRTs) were measured for sentences produced by a female talker in the presence of a competing male talker under three conditions: diotic, dichotic, and switching. In the switching condition, target and masker were presented dichotically, but their lateralization was switched after every word in the target sentence.

Results show that SRTs for the switching condition are higher (i.e. poorer) than for the dichotic condition but lower (i.e. better) than for the diotic condition. This suggests that, contrary to findings for tonal sequences, rapidly switching the target and masker speech across the ears preserves important amounts of masking, but does not fully reintroduce IM.

43 A speech-in-noise screening test for hearing loss – Evaluation of different noise types

J. Oygarden
NTNU, Norwegian University of Science and Technology, Trondheim, Norway

The hard of hearing organization in Norway (HLF – Hørselshemmedes Landsforbund) is interested to make a screening test available for people to check their hearing status. After the project is completed the test can be administered from the web-pages of HLF or can be downloaded as an app for smart-phone use. Decision was made to use the last word in the Norwegian Matrix test for this test.

To evaluate the ability of the test to discern between normal hearing persons from those with hearing impairment, speech reception threshold (SRT) in noise was measured for 4 different noise types:

Noise type 1 – Stationary speech noise with HF-reduced. This is the same noise as Noise type 4, but the frequencies above 1400 Hz is reduced by 15 dB.

Noise type 2 – Amplitude modulated speech noise with HF-reduced. Noise type 1 100 % amplitude modulated with a 16 Hz sine tone.

Noise type 3 – Reverse speech with HF-reduced. Matrix sentences played backward, silent periods longer than 200 ms are deleted. 4 recordings on top of each other.

Noise type 4 – Stationary speech noise. This is masking noise in the Norwegian Matrix test.

A pilot app was developed using Matlab software to measure SRT for words in these four noise types mentioned above. Measurements were performed on two groups. Group 1 consisted 28 audiology students with normal hearing. Group 2 consisted 70 persons with varying degrees of hearing loss. Audiometry was also performed for the frequencies of 250, 500, 1000, 2000, 3000, 4000, 6000 and 8000 Hz.

The poster will present results of correlation between SRTs for different noise types and audiometric thresholds, and receiver operating characteristic curves (ROC-curves) in terms of ability of the noise types to discern hearing impaired persons from those with normal hearing.

44 Learning effects in the Danish HINT

L. Bramsløw
Eriksholm Research Centre, Denmark

L. B. Simonsen
University of Copenhagen, Faculty of Humanities, Denmark

M. El Hichou, R. Hashem
University of Southern Denmark, Faculty of Humanities, Denmark

R. K. Hietkamp
Eriksholm Research Centre, Denmark

The benefit from hearing aid signal processing can be tested in the laboratory by means of a speech test, such as the open-set Hearing In Noise Test (HINT, Nilsson et al, 1994). The Danish HINT (Nielsen & Dau, 2011) consists of only 200 test sentences organised in 10 lists. These HINT lists are used heavily in the newly developed Competing Voices Test (Bramsløw et al, 2016) in which pairs of sentences are presented. Thus, reuse of the material will occur multiple times within visits. This can lead to learning of the sentences and possibly confound the test outcome. The change in performance over time has two components; 1) a practice effect from becoming more trained in the listening task (or alternatively, fatigue) and 2) a learning effect due to memorization of the specific sentences.

This poster presents practice and learning effects of the Danish HINT sentences from two experiments. In experiment 1, we used varying degrees of exposure at 3 visits with 3 weeks interval; selected lists were reused 1-9 times altogether. Ten elderly listeners were tested using a default adaptive HINT procedure aiming at 50% correct sentence score in speech shaped noise. Results showed a maximum within visit learning effect of 1.5 dB and a between visit learning effect of 1.1 dB. These effects can be attributed mainly to learning, as practice effect across varying lists was not statistically significant.

In experiment 2, we investigated the effect of exposing at 80% sentence score vs 50% score in two visits with 2 and 4 repetitions of lists. This test used 15 elderly listeners. While the previous learning effects were confirmed, there was no added learning due to the 80% sentence score exposure with its higher word recognition.

The implications of the results for future similar speech tests are discussed. Clever experimental design is proposed to compensate for practice and learning effects.

45 Better-ear rating based on glimpsing and its relation to speech intelligibility

E. Schoenmaker, S. van de Par
University of Oldenburg, Cluster of Excellence Hearing4all, Germany

The better ear is usually expressed as the ear that listens to the best overall signal-to-noise-ratio (SNR). For speech in speech interference, dip-listening and glimpsing play an important role, suggesting that the highest global SNR may not be the optimal indicator of the better ear. Previous work [J. Acoust. Soc. Am. 135, 2190 (2014)] has shown that the better ear can be successfully expressed as the ear that has access to the most salient glimpses.

In the current work the salience of the glimpses at the better ear is quantified and used as a predictor for speech intelligibility in a logatome recognition test with three concurrent talkers at equal presentation levels. The test included three different types of binaural stimuli. The first type contained only interaural time differences as binaural cues („ITD only"), and thus excluded the possibility to listen to a better ear. The second type was the Equal Local Azimuth stimulus („ELA") that has been presented in Schoenmaker et al [J. Acoust. Soc. Am. 139, 2589-2603 (2016)]. These stimuli did not allow for binaural unmasking, but did elicit a percept of distinct spatial talker locations. The final type of stimuli were spatialized using non-individual head-related transfer functions („HRTF") that contained all naturally occurring spatial cues. The use of these different stimulus types can provide insight into the contributions of better-ear listening, binaural unmasking and spatial stream segregation in a situation with multiple spatially separated talkers.

Results show a clear and monotonic dependence of speech intelligibility on the salience of glimpses available to the better ear in each stimulus. The performance increased for increasing angular separation of talkers. A logistic regression model showed that these effects were significant. Interestingly, our results show that for a given separation angle and salience of glimpses a similar speech intelligibility was found that was nearly independent of the type of spatial cues contained in the stimuli. Two implications are especially worth noting. First, the better-ear advantage of the full-cue HRTF stimuli is fully compensated for when the salience of glimpses at the better ear is considered. Second, binaural unmasking did not play a role in this three-talker scene, but rather did stream segregation based on perceived spatial location. This suggests that better-ear listening based on glimpsing and spatial stream segregation are the only two contributing factors to the spatial benefit in this spatial three-talker situation.

46 The AVATAR-approach: how real-life listening affects speech intelligibility and listening effort

A. Devesse, A. Dudek, A. van Wieringen, J. Wouters
KU Leuven – Department of Neurosciences, ExpORL, Belgium

Having a successful conversation does not only depend on good auditory processing abilities. During complex real-life listening situations, cognition, multitasking skills and the processing of visual (speech) cues are well known to be of particular importance. Nevertheless, current speech-in-noise (SPIN) tests take these aspects only partially into account since they are developed from a static, one-dimensional auditory perspective. This results in a large variation in outcome measurements of persons with hearing impairment and hearing aid-users.

To close the gap between well-controlled SPIN measurements and self-report approaches, we developed the AVATAR-paradigm: a comprehensive method and test set-up for the real-life assessment of auditory functioning. A unimodal SPIN-measurement is extended to a multitasking test that incorporates both auditory and visual cues as well as cognition. Realistic listening scenarios are presented on a large screen, including computer animated avatars speaking to the test person.

The present project employed the AVATAR-approach to investigate the effect of multitasking on speech intelligibility (SI) and listening effort (LE). Multiple tasks were combined into listening scenarios of increasing complexity. In the most difficult situation the test person had to execute a SPIN-task together with three secondary tasks: an auditory localization test, detection of direction of sounds passing by and short-term memory storage of visual stimuli, imposing an extra cognitive load. Results showed that, for normal hearing adults of 18-30 years old, SI was robust to the amount of tasks that had to be performed simultaneously. On the contrary, LE during speech-processing significantly increased when the listening situation became more complex. Furthermore, the impact of multitasking on LE was more prominent for older and/or hearing impaired adults. Implications of these results will be discussed, together with data on the feasibility and sensitivity of the AVATAR-approach.

47 A test word selection and optimization method for a new Swedish test of phonetic perception in noise

E. Witte, C. Möller
Audiological Research Centre, Örebro Sweden

S. Köbler
Örebro University, Sweden

J. Ekeroot
Audiological Research Centre, Örebro Sweden

The present project aims to develop a highly reliable, and clinically suitable, Swedish language test of phonetic perception in background noise for adult persons with hearing loss. The test will be a multiple choice rhyme test based on test word groups consisting of real words with minimal phonemic contrast.

As previous literature clearly points out, for a speech perception test to be highly reliable, all contrasting test items need to be as equally difficult as possible. This is often achieved through a process of test stimuli optimization. The present study uses an optimization strategy based on test word perceptual probability. This approach is based on a thorough linguistic analysis in which a set of variables derived from measures previously proven to affect access to the mental lexicon, namely word frequency, phonetic neighborhood density, and phonotactic probability, were calculated for each word in a large corpus.

In order to accomplish this, a large word list of high quality phonetic transcriptions would be needed. Unfortunately, no such word list was readily available in Swedish. There exists, however, a large lexical database (NST), consisting of many, but unfortunately often erroneous, phonetic transcriptions. Using the NST-database as a starting point, a revised and updated word list containing around 800 000 words with phonetic transcriptions reflecting natural everyday pronunciation of the central Swedish dialect has been compiled. To this list, word frequency data from a corpus of Swedish internet blogs with a size around 500 000 000 tokens have been added. In addition, the word list also contains data on spelling regularity, special orthographic characters, homographs and homophones, abbreviations/acronyms, non-Swedish words, as well as word class information.

In order to select the most suitable test words, this new list were utilized in order to identify the test word groups that have the least intra-group variation in word frequency, phonetic neighborhood density and phonotactic probabilities.

Equal perceptibility within these selected test word groups will then be attained through an acoustic optimization process, in which the sound levels of each contrasting phoneme (rather than of the whole test word, which is common procedure) are experimentally adjusted until a predefined degree of intelligibility is reached.

Our hypothesis is that this careful test word selection process will prove to decrease the amount of acoustic optimization needed to achieve equal perceptibility within the test word groups, and thereby increase both the reliability and the ecological validity of the test material.

48 Modulating speaker- and language-specific effects in speech intelligibility

S. Hochmuth, M. R. Schädler
Medizinische Physik and Cluster of Excellence Hearing4all, Universität Oldenburg, Germany

B. Kollmeier
Cluster of Excellence Hearing4all, Universität Oldenburg, Germany

Recently, the simulation framework for auditory discrimination experiments (FADE) was presented which could successfully predict speech reception thresholds (SRTs) for the German matrix sentence test in stationary, speech-shaped noise. In principle, FADE consists of a standard automatic speech recognition system using separable Gabor filter bank features and hidden Markov models as back-end. In this contribution, we investigated if FADE could also predict speaker- and language specific effects on SRTs measured with normal-hearing listeners using matrix sentence tests of different languages. Such effects were recently reported for speech materials recorded with bilingual Spanish/German and German/Russian speakers of almost accent-free pronunciation of both of their languages. The data showed that SRT differences between different speakers within one language were generally larger than differences between languages. Speaker-specific intelligibility transferred across languages, i.e., speakers that were well intelligible in one of their languages were also well-intelligible in the other language. Speaker-specific intelligibility also transferred across noise types and were observed for stationary noise, amplitude-modulated noise and multi-talker babble. A systematic language effect was observed in that Spanish had consistently higher SRTs. The present study tested FADE on these data. The model was generally able to model speech in noise perception for different talkers and also different languages. Performance was generally overestimated in stationary speech-shaped noise by about 3 dB. In modulated noise and multi-talker babble the predictions matched well the data. Across all noise types, speakers and languages, FADE could predict 86% of the observed SRT variance. FADE predictions are also compared to predictions of the intelligibility index (SII).

49 The role of short-time power and envelope power SNRs in psychoacoustic masking and speech intelligibility

T. Biberger, S. D. Ewert
Medizinische Physik and Cluster of Excellence Hearing4all, Universität Oldenburg, Germany

Effects of spectral masking in psychoacoustics can be explained with the power-spectrum model (PSM) of masking. Similarly the envelope power spectrum-model (EPSM) has been suggested to account for masking in the envelope domain. Recently, Biberger and Ewert [(2016). J. Acoust. Soc. Am. 140, 1023-1038] proposed the generalized power spectrum model (GPSM) which combines the concepts of the PSM and EPSM. The GPSM was shown to account for a broad variety of data from psychoacoustic and speech intelligibility (SI) experiments. In the suggested GPSM, the PSM path of the model uses long-time power signal-to-noise ratios (SNRs), while the EPSM path uses short-time envelope power SNRs. A recent study of Schubotz et al. [(2016). J. Acoust. Soc. Am. 140, 524-540] showed that short-time power features are important to account for SI for several spectro-temporal manipulations of speech maskers and gender combinations of target and masker speakers. Here an extension of the GPSM is suggested, where both the envelope power SNRs and power SNRs are calculated on short-time scales. In contrast to Biberger and Ewert (2016), the envelope power SNRs and power SNRs are combined by applying a maximum operation, where only the most contributing domain is considered, instead of using an additive combination of envelope power and power SNRs. The proposed model is shown to account for a critical set of psychoacoustic experiments and for SI in a variety of noise- and speech-like maskers, reverberation and spectral subtraction. Model predictions are compared to those of the extended speech intelligibility index (ESII) and the multi-resolution speech-based EPSM, demonstrating that the current approach shows the highest predictive power. The contribution of amplitude modulation masking and energetic masking in the different psychoacoustic and SI experiments is analyzed using the suggested model.

50 Common Audiological Functional Parameters (CAFPAs): an abstract representation of audiological expert data

M. Buhl
Medizinische Physik and Cluster of Excellence Hearing4all, CvO Universität Oldenburg, Germany

M. R. Schädler
Medical Physics and Medizinische Physik and Cluster of Excellence Hearing4all, CvO Universität Oldenburg, Germany

A. Warzybok, B. Kollmeier
Medical Physics and Medizinische Physik and Cluster of Excellence Hearing4all, CvO Universität Oldenburg, Germany

Correct diagnosis of hearing impairment and indication for individual treatment is a complex task influenced by many factors like the choice of audiological measurements, the most frequent cases at a clinic's location or the experience of the ENT doctor. The goal of the Common Audiological

Functional Parameters (CAFPAs) approach is to build up a supporting tool uniting expert knowledge and providing it to the community of ENT doctors.

The CAFPAs are designed as an abstract representation of audiological measurements, thereby generalizing over different measurements and representing functional principles of the auditory system. First data for CAFPAs and measurement results, for different, predefined diagnostical cases and indications for treatment, was collected by the use of a survey among experts from Hanover and Oldenburg. With this data, it is possible to distinguish different diagnostical cases on the basis of audiological measurements as well as on the basis of the newly-introduced CAFPAs. Thus, the CAFPAs serve as an abstract and compact, but interpretable representation of the measurements that do not take into account every detail. Furthermore, the survey results show that the degree of agreement of the experts on the measurement and CAFPA outcomes differed for different diagnostical cases. This property supports the idea of using statistical methods, e.g. machine learning methods, for modeling the audiological diagnostical path.

51 Computational auditory scene analysis in multi-talker environments

A. Josupeit, J. Luberadzka, V. Hohmann
Medizinische Physik and Cluster of Excellence Hearing4all, Universität Oldenburg, Germany

In many everyday situations, listeners are confronted with complex acoustic scenes. Despite the complexity of these scenes, they are still able to follow and understand one particular talker. This contribution presents auditory models that aim to solve different speech-related tasks in multi-talker settings. The main characteristics of the models are: (1) restriction to salient auditory features ("glimpses"); (2) usage of periodicity, periodic energy, and binaural features; and (3) template-based classification methods using clean speech models. The model performance is evaluated on the bases of (already existing) human psychoacoustic data [e.g., Brungart and Simpson, Perception & Psychophysics, 2007, 69 (1), 79-91]. The model results were found to be similar to the subject results. This suggests that sparse glimpses of periodicity-related monaural and binaural auditory features provide sufficient information about a complex auditory scene involving multiple talkers. Furthermore, it can be concluded that the usage of clean speech models is sufficient to decode speech information from the glimpses derived from a complex scene, i.e., computationally complex models of sound source superposition are not required, even in complex scenes.

52 A correlation metric in the envelope power spectrum domain for speech intelligibility prediction

H. Relaño Iborra, T. Dau, T. May, J. Zaar, C. Scheidiger
Hearing Systems group, Department of Electrical Engineering, Technical University of Denmark

A powerful tool to investigate speech perception is the use of speech intelligibility prediction models. Recently, a model was presented, termed correlation-based speech-based envelope power spectrum model (sEPSMcorr), that uses a correlation-based back end at the output of an audio-frequency and modulation-frequency selective auditory preprocessing (Relaño-Iborra et al., 2016). The use of the correlation back-end extended the predictive power of earlier versions of the sEPSM framework (e.g. Jørgensen et al. 2013) towards conditions of non-linear signal processing, such as phase jitter and ideal binary mask processing. Moreover, the model was shown to account for conditions with fluctuating interferers, unlike other correlation-based models.

Here, the back end of the sEPSMcorr was combined with a more realistic auditory pre-processing front end adopted from the computational auditory signal processing and perception model (CASP; Jepsen et al., 2008). The preprocessing contains outer- and middle-ear filtering and a non-linear auditory filterbank (DRNL, López-Poveda and Meddis, 2001), followed by inner hair-cell transduction, adaptation and a modulation filterbank.

The predictions were compared to measured data in conditions of additive masking noise, phase jitter distortions, reverberation and noise-reduction algorithms. The effects of the back end as well as the different preprocessing stages on the predicted results were analyzed. The modelling framework could be useful for the design and evaluation of, e.g. speech transmission algorithms or hearing-instrument algorithms.

53 Predicting effects of additive noise and hearing-instrument signal processing on consonant recognition and confusions

J. Zaar, T. Dau
Hearing Systems Group, Technical University of Denmark, Kgs. Lyngby, Denmark

The perception of consonants has been investigated in various studies and shown to critically depend on fine details in the stimuli. In the present study, a microscopic speech perception model is proposed, which combines an auditory processing front end with a correlation-based template matching back end. The model represents an extension of the auditory signal processing model by Dau et al. [(1997). J. Acoust. Soc. Am. 102, 2892–2905] towards predicting microscopic speech perception data. It was evaluated based on the extensive consonant perception data set provided by Zaar and Dau [(2015). J. Acoust. Soc. Am. 138, 1253–1267], which was obtained with normal-hearing (NH) listeners using 15 consonant-vowel combinations (CVs) mixed with white noise. Accurate predictions of the consonant recognition scores were obtained across a large range of signal-to-noise ratios. Furthermore, the model yielded convincing predictions of the consonant confusion scores, such that the predicted errors were clustered in perceptually plausible confusion groups.

The model was further evaluated with respect to perceptual artifacts induced by hearing-aid (HA) and simulated cochlear-implant (CI) processing in NH listeners. In terms of HA processing, effects of strong nonlinear frequency compression and impulse-noise suppression were measured in 10 NH listeners using CV stimuli. Regarding the simulated CI processing, the consonant perception data from DiNino et al. [(2016). J. Acoust. Soc. Am., under review] were considered, which were obtained with noise-vocoded vowel-consonant-vowel (VCV) stimuli in 12 NH listeners. Both the HA and the simulated CI processing induced strong perceptual confusions of specific consonants, whereas other consonants remained perceptually unaffected. The model predictions obtained for the two data sets showed a large agreement with the perceptual data both in terms of consonant recognition and confusions, demonstrating the model's sensitivity to supra-threshold effects of hearing-instrument signal processing on consonant perception.

Overall, the large predictive power of the proposed model suggests that adaptive processes in the auditory preprocessing in combination with a cross-correlation based template-matching back end can account for some of the processes underlying consonant perception in normal-hearing listeners. The proposed model may provide a valuable framework for the evaluation of hearing-instrument processing strategies, particularly when combined with simulations of individual hearing impairment.

54 Prediction of speech-in-noise intelligibility by hearing-impaired listeners: a re-analysis of Summers et al. (2013) auditory processing data

M. A. Huckvale, G. Hilkhuysen
University College London, UK

Data collected by Summers, Makashay, Theodoroff & Leek (2013) is re-analysed using statistical learning methods to determine the accuracy with which the intelligibility of speech in noise to hearing-impaired listeners may be predicted from psychoacoustic measurements of their hearing ability. Summers et al suggest that FM detection thresholds make some contribution to individual intelligibility scores, but we show that measures of peripheral compression and FM detection thresholds made little or no improvement to the prediction of individual listener speech reception thresholds beyond that obtained using hearing thresholds alone. A small improvement in prediction of listener SRT was obtained using auditory bandwidth measurements, but the benefit may not be useful in practice. A good estimate of individual listener intelligibility scores at different SNR values was obtained using a single intelligibility score and an average performance function slope.

Our main conclusions are as follows:

1. Variation in speech intelligibility across impaired listeners seems to be well modelled by the concept of a "listener SRT", that is a single SNR value that delivers 50% intelligibility to the listener.

2. Correlational analysis of the psychoacoustic measures shows that a great deal of variation across features within a listener may be captured by a small number of principal components.

3. Cross-validated predictions of listener SRT shows little utility for the Compression and FM threshold features once Hearing Thresholds have been taken into account.

4. Support Vector Regression seems to produce better cross-validated predictions of intelligibility than linear regression. Listener SRTs can be predicted from the hearing thresholds alone with a Mean Absolute Error (MAE) of less than 1.5dB for speech-shaped noise and less than 2.2dB for modulated noise.

5. Individual performance scores for different SNR conditions can be predicted to within an MAE of about 1Berkson from the psychoacoustic data, but considerably better predictions (of about 0.6Berkson) can be obtained if a single intelligibility score is used instead of the predicted SRT.

Reference

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

55 Performance evaluation of the short-time objective intelligibility measure with different band importance functions

A. H. Andersen
Oticon A/S | Aalborg University, Denmark

J. M. de Haan
Oticon A/S, Denmark

Z. H. Tan
Aalborg University, Denmark

J. Jensen
Oticon A/S | Aalborg University, Denmark

Methods for speech intelligibility prediction are quickly becoming popular tools within the speech processing community. Such methods can easily and objectively estimate the effectiveness of different speech enhancement schemes. The short-time objective intelligibility (STOI) measure has enjoyed particular popularity due to its simplicity and its proven ability to provide accurate predictions across a wide range of conditions. The STOI measure has a simple structure which is similar to many other intelligibility measures: 1) clean and degraded speech signals are split into one-third octave bands with a filter bank, 2) envelopes are extracted from each band, 3) the temporal correlation between clean and degraded envelopes is computed in short time segments, and 4) the correlation is averaged across time and frequency bands to obtain the final output. An unusual choice in the design of the STOI measure, is that all frequency bands are equally weighted in the final measure. This is in contrast to classical methods such as the speech intelligibility index (SII) which employs empirically determined band importance functions (BIFs), specifying the relative contribution of each frequency band to intelligibility.

In this study we investigated the use of BIFs in the STOI measure. BIFs were fitted to several datasets of measured intelligibility. This was done such as to minimize the root-mean-squared prediction error. We then performed a cross-evaluation of the obtained BIFs on all datasets, using three different performance measures: root-mean-squared-error, Pearson correlation, and Kendall rank correlation. The results show substantially improved performance when fitting and evaluating on the same dataset. However, this advantage does not necessarily subsist when fitting and evaluating on different datasets. When there are big differences between the datasets used for fitting and evaluating, poor performance may result. In contrast, the uniform BIF used in the original STOI measure leads to decent performance across all datasets. We therefore conclude that, while prediction performance of the STOI measure can be improved considerably under some conditions by the use of fitted BIFs, this should be done with caution.

56 Using models to evaluate whether spectral centroid could play a role in F0 and VTL perception in acoustic and electric hearing

E. Gaudrain
Lyon Neuroscience Research Center, CNRS, Lyon, France

Normal-hearing (NH) listeners rely on two principal voice characteristics — voice pitch (F0) and vocal-tract length (VTL) — to segregate voices in cocktail party situations. Cochlear implant (CI) listeners have been shown to have much larger F0 and VTL discrimination thresholds than NH listeners. However, the mechanisms underlying perception of these cues in CI users, but also in NH listeners, remain largely unknown.

In implants, while some studies argue that F0 can be coded temporally, other recent studies have suggested that spectral centroid (SC) could be used instead. When the F0 changes, the lower frequency channels of the implant are more or less excited, thus shifting the SC. Similarly, while some researchers argue that VTL is perceived through its effect on individual formants, others have argued that, like musical timbre, VTL perception might also rely on SC.

However both these assumptions result from observation from steady-state stimuli. In natural speech, the formant trajectories create a tremendous SC variability which may blur small F0 and/or VTL differences. Using basic auditory models, the variability of perceptual SC in natural speech was evaluated and compared to the effects of VTL and F0 variations.

57 Decoding speaker attendance from EEG-data using deep machine learning in continuous speech

T. de Taillez, B. Kollmeier, B. Meyer
University Oldenburg, Germany

Previous research has investigated the question if signals obtained from EEG can be used to predict which speaker is attended in an acoustic scene. The long-term goal is to provide solutions for hearing aid users using EEG-based speaker selection or optimization. In this work, we analyze EEG data from listeners in a two-speaker scenario and test the application of algorithms borrowed from automatic speech recognition (ASR) to estimate which speaker was attended. Specifically, a deep neural net is trained to predict the envelope of the attended speech signal. We compare our results to previous research [Mirkovic et al., 2015], in which a linear model was applied to obtain the estimate. The DNN-based approach requires shorter data segments to be analyzed for a decision, which is partially explained by the transferred information in the experiment that is four times higher compared to the linear model.

Mirkovic B, Debener S, Jaeger M, De Vos M. “Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications.” Journal of neural engineering 12.4 (2015): 046007.

58 Access to sub-lexical information affects degraded speech processing: insights from fMRI

S. Bräuer, I. M. Schepers
University of Oldenburg, Cluster of Excellence Hearing4all, Germany

G. Mueller
University of Oldenburg, Germany

J. W. Rieger
University of Oldenburg, Cluster of Excellence Hearing4all, Germany

Natural speech perception suffers from uncertainty and ambiguity due to environmental noise. One factor that greatly improves the perception of degraded speech is contextual information. Previous studies on the modulatory effect of contextual information mostly focused on the beneficial role of semantic-lexical knowledge. Several studies manipulated the semantic expectancy of the last word within a sentence (Obleser et al. 2010) or, the semantic relationship between words (Golestani et al. 2013) and whole sentences (Guediche et al. 2016). Others investigated semantic priming of the degraded stimulus material through prior exposure to undistorted written or spoken presentation of the same stimulus (Sohoglu et al. 2012, Clos et al. 2014). However, the modulatory effect of pure sub-lexical speech context on speech intelligibility and the underlying neural responses are not well understood. We conducted a functional magnetic resonance imaging (fMRI) study (N=20) to determine the underlying neuronal activation pattern during perceptual adaptation to degraded sub-lexical speech. To isolate language specific sub-lexical characteristics from semantics our stimulus material consisted of pseudowords, phonetically and phonologically balanced to the German language (Wendt et al. 2007). Subjects listened to three successively presented pseudowords in each trial. The first and third pseudowords were identical and physically degraded. The intermediate pseudoword was intelligible and either matched or did not match the degraded pseudoword. Subjects were instructed to repeat the last presented degraded word. For the assessment of perceptual clarity, we analyzed success or failure of repetition. In addition, we analyzed the blood oxygen level dependent (BOLD) activation changes of the second degraded word in dependency of the intermediate clear word. Matching clear pseudowords compared to non-matching clear pseudowords significantly improved the listener's perception of degraded speech stimuli (from 3.44% to 68.44% correct responses). Speech intelligibility was associated with activation increases in a frontal-temporo-parietal network, including bilateral posterior superior temporal cortex, angular gyrus, right supramarginal gyrus, right middle temporal cortex and left somato-motor cortices ($p<0.001$, uncorrected). To further investigate the adaptational changes in primary auditory areas we conducted a bilateral region of interest (ROI) analysis on the cytoarchitectonic-maps of the primary auditory cortex (Wild et al. 2012). All ROI's showed increased activation when prior exposure to identical clear pseudowords enabled successful speech perception. Our results indicate that the exposure to intact sub-lexical speech characteristics, independent of semantic-lexical context, supports the perception of severely degraded speech and enhances activations along a broad cortical speech perception network.

59 Behavioural and neural consequences of closed eyes during attentive listening

L. M. Schmitt, J. Obleser, M. Wöstmann
Department of Psychology, University of Lübeck, Germany

A common strategy of listeners confronted with a noisy speech signal is to close the eyes. However, the empirical basis on whether closing the eyes actually benefits attentive listening is surprisingly sparse. Interestingly, closing the eyes and attentive listening induce the same neural response: a power increase of neural alpha oscillations (8–12 Hz) in parieto-occipital cortex regions. In the present study, we tested whether this increased alpha response to eye closure predicts more successful listening. Young healthy humans ($n = 22$) heard ten successively spoken digits, which were presented at a 0.75-Hz rate and embedded in white background noise (+10 dB SNR). Successive digits were alternately spoken by a female and a male voice and participants were instructed to attend to either of these voices. At the end of each trial, participants heard three of the ten digits again (uttered by a neutral voice) and they had to decide for each digit whether it was amongst the to-be-attended digits. The experiment was performed in complete darkness with participants having their eyes open or closed in half of the trials, respectively. Across all participants, closing the eyes did not enhance the accuracy of participants' behavioural responses. However, recognition performance evenly divided listeners into profiteers (performance benefit from closed versus open eyes) and sufferers (performance decline under closed eyes). In the electroencephalogram (EEG), parieto-occipital alpha power fluctuated in synchrony with digits, peaking ~400 ms before the onset of each to-be-attended digit. Closing the eyes boosted this rhythmic alpha power modulation only in those listeners who behaviourally profited from closed eyes. Our findings demonstrate that attentive listening with open and closed eyes is accompanied by a dynamic neural adaptation of alpha oscillations to attended versus ignored speech items. Closing the eyes supports the inhibition of the visual system to improve auditory attention. However, only those individuals who increase their alpha power modulation under closed eyes can use this strategy to enhance the success of attentive listening.

60 Attention governs neural oscillatory responses to degraded speech

M. Wöstmann, S. J. Lim, J. Obleser
Department of Psychology, University of Lübeck, Germany

Listening to acoustically degraded speech increases effort and reduces speech comprehension. Neurally, degraded speech commonly evokes an increase in the power of ~10 Hz alpha oscillations in the magneto-/electroencephalogram (M/EEG). However, this prominent alpha response to speech is far from being fully understood. In the present study, we tested whether the neural alpha response is driven by degraded acoustics per se or by the degradation-driven need to enhance attention to speech in memory. To this end, we used an irrelevant-speech paradigm where young healthy listeners (n = 23) were presented with nine randomly ordered target digits followed by one distractor sentence. Participants had to maintain the serial order of digits in memory while ignoring the distractor. We parametrically degraded acoustic detail of the distractor sentence (using noise-vocoding with 1, 4, or 32 frequency channels). As expected, more acoustic detail of the speech distractor led to more distraction, measured as disrupted serial recall of target digits. Most importantly, our task design elicited a reversal of the commonly observed neural alpha response to degraded speech: While alpha power decreases when better acoustics facilitate comprehension of attended speech, here, alpha power in parietal and temporal cortex instead increased as better acoustics of the to-be-ignored speech aggravated distraction. Beyond the alpha band, low-frequency delta and theta power (1–5 Hz) in fronto-parietal cortical regions increased as well when more acoustic detail boosted the interference by distracting speech. Our results suggest that acoustically more intact, distracting speech taxes attentive processing of target speech in working memory. The neural alpha response is governed by a listener's focus of attention, which controls whether more acoustic detail of speech facilitates comprehension (of attended speech) or enhances distraction (of ignored speech). In sum, auditory attention utilises a stronger alpha response to selectively filter out irrelevant distractors from interference with target speech.

61 Closing the efferent auditory loop: development and testing of a personalized wearable ear-EEG recording device

G. Cretot-Richert
Université du Québec, École de technologie supérieure, Canada

S. Debener
Oldenburg University, Department of Psychology, Neuropsychology Lab, Germany

M. De Vos
University of Oxford, Institute of Biomedical Engineering, Department of Engineering Science, UK

M. Bleichner
Oldenburg University, Department of Psychology, Neuropsychology Lab, Germany

J. Voix
Université du Québec, École de technologie supérieure, Canada

Hearing aids and cochlear implants aim at restoring auditory cognition for many patients, from congenitally deaf children to injured individuals and senior populations. Current technologies amplify signals relevant for speech understanding and feed it to the afferent nervous system. Yet, the healthy human auditory system is more intricate, it can isolate weak sounds in highly complex and noisy acoustic environments. This requires some level of top-down contribution from the efferent nervous pathway influenced by the listener's emotions and intent. Speech perception is enhanced through this auditory feedback loop, however current hearing devices do not mimic this fine-tuning ability. Future hearing technology may benefit from brain-computer interface (BCI) technology to achieve this goal. Electroencephalography (EEG) has been widely used in the development of non-invasive BCIs because of its good temporal resolution and established technology. However, research-grade EEG has been limited to controlled environments since conventional recording devices are bulky and sensitive to movement. In addition, they typically require caps to position arrays of sensors on the scalp, which are unsuitable for daily-life settings. To overcome this limitation a portable and unobtrusive ear-EEG recording system was developed and tested against laboratory-conscribed scalp-EEG. The ear-EEG system senses brain-electrical signals from

in and around-the-ear, using custom-fitted earpieces. Because the ear canal is usually hair-free and asymmetrical, this earpiece solution is less susceptible to motion. The research goal was to assess if ear-EEG could produce similar results to conventional EEG apparatus.

Miniaturized wet Ag/AgCl electrodes were installed in a custom-fitted EERS/SonoFitTM in-ear audio platform. This in-ear audio platform was coupled with a behind-the-ear piece forming a 5 mini-electrode interface. Event-related potentials (ERPs) obtained from an auditory oddball and a mismatch negativity paradigm were collected while recording brain activity simultaneously with this setup, dubbed "EARtrode", and scalp-EEG technology. Although EARtrode's signals had lower amplitudes, resulting signal-to-noise ratio and condition effect size were similar for both methods. As a consequence, EARtrode is a promising candidate for future small, mobile, and unobtrusive BCI platforms. In the long term ear-EEG systems could be merged with hearing aids to build next-generation devices that dynamically adapt to the listener's intentions and state changes. Our next step aims at testing new applications to explore ear-EEG's potential. Specifically, an ear-centered smartphone-controlled EEG platform could be used to measure levels of vigilance and cognitive load. Ear-EEG BCIs may evolve into truly assistive technology if they can be shown to reliably decode user mental states.

62 Is speech only noise for newborn infants? Electrophysiological responses to events detected within continuous speech

J. Simon
Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary
Department of Cognitive Science, BME Faculty of Natural Sciences, Budapest, Hungary

A. Kovács
Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary
Department of Telecommunication and Media Informatics, Budapest University of Technology and Economics, Budapest, Hungary

G. Háden, I. Winkler
Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

Converging evidence shows that newborns prefer speech over other sounds of similar complexity. It is also widely accepted that infants are able to discriminate phonemes within an oddball paradigm. However, little is known about their ability to extract and process moderately salient events embedded in the variance of continuous speech. One methodological problem for studying this issue with event-related brain potentials (ERP) is how one can define salient events within speech with a precise onset. We tested a possible solution by extracting speech events using a skewness based event detection algorithm, which is sensitive to energy changes between frequency bands. The events were classified into four groups: high-frequency, low-frequency, stop-consonant and phrase-onset events. Eight sentences were presented in two speech modes (infant directed [ID] and adult directed [AD]), 25 times, each to adults and newborn infants. Significant (compared to zero; $p<0.01$ in 15 consecutive time points) ERP responses were obtained for all four event categories in adults, while only the ERP to stop consonants failed to elicit a significant response in newborn infants. When the trials were split by speech mode, in newborns, significant ERPs were obtained with ID speech and the difference between the ID- AD-responses was also significant. In contrast, no significant differences were found between the responses to ID- and AD-speech events in adults. These results suggest that ID speech enhances the processing of speech events in the newborn brain. Further studies are needed to assess the limitations of the method and to determine what kind of speech events can be studied with this approach.

63 Does listening effort modulate speech envelope entrainment?

L. Decruy, J. Vanthornhout, T. Francart
KU Leuven, Leuven, Belgium

Studies have shown that cortical neural activity is entrained to the amplitude envelope of running speech. Given that the envelope is a primary cue for speech intelligibility, the degree of envelope entrainment may reflect different levels of speech intelligibility. Consequently, envelope entrainment in function of signal-to-noise ratio (SNR) may provide an objective correlate of speech intelligibility which can be used for the audiological assessment of populations such as young children. In our study, we measured envelope entrainment in function of SNR by calculating the correlation between the actual and reconstructed envelope from EEG-signals.

Interestingly, several subjects showed a decrease in envelope entrainment at SNRs where they achieved a high speech intelligibility score. This decrease contradicts our hypothesis since we assumed maximal entrainment when speech is completely intelligible. On the other hand, studies have found that high-level processes can modulate envelope entrainment. Taken this into account, listening effort may explain the decrease in entrainment on high SNRs. In challenging situations, a person will invest extra effort to restore the degraded speech signal, while having a conversation in quiet requires almost no effort. The most widely used method to measure listening effort is the dual task paradigm. This method relies on the theory that cognitive resources are limited and suggests that performance on a secondary task will deteriorate when the primary task is very demanding. Despite this theory, dual-tasks often involve two different sensory modalities which results in the allocation of different resources instead of shared. Furthermore, several studies use the target of the primary task, for example a word, as input for the secondary task. In this case, the performance on the latter task will not only be influenced by the effort needed in the primary task but also by the extent in which the word was heard. Therefore, we developed a new dual task.

Thirteen normal hearing and 2 hearing impaired, young subjects participated in our study. The primary task of our dual task involved a speech-in-noise test where the Flemish Matrix sentences were presented at fixed speech intelligibility levels. Our secondary task was a verbal-memory test in which response time was measured. To compare with our objective measure, we presented the same sentences on the same levels during an EEG-experiment. Additionally, we also investigated the reproducibility of our objective measure.

64 Attentional effects on the processing of syntactic violations during listening two simultaneous speech streams

O. Szalárdy, B. Tóth, D. Farkas
Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

A. Kovács
Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, Hungary

G. Urbán
Department of Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology and Economics, Budapest, Hungary

G. Orosz
Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

L. Hunyadi
Department of General and Applied Linguistic, University of Debrecen, Hungary

T. Szabó
Faculty of Information Technology and Bionics, Pázmány Péter Catholic University

B. Hajdu, I. Winkler
Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

The notion of automatic syntactical analysis for linguistic stimuli received support from some event-related studies. Here we provide a stronger test of this issue by presenting to listeners two concurrent continuous speech streams and manipulating in a fully crossed design two variables that can potentially affect speech processing: the direction of attention (focused vs. divided) × task (lexical – detecting numerals vs. syntactical – detecting syntactic violations). By recording EEG, we could thus compare between the event-related potentials (ERP) elicited by syntactic violations and numerals as targets (task-relevant events in the attended speech stream) with those for distractor (task-relevant events in the unattended speech stream) and attended and unattended task-irrelevant events with attention focused on one or divided between the two streams. Both task-relevant and task-irrelevant syntactic violations elicited the ELAN or possibly the N400 ERP component for the attended but not for the unattended speech stream, irrespective of the direction of attention. P600 was only elicited by target syntactic violations. Numeral targets elicited the N2b and P3 irrespective of the direction of attention, whereas none of the non-target numerals elicited either of these ERPs. The N2b am-

plitude was associated with the participant's performance in recognizing information from the speech material. The results provide no support for the notion of automatic syntactic analysis, because unattended syntactic violations failed to elicit any detectable ERP response.