SPEECH IN NOISE WORKSHOP

10-11 January 2019

Ghent, BE

Abstracts

The 11th Speech in Noise Workshop is organised by selfless volunteers from the University of Ghent, Belgium:

- *Sarah Verhulst*
- *Sarineh Keshishzadeh*

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

Contact information: info@spin2019.be

Front cover: Ghent by Ronan Shenhav (CC BY-NC 2.0).

The Speech in Noise Workshop is generously supported by:

Programme

Thursday, January 10

08:30 – 09:00 Registration, Poster setup & Coffee

09:00 – 09:15 Welcome / Introduction

09:15 – 09:45 “Interaction of hearing impairment, language, and talker on speech recognition” — Sabine Hochmuth, Anna Warzybok — *Oldenburg University, DE* 5

09:45 – 10:15 “Discovering the building blocks of hearing: a data-driven approach” — Lotte Weerts, Claudia Clopath, Dan Goodman — *Imperial College London, UK* 6

10:15 – 10:45 “End-to-end speech enhancement models using deep learning” — Deepak Baby, Sarah Verhulst — *Dept. Information Technology, Ghent University, BE* 6

10:15 – 10:45 Coffee, Picture and more Poster setup

11:15 – 11:45 “Speech intelligibility prediction – The story continues” — Richard Hendriks — *Delft University of Technology, NL* 7

11:45 – 12:15 “Meta adaption at the auditory-nerve level and its implications for speech intelligibility” — Jacques Grange, Ray Meddis, John Culling — *Cardiff University, UK* 8

12:15 – 12:30 “Tribute to the life and work of Prof. Ray Meddis” — Enrique Lopez-Poveda — *University of Salamanca, ES* 9

12:30 – 13:30 Lunch

13:30 – 14:30 **Keynote lecture**
“Cochlear synaptopathy and speech-in-noise deficits in normal hearing listeners” — Stéphane F. Maison — *Department of Otolaryngology, Harvard Medical School - Eaton-Peabody Laboratories, Massachusetts Eye & Ear Infirmary, US* 9

14:30 – 15:00 “Evidence for age-related cochlear synaptopathy in humans unconnected to speech-in-noise intelligibility deficits” — Enrique Lopez-Poveda, Peter T. Johannesen, Byanka C. Buzo — *University of Salamanca, ES* 11

15:00 – 15:15 Coffee / Drinks

15:15 – 17:30 Poster session 1 – even numbered posters 17

19:30 – ... Dinner at Bar Mirwaar (Burgstraat 59)

Friday, January 11

09:00 – 11:30 Poster session 2 – odd-numbered posters 17

11:30 – 12:00 "Does cortical entrainment exist? What we can learn from studying perception of naturalistic speech" — Anna Maria Alexandrou — *Aalto University, FI* 12

12:00 – 12:30 "Adaptive neural states and traits at the cocktail party" — Jonas Obleser — *Dept. of psychology, University of Luebeck, DE* 13

12:30 – 14:00 Lunch

14:00 – 14:30 "Do speakers make an active use of the visual modality when communicating in noise?" — Maëva Garnier, Lucie Ménard, Boris Alexandre — *Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, FR* 13

14:30 – 15:00 "School-age children's development in sensitivity to voice gender cues is asymmetric" — Leanne Nagels, Etienne Gaudrain, Debi Vickers, Petra Hendriks, Deniz Başkent — *Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, Groningen, NL* 14

15:00 – 15:30 Coffee break (take down all posters)

15:30 – 16:00 **Recipient of the Colin Cherry Award 2018**
"Auditory processing for speech in noise is enhanced in hearing aid users"
— Anna Exenberger — *UCL, London, UK* 15

16:00 – 16:30 "Heterogeneity in speech-in-noise recognition by CI listeners" — Chris James — *Cochlear France SAS, Toulouse, FR* 16

16:30 – 17:00 **Announcement of the Colin Cherry Award 2019 recipient, Discussion**

Talks

Thursday 10 January, 09:15—09:45

Interaction of hearing impairment, language, and talker on speech recognition

S. Hochmuth, A. Warzybok
Oldenburg University, Germany

Hearing impairment and age affect speech recognition particularly in acoustically challenging conditions. Studies related to better understanding and improvement of speech recognition in diverse acoustic conditions in different languages /for different talkers considering listeners with different hearing status will be a leitmotiv of this talk.

In order to better understand the underlying mechanisms of speech intelligibility accurate internationally comparable tools are needed. With this aim the Matrix sentence test has been developed in about 20 languages. Although the methodology of the test development was controlled across languages, differences in speech recognitions thresholds in noise of up to 5 dB are noticed for normative data of matrix tests across languages. The first part of the talk will focus on studies investigating the impact of language and talker on speech intelligibility in noise and will ask for acoustic-phonetic cues that might be suitable to predict those differences. Cross-language and –talker comparisons will be shown for German, Spanish, Russian and American English.

The second part of the talk will consider multicenter and multilingual studies with hearing-impaired listeners in order to test the reliability, sensitivity, and specificity of the matrix test for diagnostic purposes. Measurements with hundreds of participants conducted in different countries including Germany, Russia, Italy, Poland, USA or China will be shown. Furthermore, influence of hearing impairment on speech intelligibility in different masking conditions varying from very well controlled laboratory conditions with stationary noise to realistic noise like cafeteria noise will be demonstrated. Here, differentiation of listeners with various degree of hearing impairment using different maskers will be discussed.

Finally, the application of matrix tests and comparisons to existing standard audiometric tools will be shown in terms of hearing rehabilitation with hearing devices like hearing aids and cochlear implants.

Discovering the building blocks of hearing: a data-driven approach

L. Weerts, C. Clopath, D. Goodman
Imperial College London, UK

Experimental approaches to study hearing typically require simple stimuli to allow for controlled experiments. In order to improve our general understanding of features that are important for hearing in more complex environments, we propose a data-driven approach to determine good basic auditory features for speech processing. More specifically, we introduce a neuro-inspired feature detection model that relies on a modest amount of parameters. We first show that our model is capable of detecting a range of features that are thought to be important for noise-robust speech processing, such as amplitude modulations and onsets. Additionally, we propose a new methodology to identify important features within the parameter space of our model. This analysis leverages both information theory (in particular the Information Bottleneck principle) and supervised machine learning. The validity of our methodology is confirmed by comparing our results with psychoacoustic studies. Altogether our analysis framework of this new class of feature detectors may improve our current understanding of human hearing in challenging environments, both in terms of fundamental science and reproducing this ability in machine hearing systems.

End-to-end speech enhancement models using deep learning

D. Baby, S. Verhulst
Dept. Information Technology, Ghent University, Belgium

Suppressing background noise from recorded speech (a.k.a. speech enhancement) has widespread applications in mobile communication, voice activated systems and hearing aids. Recently, deep neural network (DNN)-based approaches gained success in speech enhancement as they can efficiently learn the speech and noise statistics from a training dataset containing noisy speech and clean speech pairs. Popular neural network-based speech enhancement systems operate on the magnitude spectrogram and ignore the phase mismatch between the noisy and clean speech signals. However, it is well established that speech quality can be significantly improved when the clean phase spectrum is known. This talk therefore concentrates on recent approaches for DNN-based speech enhancement in the time-domain. Such models are also known as end-to-end models since the time-domain speech signal is not transformed into any other domain before feeding it to the DNN.

In particular, we will discuss end-to-end models using the deep convolutional neural networks (DeepCNN) that map the time-domain noisy speech signal to the underlying clean speech signal. In addition, conditional generative adversarial networks (cGANs) have been recently shown promise in improving DeepCNN-based speech enhancement. This talk will introduce the DeepCNN approaches and their GAN-variants, the model architecture, loss functions and training strategies.

Thursday 10 January, 11:15—11:45

Speech intelligibility prediction – The story continues

R. Hendriks
Delft University of Technology, The Netherlands

Developed about a century ago, speech intelligibility prediction has been dominated for a long time by the Articulation Index (AI), and, later on by its successor, the Speech Intelligibility Index. Although effective in some situations, these measures have difficulties to account for distortions introduced due to e.g., non-stationary noise and enhancement algorithms. Therefore, during the last decade, we have seen a significant increase in the interest on speech intelligibility measures, which led to the introduction of many new measures. Although the general trend is that these measures lead to improved prediction in certain specific situations, they typically only work well for a very narrow set of conditions and are often developed based on experience, hampering generalizations to new environments and processing conditions. Moreover, while the AI was quite close to a measure on the maximum information rate that may be transmitted from a talker to a listener, the majority of the more recent intelligibility metrics have lost this information theoretical interpretation.

In this presentation we use information theoretical concepts to model the transfer of information from the talker to the listener. Based on a rather simple model of communication, we present a new intelligibility metric that expresses the intelligibility as the estimate of the information shared between a talker and a listener in bits/sec. This measure, named Speech Intelligibility in Bits (SIIB), shows a very high correlation with speech intelligibility under a wide range of processing conditions.

Meta adaption at the auditory-nerve level and its implications for speech intelligibility

J. Grange
Cardiff University, UK

R. Meddis
University of Essex, UK

J. Culling
Cardiff University, UK

Speech intelligibility relies on the faithful coding of the temporal modulations that carry speech information. For such coding to be possible, the auditory periphery must adapt to changes in contextual sound level because the dynamic range of auditory nerve (AN) fibres is limited to ~30 dB, as found in conventional (adapted-rate) measures of rate-level functions (RLFs). Such meta adaptation is defined as the shift of RLFs to higher probe levels as contextual level increases; it was found in small mammals at the AN and IC levels (Dean et al., 2005; Wen et al., 2009).

A first, modelling study compared predictions by a computer model of the auditory periphery (MAP) to previous modelling of meta adaptation at the AN level. While previous studies found the need for phenomenological modelling of meta adaptation at the inner hair-cell (IHC)-AN synapse level, the physiologically-inspired MAP model readily accounts for it. RLF predictions were extracted from the modelled response of 500-AN fibres over a 50 ms test period and immediately following a 400 ms precursor that set context level. Test and context levels were independently varied in the 0-100 dB range for tone pips or noise bursts. In the physiologically-tested 40-80 dB context-level range, meta adaptation in excess of 0.5 dB/dB was predicted when both medial olivocochlear (MOC) and acoustic reflexes were enabled. Disabling the acoustic reflex reduced it to ~0.25 dB/dB. Disabling also the MOC reflex reduced it a little further, but did not remove it. Firing rates measured during a meta-adaptation paradigm are inherently not adapted rates: the 50 ms test-burst duration is so short that the measurement encompasses a fast stimulus-driven mechanism of neurotransmitter release in the cleft with a slower IHC vesicle-store replenishment mechanism and the early part of the slower modulation of basilar membrane vibrations by efferent reflexes. Anaesthesia used in small-mammal electrophysiology is often assumed to suppress efferent reflexes. Therefore, existing measures of meta adaptation do not reveal its full potential. Modelling efferent reflexes demonstrates the potential for normal-hearing meta adaptation reaching close to an ideal 1 dB/dB.

A second, psychophysical study measured speech intelligibility in noise with normal-hearing listeners attending to vocoded speech based on MAP-predicted AN firing patterns. Incorporating MOC and acoustic reflex information in the vocoder significantly improved speech intelligibility. This finding supports a key role of full meta adaptation in the optimal neural coding of speech modulations in quiet or in noise.

Thursday 10 January, 12:15—12:30

Tribute to the life and work of Prof. Ray Meddis

E. Lopez-Poveda
University of Salamanca, Spain

The hearing community was deeply saddened to learn about the recent passing of Ray Meddis. He was a dear colleague to many of us and his work has had a tremendous impact on auditory modeling and our understanding of hearing loss. With this talk we would like to honor his memory.

Thursday 10 January, 13:30—14:30

Cochlear synaptopathy and speech-in-noise deficits in normal hearing listeners — **Keynote**

S. F. Maison
Department of Otolaryngology, Harvard Medical School - Eaton-Peabody Laboratories, Massachusetts Eye & Ear Infirmary

Animal studies have well established that hair cell loss following noise exposure or aging is often preceded by loss of synapses between the sensory cells and the auditory nerve fibers. A recent temporal bone study provides histopathological evidences that primary neural degeneration greatly exceeds inner hair cell loss in aging human ears as well (Wu et al., 2018). The silencing of these neurons, especially those with high thresholds and low spontaneous rates, degrades auditory processing and may contribute to difficulties understanding speech in noise. In mice, cochlear synaptopathy has been diagnosed by supra-threshold amplitude of ABR wave 1 (Kujawa and Liberman, 2009), the summed activity of cochlear neurons. The fractional reduction in responses to moderate level tone-pips is correlated with the fractional reduction in synaptic counts (Sergeyenko et al., 2013). Cochlear synaptopathy is also correlated with measures of middle-ear muscle reflex (MEMR) strength, possibly because the missing high-threshold neurons are important drivers of this reflex (Valero et al., 2015, 2018).

We recruited 165 normal hearing subjects (≤ 25 dB HL from 0.25 – 8 kHz) between the ages of 18 and 63, with no history of ear or hearing problems, no history of neurologic disorders and unremarkable otoscopic examinations. All subjects were native speakers of English and passed the Montreal Cognitive Assessment. Word recognition scores

were assessed at 55 dB HL in quiet and in difficult listening situations using the NU-6 corpus (with competing white noise at 0 SNR or with a 45% or 65% compression with 0.3 s reverberation). Performance on a modified version of the QuickSIN was measured as well. Outer hair cell function was assessed using DPOAEs from 0.5 kHz to 16 kHz while cochlear function was assessed by electrocochleography (ECochG). EcochG waveforms were obtained in response to 100-µs clicks delivered in alternate polarity at 125 dB pSPL with a repetition rate of 9.1 Hz in absence or in presence of a forward masker consisting of an 8-16 kHz noiseband of 90-ms duration presented at 15 dB SL or 35 dB SL. Finally, one EcochG waveform was obtained from all subjects at a rate of 40.1 kHz. The total noise dose for all ECochG measurements was well within OSHA and NIOSH standards.

MEMR effects were assessed in a second session using a custom method similar to that of Keefe and colleagues (Keefe et al. 2010). This approach measures changes in ear-canal sound pressure to a click probe evoked by an ipsilateral noise elicitor. Specifically, we use a pair of 100-µs clicks at 95 dB SPL separated by a 500-msec elicitor (noise burst with a 2.5 ms ramp) presented 30 ms after the first click and preceding the second by 5 ms. This click-noise-click complex was repeated every 1535 ms, leaving 1 s of silence between noise bursts to allow relaxation of the MEMs. Four complexes were presented at each elicitor level, and elicitor level was raised in 5 dB steps from 40 to 95 dB SPL. For each click-noise-click complex, the spectral difference between the two click waveforms was computed.

Our results show that middle-ear reflex thresholds and electrocochleographic measures of neural health are correlated with speech-recognition performance whereas measures of hair cell function are not, consistent with selective neural loss. Furthermore, forward masking has a differential suppressive effect: both SP- and AP- amplitudes decrease to a greater extent with masking levels in subjects who obtained the poorest word recognition scores when compared to those who did best. These results further support the idea that cochlear synaptopathy may lead to deficits in hearing-in-noise, despite the presence of normal thresholds at standard audiometric frequencies.

References

Keefe DH, Fitzpatrick D, Liu YW, et al. (2010). Wideband acoustic-reflex test in a test battery to predict middle-ear dysfunction. Hear Res 263:52-65.

Kujawa SG, Liberman MC (2009) Adding insult to injury: cochlear nerve degeneration after “temporary” noise-induced hearing loss. J Neurosci 29: 14077-85.

Sergeyenko Y, Lall K, Liberman MC, et al. (2013) Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. J Neurosci 33:13686-94.

Valero MD, Hancock KE, Liberman MC (2015) The middle ear muscle reflex in the diagnosis of cochlear neuropathy. Hear Res 332:29-38.

Valero MD, Hancock KE, Liberman MC (2016) The middle ear muscle reflex in the diagnosis of cochlear neuropathy. Hear Res 332:29-38.

Wu PZ, Liberman LD, Bennett K, et al. (2018) Primary neural degeneration in the human cochlea: evidence for hidden hearing loss in the aging ear. Neuroscience, in press.

Research supported by a grant from the NIDCD (P50 DC015857)

Evidence for age-related cochlear synaptopathy in humans unconnected to speech-in-noise intelligibility deficits

E. Lopez-Poveda, P. T. Johannesen, B. C. Buzo
University of Salamanca, Spain

Cochlear synaptopathy (or the loss of primary auditory synapses) remains a subclinical condition of uncertain prevalence. Here, we investigate whether it occurs in humans, and whether it contributes to suprathreshold speech-in-noise intelligibility deficits. For 94 human listeners with normal audiometry (aged 12-68 years; 64 female), we measured click-evoked auditory brainstem responses (ABRs), self-reported lifetime noise exposure, and speech reception thresholds (SRTs) for sentences (at 65 dB SPL) and words (at 50, 65 and 80 dB SPL) in steady-state and fluctuating maskers. Based on animal research, we assumed that the shallower the rate of growth of ABR wave-I amplitude versus level, the higher the risk of suffering from synaptopathy. We found that wave-I growth rates decreased with increasing age but not with increasing noise exposure. SRTs were not correlated with wave-I growth rates, and mean SRTs were not statistically different for two subgroups of participants (N=14) with matched audiograms (up to 12 kHz) but different wave-I growth rates. Altogether, the data are consistent with the existence of age-related but not noise-related synaptopathy. In addition, the data dispute the notion that synaptopathy contributes to suprathreshold speech-in-noise intelligibility deficits.

We thank Filip Rønne, Niels H. Pontoppidan, and James M. Harte for useful discussions. BCZ was supported by a postdoctoral scholarship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (Brasil). Work supported by the Oticon Foundation, Junta de Castilla y León (grant SA023P17), European Regional Development Fund and the Spanish Ministry of Economy and Competitiveness (grant BFU2015-65376-P) to EAL-P.

Does cortical entrainment exist? What we can learn from studying perception of naturalistic speech

A. M. Alexandrou
Aalto University, Finland

The popular theoretical framework of cortical entrainment postulates that speech comprehension crucially depends on the continuous alignment of low-frequency cortical oscillatory activity with the amplitude envelope of perceived acoustic speech signals. This alignment has been suggested to represent a neural sampling mechanism that packages the incoming speech into discrete chunks, which are then transmitted to higher-order cortical regions for subsequent processing and extraction of meaning from an utterance.

Empirical evidence for cortical entrainment mostly stems from tightly controlled experimental paradigms focusing on repeated perception of isolated sentences or of read-aloud texts. However, these kinds of stimuli do not reflect natural speech as we encounter it in real life: spontaneously produced, real-life speech demonstrates variable speaking rate, and is characterised by significant disfluencies in the forms of interruptions, repetitions, filler words and revisions.

In my talk, I advance the view that naturalistic experimental paradigms, utilising spontaneously produced speech as stimuli and suitable frequency-domain methodological tools, should be used to address an important question that remains open: whether cortical entrainment is observed during speech perception and comprehension in real-life communicative situations, as opposed to tightly controlled experimental settings.

I also present evidence that the phenomenon currently labelled as cortical entrainment might be confounded by a regular repetition of evoked responses, based on the analysis of the acoustic structure of isolated sentences, read-aloud texts and spontaneously produced speech.

Finally, I propose different alternative viewpoints of what spontaneously produced, real-life speech could teach us about cortical entrainment and language comprehension, in general.

Adaptive neural states and traits at the cocktail party

J. Obleser
Dept. of psychology, University of Luebeck, Germany

Challenging listening situations do pose multiple challenges to our behavioural goal of communicating successfully. What's more, for us as neuroscientists, there is a lot to learn from these situations: How do neural brain states affect the percepts evoked by ambiguous or noisy sensory input? What are possible neural implementations of spatial and temporal filters for solving the 'cocktail party problem'? With some tentative answers to these questions from M/EEG as well as fMRI at hand, my laboratory is currently exploring (i) the degree to which aging individuals differ in these neural states and traits, and (ii) whether this can explain some of the variation in listening success at the cocktail party.

Do speakers make an active use of the visual modality when communicating in noise?

M. Garnier
Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, FR

L. Ménard
Departement de Linguistique, Laboratoire de Phonetique, Center for Research on Brain Language and Music, Universite du Quebec a Montreal, CA

B. Alexandre
Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, FR

Background — It is now well known that seeing speech improves its perception, especially when speech is perturbed in the acoustic domain, for example by a noisy background. However, it is not clear yet how the visual modality is exploited in speech production, and whether speakers can make active use of the visual channel (consciously or not) to improve their intelligibility in noisy conditions.

Method — Six native speakers of Canadian French produced speech in quiet conditions and in 85 dB of babble noise, in three situations: interacting face-to-face with the experimenter (AV), using the auditory modality only (AO), or reading aloud (NI, no interaction). The audio signal was recorded with the three-dimensional movements of their lips and tongue, using electromagnetic articulography.

Results — All the speakers reacted similarly to the presence vs absence of communicative interaction, showing significant speech modifications with noise exposure in both interactive and non-interactive conditions, not only for parameters directly related to voice intensity or for lip movements (very visible) but also for tongue movements (less

visible); greater adaptation was observed in interactive conditions, though. However, speakers reacted differently to the availability or unavailability of visual information: as expected, four of them enhanced their visible articulatory movements with NE more in the AV condition than in the AO condition. However, one participant showed the opposite behavior. The final participant applied an intermediate strategy, enhancing acoustic cues more in the AO condition and amplifying lip protrusion cues and visible inter-vowel contrasts more in the AV condition. These results support the idea that the Lombard effect is at least partly a listener-oriented adaptation. However, to clarify their speech in noisy conditions, only some speakers appear to make active use of the visual modality.

School-age children's development in sensitivity to voice gender cues is asymmetric

L. Nagels
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, Groningen, The Netherlands

E. Gaudrain
Auditory Cognition and Psychoacoustics, CNRS, Lyon Neuroscience Research Center, Lyon, France

D. Vickers
Speech, Hearing and Phonetics Science Research Department, University College London, 2 Wakefield Street, WC1N 1PF London, UK

P. Hendriks
Center for Language and Cognition Groningen, University of Groningen, The Netherlands

D. Başkent
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, Groningen, The Netherlands

Speakers' perceived voice gender is characterized by a number of acoustic features, but it is primarily defined by fundamental frequency (F0), related to the speaker's glottal pulse rate, and vocal tract length (VTL), related to the speaker's size. At a young age, children are already able to distinguish voices based on clear voice gender differences, but they are still less sensitive than adults to less salient differences in voice cues, such as prosody or accents. It remains unclear how children's sensitivity to differences in voice cues develops over time. In the present study, we investigated whether children's sensitivity to F0 and VTL differences and their weighting of F0 and VTL cues for voice gender categorization develop in parallel or if there is an asymmetry. We hypothesized that there may be a potential hierarchy in development as children generally focus more on global, dynamic acoustic cues, such as variations in F0, but may require more experience to make use of fine-grained, static acoustic information, such as VTL (Nittrouer,

Manning, & Meyer, 1993). Furthermore, we investigated how children's sensitivity to F0 and VTL relates to their weighting of F0 and VTL for voice gender categorization, as categorization may depend more on underlying representations of voice gender rather than sensitivity.

We tested fifty-eight children between the ages of 4 to 12 and fifteen adults. In experiment 1, we measured participants' just noticeable differences (JNDs) in F0 and VTL via a 3-AFC procedure. In experiment 2, we studied participants' weighting of F0 and VTL cues for voice gender categorization by manipulating the F0 and VTL parameters of an originally female speaker's voice. Our results of experiment 1 showed that children's sensitivity to VTL becomes adult-like around the age of 8, but their sensitivity to F0 still differs from adults at the age of 12. On the contrary, children's weighting of F0 for voice gender categorization was more adult-like than their weighting of VTL but continued to differ from adult's weighting at all ages. After correcting for a general effect of age, children's weighting of F0 and VTL cues was weakly to moderately related to their sensitivity to F0 and VTL. Hence, there is asymmetry in children's development of sensitivity and weighting of voice cues related to voice gender, but both abilities seem to develop at a different pace, implying that perceptual sensitivity and categorization of voice cues rely on different underlying mechanisms.

Auditory processing for speech in noise is enhanced in hearing aid users

A. Exenberger
UCL, London, UK

Hearing impaired people typically experience more difficulty when listening to speech in noisy situations than normal hearing people. However, it is still unclear where this increased effort originates. Electrophysiological measures were used to examine differences between auditory processing and lexical processing. Neural entrainment, a measure of synchronization between an auditory stimulus and neural brain activity, has previously been shown to be greater for intelligible speech and when a listener is focusing attention. It can thus be expected that relative to a normal hearing group, a hearing impaired population would show reduced neural entrainment in connection with poorer speech recognition performance and higher listening effort. However, with our study we demonstrated the opposite; in a speech-in-noise task with different levels of multi-talker babble noise, the hearing impaired group had significantly higher entrainment than the normal hearing group, despite similar or even poorer levels of speech recognition. Based on our findings, we suggest that higher entrainment in the hearing impaired group might reflect that more auditory processing is required to understand speech in noise.

Heterogeneity in speech-in-noise recognition by CI listeners

C. James
Cochlear France SAS, Toulouse, France

Adult cochlear implant (CI) users exhibit a very large range of speech recognition outcomes, and in addition maybe differentially affected by the same level of background noise. We also notice that CI users take varying amounts of time to acquire speech recognition in quiet, and equivalent levels of performance in noise: For example, some CI users achieve levels of performance within one day, or at least within one month of activation equivalent to normal-listener acoustic 'vocoder' simulations, while others require months of experience or never achieve open-set speech recognition (James et al., Ear and Hearing 2018)

Optimal electrode position can have some effect on performance. However subject-related factors such as duration of deafness and early-onset or congenital hearing loss also affect performance, both in the short and long term. Most importantly etiologies which may endanger neural survival and/or distort current paths appear to most severely limit speech recognition outcomes.

I will consider what kinds of effect these factors may have on our conception of speech recognition via cochlear implants; on the 'effective channels' analogy, and our attempts to improve the performance of CI users in noise via changes in sound processing and coding.

Posters

02 Neurophysiological and subjective measures of listening effort

A. H. Winneke
Fraunhofer IDMT-HSA, Oldenburg, Germany

M. Jäger
University of Oldenburg, Department of Psychology, Oldenburg, Germany

M. Krüger, M. Schulte
Hörzentrum Oldenburg GmbH, Oldenburg, Germany

Speech perception in suboptimal environments requires extra investment of neurocognitive resources. This is coined listening effort and is a commonly reported problem. Yet, the neurophysiological processes underlying listening effort are not clearly understood. The goal of this study was to investigate the relationship between subjective and objective measures of listening effort when listening to speech in a noisy background. Various approaches to asses listening effort have been devised and there is an ongoing debate on whether objective, physiological measures are associated with the subjective percept of effortful listening.

Using a modified version of the adaptive categorical listening effort scale (ACALES; Krüger et al., 2017) we asked 20 normal-hearing adults to rate their subjective listening effort when listening to the German Oldenburger sentence test (OLSA) in five fixed SNRs (-5, -2.5, 0, +2.5, +5 dB SPL) and for two types of noise (OlNoise, International Female Fluctuating Masker (IFFM)) presented at an intensity of 65 dB SPL. While participants were performing the task, an electroencephalogram (EEG) was recorded using unobtrusive behind the ear EEG electrodes (cEEGrids, Debener et al., 2015). EEG data analysis focused on alpha band activity (8-12Hz) which has been suggested to reflect an inhibitory function of processing task-irrelevant noise signals (Obleser et al., 2012, Strauß et al., 2014) and can be considered as an index of listening effort. Also, in a subset of participants (N=15) speech intelligibility at above mentioned SNRs was measured.

For all SNRs and both noise types speech intelligibility ranged between 94 and 100%, while ACALES ratings indicate that with increasing SNR the perceived listening effort. The EEG data reveal that alpha power spectral density values are higher during poor SNRs as compared to advantageous SNRs. Further analyses show positive correlations between subjective ratings and EEG alpha power. The results confirm that listening to speech in noise induces listening effort even in positive SNRs as well as a linear relationship between subjective listening effort scores and EEG alpha power spectral density with more effort related to more alpha power. These findings support the proposal of al-

pha as an indicative parameter for detecting changes in listening effort objectively. The speech intelligibility data confirm that measures of listening effort are more sensitive than SI measures for positive SNRs. Finally, the study shows that unobtrusive behind the ear electrodes can be used to measure changes in listening effort revealing the potential to use this technology outside the lab.

03 Eye gaze steering works miracles for hearing aid users in noisy enviroments

R. K. Hietkamp, S. Rotger-Griful, S. B. Lange, C. Graversen, T. Bhuyian
Eriksholm Research Centre, Denmark

A. Favre-Félix
Eriksholm Research Centre, Denmark and Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs.Lyngby, Denmark

T. Lunner
Eriksholm Research Centre, Denmark and Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs.Lyngby, Denmark and Swedish Institute for Disability Research, Linnaeus Centre, HEAD, Linköping University, Linköping, Sweden

People with hearing loss experience problems especially in noisy environments such as restaurants or family dinners. Technical solutions for these problems include external microphones, beamforming features and noise reduction schemes. These solutions provide some benefit but they do not offer the listener the possibility to steer the sound by attention. One way of distinguishing where the attention of the listener lies is by estimating the eye gaze position, which then can be used as input to a sound enhancing system.

In a previous study with hearing aid users, Favre-Felix et al (2018) identified eye gaze through electrooculograpy (EOG) in off-line recordings and used it to enhance the attended speaker in a three-speaker spatial setup. The results showed that listening performance increased, even though the target estimation of the algorithm was only 65%. But even if eye gaze steering would work perfectly all the time, would it be desirable for the hearing aid user in daily life?

This question was addressed in the study presented here: How do hearing aid users perceive the extra benefit from devices that enhance attended speakers by means of eye gaze steering compared to conventional hearing aids? Nine hearing aid users with mild to moderate hearing loss, but otherwise diverse with respect to age, gender, hearing aids experience and aetiology, were equipped with a golden standard device (Vicon motion tracker and Dikablis eye-tracker systems) that identifies eye gaze position in a virtual environment. The eye gaze position was then used to enhance an attended speaker in a competing talker situation (2 speakers in babble background noise), pre-

sented as video recordings, live size, using a 88'' 4K screen and 3 loudspeakers. The users' reaction to the effect of the equipment was recorded on camera. Furthermore, the end-user benefit was measured by means of speech comprehension scores and subjective evaluation of speech intelligibility and listening effort.

The poster presents the test setup, its ecological validity and the main findings.

Research supported by European Community's EU Horizon 2020 Programme under grant agreement no. 644732 (Cognitive Control of a Hearing Aid, COCOHA).

References

Favre-Felix, A., Hietkamp, R., Graversen, C., Dau, T., & Lunner, T. (2018). Steering of audio input in hearing aids by eye gaze through electrooculography. Proceedings of the International Symposium on Auditory and Audiological Research, 6, 135-142. Retrieved from https://proceedings.isaar.eu/index.php/isaarproc/article/view/2017-16

04 The perception of dynamic pitch in speech and non-speech

H. S. Jeon
University of Central Lancashire

A. Heinrich
University of Manchester

Pitch in speech varies continuously and delivers information on linguistic structure, as well as paralinguistic information on a speaker's identity and emotional state. And although pitch plays a significant role in communication, we have little understanding on how dynamic pitch is perceived.

The focus of this study was to investigate how the perception of pitch height is affected by the direction of the pitch movement (rise-fall forming 'peak' vs. fall-rise 'valley') and F0 turn shapes (sharp vs. plateau). Moreover, we aimed to understand whether pitch contours are perceived differently in speech compared to non-speech, i. e., whether linguistic information interacts with the auditory information. This question arose because in speech perception the shape of pitch movement affects the perceived height and timing of the pitch event. Two examples are the findings that a F0 plateau following a rise sounds higher than a sharp peak with the same maximum F0 and that pitch perception may differ for the same amount of change in F0 depending on whether it is rising or falling. Whether these perceptual effects are unique to speech sounds has not been explored.

We will discuss results of an experiment that used 3 stimulus types [speech, nonsense, complex tone], 2 directions of pitch movement [peak, valley] and 2 turning point types [sharp turn, plateau] (all crossed). The speech stimuli were four English sentences: 'is Lemmy near Nelly?', 'is Nelly near Lemmy?, 'does Mona know Nina?', and 'does Nina

know Mona?'. From them, duration and intensity contours were extracted and embedded in nonsense and complex tone stimuli. Nonsense stimuli were reiterated speech. Complex tones were harmonic complexes with energy between 200 Hz and 6000 Hz. All stimuli had a reference line F0 (F0 at start and end of stimulus) of 200Hz. All stimuli had two F0 turns; the first one always formed a sharp turn and the second one was either a sharp turn or a 100 ms plateau. The difference between the reference line and the first turn was always 5.4 semitones and that between the reference line and the second turn varied between 3.4 and 7.4 semitones. All stimuli were resynthesised with Praat. Young native English speakers with normal hearing listened to stimuli of all three types and judged which turning point sounded higher for the 'peak' stimuli or lower for the 'valley' stimuli.

05 Automatic evaluation of children reading aloud in babble noise of classroom context

M. Daniel
Lalilo, Paris, France

L. Gelin
IRIT, Universite de Toulouse, CNRS, Toulouse, France

Learning how to read is the most important step in the intellectual development of a child, who then acquires a skill that will impact his/her whole life. The learning process passes through oralisation, but it is time-consuming for teachers to regularly assess their student's reading level, as it needs to be done individually with each of the 20-to-30 children in their class. We are developing, as a part of a pedagogical assistant for Kindergarten to 2nd grade teachers, a tool to automatically score the fluency and the pronunciation of students reading aloud, based on a speech recognition HMM-GMM system. Evaluating the performance of 5-7 years old children in the classroom environment is very challenging : lack of data, characteristic non-reader children's prosody and 'babble' noise are the 3 main challenges we encounter. This poster presents the different strategies to address those challenges. The development of an efficient process to quickly collect and annotate a large amount of children's voice recordings enables us to improve our acoustic models and networks. The slow-reading and underdeveloped enunciation of young children is addressed with personalized language and grammar models. The 'babble' noise, that is typical in classrooms, is hard to filter due to variable and often poor quality of audio gear of schools and the targeted child's voice sometimes being drowned out by noise. Further work will consist of training a neural network that automatically rejects the recordings containing too much noise: if so, the child is asked to speak louder, or, in extreme cases, the reading aloud exercise is stopped.

06 A model predicting the effect of hearing impairment on binaural speech intelligibility in noise

T. Vicente, M. Lavandier
Univ Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, Rue M. Audin, 69518 Vaulx-en-Velin Cedex, France

J. M. Buchholz
Department of Linguistics – Audiology, Australian Hearing Hub, Macquarie University, 2109 NSW, Australia

An update to the preliminary model proposed by Lavandier et al. [Acta Acust. united Ac. 104, 909-913 (2018)] will be presented that allows predicting binaural speech intelligibility in noise for normal hearing (NH) and hearing impaired (HI) listeners as a function of the listener's audiogram and the noise level.

The model inputs are the masker and target signals as well as the audiogram at the listener's ears. An internal noise is used to model hearing impairment and considered as another masker by the model. The model applies a short-time frequency analysis of the signals, to predict the binaural masking level difference (BMLD) and the signal-to-noise ratio (SNR) at the better ear. If the target or masker level is lower than the internal noise level at a given ear, the BMLD is set to 0. The maximum between the masker and internal noise levels is considered to compute the SNR at the better-ear. The BMLD and better-ear SNR are summed, integrated across frequency with a SII-weighting and averaged across time to obtain a binaural ratio. The relative difference between binaural ratios are compared to speech reception threshold differences measured in listening tests.

Three experiments were considered here, which used headphones to simulate an anechoic room, enrolled NH and HI listeners, and involved 2 vocoded speech maskers either collocated with the target in front of the listener or spatially separated. The separated conditions were artificially designed: the left masker was played only on the left channel of the headphones and the right masker only on the right channel. In the first and third experiments, the NH maskers were played at 60 dB SPL and a linear amplification was applied on the HI stimuli depending on their hearing loss; while in the second experiment, masker and target were filtered in order to equalize the audibility of the stimuli across listeners, then they were played at four different sensation levels. The first experiment also involved two unmodulated speech-shaped noises. In the third experiment, three types of separated conditions were tested: the first was similar to the one used in the other experiments, the second used realistic binaural cues and the third involved stimuli with interaural time differences removed. The proposed binaural model was able to predict the experimental data satisfactorily in all experimental conditions for NH and HI listeners.

07 Consequences of cochlear synaptopathy in noise-exposed adults

T. Schoof, T. Green, S. Rosen
University College London

Evidence for cochlear synaptopathy using non-invasive measures in humans remains limited. However, previous research has predominantly focused on young adults with normal hearing thresholds whose suspected synaptopathy has been based on self-reports of noise exposure history. It is quite likely that measurable synaptopathy only becomes evident later in life through the combined effects of ageing and noise exposure.

We are investigating the incidence of cochlear synaptopathy in humans by focusing on normal-hearing middle-aged adults (45-60 years old) with self-reported noise exposure history and comparing their data to a young control group (18-35 years old) without much noise exposure history. Since data collection is still in progress, we will focus on data from the young control group in this presentation.

To assess cochlear synaptopathy, we are measuring ABR wave I amplitudes in response to clicks presented at three different levels (95, 105, 115 dB peSPL), and envelope following responses (EFRs) in response to a 2.8 kHz transposed tone amplitude modulated at 176 Hz at three different modulation depths (0, -4, and -8 dB). To quantify the degree of cochlear synaptopathy, we will compute the slope of the click ABR wave I amplitude across levels and the spectral magnitude of the EFR at F0 across modulation depths.

To examine the potential effects of cochlear synaptopathy on speech perception in noise, we are measuring speech reception thresholds (SRTs) for both consonants and sentences at two different levels (40 and 80 dB SPL). SRTs are measured in two conditions, one where both the target and background noise are diotic (N0S0), and another where the target signal is inverted in polarity in one ear (N0Sπ) leading to a phase disparity across the ears. We compute the binaural intelligibility level difference (BILD) by taking the difference between the N0S0 and N0Sπ conditions. Of particular interest is the relationship between our electrophysiological measures and our speech-in-noise measures.

Given that cochlear synaptopathy is thought to selectively affect high-threshold low spontaneous rate fibres, we hypothesise that clicks presented at lower levels and transposed tones with shallower modulation depths will not be encoded as robustly in individuals with cochlear synaptopathy. We therefore predict that the middle-aged group will show shallower ABR and EFR slopes compared to the young controls. In addition, we expect that individuals with shallower ABR and EFR slopes will perform more poorly on the speech-in-noise task, particularly at higher stimulus levels.

08 On the use of response time in a single-task paradigm to evaluate listening effort in noisy and reverberant conditions

C. Visentin, N. Prodi
University of Ferrara, Ferrara, Italy

Everyday communication most often takes place in the concomitant presence of reverberation and background noise, and many of the real maskers are fluctuating in character. Such adverse conditions impair the speech intelligibility and make the listening process effortful. The present work investigates the combined effects of noise fluctuations and reverberation on listening effort, with reference to a speech reception task and a group of young adults with normal hearing. The behavioral measure of response time (RT) in a single-task paradigm is used to evaluate the listening effort, and a slowing down of the measure is interpreted as an increase in the allocation of cognitive resources.

Speech-in-noise tests were presented to 79 participants in reverberant conditions, created by convolving the anechoic speech and background noises with simulated binaural impulse responses. Speech reception was measured in presence of continuous spatially diffuse stationary and fluctuating noise (ICRA), over a wide range of signal-to-noise ratios (SNRs); three reverberation conditions were considered (anechoic, 0.30 and 0.65 s). The experiment was presented in a closed-set format; for each participant, data on speech intelligibility (SI) and manual RT (time elapsed between the offset of the audio playback and the response selection on a touchscreen) were collected.

The SI results showed a benefit in speech reception under fluctuating noise in both reverberant conditions, due to the “listening in the dips” phenomenon. The RTs were sensitive to the effect of SNR and reverberation, slowing down with both decreasing SNR and increasing reverberation; noticeably, the effect was more pronounced in the presence of fluctuating noise. Moreover, the RTs were sensitive to the effect of noise type only for the less reverberant condition, where faster RTs were found for the fluctuating noise. When SI was fixed, it was found that RTs in fluctuating noise became gradually longer compared to stationary noise, with increasing reverberation; the opposite trend was observed in anechoic conditions.

The pattern of the data indicates that adding reverberation to background noise decrements the processing speed in a speech reception task, and even more so in presence of noise fluctuations. The outcomes at fixed SI suggest that measures of RT in anechoic conditions may return results that are not fully representative of the listening effort experienced in realistic conditions. The present findings support the need for considering both noise and reverberation when predicting listening effort in real-life conditions.

09 Are musicians at an advantage when processing speech on speech?

E. C. Kaplan, A. Wagner, D. Baskent
University of Groningen, the Netherlands

In the current study, we explored whether understanding speech in the presence of background talkers may be enhanced through musical training. Earlier studies have shown that musical training can grant normal-hearing listeners an advantage on auditory tasks, not only when these relate to music, but also for speech comprehension, in particular in noise or in the presence of background talkers (Başkent & Gaudrain, 2016; Swaminathan et al., 2015). Taken together, however, various studies addressing a 'musician advantage' provide inconclusive results (for a review see: Coffey, Mogilever, & Zatorre, 2017), which can be partly explained through the use of different measures (e.g. behavioral versus physiological).

The present study combined an off-line and on on-line measure of speech perception to investigate how automatic processes can contribute to the potential perceptual advantage of musicians. We first used a sentence-recall task (offline task), in which participants recall and repeat target Dutch sentences that are masked by sentences from two different talkers with varying target/masker ratios. Following, we used a visual-world paradigm employing similar stimuli, using eye tracking (Cooper, 1974; Salverda & Tanenhaus, 2017). Here, while listening to sentences, participants visually search for the image of the target word among images of a phonological competitor (ie. ham-hamster) and two unrelated distractor images. This task provides an online measure of speech processing as it captures the time course of gaze fixations to the target and/or the competitor word, as well as the changes in event related pupil dilations. The online measure indicates how quickly participants integrate the acoustic information in the signal when they are accessing the mental lexicon and the extent of the mental effort involved in processing of linguistic information. Results indicate that there is an overall effect of musicianship. Both results will be presented in a comparison between musicians and non-musicians.

11 Linking audiovisual integration to audiovisual speech-in-noise performance

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

The process of audiovisual (AV) integration may allow for improved processing of simple and complex stimuli such as speech. In fact, research has shown that redundant visual information improves speech comprehension, especially in adverse listening conditions. However, listeners differ greatly in understanding speech in noise and the benefit they obtain from additional visual cues. This variability has not yet been fully accounted for by measures of age, auditory - or cognitive abilities. Presumably, further variables such as AV integration capacities, i.e. the extent to integrate auditory and visual input, play a role, as real-life communication does not occur only in the auditory domain but activates different modalities simultaneously, making speech essentially multisensory in nature. Paradoxically, classical speech reception threshold (SRT) tests assess speech intelligibility solely in auditory-only (AO) conditions, which might not reflect a realistic listening scenario.

To address these issues, we focus here on the relation between AV integration capacities and AV speech-in-noise performance. AV integration capacities can be quantified based on illusory percepts induced by incongruent, audiovisual conditions. We use the susceptibility to the McGurk effect and the sound-induced flash illusion (SIFI) as measures for the strength AV integration. In order to determine the temporal window of integration, i.e. the time interval in which the inputs are likely to be integrated, we vary stimulus onset asynchronies (SOAs) between the stimuli between 70 and 420 ms. To assess AV speech intelligibility and the gain from adding visual information in terms of SRTs, we employ a newly developed audiovisual version of the well-established Oldenburg sentence test (OLSA), the AV-OLSA, using AV and AO conditions in different noise types.

Measuring 25 normal-hearing, elderly individuals (60 - 80 years), we aim to investigate:

(1) how inter-individual differences in the strength and window of AV integration relate to the variability in AV speech intelligibility in noise and in the benefit obtained from additional visual information, i.e. audiovisual gains.

(2) how the two tests of AV integration (SIFI, McGurk) relate to each other as they reflect distinct types of changes in perception: In SIFI, visual perception is modulated by auditory stimuli, whereas in the McGurk effect, auditory perception is changed by visual stimuli.

Furthermore, these data shall be compared to a group of mild-to-moderately hearing-impaired elderly individuals in order to quantify the separate contribution of hearing loss from age on changes in AV integration and AV gains. Results are presented at the conference.

12 Hidden hearing loss and selective attention in the brainstem

M. Saiz Alia, T. Reichenbach
Imperial College London, UK

Cochlear synaptopathy or hidden hearing loss can be caused by noise exposure and ageing. It refers to the damage of higher-threshold auditory-nerve fibres and may account for the differences in the ability of normal hearing threshold listeners when communicating in challenging environments (Bharadwaj et al., 2014). However, it remains unclear if the condition actually occurs in humans, how it can be best diagnosed, and how exactly it impacts speech-in-noise processing.

Recently we proposed a method for measuring the brainstem's response to natural non-repetitive speech and employed it to show that the auditory brainstem already plays a role in selective attention to speech (Forte et al., 2017). We thereby observed individual differences in the modulation of the brainstem response by selective attention: some subjects showed large attentional modulation while others exhibited only little modulation. We therefore wondered if the strength of the attentional modulation correlates with hearing ability and if it relates to cochlear synaptopathy.

We approached this issue through a computational model and experimental measurements. First, to explore the effects of hidden hearing loss, we developed a realistic computational model of the auditory-brainstem response (ABR) to speech based on an existing model (Zilany et al., 2014). We employed it to investigate the neural response to continuous speech at different stages in the brainstem, and to explore the effects of hidden hearing loss. We found significant responses and characteristic latencies for neural signals generated at the level of the auditory-nerve fibres, the cochlear nuclei and the inferior colliculus (IC). The latency of the response of the IC matched the latency that we found experimentally, suggesting that the scalp-recorded brainstem response to speech is dominated by the IC.

Secondly, we assessed young healthy listeners for speech-in-noise comprehension, lifetime noise exposure, the middle ear muscle reflex, binaural hearing and different brainstem measures, including the brainstem response to continuous speech and its modulation by selective attention. We found that there was considerable variability in all measures across the participants. However, only few of the objective measures were able to explain the differences in speech-in-noise comprehension between the participants. Interestingly, the modulation of the brainstem response by selective attention correlated with the performance in the speech-in-noise task. Our findings suggest that the attentional modulation in the brainstem response can inform on hearing ability and potentially on hidden hearing loss.

References

Bharadwaj et al. (2014). Frontiers in systems neuroscience, 8, 26.

Forte et al. (2017). eLife, 6, e27203.
Zilany et al. (2014). The Journal of the Acoustical Society of America, 135(1), 283-286.

13 Setting the scene: speech understanding and listening effort in virtual scenarios

A. Devesse, A. van Wieringen, J. Wouters
KU Leuven, ExpORL, Belgium

Background – The assessment of speech understanding in ecologically valid listening scenarios has gained a lot of interest over the past few years. To answer the need of ecological test paradigms, we developed AVATAR, a comprehensive method and test set-up for the real-life assessment of auditory functioning. Compared to clinical speech-in-noise tests, which are most of the time static and one-dimensional, AVATAR aims to mimic the dynamic visual and auditory aspects of everyday listening scenarios. Hence, we aim to capture the auditory challenges people face during daily communication and take cognitive factors like listening effort into account.

In AVATAR, listeners are immersed in different virtual scenarios: a restaurant, living room and public transport environment, each including virtual humans uttering speech auditory-visually. Listening effort is measured by means of an extended dual-task paradigm, with secondary tasks on both auditory localization and visual memory. While previous measurements have shown that the restaurant scenario allows to assess both speech intelligibility and listening effort effectively, the living room and public transport scenarios have not been evaluated yet. The aim of this methodological study is to compare outcome measures in all three scenarios, for different age groups. Additionally, we want to investigate the effect of informational versus energetic masking on speech understanding and listening effort.

Methods – Young (N = 10) and middle-aged (N = 10) normal hearing, Dutch speaking adults will participate in the study. All perform an auditory-visual speech-in-noise task in the restaurant, living room and public transport scenario. Speech is presented in both a multitalker babble noise (energetic masker) and competing Swedish talker (informational masker). Next, secondary tasks are added to the speech-in-noise task to obtain a measure of behavioral listening effort. Finally, participants fill in a questionnaire on subjective listening effort and motivation for each scenario and noise type.

Results – First data on young adults suggest speech understanding is better when presented in a competing talker noise compared to multitalker babble, while the amount of listening effort seems equally high in both noise types. The type of scenario does not affect speech intelligibility, nor listening effort. So far, no clear correlations are found between the behavioral outcome measures and subjective listening effort or motivation.

Funding - a TBM-FWO grant from the Research Foundation-Flanders (grant number T002216N).

14 Effects of limited attenuation and signal replay on ideal binary masked speech with very low mixture SNRs

S. Graetzer, C. Hopkins
Acoustics Research Unit, University of Liverpool, United Kingdom

This study concerns the effects of limited attenuation and signal replay on intelligibility when speech is mixed with white Gaussian noise at low signal-to-noise ratios (SNRs) and subsequently enhanced with an Ideal Binary Mask (IBM). Such masks require a priori knowledge of both the target signal and the masker. The standard IBM takes values of zero and one, and is derived by comparing the instantaneous or 'local' SNR in each time-frequency bin against a pre-set threshold ('Local Criterion' or LC), e.g., 0 dB. Speech produced by four speakers of British English was mixed with white Gaussian noise at SNRs as low as -25 dB. These signals were subsequently enhanced using IBMs with LC = 0 dB or LC = SNR. The standard IBM was compared with an alternative IBM that took values of 0.2 and one, i.e., using limited signal attenuation. To investigate the effect of replay on the intelligibility of enhanced speech involving very low mixture SNRs, each signal was presented three times consecutively to normal-hearing listeners. The results indicate the importance of mask density for speech signals mixed with white Gaussian noise at low SNRs, where density is measured as the number of ones in the mask. In this study, some masks were sparse due to high Relative Criterion (RC) values, where RC is defined as LC – (global) SNR. There were benefits of limited attenuation at low SNRs for LC = 0 when the masks were sufficiently dense (> 5%), but these tended not to occur for LC = SNR. For LC = SNR, a second and third presentation resulted in increases in intelligibility scores, whereas for LC = 0, a third presentation was only beneficial when masks were sufficiently dense.

15 Modelling binaural speech intelligibility against a harmonic masker

L. Prud'homme, M. Lavandier
Univ Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, Vaulx-en-Velin, France

V. Best
Dept of Speech, Language and Hearing Sciences, Boston University, USA

Speech intelligibility models are able to predict intelligibility of a target voice in the presence of non-stationary noise interferers. However, there is currently no model able to accurately predict intelligibility in the presence of competing speech maskers. Our aim is to develop a binaural intelligibility model able to do so. As a first step, we need to accurately predict energetic masking for speech maskers, which will then allow us

to quantify informational masking. Contrary to a noise signal, a speech signal has a harmonic structure that allows for F0 segregation. F0 segregation can be due to either spectral glimpsing or harmonic cancellation, but it is unclear what the relative contributions of these two mechanisms are to F0-based release from masking. In this work we have modified the model of Collin and Lavandier (2013) in order to take into account spectral glimpsing, and we are currently working on including harmonic cancellation. The model is being applied to two data sets. Leclère et al. (2017) measured SRTs for harmonic maskers that varied in their fundamental frequencies, temporal envelope and spatial position. Deroche et al. (2014) also measured SRTs for harmonic maskers that varied in their fundamental frequencies and degree of harmonicity. These two data sets represent a step between noise and speech masking as they include F0 differences but no informational masking. Comparison of model predictions to the two data sets will allow us to establish to what extent the model can predict F0 segregation for harmonic maskers, and determine the relative roles of spectral glimpsing and harmonic cancellation.

16 The effect of cognitive noise on duration, intensity and pitch discrimination of a synthesised vocoid

F. Chiu, L. L. Rakusen, S. L. Mattys
Department of Psychology, University of York, United Kingdom

Research on sub-optimal listening conditions has been concerned primarily with speech-in-noise, where the acoustic signal is affected by energetic and/or informational masking. Less is known about "cognitive noise". Cognitive Noise (CN) is a load-based adverse condition where listening takes place alongside a concurrent non-auditory attentional or mnemonic task. While the concurrent task creates no acoustic interference to the speech signal itself, it is thought to place demands on cognitive resources and deplete resources available for speech perception. The effect of CN on sentence comprehension, word segmentation, and phoneme identification has been documented. Its effect on basic hearing processes is poorly understood, however. Our study investigates the effect of CN on low-level auditory perception, specifically, the discrimination of duration, intensity, and pitch cues.

Ninety-six participants were randomly assigned to one of three discrimination tasks: Duration, Intensity, and Pitch (n = 32 in each). Discrimination was assessed by performance on a 3I-2AFC test, which provided just-noticeable differences (jnd) for each of those three dimensions adaptively. The stimuli were derived from a base Klatt-synthesised vocoid 500-ms stimulus resembling /ɑ/, with F0 150 Hz, F1 836 Hz, F2 1152 Hz, F3 2741 Hz, played at 60dB. Deviant stimuli varied in Duration (500–800 ms), Intensity (60–70 dB), or Pitch (150–153 Hz) across 60 equidistant steps. CN was imposed via a secondary visual n-back task implementing two types of load, Rhyme and Image. Rhyme CN stimuli were written monosyllabic nonwords; Image CN stimuli were un-nameable,

meaningless characters In the Rhyme condition, participants pressed a key every time they saw a nonword that rhymed with the preceding nonword (1-back, low CN) or with the nonword two steps earlier in the sequence (2-back, high CN). In the Image condition, they responded to image repetition either consecutively (1-back) or separated by one intervening image (2-back).

CN significantly increased jnds for Duration and Intensity, but not for Pitch. This pattern held for both Image and Rhyme CN. The results show that CN, just as physical noise, affects the precision with which the primary dimensions of sound are processed. The apparent encapsulation of pitch from cognitive load suggests that this dimension is processed more automatically than duration and intensity. Consequences for theories of speech perception in adverse conditions will be discussed.

17 Auditory emotion recognition: Insight from affective speech, music and vocalisations

J. Kirwan, A. Wagner, D. Başkent
Research School of Behavioural and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, Netherlands; Department of Otorhinolaryngology, University Medical Center Groningen, University of Groningen, Netherlands

The communication of emotion is an essential part of our daily interaction, and recognising emotions can elicit a reaction in the observer, such as the well-known fight-or-flight response. An index of this response can be observed in pupil dilations, which have been demonstrated to dilate more for emotional stimuli in comparison to neutral across several stimulus types, including pictures, environmental sounds and even music. However, the pupillary response has also been shown to reflect mental effort, attention, and other cognitive processes. In this study, we aimed to understand further if the pupil reflects an automatic response to emotions, or if it is indicative of the cognitive processes involved in the labelling of emotions.

We recorded pupil dilations as participants listened to affective speech, as well as other auditory stimuli in the form of music and vocalisations. Participants were instructed to provide behavioural responses to the perceived valence, arousal and emotion category of the stimulus. Our three stimulus types differ in their emotional salience, for instance, crying is a clear indicator of sadness whereas a short minor musical piece may require the listener to use their cognitive abilities to recognise the emotion. We expect our semantically neutral speech to elicit a response somewhere in-between these two types. Therefore, we can trigger different emotional responses in the pupil that may be indicative of autonomic arousal in response to crying and cognitive processing in response

to music. By investigating the pupil's response in this way, we explored the existence of a signature in the pupil to emotional stimuli, and this was shown to be somewhat generalisable across stimulus types, but the response also contains unique features to each stimulus set; speech, music and vocalisations.

18 Differences in pitch in native and non-native speech in noise

K. Marcoux, M. Ernestus
Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

When speaking in noise, the acoustics of our speech change and we produce Lombard speech. Extensive research on native speakers has shown that, compared to speech produced in quiet, Lombard speech has, among other properties, a higher Fundamental Frequency (F0, pitch), higher amplitude (loudness), and a decrease in spectral tilt (shift in energy) (e.g. Summers et al. 1988). Lombard speech is typically viewed as a reflex, shown by speakers across different languages. Non-natives, however, are confronted with several added difficulties when producing Lombard speech, including experiencing a higher cognitive load and potential differences in pitch ranges between the two languages (Jenner 1976; Wen, Mota, and McNeill 2015). Considering these added challenges, we wanted to investigate whether native and non-native Lombard speech differs in terms of F0.

We recorded 30 Dutch natives reading stimuli in Dutch and English and 9 American-English natives in English, in quiet and noise (hearing 83 dB SPL Speech-Shaped Noise through headphones). We additionally manipulated the focus in the sentences by means of contrastive question answer pairs, creating early and late focus answer sentences. This resulted in 36 question-answer pairs in four conditions (quiet early-focus, quiet late-focus, noise early-focus, and noise late-focus), for instance, "No, the tall (focus) woman drove to the pub in town" (early-focus) and "No, she drove to the pub (focus) in town." (late-focus).

The recordings were segmented at the sentence level and Praat returned the median F0 value. Using the lme4 package in R, we conducted linear mixed effects models with language or nativeness, noise, and focus as fixed effects and random slopes for participant and stimuli as random effects. Our preliminary analyses illustrated that there was an effect of noise for the Dutch speakers, leading to an increase in F0 in their production of English and Dutch Lombard speech as compared to their speech produced in quiet. This effect of noise was only present for the American-English speakers in the late-focus condition.

The data on native and non-native Lombard speech suggest that non-native speakers also produce Lombard speech, supporting the hypothesis that Lombard speech is a reflex. Against the reflex hypothesis, however, are our findings that, rather than being a universal phenomenon, Lombard speech interacts with the phonetic properties of the language. This indicates that it is more language specific than originally expected and that it has to be acquired by second language learners.

19 Task-dependent decoding and encoding of sounds in natural auditory scenes based on modulation transfer functions

L. Hausfeld, G. Valente, F. De Martino, E. Formisano
Dept. of Cognitive Neuroscience, Maastricht University, The Netherlands

Recent studies used decoding and encoding analyses to examine the neural processing of natural continuous sounds as measured with EEG and MEG in multi-speaker environments. A robust finding is that the (attended) sound envelope is strongly represented by these cortical signals. However, it remains unclear whether and which acoustic features of the sound are more strongly represented in cortex than others.

In this study, participants (N=17) are presented with natural listening situations including two speakers and music while they perform selective attention tasks and EEG signals are acquired. We use a decoding and encoding approach based on sound descriptions by modulation transfer functions (MTF) and, more specifically, temporal modulations of 2-32Hz (log-scale).

Temporal response functions (TRFs) are estimated to reconstruct sounds from EEG data for each modulation (decoding) and to predict EEG responses based on MTFs (encoding).

Decoding results show better reconstruction when based on temporal modulations compared to decoding based on sound envelopes for all temporal rates. Attention effects for speech and music decoding were strongest during faster rates (4-8Hz).

Encoding results based on models of temporal rates performed similar to envelope-based encoding. For speech, encoding models suggest strong differences between attended and unattended sounds at specific rates and delays. For music, models were independent of attentional state.

These results shed more light on previously reported envelope-based descriptions of EEG data and point towards distinct processing of attended and unattended speech sounds as characterized by temporal modulations and delays. Remaining analysis including spectral modulations will provide additional insights.

20 The effects of ceiling height and absorber placement on speech intelligibility in simulated restaurants

J. F. Culling
School of Psychology, Cardiff University, United Kingdom

The intelligibility of speech was measured in simulated rooms with parametrically manipulated acoustic features. In experiments 1 and 2 binaural room impulses were generated using a simple ray-tracing model for rectangular spaces. In order to simulate more complex geometries, including representations of furniture and room occupants, experiments 3 and 4 used CATT AcousticTM. The rooms were designed to simulate restaurant environments with either three or nine occupied tables. In Experiment 1, rooms of equal total absorbance were compared, but with most absorption located either on walls or on the ceiling. Wall absorption produced shorter reverberation times and improved speech reception thresholds (SRTs). In experiment 2, rooms differed in ceiling height. Lower ceilings produced shorter reverberation times but poorer SRTs. Both total absorbance and reverberation time were thus poorly correlated with speech intelligibility.

A psychoacoustic model of spatial release from masking (Jelfs et al., 2011) produced very accurate predictions of SRTs, based on the binaural room impulse responses for each experiment. Experiment 3 also varied ceiling height, but in combination with the effect of ground-level acoustic clutter, formed by furniture and room occupants. As predicted by the model, both high ceilings and acoustic clutter produced better SRTs. Experiment 4 compared acoustic treatments of the ceiling in the presence of the acoustic clutter. As predicted by the model, continuous acoustic ceilings were more effective at improving SRTs than suspended panels, and suspended panels were more effective if they were acoustically absorbent on both sides. The results suggest that the most effective control of reverberation for the purpose of conversational intelligibility is provided by absorbers placed vertically and close to the room occupants.

21 SpiNNak-Ear — Auditory pathway simulation on neuromorphic hardware

R. James, J. Garside
School of Computer Science, University of Manchester, UK

The first contribution of this work to the Speech In Noise (SpiN) workshop is to confuse matters by introducing a second ‘SpiN’ acronym. The SpiNNak-Ear system is an auditory pathway simulation tool implemented on the Spiking Neural Network architecture (SpiNNaker) neuromorphic platform.

A SpiNNaker machine is neuromorphic (brain inspired) in its design, where computational nodes (ARM microprocessors) are low power and spread across a vast network - much like networks of biological neurons. It was designed for applications of simulating large scale spiking neural networks; however, its massively parallel computational architecture is suitable for a range of applications that are not exclusively based around modelling spiking neurons.

SpiNNak-Ear takes advantage of the inherent parallel processing across the programmable computational nodes in the SpiNNaker system, achieving real-time binaural simulation of the early auditory pathway (from pinna to Auditory Nerve [AN]) to a biologically realistic scale (30,000 AN fibres).

Large scale simulation of neural cell populations in subsequent regions of brainstem nuclei are performed on the same SpiNNaker hardware. The scope for continuing this process into modelling the remainder of the binaural nuclei in the auditory pathway and associated cortical regions on this platform is achievable across a large scale machine of up to one million microprocessor cores.

The system presented here has the capabilities for modelling the auditory brain at a biologically realistic scale without incurring the inherent performance penalties of traditional computer architectures. It has advantages over alternative high-performance computational platforms by using a unique, brain inspired core-to-core 'multi-cast' communication protocol. This can be utilised when simulating the numerous inter-nuclei descending projections that feature in the auditory pathway.

The SpiNNak-Ear platform can aid future auditory research in explaining the complex in-vivo auditory neuron response, the development of future hearing prostheses, and in gaining a better understanding of the biological feedback pathways and their potential role in extracting salient stimuli from a noisy environment.

22 Investigating the role of linguistic information in perception of voice cues

F. Arts, E. Gaudrain, T. N. Tamati, D. Başkent
Department of Otorhinolaryngology, University Medical Center Groningen, Groningen, The Netherlands

Speech perception is strongly dependent on perception of talkers' voices. Through voices, we identify individual speakers, and in situations with multiple talkers (i.e., cocktail party situations), voices help us to discriminate between different speakers.

Talker voice perception is a major challenge for cochlear implant (CI) listeners due to the spectro-temporally degraded signal the CI delivers, caused by limitations related to electric stimulation. Limitations in voice perception may contribute to difficulties perceiving and understanding speech.

Talker voice perception relies upon vocal source and articulatory cues, which convey indexical information about the speaker, but linguistic information also influences perception of talkers' voices. Acoustic-phonetic variability that affects phonetically relevant properties of speech impairs linguistic processing, and slower processing enables listeners to better attend to voice characteristics. Furthermore, perceptual vocal variability may be stored in linguistic memory representations, resulting in varying sensitivity to talker voice details depending on the type of linguistic information. Although the interaction of linguistic information and talker voice perception has been established, how acoustic-phonetic properties of speech and different linguistic information interact with perception of individual voice cues remains unknown.

The current research investigates the role of linguistic variability (e.g., words, nonwords) in the perception of individual vocal characteristics. Just-Noticeable-Differences (JNDs) were obtained for F0 and formant frequency distributions related to vocal tract length (VTL), presenting normal-hearing (NH) listeners with artificial F0 and VTL manipulations in a 3AFC paradigm. JNDs were compared for easy and hard words based on frequency and neighborhood density (NHD), and for easy and hard nonwords based on NHD and phonotactic probability. Furthermore, token identity was varied, presenting listeners with three equal or different stimuli.

Preliminary findings show that voice cue perception is affected by language-specific acoustic-phonetic details rather than top-down lexical characteristics, and suggest that voice cue perception in nonwords depends on phonotactics. This confirms the dependence of voice cue perception on acoustic-phonetic features in speech. Similar thresholds for easy and hard words were observed, strengthening the idea that linguistic variability affects voice cue perception only if this variability alters acoustic-phonetic details of speech. Better insight in the nature of talker voice perception may be a next step towards improvement of talker voice perception in CI listeners, which may eventually improve speech perception in this clinical group.

23 Brain monitoring of distraction from speech in noisy context using EEG

E. Eqlimi, D. Botteldooren
WAVES, Dept. of Information Technology, Ghent University, Belgium

A. Bockstael
École d'orthophonie et d'audiologie, Université de Montréal, Canada

Speech is one of the most important forms of communication, yet it is often embedded in background noise, such as babble, traffic, and other environmental sounds. This background impairs speech intelligibility but can also distract attention away from the narrative. The main goal of this study is to identify neurological EEG biomarkers that may indicate whether participants are paying attention to the information that is pre-

sented, or not. Prior work has used trained computer models to identify which of two interfering speech segments is attended to. The current research focusses on non-speech background sound and in particular on the possibility to identify distraction by salient events such as a phone ringing, a car horn's honk, or an emergency siren.

Five-minute meaningful speech fragments were presented in a very low level of pink noise and in three types of background noise. The level of the background sound was low enough not to cause energetic masking. Participants were instructed to pay attention to the lecture and were informed they would be questioned about its content. The background sounds were also presented separately during a resting period where participants were asked not to pay attention to the sound. In total 23 participants were exposed to the stimuli while their 64-channels EEG was recorded. For exclusion purposes, a full battery of audiological tests was also performed. After cleaning artifacts, the single-trial EEG was analyzed using a wavelet spectrogram. The overall spectral power as well as the response evoked by assumed distractor (e.g. phone ringing) were assessed.

In agreement with previous research, significant differences were found between participants. However, there were also some consistent trends. When comparing the attentive listening to speech in background noise to inattentive exposure to background: (a) an increase in low-frequency fluctuating power (especially in the frontal channels) was observed which might be attributed to an amplitude following response; (b) an increase in gamma power in F4 (right) in comparison with F3 (left) occurs, which could be related to linguistic processing of speech. Based on the hypothesis of increased inhibition and gating out of disturbing events during attentive as well as inattentive listening, an increase in frontal alpha power could be expected during phone ringing. This was only confirmed in the background-only samples. Between people, more alpha power in the frontal area seems related to lower amplitude following responses to the disturbing event.

24 The combined predictive value of multiple cognitive abilities for speech-in-noise perception by older adults

A. Heinrich
University of Manchester, Manchester, UK

S. Knight
University College London, London, UK

There is a broad consensus that cognition is important for speech-in-noise (SiN) perception, but its exact role remains to be understood. Even for the simplest case where the relationship between cognition and SiN performance is tested for different cognitive components in isolation, it is not clear exactly which cognitive components contribute in any given listening situation. Moreover, it is at least possible that a combination of cognitive abilities may provide a truer description of cognitive contributions. For example, compensatory mechanisms may lead to listening being accomplished in different ways: listeners with relatively poor abilities in one relevant cognitive domain may compensate with relative strength in another. Alternatively, cognitive abilities may interfere with one another – for example, a large vocabulary combined with good working memory may be deleterious if listeners are storing phonetically similar competitor words instead of the target.

In this study, 50 older adults (ages = 61-86, mean = 70; age-normal hearing) performed tasks designed to assess a range of cognitive abilities (simple/complex working memory, verbal knowledge, reading comprehension, inhibition). They also performed three SiN tasks (isolated words, and final words in low-and high-predictability sentences), presented in speech-modulated noise at two signal-to-noise ratios. Individual measures of hearing (PTA 0.25-8kHz) were obtained. A series of multiple regression models were fitted to assess the contribution of the tested cognitive abilities to SiN perception, alone and in interaction, for each listening condition and while accounting for PTA.

Results showed interactions between cognitive task scores in their ability to predict SiN performance, but only at the lower (more challenging) signal-to-noise ratio. Many of these interactions followed a pattern that was consistent with a compensatory mechanism: if participants had relatively poor scores in one cognitive task, then increasingly high scores in another cognitive task were associated with improved SiN perception. We also found some evidence for interference, where increasingly high scores on a cognitive task were associated with poorer SiN perception if participants were already performing well on another cognitive task.

These findings reveal a complex picture of the relationships between intelligibility and cognition, which may help us understand some of the inconsistencies in the literature regarding cognitive contributions to SiN perception. In particular, they suggest that: 1) more than one cognitive ability predicts SiN performance; 2) these cognitive abilities may interact to predict SiN performance; and 3) these interactions are not always advantageous to SiN listening.

Supported by BBSRC grant BB/K021508/1.

25 Computational model for the modulation of speech-in-noise comprehension through transcranial electrical stimulation

M. A. Kegler, T. Reichenbach
Imperial College London, UK

Background — Transcranial electrical stimulation (TES) can non-invasively modulate neuronal activity in humans. Recent studies have shown that TES with an alternating current that follows the envelope of a speech signal can modulate the comprehension of this voice in background noise (Wilsch et al., 2018). However, how exactly TES influences cortical activity and influences speech comprehension remains poorly understood. Here we present a computational model for speech coding in a spiking neural network and employ it to investigate the effects of TES on the coding of speech in noise.

Methods — Based on previous work, we established a computational model of a spiking neuronal network that encodes natural speech through entraining network oscillations in the theta and gamma frequency range (Hyafil et. al., 2015). We used the network's spiking output to classify speech in different levels of background babble noise. We then investigated the effect of different external currents on the network dynamics as well as on the neural output and the associated speech coding. Finally, we analysed the behaviour of the computational model and its speech classification performance in different conditions to optimize the stimulation paradigm for enhancement of natural speech processing.

Results — The computational model generated coupled oscillations in the theta and the gamma frequency range. In agreement with experimental results, the slower theta oscillations reliably predicted the onsets of syllables and provided a temporal reference frame for the faster activity in the gamma band that encoded phonemes. Classifying speech in different levels of background noise yielded results comparable to normal human performance, with a 50% speech recognition threshold at approximately -1 dB SNR. Simulating the effect of simultaneous external current with a range of different temporal patterns and stimulation intensities we were able to identify the parameters that impeded as well as enhanced the neural coding of speech in noise.

Conclusions — The developed model provides an insight into the neural mechanisms through which speech in noise can be processed in the auditory cortex and how TES can enhance this processing. Moreover, our computational model allows to optimize the temporal pattern of the stimulation for improving speech-in-noise comprehension.

References

Hyafil et al. (2015) Elife, 4, e06213.

Wilsch et al. (2018) NeuroImage, 172, 766-774.

This study was supported by the EPSRC Centre for Doctoral Training in Neurotechnology for Life and Health. We thank Alexandre Hyafil, Shabnam Kadir and Milos Cernak for helpful comments and fruitful discussions.

26 Using multimodal imaging techniques to study effects of auditory training in older adults

G. Mai
Dept. of Experimental Psychology, University College London, United Kingdom

I. Tachtsidis
Dept. of Medical Physics and Biomedical Engineering, University College London, United Kingdom

P. Howell
Dept. of Experimental Psychology, University College London, United Kingdom

More difficulty in speech-in-noise (SPiN) perception is experienced by older, than young, adults. Targeted auditory training (AT) is beneficial for improving older adults' SPiN performances. However, change in brain activity in different neural mechanism consequent on AT remains unclear. The present study uses multimodal techniques that combine functional near-infrared spectroscopy (fNIRS) and systemic physiological measurements to look into neural and physiological changes resulting from AT in older adults (native English speakers over 60 years of age).

Participants receive a pre-training and a post-training fNIRS scanning session. Specifically, they complete a SPiN recognition task under six-talker babbles while the speech reception threshold (SRT) is measured. Subsequently, a neuroimaging experiment is conducted, in which they are instructed to actively attend to short sentences in noisy environments as in the SPiN task but at a fixed SNR level for pre- and post-training sessions. FNIRS signals are recorded over the temporo-frontal areas. Activities are measured at the posterior superior temporal gyrus (pSTG) that reflect responses to speech intelligibility and left inferior frontal gyrus (LIFG) that reflect listening efforts. Connectivity analyses between these two regions are also conducted. Systemic physiology of skin conductance (EDA), heart rate variability (HRV) and respiration rate are simultaneously

recorded to provide further measures of listening effort and as control for physiological confounds in fNIRS signals. Between the two scanning sessions, participants receive an adaptive online take-home SPiN perception training that they complete over four weeks.

Differences in behavioral, neural and physiological signals between the pre- and post-training are used to quantify the effects of AT. We anticipate that results of the present study can elucidate the neural mechanisms of how AT contributes to SPiN perception and shed light on future clinical implications for auditory rehabilitation in aging populations.

27 The role of talker acoustics for intelligibility and effort in adverse listening conditions

M. Paulus, P. Adank, V. Hazan
University College London, United Kingdom

A. Wagner
University Medical Center Groningen, the Netherlands

A current focus in hearing research is the use of measures of listening effort to complement speech audiometry (Pichora-Fuller et al., 2016). Central to research into listening effort is the emphasis on the characteristics of the listener such as nativeness or hearing status. In this study, we investigated the effect of talker acoustics on listening effort by measuring the relative importance of various acoustic-phonetic dimensions in adverse listening conditions. Previous research related talker-specific acoustic-phonetic features to intelligibility in noise (Hazan & Markham, 2004), but has not yet accounted for listening effort.

Based on an anechoically recorded corpus of sixteen Southern British English speakers, we conducted listening experiments in combination with pupillometry. We presented temporally or spectrally distorted speech (using time-compression and noise-vocoding, respectively). Furthermore, undistorted speech was presented in clear and with speech-shaped masking noise. Intelligibility scores were obtained based on keywords recognised correctly. Listening effort was measured by tracking pupil size changes over time.

Our results are in line with previous studies measuring talker intelligibility in noise. Pupil dilation was increased for both, distorted and masked speech. We found that speaking rate was a common predictor of intelligibility in both distorted conditions. We furthermore observed that speaking rate was related to changes in pupil size, indicating reduced listening effort and sustained retrieval processes for slower talkers. The current results are interpreted in the context of models of listening effort such as FUEL (Pichora-Fuller et al., 2016).

28 Effects of temporally fluctuating maskers on speech production and communication

J. Saigusa, V. Hazan
University College London

Several decades of research have been devoted to investigating how speech production in noise manifests as a function of loudness, task type, spectral content of noise, and other properties. Recently, work on how listeners track speech in noise has found that neural oscillators entrain with envelope modulations of the attended stream, and that speakers produce more pronounced modulations when speaking in noise. However, to date there has not been any investigation into the possible effects of varying noise types on amplitude modulations and the ability of talkers to adapt them to the noise environment.

This aim of this study is to expand on the finding that speakers produce more pronounced modulations when speaking in noise, and more generally that speakers can adapt to temporal fluctuations in the masker. Pairs of normally hearing adults between the ages of 18-35 are audio recorded while completing a sentence repetition task (N=40). The talkers are seated in adjacent booths and communicate via headset in a 'virtual room' that simulates the acoustics of a real room, including reverberation and sound/speaker locality (www.phon.ucl.ac.uk/resource/audio3d/). One talker ('Talker A') reads Harvard sentences to Talker B, who repeats back what they heard in 4 noise conditions (speech-shaped noise modulated by 1 Hz, 4 Hz, and 8 Hz square waves for an 'on-off' effect) presented at 80dB, as well as in a quiet condition. Various global acoustic-phonetic measures are taken of Talker A's speech, including f0 median and range, articulation rate, and mean energy, as well as some temporal measures. These include a comparison of speech energy in masker 'on' periods to energy in masker 'off' conditions, and calculating the modulation spectrum of Talker A's speech. It is predicted that speech produced in the presence of 1 and 4 Hz maskers will show more pronounced modulations in the amplitude envelope at 1 and 4Hz as talkers adjust to speak in the 'gaps', whereas 8 Hz, which is faster than a normal syllable rate, will not show this effect. The results will further understanding of how and to what extent talkers interact with temporally fluctuating noise.

29 Effects of energetic and informational masking on interactive speech communication in younger and older adults

O. Tuomainen, L. Taschenberger, V. Hazan
University College London, UK

Our ability to communicate successfully with others can be strongly affected by the presence of noise and other voices in the environment, and older adults (OAs) can be more greatly affected than young adults (YAs) in these situations. The interference (or masking) that is caused by one sound on the perception of another depends on the characteristics of the masker: whether energetic (EM) or informational masking (IM, e.g., speech-on-speech masking), and the weighting between the effects of EM and IM can vary with the age of the listener, some studies showing increased effect of IM in older adults. However, these results are mostly based on perception tests using simple pre-recorded sentences that are void of communicative intent and, therefore, less representative of our everyday communicative situations.

The focus of this study is to investigate the impact of EM and IM on interactive speech communication between groups of YA and OA talkers. We audio record pairs of normally hearing YA and OA female talkers (18-30 and 55-75 years; N=30) while they carry out a "spot-the-difference" picture description task (diapix) with another volunteer from the same age range. The picture task is carried out in four listening conditions affecting both participants: both speakers in i) quiet, ii) IM that is semantically related to the picture description task iii) IM that is semantically unrelated to the task (both 3-talker maskers: male, female and a child), and iv) EM (speech-shaped-noise, SPSN, for the male, female and a child talker). To simulate acoustics in real rooms, we use real-time virtual audio environment software (www.phon.ucl.ac.uk/resource/audio3d/) with spatially separated sound sources, for each of the three talkers in the maskers, presented over headphones (at 72 dB in the noise conditions). To assess the impact of EM/IM on communication we measure i) communication efficiency (time it takes to find differences in the picture task) and ii) speaking effort (acoustic-phonetic features of their speech: f0 median, mean energy in 1-3 kHz range, articulation rate). We expect greater difficulty (less efficient communication, more effort) in adverse conditions by OA talkers, with greatest difficulty imposed by task-relevant IM condition. In order to assess if communication efficiency in noise is influenced by differences in cognitive factors between YA and OA talkers, such as selective attention, we measure distractibility via a secondary go/no-go auditory detection task, and expect secondary task accuracy to be associated with task performance in older talkers.

30 Studying effects of vibrotactile stimulation on the neural tracking and intelligibility of continuous auditory speech

L. Riecke, S. Snipes, S. van Bree, A. Kaas, L. Hausfeld
Maastricht University, The Netherlands

Viewing a speaker's lip movements can improve the brain's ability to 'track' the amplitude envelope of the speech signal and facilitate intelligibility. We hypothesize that such neural and perceptual benefits can also arise from tactually sensing the speech envelope on the skin. To test whether a tactile speech envelope can improve neural tracking of degraded (envelope-reduced) auditory speech and speech recognition, we present continuous audio-vibrotactile speech at various asynchronies to the ears and index fingers of normally-hearing listeners while simultaneously assessing auditory cortical speech-envelope tracking (using electroencephalography and a stimulus-reconstruction approach) and speech-recognition performance.

Our results so far indicate that tactile stimuli carrying speech-envelope information may benefit cortical tracking, but not intelligibility, of degraded auditory speech. The speech-tracking benefit is strongest when the tactile input leads the auditory input by ~50ms and it seems to be primarily driven by early (~100ms post tactile input) cortical responses in the delta (1-4Hz) range. These observations provide preliminary insights into how the human auditory system integrates continuous audio-tactile speech and how this affects intelligibility of degraded auditory speech, which may be relevant for understanding the Tadoma method.

31 Brain's temporal response function: Can it be improved with behavioural account of attended stream?

M. P. Huet
University of Lyon, CNRS, Inserm, CRNL, Lyon, France

C. Micheyl
Starkey, Créteil, France

E. Gaudrain
University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology, Groningen, The Netherlands

E. Parizet
University of Lyon, INSA, LVA, Villeurbanne, France

During the past decade, there has been growing interest in the neural correlates of selective attention to speech. In these studies, listeners were instructed to focus their attention toward one of two concurrent speech streams. However, in everyday-life situations, listeners are unlikely to maintain undivided attention on a single talker, and instead, can switch rapidly between different voices. To study this phenomenon, we have developed a behavioural protocol that provides information about which of two competing voices is listened to at different time points, thus reflecting the dynamic nature of concurrent speech perception.

A corpus of short stories was extracted from an audiobook. After listening to two simultaneous stories — a target and an interferer — the participants have to find, among a set of words, those present in the target story. The participant's responses are then used to estimate, retrospectively, when they were listening to the target, or to the interferer.

Neural data, recorded with EEG, and behavioural measures are combined to extract the brain's temporal response function in response to these stimuli. To modulate how many switches between the two voices occur during the course of the stories, the interferers were uttered by the same talker as the target stories, but the voice parameters were manipulated to parametrically control the similarity of the two voices from clearly dissimilar to almost identical.

We will discuss the results in terms of attentional selection and voice confusion, and suggest possible applications of this dynamic behavioural test of selective auditory attention.

32 Switching attention and integration of binaural information: Examining the effects of masker types and binaural listening on perception of interrupted speech in noise

S. Koifman, S. Rosen
University College London, UK

Over the last few years we have been investigating the more general utility of a task which we have shown to be highly sensitive to the effect of aging for speech maskers when compared with a standard speech in noise task. In this task, the target speech is interrupted at a fixed modulation rate (5 Hz), with successive segments of the target being switched from ear to ear. An adaptive procedure is implemented by varying the duty-cycle (DC), which is the proportion of time the signal is 'on' in each modulation period, in order to find the proportion of speech required to understand 50% of the keywords (i.e., the speech reception duty-cycle threshold, SRdT). A masker is interrupted in the same way, and alternated between the two ears out-of-phase with the target speech, resulting in alternated segments of both target and masker signals between the two ears, with only one stimulus present in each ear at any given time. This task appears to exploit some higher-level cognitive aspects of listening not probed by simpler tasks, and is believed to necessitate the listeners' ability to switch attention and integrate short-term auditory information between the two ears.

Here, we aim to examine several parameters that might affect performance in the task, namely: the effect of masker types (speech vs. non-speech), the extent to which listeners are actually obtaining information from both ears as opposed to attending to one ear only, and also the influence of speech material.

To examine the effect of masker type, listeners were presented with simple sentences in three types of maskers; unrelated connected speech, and two non-speech maskers which were extracted from the original speech maskers and varied in their amount of "speech-like" characteristics (from high to low): (1) a single band vocoded speech with natural mix of periodicity and aperiodicity [1]; (2) amplitude modulated speech-shaped-noise. To examine the ability to make use of alternated stimuli, we compared the listeners' performance in two listening configurations: (1) binaural in which the stimuli are fully preserved when segments of the stimuli from both ears are combined, (2) and a monaural configuration where only the information in one ear is presented. Lastly, the influence of speech material was explored by comparing performance with CRM-like sentences.

A fuller understanding of the abilities exploited by this task may make it useful in helping to disentangle the reasons why various groups of people experience difficulty in listening in noisy situations.

[1] Steinmetzger, K., Rosen, S. (2015). The role of periodicity in perceiving speech in quiet and in background noise. J. Acoust. Soc. Am., 138(6), p.3586–3599.

This work was supported by Action on Hearing Loss, UK.

33 Try harder! The influence of evaluative feedback on the pupil dilation response, saliva-cortisol, and saliva alpha-amylase levels during listening

A. A. Zekveld, H. van Scheepen, N. J. Versfeld
Otolaryngology-Head and Neck Surgery, Ear & Hearing, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health, Netherlands

C. E. Teunissen
Clinical Chemistry, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health, Netherlands

S. E. Kramer
Otolaryngology-Head and Neck Surgery, Ear & Hearing, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health, Netherlands

The pupil dilation response is sensitive to listening effort, but is also sensitive to emotions such as those evoked by the social and emotional significance of a task. Standard listening tests that ask listeners to repeat auditory stimuli usually do not take social aspects such as feedback and (threat of) evaluation by others into account. As a result, it is currently unclear to what extent evaluative feedback may influence listening effort as assessed by the pupil dilation response. We aimed to assess this effect using an adapted speech reception threshold (SRT) task. Besides the pupil dilation response, we acquired two physiological biomarkers sensitive to stress: cortisol and alpha-amylase levels as determined in saliva samples.

We included 34 participants with normal hearing (mean age = 52 years, age range 25-67 years) and 29 age-matched participants with mild-to-moderate hearing loss (mean age = 52 years, age range 23-64 years). Half of the participants performed a standard SRT test without feedback, and the other half performed an SRT test in which 1) written feedback was provided after each trial, 2) a performance indicator showed the actual performance level and an unrealistically high target performance level, and 3) the experimenter provided relatively cold, evaluative feedback twice during the test. The SRT conditions targeted 50% and 71% correct reception of the sentences. Pupil size was recorded during listening and saliva samples were obtained before, during and after the test.

As expected, the participants with hearing loss performed poorer on the SRT test than listeners with normal hearing, and lower (better) SRTs were obtained for the 50% as compared to the 71% intelligibility condition. The participants receiving feedback had lower (better) SRTs in the 71% intelligibility condition and higher peak pupil dilation in both intelligibility conditions as compared to the participants who performed the standard SRT test. No effect of hearing status on the pupil dilation response was observed. We will furthermore present the influence of feedback on the subjective ratings, and the cortisol and alpha-amylase levels. We will discuss our findings that indicate that social evaluation can influence listening performance and several physiological measures.

34 Coping with noise at multiple levels: Auditory cortical and lexical effects of maskers for native and non-native listeners

J. Song, L. Martin, P. Iverson
University College London, United Kingdom

Non-native listeners (L2) have more difficulty than native speakers (L1) in noise, but L2 listeners appear to track the acoustics of a target speaker better than do L1 listeners, as measured by cortical entrainment to the speech envelope. In contrast, L1 speakers appear to have more flexibility in lexical processing (i.e., N400) depending on the listening situation. The present study investigated these issues within listening conditions that vary the demands on peripheral and central processing: babble vs. single-talker maskers and a masker collocated with the target or 45° away. Native English and Korean subjects listened to English sentences, and pressed a button when they heard catch trials that did not make sense. EEG recordings were used to measure cortical entrainment to the speech envelope and N400.

The results confirmed previous work that non-native listeners have greater cortical entrainment for target speakers than do native listeners. However, their target-speech entrainment was differentially modulated by masker type, with greater entrainment in the no-masker condition, whereas entrainment by native listeners was less affected by the masker. The N400 effect for target sentences was smaller for non-native than native listeners overall. However, native listeners varied lexical processing depending on the masker, having larger N400 responses when there was a single-talker masker. The results thus suggest that noise can produce different effects for L1 and L2 listeners, and that L1 listeners may be better able to use more processes to cope with informational masking (e.g., searching more thoroughly for a lexical competitor).

35 Effect of sentence complexity, amplification, and hearing loss on the psychometric function obtained from adaptive speech-in-noise tests

C. Lesimple, J. Tantau, B. Simon
Bernafon AG

Hearing aid users expect their hearing devices to provide improved understanding in speech-in-noise situations. This requirement can be validated with adaptive speech tests using sentences as test material. Available speech corpora show differences in terms of complexity and predictability which impact intelligibility and processing demands. Speech material can be classified in two main categories: a) matrix sentences with a fixed grammatical structure which provides a uniform test material once the training phase is done, and b) everyday sentences that introduce more variability in terms of structure and word predictability. Both tests determine the speech reception threshold (SRT) with an adaptive procedure to find the signal-to-noise ratio (SNR) at which 50% of the words are recognized.

The current study investigates the effect of sentence complexity on amplification benefit as a function of the tested SNR. SRTs were measured with a matrix sentence test (OLSA) and an everyday sentence test (GOESA) in three listening conditions: unaided, aided with a RITE device, and aided with an ITE device. Test results, from 16 experienced hearing aid users with a mild to moderately-severe hearing loss, were fitted with a logistic function to extract the SRTs and the slope (s50) of the psychometric function at 50 %. The difference between aided and unaided fitted psychometric curves was used to visualize the differences between listening conditions and sentence complexity over a large test SNR interval. Test material, unaided SRTs, unaided s50, age, and subjective satisfaction measured with the APHAB were selected as fixed effects in a mixed effect regression to predict the amplification benefit.

Results indicate an overall SRT shift towards more positive SNRs with the GOESA as compared to the OLSA test material. This shift suggests that sentences with a more complex structure and unpredictable vocabulary increase the test difficulty. The different SNR range achieved with each test could further impact the aided scores due to the fact that SNR dependent features of hearing aids (like directionality) will provide more benefit at lower SNRs. The results from this test provide indications about factors that might explain variations in the performance of different hearing aids or signal processing algorithms when measured with different speech tests.

36 Quantifying listening effort: who tells what?

A. Bockstael
Université de Montréal, Canada

T. Koelewijn
VU University Medical Center, The Netherlands

A. Khan
Université de Montréal, Canada

A. Guillemette
Institut Universitair de gériatrie de Montréal, Canada

J. P. Gagné
Université de Montréal, Canada

Listening effort becomes more integrated in the assessment of speech comprehension in noise. Measuring listening effort assesses the cognitive resources and exertion related to achieve a certain level of speech intelligibility, and is as such a complementary parameter to speech intelligibility scores (i.e. the number of syllables/words/sentences identified correctly) to quantify speech comprehension.

Listening effort is measured indirectly and therefore challenging, and different measurement paradigms exist. In general, listening effort can be assessed behaviorally, physiologically and through self-report. The different paradigms (to a certain extent) assess different aspects of listening effort, and each of them has its proper strengths and advantages. To-date it is unclear how the outcome from different paradigms should be combined into an overall view on listening effort.

This project investigates how listening effort measured by a dual task paradigm (behavioral measure) and by pupillometry (physiological measure) correlate, and how these two paradigms can be combined to quantify listening effort efficiently and effectively. Outcomes of the dual-task paradigm and pupillometry are both well-established measures for listening effort, however they have never been measured in parallel. In the dual task paradigm, participants are asked to perform two task simultaneously, a primary speech-in-noise task and a secondary tactile task. In this project, speech-in-noise has been assessed for young normal-hearing participants for speech (sentences) presented in multitalker background noise at three signal-to-noise ratios: -5 dB, 0 dB and 5 dB. Speech intelligibility in noise has been carried out alone (single task) and in combination with the tactile task (dual task). Pupillometry has been carried out for all signal-to-noise ratios, and for both single and dual task conditions.

Pilot data from 9 participants show very promising results. For pupillometry, especially the slope of the pupil diameter varying over time appears to be sensitive to the listening conditions. Even with the small sample size, a significant increase in pupil diameter slope is seen during the listening phase when the sentences are presented, compared

to the baseline conditions before and after stimulus presentation. In addition, this significant increase appears to be more pronounced in the dual task conditions when listening task and tactile task have to be carried out simultaneously, compared to the single task paradigm where only the speech intelligibility task is done.

37 Exploring speech envelope enhancement for auditory intervention in children with dyslexia

T. Van Hirtum, A. Moncada-Torres
KU Leuven - University of Leuven, Dept. of Neurosciences, ExpORL

P. Ghesquière
KU Leuven – University of Leuven, Faculty of Psychology and Educational Sciences, Parenting and Special Education Research Unit

J. Wouters
KU Leuven - University of Leuven, Dept. of Neurosciences, ExpORL

It is well established that the dominant core deficit in dyslexia is phonological, but there is growing evidence that subtle speech perception deficits precede the phonological difficulties. Recent theories attribute impaired speech perception in dyslexia to altered processing of dynamic features of the speech envelope, such as slow amplitude fluctuations and transient acoustic cues. Therefore, if speech perception deficits in dyslexia indeed stem from faulty speech envelope tracking, then consequently enhancing the envelope might improve speech perception by persons with dyslexia.

In our previous study we implemented an envelope enhancement strategy (EE) to amplify specific amplitude transition in the envelope, without affecting other parts of the speech signal (Koning and Wouters 2012) in an adult population. We hypothesized that emphasizing these challenging dynamic features might strengthen information processing of syllable onsets and phoneme discrimination and in turn ameliorates speech perception for adults with dyslexia. We found that EE instantaneously improved atypical speech perception in adults with dyslexia.

In the present study, the objective is to generalize these results to a younger population. Therefore, we tested children, age 9 to 12 years, with and without dyslexia using a sentence repetition task in a speech-weighted background noise. We tested speech perception in four different conditions: natural speech, vocoded speech and their enhanced versions. These conditions were used to assess both the nature of the speech perception deficit and the effect of the EE-algorithm on speech perception. Additionally, cognitive test of phonological awareness, language skills, verbal short term memory and working memory were administered to investigate possible confounding effects. The preliminary results of this study will be discussed at the conference.

38 'Normal' hearing thresholds and figure-ground perception explain significant variability in speech-in-noise performance

E. Holmes
University College London, United Kingdom

T. Griffiths
Newcastle University, United Kingdom

Speech-in-noise (SIN) perception is a critical everyday task that varies widely across individuals and cannot be explained fully by the pure-tone audiogram. One factor that likely contributes to difficulty understanding SIN is the ability to separate speech from simultaneously-occurring background sounds, which is likely not well assessed by audiometric thresholds. A basic task that assesses the ability to separate target and background sounds is auditory figure-ground perception. Here, we examined how much common variance links speech-in-noise perception to figure-ground perception, and how this relationship depends on the properties of the figure to be detected.

We recruited 96 participants with normal hearing (6-frequency average pure-tone thresholds < 20 dB HL). We presented sentences from the Oldenburg matrix corpus (e.g., "Alan has two old sofas") simultaneously with multi-talker babble noise. We adapted the target-to-masker ratio (TMR) to determine the participant's threshold for reporting 50% of sentences correctly. Our figure-ground stimuli were based on Teki et al. (2013; PMID 23898398) in which each 50 ms time window contains random frequency elements. Figure frequencies either remained fixed or changed over time, mimicking the formants of speech. Participants had to discriminate gaps that occurred in the "figure" or "background" components—a task that cannot be performed based on global stimulus characteristics. We adapted the TMR to determine the participant's 50% threshold for discriminating gaps in the figure-ground stimuli.

Average audiometric thresholds at 4-8 kHz accounted for 15% of the variance in SIN performance, despite recruiting participants with hearing thresholds that would be considered clinically 'normal'. Figure-ground performance explained a significant portion of the variance in SIN performance that was unaccounted for by variability in audiometric thresholds. Performance with different figure-ground stimuli explained different portions of the variance, demonstrating they index different reasons why people find SIN difficult.

These results in normally-hearing listeners demonstrate that SIN performance depends on sub-clinical variability in audiometric thresholds. In addition, the results show that we can better predict SIN performance by including measures of figure-ground perception alongside audiometric thresholds. Importantly, the results support a source of variance in speech-in noise perception related to figure-ground perception that is unrelated to audiometric thresholds. Given previous work demonstrates cortical contributions

to both speech-in-noise and figure-ground perception, this shared variance likely arises at a central level. Overall, these results highlight the importance of considering both central and peripheral factors if we are to successfully predict speech intelligibility when background noise is present.

39 Towards improving the speech recognition of cochlear implant users by identification and deactivation of less informative channels

L. Zamaninezhad
Medical Physics and Cluster of Excellence "Hearing4all", Carl-von-Ossietzky Universität Oldenburg, Germany.

V. Hohmann
Excellence "Hearing4all", Carl-von-Ossietzky Universität Oldenburg, Germany.

T. Jürgens
Institute of Acoustics, University of Applied Sciences Lübeck, Germany and Medical Physics and Cluster of Excellence "Hearing4all", Carl-von-Ossietzky Universität Oldenburg, Germany.

Cochlear implants (CIs) have improved the speech perception of patients suffering from severe to profound hearing loss significantly. However, CI users' speech perception is still not satisfactory when it comes to challenging acoustic scenarios, e.g., in presence of noise. Poor spectral resolution due to channel interaction is hypothesized to be one of the reasons behind CI users' degraded speech-in-noise perception. Therefore, mitigating channel interaction by deactivating channels with poor speech information could be beneficial for CI users. However, it is not straight forward to identify these channels. Using model simulations, this study investigates, whether speech perception in CI users may be improved by identifying and deactivating channels that deteriorate speech recognition.

Speech in general is a redundant signal: The speech information in different frequency channels is correlated. In addition to the natural across-channel correlation of the speech signal, in electric listening, the spread of electric field highly influences the channel cross correlation. Since the spread of electric field is individual, it is hypothesized that an individual procedure is required to identify independent channels that carry relevant speech information. To assess this hypothesis, a computer model of speech intelligibility for CI users was employed. The model simulates the spread of electric field in the cochlea and outputs an internal representation (IR), i.e., the post-processed spiking pattern in different auditory channels as a function of time. Individual spread of electric fields were simulated according to the clinical data of 14 CI users. To identify independent channels, the across-channel amplitude modulation correlation (AMCor) matrices based on IRs were obtained. AMCor matrices have been successfully applied

in acoustic hearing to guide speech separation (Anemüller and Kollmeier, 2000). The results showed individual patterns across AMCor matrices that vary substantially with the individual pattern of field spread. This confirms the hypothesis that a selection of a subset of channels for maximizing speech information while minimizing the influence of field spread would lead to a highly individual channel selection, and that this selection may be guided by the AMCor patterns. Whether this channel selection strategy would lead to an improvement in speech reception in individual CI patients remains to be shown.

References:

Anemüller, J., Kollmeier, B., "Amplitude Modulation Decorrelation for Convolutive Blind Source Separation". Proceedings of the second international workshop on independent component analysis and blind signal separation, June 19-22, 2000, Helsinki, Finland, pp. 215-220.

40 Hearing aid use, audiovisual integration, and speech performance: probing the interplay

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

Perception is based on information arriving from different sensory modalities, such as audition and vision, and our brain has the capability to combine these different inputs to form a coherent percept. This process of multisensory integration may lead to an enhanced processing of stimuli and thus, prove beneficial for the interaction with the environment. A plethora of multisensory processes in general and audiovisual phenomena in particular have been studied at both the behavioral and neural level in different populations. For instance, studies have repeatedly shown altered multisensory processing in the elderly. This has been extended to cochlear implant (CI) patients, with a number of studies pointing towards enhanced audiovisual integration abilities of CI patients compared to normal-hearing individuals. In contrast, the effects of hearing aid usage on audiovisual processing in milder hearing impairment have been largely neglected.

We aimed at filling this gap by a recent study, where we found differences in audiovisual integration capacities (measured with the sound-induced flash illusion, SIFI) as a function of hearing aid experience in elderly individuals with mild hearing impairment (PTA 26-40 dB HL). Specifically, hearing aid users displayed greater audiovisual integration than their age-matched peers with the same hearing impairment who were not using hearing aids. A parallel study showed increased McGurk illusions in elderly, mild-to-moderately hearing-impaired listeners compared to age-matched normal-hearing individuals.

A recurring question is how those results may link to audiovisual speech understanding. Typical speech performance tests, as used to evaluate hearing aid benefits for instance, assess outcomes in auditory-only conditions. To address this issue, we focus here on the interplay between audiovisual integration and potential effects of hearing aids. We investigate mild-to-moderately hearing-impaired elderly individuals without hearing aid experience in an interventional, longitudinal 6-month study, whereby a treatment group fitted with hearing aids is compared to a waiting control group. Investigating their audiovisual integration (SIFI, McGurk), cognitive performance, as well as potential changes in brain structure using MRI, and relating those to audiovisual speech performance, we hope to help clarify how hearing aid experience affects audiovisual integration, possible cross-modal re-organization in the brain, and potentially associated benefits in audiovisual speech perception.

First results will be presented at the conference.

41 Relating speech perception in noise to temporal-processing auditory capacities

L. Cabrera
CNRS – Université Paris Descartes, France

Temporal cues (e.g., amplitude modulation, AM) play a crucial role in speech intelligibility for adults. This study explored whether speech-in-noise (SIN) intelligibility relates more to sensory or cognitive factors involved in AM processing. AM masking and temporal integration were measured using non-speech sounds to better characterize sensory (AM encoding) and cognitive (memory and decision) mechanisms.

Twenty-two adults with normal-hearing completed three 3I-3AFC adaptive tasks. The first task assessed AM sensitivity using pure tone carriers and three modulation rates (4, 8, 32 Hz). The second task assessed AM masking by comparing AM detection thresholds at the same three modulation rates using three carriers varying in their inherent AM fluctuations: tones, narrowband noises with small inherent AM fluctuations and noises with larger fluctuations. The third task assessed temporal integration, the effect of increasing the number of AM cycles (between 2 and 8 cycles) on AM detection using tones modulated at 4 or 32 Hz. Finally, a fourth XAB task was designed to measure identification thresholds in speech-shaped noise using fricative and stop consonants. Four phonetic contrasts were tested here: changes in voicing or changes in place of articulation for either fricative or stop consonants.

Results showed that AM detection thresholds were affected by AM rate, large carrier fluctuations and number of AM cycles. Regarding SIN, thresholds were significantly better for changes in place of articulation for fricatives compared to stops and also compared to changes in voicing for fricatives. Preliminary regression analyses suggested that for

stop consonants (either changes in voicing or place), better AM detection thresholds obtained in the different 3-AFC tasks contributed to explain up to 58% of the variance in the SIN thresholds. For fricative consonants, AM detection thresholds did not contribute to explain the variance in the SIN thresholds.

These results suggest that SIN perception is related to some extent to temporal processing, and better AM processing might contribute to better SIN perception for adults. Computational modelling will help to better disentangle the relationship between sensory and non-sensory processing of AM in SIN.

42 Hard to Say, Hard to See? Speech-in-noise discrimination at different levels of sensorimotor proficiency

I. Lorenzini
Laboratoire Psychologie de la Perception, Université Paris Descartes, France; Scuola Normale Superiore di Pisa, Italy

A. M. Chilosi
IRCCS Stella Maris, Pisa, Italy

Sensorimotor processing strengthens speech-in-noise perception (e.g. Skipper, Devlin & Lametti, 2017). Yet, sensorimotor knowledge can only give its contribution as long as it is accumulated through experience with speech production. Relevantly, full speech motor control is only reached at the end of childhood (e.g. Cheng et al., 2007). Thus, during typical ontogenesis, it should be possible to detect the effects of the gradual emergence of sensorimotor knowledge on speech-in-noise perception. Moreover, in the case of developmental speech production disorders, the more severe consequences of the lack of such information should be observable.

This research tested such hypotheses. The participants (98) were as follows: 32 adults; 25 typically developing preschoolers; 23 typically developing school-aged children; 18 children with Childhood Apraxia of Speech.

The perception experiment assessed the in-noise discrimination (Gaussian white noise, 0dB) of non-sense CV syllables varying along the degree of motor proficiency required by the production of the C: one half of the trials contrasted 'easy-to-produce' consonants (plosives and nasals), the other half 'difficult-to-produce' consonants (affricates).

The production assessment measured speech sensorimotor skills by means of the Maximum Performance Rate Task. This was run within an auditory masking paradigm (white noise, 70 dB) preventing the participants to perceive their own auditory feedback (thus forcing them to rely on proprioception of the speech movements).

The tasks were analyzed in Mixed-Effect Models; a correlational analysis was run on the two (Kendall's Tau). As to perception, the adults obtained high scores with all of the trial-types. Conversely, the typically developing children displayed an advantage for the discrimination of the 'easy-to-produce' Cs, close to significance in the school-aged group ($z=1.8$; $p=0.07$) and significant in the preschoolers ($z=2$; $p= 0.02$). Speech-in-noise discrimination was uniformly below chance levels in the Apraxia group.

Furthermore, specific correlational patterns emerged. In the adults, production skills only correlated with discrimination accuracy for the 'difficult' consonants ($z=3$; $p=0.002$) while, in both groups of typically developing children, production skills only correlated with the ability to discriminate the 'easy' consonants ($z=2$; $p=0.02$ and $p=0.006$). In the Apraxia group, the ability to discriminate the 'easy' consonants only correlated with the duration of the therapy targeting speech production ($z=4$; $p<0.0001$).

Such results support the hypothesis of a gradual emergence of the sensorimotor contribution during typical ontogenesis and argue for the perceptual relevance of the lack of sensorimotor information in Apraxia. This was the first study to analyze speech-in-noise perception in this disorder.

43 Speaker and speech dependence in a deep neural networks speech separation algorithm

L. Bramsløw, C. Grant
Eriksholm Research Centre (Oticon A/S), Denmark

G. Naithani, T. Virtanen
Laboratory of Signal Processing, Tampere University of Technology

N. Pontoppidan
Eriksholm Research Centre (Oticon A/S), Denmark

Hearing aid users are challenged in listening situations with noise and especially speech-on-speech situations with two or more competing voices. Specifically, the task of segregating two competing voices is very hard, unlike for normal-hearing listeners.

Recently, deep neural network (DNN) algorithms have shown great potential in tasks like blind source separation of a single-channel (monaural) mixture of multiple voices. The idea is to train the algorithm on relatively short samples of clean speech, thus learning the characteristics of each voice. Once trained for those specific voices, the network can then be applied to mixtures of new speech samples from the same voices.

The current implementation of the DNN has shown a benefit for hearing impaired listeners (Bramsløw et al., 2018) using this voice-specific training on the Danish HINT sentence material (Nielsen and Dau, 2011), but the network may also provide a benefit when applied to new voices.

New speech material has been recorded, both HINT and continuous speech, using three new male and three new female voices. The present study investigated the effect of changing targets and maskers in voice-specific DNN's using objective metrics as predictors of speech separation performance. Furthermore, the effect of training on sentence material and testing on continuous material and vice versa, was evaluated.

44 The influence of motives on the perceived communication success – A qualitative investigation

R. L. Fischer
Sivantos GmbH, Germany

K. Wiedenbrüg
Justus-Liebig-Universität, Germany

R. Hannemann
Sivantos GmbH, Germany

Hearing aid fitting usually targets on speech intelligibility improvement. However, in real life not only speech understanding, but communication matters. According to literature there are different purposes to communication. Delia and Clark (1979) assume three goals of social interaction. Firstly, people communicate because of the problem-solving aspect of communication. If they want to reach a goal, which is not achievable by their own, they have an instrumental goal for communication. Second, there are relational goals, where people communicate to establish and maintain relationships. The last objectives are those to evoke, maintain or pattern a specific impression (identity goals).

The aim of the present work was to gather first insights whether – besides acoustical demands – communication motives alternate the expectation of hearing impaired regarding their hearing aids. Under the assumption that the current hearing aid will provide the same processing in an acoustically similar situation, we aimed to analyse whether the motives cause differences in the perceived communication success. Consequently, differences in the signal processing depending on the active communication motive might be indicated.

For this purpose, a semi-structured interview with experienced hearing aid users (N=7) was conducted. The investigator asked them to imagine situations that were narrated by her. The stories were varied to suggested certain acoustical demands for the imagined situation (e.g. quiet or background noise) and to consecutively activate the three communication motives in each of them (instrumental, relational, identity). For every combination of the acoustical situation and the activated motive, the participants had

to rate on a 7-point Likert-scale to which extent a) it is possible to exchange information, b) communication is exhausting, and c) the surrounding is burdening. Those subscales are facets of the perceived communication success and mirrored therefore the underlying construct.

Data analyses revealed, as expected, an influence of acoustical demands as background noise and number of communication partners on the perceived communication success. Beyond this, a trend significant effect for one communication motive was unfold: The participants perceived the surrounding as more burdening when they imagined a situation, where they were aiming to evoke and maintain a certain personal impression on their communication partners (identity goal) towards a situation where they were seeking for help by them (instrumental goal). These results indicate that communication motives alternate the perception of an acoustically similar situation and might be a source for additional insights into different perceptual needs.

45 Machine learning for Audio Scene Analysis

D. Greenberg, K. Al-Naimi, G. White
EAVE, London, United Kingdom

A user centric circum-aural headset with an internal and an external microphone that enhances user experience in very harsh environmental noise. The system continually measures the dB SPL noise level at the user's ear (dosimetry), monitors the quality of the seal at regular time intervals, and, enhances speech in noise by selecting the optimum pre-set (default or user) for best user experience using machine learning for Audio Scene Analysis (ASA). The headset uses the classification of the user's environment to control filter shapes, gain and enhancements for that specific scene.

The dosimetry data is stored locally on each headset and on a server on the IoT cloud. The user is notified when the daily noise exposure dosage is reached. The supervisor is also able to monitor workers for the noise exposure and issue a change of activities before any permanent hearing damage occurs.

A good seal around the ear is required for hearing protectors to attenuate sound. Therefore, an initial test of the seal at power on, and periodic checks would monitor and if necessary, alert the user of a bad seal; such that the user would be able to rectify the poor seal.

46 Effects of hearing-aid amplification on consonant audibility and forward masking

B. Kowalewski, J. Zaar, M. Fereczkowski, E. MacDonald
Hearing Systems, Dept. Electrical Engineering, Technical University of Denmark, Denmark

O. Strelcyk
Sonova US Corporate Services, United States

T. May, T. Dau
Hearing Systems, Dept. Electrical Engineering, Technical University of Denmark, Denmark

Dynamic range compression (DRC) is a widely-used hearing-aid compensation strategy. The speed of gain reduction and recovery in a compressor are dictated, respectively, by its attack and release time constants. It has been hypothesized that fast-acting compression, characterized by release times shorter than 200 ms, can provide superior speech audibility and improve the rate of recovery from forward masking in hearing-impaired (HI) listeners. On the other hand, it has been reported that fast-acting compression can lead to distortions of the temporal envelope of the stimuli and degrade speech recognition.

Here, the effects of DRC on HI listeners' consonant identification in quiet and in interrupted noise were investigated. Several input levels of speech and two compression conditions were considered that differed only in terms of the release time: fast-acting (10 ms release time) and slow-acting (500 ms release time).

A benefit of fast-acting compression was observed at the lowest speech input level in quiet and at medium speech levels in noise. No detrimental effects of fast-acting compression on recognition were found at any of the tested speech levels.

Additionally, the two compensation strategies were evaluated in terms of objective measures such as the output gain, envelope distortion index (EDI) and a metric of consonant audibility. The average amount of temporal envelope distortion was found to be minimal, consistent with the results of the perceptual evaluation. Consonant audibility was found to account for a large part of the variance in the individual performance scores. However, the listeners seemed to differ in how efficiently they use the audible information to correctly identify the consonants.

The results provide more evidence for beneficial effects of fast-acting DRC, at least in a limited class of acoustic scenarios.

47 Neural processing of speech in children with sensorineural hearing loss

A. Calcus, S. Rosen, L. Halliday
University College London

Previous research has shown that even a mild (21-40 dB HL) or moderate (41-70 dB HL) sensorineural hearing loss (MMHL) can impair cortical processing of speech sounds, as evidenced by differences in event-related potential responses (P1-N1-P2-N2 and MMN) between children with MMHL and chronological age-matched normally hearing (NH) controls (Koravand et al., 2013). However, to date no studies have examined speech processing at the subcortical level in children with MMHL. Moreover, the effects of amplification on the neural encoding of speech are still poorly understood, with previous data suggesting a significant benefit at the subcortical (Anderson et al., 2013) but not the cortical level (Billings et al., 2007).

The aims of this project were to (i) investigate the cortical and subcortical processing of speech sounds in children with MMHL and (ii) evaluate the effects of amplification on the neural processing of speech in this group. To do so, cortical and subcortical EEG activity evoked by speech stimuli (/ba/-/da/) were simultaneously recorded in 18, 8- to 16-year-old children with MMHL and 16 age-matched NH controls. Subcortical processing was assessed using the frequency following response (FFR), an EEG component evoked at the subcortical level which reflects the encoding of the fundamental frequency (F0) and first few harmonics of complex auditory signals such as speech. For the MMHL group, stimuli were presented both unamplified (70 dB SPL), and with a frequency specific gain (without compression) based on their individual audiograms.

Results revealed that children with MMHL had smaller cortical responses than NH controls in both unamplified and amplified conditions, and did not show an MMN. In contrast, at the subcortical level, they showed a smaller FFR than NH controls in the unamplified condition only. With simulated amplification, children with MMHL demonstrated an FFR that was comparable to that observed in NH controls. Our findings suggest that the neural processing of unamplified speech may be impaired at both subcortical and cortical levels in children with MMHL. However, consistent with previous studies in adults, amplification appears to benefit auditory processing at subcortical but not cortical levels in children with MMHL. This might be explained by increasing multi-sensory integration at successive levels of the auditory system: whereas the inferior colliculus processes unimodal information, the auditory cortex processes multimodal information. Alternatively, this could reflect the later maturation of the auditory cortex compared to the inferior colliculus. MMHL may have a bigger impact upon cortical than subcortical processing.

48 Speech intelligibility with symmetrically-placed interferers for German- and Mandarin- speaking listeners in anechoic and reverberant conditions

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

Although the semantic information is also expressed by pitch contour in tonal languages, it is not clear whether tonal language cochlear implant (CI) listeners perform worse than western CI listeners if the CI coding strategy only delivers envelope information. To investigate possible language-specific effects in speech intelligibility (SI) and spatial release from masking (SRM), referring to the effect that listeners benefit when the target speaker is spatially separated from interfering sources in comparison to co-located target and interferers, speech reception thresholds (SRTs) were obtained. Co-located and symmetrically placed maskers (±60°) were used in connection with the German matrix test (OLSA) and the newly developed Mandarin Chinese matrix test (CMNmatrix, Hu et al., 2018, IJA), for both vocoded and non-vocoded signals in two groups: native German-speaking (GS) listeners and Mandarin-Chinese-speaking (CMNS) listeners. SRTs were tested using headphone presentation in a sound attenuated booth with either a stationary masker or a fluctuating nonsense speech masker. The co-located and separated conditions in three different rooms (anechoic, 0.6 s, and 3 s reverberation time) were simulated by using virtual acoustics and headphone auralization. For the anechoic room, an artificial infinite interaural level difference (ILDinf) condition, where the acoustic crosstalk was removed, was additionally tested.

For the anechoic room, the results were comparable to those reported in (Hu et al., JASA, 2018): both groups greatly benefited from spatial separation and ILDinf for the non-vocoded signals, while no binaural benefit was observed in noise vocoder simulated BiCI listeners for the ±60° spatial configuration. CMNS listeners showed a slightly reduced binaural benefit in both ±60° and ILDinf conditions relative to the co-located condition when compared to GS listeners. As expected, speech intelligibility decreased in reverberant conditions compared to the anechoic condition. Increasing reverberation time showed less effect on the SRM in CMNS listeners than in GS listeners. The potential role of pitch contour perception and the role for CI processing are discussed.

49 Class-like speech audiometry for the clinical evaluation of school aged CI users

S. Krijger
Department of Otorhinolaryngology, Ghent University, Belgium

I. Dhooge
Department of Otorhinolaryngology, Ghent University and Ghent University Hospital, Belgium

M. Coene
Language and Hearing Center Amsterdam, Free University Amsterdam, The Netherlands

P. Govaerts
The Eargroup, Deurne-Antwerp, Belgium

Increasingly more children with cochlear implants (CI's) are being educated in mainstream schools. In Belgium, already 45–74% of the deaf students with a CI are enrolled in regular elementary and secondary schools. Speech perception in secondary schools is particularly difficult for CI users due to poor room acoustics, high levels of background noise, complex language and the fact that different courses are taught by different teachers. Despite these challenges affecting CI users in school, clinical evaluation of speech perception is currently administered with easy speech materials in a sound treated room and will therefore not accurately predict the speech performance in class. For this reason a class-like speech audiometry test was developed.

23 children with cochlear implants (mean age =13,7; SD = 1,8) and a control group of 28 normal hearing children participated in this study. A test situation was created in a reverberant room with 5 signal speakers and 6 noise speakers. The signal speakers were positioned as detailed in Valente (2012) to mimick the presence of students and teacher in class (1 meter from the listener, 5 different angels). The noise speakers produced a diffuse multitalker noise field of 65 dBSPL. In a randomized procedure the Linguistically controlled Sentences (LiCoS) (Coene et al., 2016) were administered from the 5 signal speakers with an adaptive procedure to obtain the 50% SRT of each speaker. An additional trial was performed in which speech was presented from all 5 speakers randomly.

Mean SRT's were calculated for each signal speaker (L1-L5) and for the randomized trial using all speakers (Lrandom). Children with CI scored significantly worse in all conditions compared to their normal hearing peers. A pairwise within subject comparison between the six conditions (one-way repeated measures ANOVA) showed that two conditions were perceived more difficult by the control group ($F(2.37, 63.96) = 16.12$, $p<.05$, Bonferroni corrected): L1 (Speaker in front, SRT+3,6 dB) and Lrandom (SRT+5,6 dB). This effect was not found in the CI group. A class-like speech audiometry model was created that accurately mimicked the listening difficulties occurring in typical classrooms. This ecological model can be used in addition to current existing clinical evaluation of speech perception. Further research is necessary to get more insight into the directivity effects and binaural benefits found in this model.

50 Mechanisms of spectro-temporal modulation detection and discrimination in normal-hearing and hearing-impaired listeners

E. Ponsot
Laboratoire des systèmes perceptifs, ENS, Paris, France

L. Varnet
Speech Hearing and Phonetic Sciences, UCL, UK

S. A. Shamma, N. Wallaert, P. Neri
Laboratoire des systèmes perceptifs, ENS, Paris, France

Speech in noise understanding relies on the ability of our auditory system to extract relevant spectro-temporal modulations from noise: sensitivity of hearing-impaired listeners for detecting elementary spectro-temporal modulations is found to predict their speech-in-noise performances. However, we do not have a full computational understanding of this connection. Here, we used a two-fold approach combining psychophysics and modeling to probe the mechanisms underlying spectro-temporal modulation processing in both normal-hearing and hearing-impaired listeners. We ran several psychophysical experiments using a newly developed methodological framework based on reverse correlation deployed in the spectro-temporal modulation space and used system identification tools as well as current auditory models to determine the potential architecture of this processing and its underlying components. Both normal-hearing and hearing-impaired listeners were asked to detect or discriminate (upward vs downward) elementary spectro-temporal modulations called 'ripples' that were embedded in 'ripple noise', i.e. noise made of ripples of other spectro-temporal modulations with random energy. First, our results indicate that listeners rely on finely tuned but not fully directional band-pass filters to extract the target modulations. Second, they show that hearing-impaired listeners with similar audiometric loss exhibit a large variety of computational strategies. In order to further understand this variability, we used modeling tools to determine the impact of the different stages as well as their combined contribution in the processing. Overall, this interdisciplinary approach paves the way toward a computational characterization of human spectro-temporal modulation processing and should therefore contribute to a better understanding of supra-threshold auditory mechanisms and their deficits in general.

51 Characterizing early markers of degraded speech encoding: temporal fine structure and envelope cues

T. Wartenberg, S. Verhulst
Hearing Technology @ WAVES, Dept of Information Technology, Ghent University

Speech contains rapidly varying temporal fine-structure (TFS) information as well as slower temporal envelopes (TENV). Both features are linked to speech intelligibility, although the presence of TFS is thought to boost speech reception in adverse listening conditions. Since the respective role of cochlear synaptopathy and outer-hair-cell (OHC) deficits to degraded speech reception is unclear, it is important to clarify which ENV and TFS information is available at the level of the auditory midbrain in the normal and impaired auditory system. Starting from the earliest neuronal correlate of an speech, we aim to understand how TFS and TENV cues are represented in a computational model of the (impaired) human auditory periphery (Verhulst, Altoè, & Vasilkov, 2018). The outcome of this study can guide the development of hearing restoration strategies tailored to listeners with synaptopathy and/or OHC loss.

Speech was decomposed into TFS or TENV chimaera and stimuli were either band-pass, high-pass or low-pass filtered to target different aspects of cochlear sound encoding. An additional TFS stimulus, based on the zero crossings of the fundamental waveform and the first harmonic, was added to prevent a reconstruction of TENV cues in the signal decomposition. Models with different combinations of synaptopathy and outer-hair-cell deficits were considered. To study which aspects of the stimuli are preserved after (impaired) cochlear processing, cross-correlations were performed between the stimuli and stimulated population auditory-nerve (AN) responses. Additionally, the relationship between simulated AN responses to the unmodified and (filtered) modified stimuli was investigated to clarify the respective role of TFS/TENV and cochlear frequency regions to speech representation in the (impaired) auditory system.

52 Assessing and reducing listening effort of listening to speech in adverse conditions

A. J. Hall, J. Rennies-Hochmuth, A. H. Winneke
Branch Hearing, Speech and Audiotechnology, Fraunhofer IDMT, Oldenburg, Germany

Normally hearing listeners are skilful speech-perceivers, even in challenging listening environments. However, the neural processes required to compensate for missing or masked speech information can lead to tiredness and fatigue, even in everyday environments such as call-centres or shopping malls. This study utilises electroencephalography (EEG) to measure this cognitive compensation – termed listening effort (LE) – and its neurophysiological correlates, by comparing unprocessed speech to speech enhanced with AdaptDRC.

AdaptDRC is a near-end-listening enhancement (NELE) algorithm that alters speech signals for playback, dependent on environmental noise, and significantly improves speech intelligibility (Schepker et al., 2013). AdaptDRC reduces the subjectively rated effort for listening to speech in noise, even at 100% intelligibility (Rennies et al., in press) but we do not yet have corresponding neurophysiological objective data.

In this study, we recorded EEG while participants performed a LE task (N=30; normal hearing adults). Participants listened to unprocessed or AdaptDRC-processed OLSA sentences (Wagner et al., 1999) in cafeteria or speech-shaped noise at five signal-to-noise ratios (-10, -5, 0, +5, +10dB), then rated the effort required to understand the speech using a modified adaptive categorical LE scale (ACALES; Krueger et al., 2017). We also measured speech intelligibility: at intermittent trials, participants were prompted to repeat the sentence aloud, to ensure that participants were actively listening to the speech. Further, we measured participants' hearing (pure tone audiometry) and cognitive abilities (working memory, selective attention, inhibition), to explore the relationship between individual differences in these abilities and subjective and objective LE.

Preliminary data analyses (N=12) indicate that, independent of noise type, speech intelligibility is at ceiling (98%) at SNRs of 0, +5 and +10 dB. At these SNRs, subjective LE decreases with increasing SNR (r=-.51) and AdaptDRC speech was rated lower than unprocessed speech. EEG analysis will identify neurophysiological markers of compensatory LE changes, and their relationships with subjective ratings. We focus on spectral power within the alpha frequency band (8-12 Hz), as work has shown a relationship between alpha spectral density and the suppression of task irrelevant information.

Thus, this experiment provides insight into the neurocognitive correlates of LE, the compensatory processes required for successful speech perception in sub-optimal conditions, and the benefits of speech enhancement technologies. The continued development and implementation of NELE technology in public and workplace environments will aid speech perception in suboptimal conditions and may also improve listener experience at low levels of noise.

53 Comparing approaches towards robust voice activity detection in noise

F. Fuhrmann, C. Leitner, F. Graf
Joanneum Research Digital, Austria

We examine several approaches towards robust Voice Activity Detection (VAD) in low Signal-to-Noise Ratios (SNRs) automatic speech recognition (ASR) applications. The aim is to derive an optimal solution for the VAD component to be applied in real-world speech interaction systems (e.g., human-machine-interaction, home automation, etc.). Generally speaking, the VAD component of a speech interaction system decides which segments of the incoming audio stream are forwarded to the ASR component for analysis.

Here, we compare several VAD methods for the application in various low-SNR scenarios. We evaluate three distinct VAD methods, examine their combination, and analyze the influence of context – prior speaker information as well as changing noise characteristics – on the methods. More precisely, the first method is based on an energy threshold in the frequency domain (subSNR), the second method evaluates a model of spectral peak frequencies extracted from isolated vowel segments (peakSig), and the third uses a supervised machine learning approach applying a Support Vector Machine (svmVAD).

The used speech data consist of clean speech spoken by 52 subjects. In total, each subject contributes the same 64 utterances, for additive noise data we used 8 different industrial noise recordings. For each speaker, we reserve 1/4 of the utterances for testing and build individual training data – depending on the contextual information – with the remaining data of all speakers. With 8 noise types, 4 SNRs (-5, 0, 5, 10 dB), and two different sets of training utterances for each speaker we hence created 64 datasets. We then averaged the resulting evaluation metrics (F1 for VAD performance and WER for ASR performance) over all speakers to yield an estimate for the performance of the system under test.

Results show that svmVAD achieves best performance, over all SNRs and noise types, for both metrics. SubSNR reports worst performance, especialy for very low SNRs. Moreover, contextual information (i.e., updating the noise prints over time) is most beneficial for subSNR, since it directly incorporates the noise properties. Next, prior speaker information (i.e., training and test data from the same speaker) does not improve performance figures; hence the properties used by the peakSig and svmVAD seem to be speaker-independent. Finally, combining subSNR and peakSig improves performance; nevertheless svmVAD performance could not be reached. In conclusion, we observed that svmVAD outperforms the signal processing methods in all regards, suggesting an application of this method in real-world speech interaction systems.

54 Auditory and non-auditory factors contributing to the benefit of amplification

M. A. S. Tahden, A. Gieseler, C. Thiel
Cluster of Excellence "Hearing4all" and Department of Psychology, University of Oldenburg, Germany

K. C. Wagener
Cluster of Excellence "Hearing4all", University of Oldenburg, Hörzentrum Oldenburg GmbH, and HörTech gGmbH, Germany

T. Brand
Cluster of Excellence "Hearing4all" and Department of Medical Physics and Acoustics, University of Oldenburg, Germany

H. Colonius
Cluster of Excellence "Hearing4all" and Department of Psychology, University of Oldenburg, Germany

Among individuals with hearing impairment there exist large differences in the benefit of amplification with hearing aids in understanding speech in noise. For rehabilitation success, explaining this variability is fundamental. To address the issue, we investigated n=92 elderly individuals with a mild-to-moderate hearing loss who completed a test battery comprising auditory and cognitive tests, a test measuring the capacity of audiovisual integration, and a questionnaire. The speech intelligibility based benefit of amplification was defined as the difference between unaided and aided 80%-SRT measurements (Goettingen sentence test [1], ICRA5-250 noise [2], amplification applied via the master hearing aid [3]). While n=55 participants were already aided with hearing devices before entering the study (hearing aid users, HA-U), the remaining n=37 individuals were completely inexperienced in using hearing aids (hearing aid non-users, HA-NU).

Preliminary results from statistical learning methods suggest that influential factors differ as a function of the amount of balancing HA-U and HA-NU with respect to the degree of hearing loss: The more the groups are matched regarding hearing performance, the more relevant are cognitive measures as well as audiovisual integration. Beyond controlling for the degree of hearing loss, however, further possible confounder variables should be taken into account, such as self-reported hearing problems, subjective listening effort, and the influence of hearing problems on quality of life. Therefore, we apply propensity score matching [4], a well-known matching method in fields such as epidemiology and medicine, making both hearing aid users and non-users more comparable.

References:

[1] Kollmeier, B., and Wesselkamp, M. (1997). Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. The Journal of the Acoustical Society of America 102, 2412-2421.

[2] Wagener, K. C., Brand T., and Kollmeier B. (2006). The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. International Journal of Audiology 45(1), 26-33.

[3] Grimm, G., Herzke, T., Berg, D., and Hohmann, V. (2006). The master hearing aid: a PC based platform for algorithm development and evaluation. Acta acustica united with Acustica 92, 618-628.

[4] Rosenbaum, P. R., and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

55 Prediction of speech intelligibility with deep neural networks and automatic speech recognition: Influence of training noise on model predictions

J. Roßbach, B. Kollmeier, B. T. Meyer

Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4all, University of Oldenburg, Germany

Accurate models of speech intelligibility (SI) can help to optimize speech enhancement algorithms, and reference-free SI models could potentially also serve as model-in-the-loop for real-time monitoring of SI in listening devices. Such models have to work without a speech or a noise reference. Spille et al. [(2018) Comp. Speech & Lang. doi:10.1016/j.csl.2017.10.004] created a model for predicting the speech reception threshold in noise based on a deep neural network (DNN) and automatic speech recognition (ASR). This model was blind to the speech signals (since it was based on a speaker-independent ASR system), but used the same noise signals for training and testing. This bears the risk of overfitting the model to the specific noise signals.

To investigate if overfitting plays a major role in this context, we modified the training procedure of the original model. Instead of using the same noise signal for training and testing we used the same noise source but different noise signals, which should be especially challenging when the source of the noise is a competing talker. This modification is one step in the direction of creating SI models that do not require a speech or noise reference. To test the DNN-based model, Spille et al. (2018) used eight different noise types that ranged from speech-shaped noise to a single talker. Six of the noises were derived from the international speech test signal (ISTS) that has a length of 60 seconds. To achieve a sufficient amount of noise samples in the current study, we created a new noise signal which resembles the ISTS. On this basis, new noises were generated, each with a length of 11 hours. For the training and testing procedure of the DNN the noise signals were split in two parts. Approximately 80% of each noise were used for training and 20% for testing. The results of our modified model are similar to the results of Spille et al. (2018): The predictions for 50% speech reception threshold are with an RMS error below 2.5 dB close to the results of the original model with an RMS error of 1.9. Both models outperform the baseline models such as the SII with an RMS error of 7.9 dB, the ESII (5.6 dB), the STOI (9.2 dB) and the mr-sEPSM with an RMS error of 3.5 dB.

56 A Theory-based treatment of the potential contribution to anti-masking by inhibition from type-II neurons in the dorsal cochlear nucleus: modeling and simulation

T. C. Liu, Y. W. Liu
National Tsing Hua University, Hsinchu City, Taiwan

The medial olivocochlear (MOC) reflex pathway, which receives excitatory input from the cochlear nucleus and gives inhibitory feedback to the cochlea, has been said to modify the cochlear amplification gain and enhance the audibility of non-stationary tones in noise. This enhancement of audibility is referred to as anti-masking. In the present research, we simulate the potential role of inhibition from dorsal cochlear nucleus (DCN) to ventral cochlear nucleus (VCN) in anti-masking. The tuberculoventral (TUB) cells in the DCN, being categorized as Type II, are known to be insensitive to noise stimuli; therefore, we hypothesize that their frequency-specific inhibition to the T-stellate (TS) cells would reduce the firing rate of MOC interneurons when a tone is present. In contrast, when only broadband noise is present, this TUB inhibition network is not activated so it does not affect the function of the MOC pathway. Based on these assumptions, an integrated computer model is built, comprised of (i) a nonlinear model of cochlear mechanics, (ii) the excitatory projection from the auditory nerve fibers (ANFs) and the inhibitory projection from the D-stellate (DS) cells to both the TS and the TUB cells, (iii) the inhibitive projection from the TUB to the TS cells, and (iv) a heuristic equation describing the dynamics of OHC gain modification. Simulations of the neural networks were conducted by the leaky integrate-and-fire method. When the system is stimulated with a stationary tone in noise, the spatial excitation pattern of ANFs exhibits higher contrast between the characteristic-frequency place and adjacent places when the TUB-TS inhibition is present. The results may suggest two things: first, the perceptibility of low-level tones in noise is enhanced. Secondly, because of the inhibition from TUB to TS, anti-masking is possible even when the tone is stationary.

57 Neural network dynamics of speech comprehension – the role of the angular gyrus

A. U. Rysop
Research Group "Modulation of Language Networks", Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

L. M. Schmitt, J. Obleser
Department of Psychology, Universität zu Lübeck, Lübeck, Germany

G. Hartwigsen
Research Group "Modulation of Language Networks", Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Speech comprehension is often challenged by acoustically adverse listening conditions (e.g. background noise or compromised signal quality). When perceiving speech in noise, successful speech comprehension relies on additional factors, such as semantic context. Previous work has shown that comprehension of sentences is facilitated by rich semantic context (e.g. "The ship sails the sea" vs. "Paul discussed the sea"). This contextual comprehension benefit is found to be largest at intermediate levels of intelligibility and is accompanied by enhanced engagement of left angular gyrus (AG). Furthermore, there is converging evidence for the recruitment of domain-general adaptive control networks when listening conditions become challenging. For instance, it has been shown that the extent to which cingulo-opercular regions are recruited has substantial predictive value for the behavioural outcome. However, it remains unclear how domain-specific speech networks and domain-general control networks interact and influence each other during successful speech comprehension.

To address these questions, we conducted an event-related fMRI experiment. Stimuli were derived from the German version of speech in noise (SPIN) sentences (Kalikow et al.), with the sentence-final word being embedded in either high or low semantic context. Sentences were presented auditorily at six levels of intelligibility tailored to each subject individually using an adaptive staircase procedure. During continuous scanning, participants performed an overt sentence repetition task.

At the behavioural level, we found significantly higher accuracy scores for sentences with high opposed to low semantic context. Univariate fMRI group-analyses revealed broad bilateral activation in superior temporal, as well as pre- and postcentral regions, together with left inferior frontal gyrus for increasing intelligibility. In contrast, regions that showed greater activation for decreasing intelligibility encompass bilateral AG, precuneus, superior and inferior frontal regions. Furthermore, we found significant context x intelligibility interactions in fronto-temporo-parietal regions: left AG, left supramarginal gyrus and posterior portions of the middle and inferior temporal lobes showed a

stronger increase in activation for high predictable sentences, left inferior frontal regions as well as bilateral insulae exhibited a stronger increase for low predictable sentences. These preliminary results provide further evidence for a strong context-dependent recruitment of left AG in degraded speech comprehension.

As a next step, task-specific changes in effective connectivity between left AG and other nodes within the speech-specific network as well as across domain-specific and domain-general networks will be investigated to provide deeper insight into the functional dynamics of speech comprehension under adverse listening conditions at a network level.

58 Listening effort during non-native speech perception in noise

G. Borghini, V. Hazan
Department of Speech Hearing and Phonetic Sciences, University College London, United Kingdom

Current evidence demonstrates that the presence of a background noise is much more detrimental for non-native than for native listeners in terms of performance. Moreover, we know that listening effort is increased for non-native compared to native listeners, even when the intelligibility levels are equated across the two groups.

Here, I discuss results from a pupillometry study exploring listening effort during a speech perception task in noise. The listening task featured lists of 12 semantically related and unrelated sentences. In order to investigate the contribution of semantic context to listening effort, results were analysed by grouping sentences into three parts (beginning, middle and end section). Participants included 21 Italian learners of English, and 18 native English controls. An adaptive procedure was used to match the intelligibility levels across participants. An intelligibility level of 80% was targeted.

Preliminary analyses consistently confirmed that non-native listeners showed a significantly greater pupil response compared to native listeners when attending to speech in noise. Additionally, we only found a significant reduction in the pupil peak dilation when native participants were attending to related compared to unrelated sentences. This benefit from context was not replicated for the non-native listeners. Results suggest that non-native listeners do not benefit in terms of reduction in the listening effort as native listener do when they are provided with a consistent semantic context.

59 Static and dynamic multitalker listening – the contribution of different types of attention

H. Meister
Jean Uhrmacher Institute, University of Cologne, Cologne, Germany

Typical of daily listening are situations in which several talkers speak simultaneously. These situations can be “static” when the aim is selectively attending to one talker, or “dynamic” when the talker of interest changes in a potentially unpredictable way.

In these situations, different types of attention play a role. On the one hand, it may be important to selectively focus on one talker and ignore the competing background voices. On the other hand, temporarily dividing and switching attention may also be necessary.

The aim of this project is to investigate attention mechanisms in static and dynamic multitalker listening conditions. Specifically, the results of young and older listeners will be presented and related to the outcome of neuropsychological tests on executive functions and working memory. An attempt is made to disentangle the different attention types and to describe dimensions that might be particularly problematic in older listeners.

Supported by grants of the “Deutsche Forschungsgemeinschaft” (DFG ME 2751/3-1).

60 Development of speech in noise and reverberation test with multichannel auralizations.

A. Kuusinen
Aalto University School of Science, Dept. of Computer Science, Espoo, Finland

V. Sivonen
Helsinki University Hospital, Hearing Centre, Helsinki, Finland

T. Lokki
Aalto University School of Science, Dept. of Computer Science, Espoo, Finland

A. Aarnisalo
Helsinki University Hospital, Hearing Centre, Helsinki, Finland

In recent years, a major step forward in improving hearing diagnostics in Finland has been the development and implementation of a new Finnish sentence-in-noise test. This test has greatly improved the accuracy of hearing diagnostics and quality control of hearing rehabilitation in Finland. However, the test stimuli are presented over headphones or via loudspeakers in a sound booth with little to no reverberation. In contrast, people often report on difficulties specifically in understanding speech in every day

reverberant conditions. Advanced spatial sound technologies may provide means to study speech perception in a variety of acoustical conditions and may help to further understand speech perception in noise and in reverberation considering both normally hearing listeners and listeners with various degree of hearing impairments. The aim of this project is to investigate how advanced spatial sound technologies could be used in hearing diagnostics and to develop novel tools with more complex and authentic sound scenes. In hearing diagnostics and rehabilitation, faithful representation of the spectral characteristics of complex sound scenes is of paramount importance.

A recently developed auralization method (Spatial Decomposition Method) with minimal spectral coloration artefacts may be well suited for bringing advanced spatial sound technology into clinically feasible environments. Here, we present preliminary results of our study, where we have used the Finnish sentence-in-noise test with multichannel auralizations of two real spaces (with approx. one and two second reverberation times) in order to compare the speech reception thresholds (SRTs) between anechoic and reverberant conditions. Currently the test has been taken by 29 test subjects, including 13 normal hearing and 16 hearing impaired listeners. The results do not indicate significant differences in SRTs between anechoic and reverberant conditions, but considering the hearing impaired, there seems to be a small improvement in SRT when measured in reverberation time of approx. one second. The tests are on-going and we are currently collecting results of patients with more severe hearing impairments and patients with hearing aids.

61 Intelligent games in audiological rehabilitation: the Pirates game

S. Magits, S. Denys, T. Francart, J. Wouters, A. van Wieringen
KU Leuven, Department of Neurosciences, Research Group Experimental ORL, Leuven, Belgium

Hearing screening using the digit triplet test (DTT) has proven to be an efficient, reliable and fast screening method (Jansen, 2013) with considerable advantages over pure-tone thresholds audiometry. However, testing in young children has been difficult due to their limited attention span. Simply reducing the number of trials will involve loss of precision. Instead, presenting the DTT as a serious game, which taps into the child's fantasy, will be more engaging. Children will be more motivated, have a higher attention span, and therefore a more reliable score can be obtained.

We have developed the Pirates DTT Game where children are encouraged to open treasure chests by entering a three-digit code. Currently, the Pirates DTT Game is validated in normal-hearing young children (first grade – 6y). Therefore, we compare outcomes on the standard DTT procedure with performance on the game-based Pirates DTT. Speech reception thresholds, test stability and test reliability are compared for the two screening methods. Preliminary results show that the standard and game-based screening

provide similar results in adults, thereby validating the game-based procedure. In children, the game-based screening enhances their sustained attention and motivation, thereby providing more reliable outcomes. These results show that intelligent games are promising tools when adapting validated SPIN tests to the interest and attention span of young children.

This work was supported by a TBM-FWO grant from the Research Foundation-Flanders (grant number T002216N).

62 A new measure to predict the a priori performance of automatic transcription systems on reverberated speech

S. Ferreira
IRIT-UPS, Authôt, France

J. Pinquier, J. Farinas, J. Mauclair
IRIT-UPS, France

R. Stéphane
Authôt, France

Advances in the field of Automatic Speech Recognition (ASR) make it possible to use this technology in less and less controlled environments. However, reverberation remains a challenge. The purpose of this study is to predict the a priori reverberation impact on the ASR performance. We analysed the statistical behaviour of the vocal excitation of the vowels and created a measure based on this observation. This measure is called Excitation Behaviour (EB). To find the link between the EB and the Word Error Rate (WER) obtained by the transcription system, a regression model was calculated. To evaluate the performance of the regression model we observed the mean prediction error. We also used the same protocol on two other measures: Speech-to-Reverberation Modulation energy Ratio (SRMR) and Spectral Decay Distribution (SDD). The speech corpus used is the Wall Street Journal (WSJ0 and WSJ1) corpus. Speech was artificially reverberated using Room Impulse Response (RIR) from the REVERB challenge corpus. Recorded RIRs allow to simulate 7 different reverberation conditions: 3 rooms with different sizes (Small, Medium and Large) with 2 types of distances between a speaker and a microphone array (near=50cm and far=200 cm), and the last condition is without reverberation. ASR system was trained with the subset train_si284, the regression model was trained with dev93 and eval92 was used to test the prediction. It shows that the EB obtains an average prediction error of 13.88 while the SRMR obtains 17.87 and the SDD 17.48. Using 20 utterances (approximately 2m20s), the average prediction error decreases to 7.63 for the EB, 13.08 for the SRMR and 13.02 for the SDD. EB measure is better correlated to ASR performance than other reverberation measures.

63 Using automatic speech recognition for the prediction of impaired speech identification

L. Fontan
Archean LABS, Montauban, France

I. Laaridh, J. Farinas, J. Pinquier
IRIT - Université de Toulouse, France

M. Le Coz
Archean LABS, Montauban, France

C. Füllgrabe
School of Sport, Exercise and Health Sciences, Loughborough University, United Kingdom

Age-related hearing loss (ARHL) is a very prevalent hearing disorder in adults that negatively impacts on the ability to understand speech, especially in noisy environments. The most common rehabilitation strategy is to fit hearing aids (HAs). Their benefit is generally assessed by measuring speech-identification performance with and without HAs. However, such so-called "speech audiometry" can be fairly lengthy, and its results are likely to be influenced by the patient's level of fatigue, cognitive state and familiarity with the speech material used for the assessment.

In order to overcome these issues, the feasibility of using objective measures based on automatic speech recognition (ASR) to predict human speech-identification performances was recently investigated (Fontan et al., 2017; Fontan et al., in preparation; Kollmeier et al., 2016).

Here, we present the results of a series of experiments, that combined ASR and an ARHL simulation to predict human performances for various tasks ranging from phoneme discrimination to sentences identification. More specifically, signal processing techniques (Nejime & Moore, 1997) were used to process the speech tokens to mimic some of the perceptual consequences of ARHL on speech perception (i.e., elevated thresholds, reduced frequency selectivity and loudness recruitment), and the processed speech tokens were then fed to an ASR system for analysis. To provide "proof-of-concept", our first experiments focussed on the prediction of unaided speech perception in quiet, while subsequent experiments investigated the applicability of the ASR system to aided and unaided speech perception in noise.

Fontan, L., Cretin-Maitenaz, T., & Füllgrabe, C. (In preparation). Automatic speech recognition predicts speech perception in older hearing-impaired listeners.

Fontan, L., Ferrané, I., Farinas, J., Pinquier, J., Magnen, C., Tardieu, J., Gaillard, P., Aumont, X., & Füllgrabe, C. (2017). Automatic speech recognition predicts speech intelligibility and comprehension for listeners with simulated age-related hearing loss. Journal of Speech, Language, and Hearing Research, 60, 2394-2405.

Kollmeier, B., Schädler, M. R., Warzybok, A., Meyer, B. T., & Brand, T. (2016). Sentence recognition prediction for hearing-impaired listeners in stationary and fluctuation noise with FADE: Empowering the attenuation and distortion concept by Plomp with a quantitative processing model. Trends in Hearing, 20, 233121651665579.

Nejime, Y., & Moore, B. C. J. (1997). Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise. Journal of the Acoustical Society of America, 102, 603-615.

64 Group conversations: How speech, movement, and gaze behaviours of hearing impaired triads change when conversing in noisy environments

L. V. Hadley, W. M. Whitmer, G. Naylor
University of Nottingham, Hearing Sciences, United Kingdom

Most conversations do not occur in perfect quiet. Gossiping in a café, talking in a restaurant, or chatting in a car all require people to ignore the background noise to concentrate on what their partners are saying. While these situations are challenging for people with normal hearing, people with hearing impairment have much more difficulty, and may even shun such environments as they know they will fail to keep up. However, interlocutors can draw on a variety of strategies to aid successful communication. Prior studies of isolated speaking and listening have indicated that strategies such as increasing vocal intensity, adjusting head orientation, or directing gaze to the speaker are beneficial in such situations. However, it is not clear whether these strategies are used in a natural interaction context, which differs by including rapid alternation of speaking and listening, the availability of multimodal communicative cues, and the possibility for mutual adaptation.

We therefore investigate how hearing impaired individuals have conversations in noise and make themselves understood in an ecologically valid context. Specifically, we report the fine-grained dynamics of natural conversation between unfamiliar triads of age- and hearing impairment- matched interlocutors (n = 33), addressing how different levels and types of background noise affect speech, movement, and gaze. We investigate behaviour in both speech-shaped noise and 8-talker babble (54dB-78dB) to identify the importance of the information content of competing noise, and we also compare the behaviour of triads with previously collected data from dyads (n = 30) to identify strategies that generalise across interaction types. We show that many potentially beneficial behaviours are not used optimally, including increases in vocal intensity. We also show that interlocutors prioritise gaze cues over beneficial head orientations regardless of interaction type. Understanding these conversation behaviours could inform broader models of interpersonal communication, as well as being used to develop new communication technologies that take advantage of the behaviours that individuals naturally use.

65 Evaluation of different hearing aid couplings in every day life with Ecological Momentary Assessment

N. Schinkel-Bielefeld, U. Giese
Sivantos GmbH, Erlangen, Germany

One of the biggest challenges for hearing impaired people is understanding speech in noisy environments. What may help them are hearing aids with state-of-the-art noise reduction and directional microphone systems. The benefit of these algorithms is biggest when using a closed coupling where little direct sound mixes with the amplified sound. Also, in comparison to open couplings, closed couplings can provide more gain at high frequencies due to better feedback stability. These two factors both help to improve speech understanding. On the other hand, occlusion effects may occur, and patients may perceive their voice as being too loud and too boomy. Thus, open coupling is often preferred and in fact about two thirds of all fittings of Signia hearing aids are done with an open or semi-open coupling.

Here we investigated how the satisfaction of hearing aids compares for different couplings in everyday life. Subjects were fitted with Signia Pure 13 Nx hearing aids and open domes, closed click sleeves and click molds with a vent of 0.8 mm. Each coupling was tested for one week at home and in randomized order.

We used the method of Ecological Momentary Assessment (EMA). This makes sure we cover all relevant situations in every day life and answers are not distorted by memory biases. Also, this method is very context sensitive allowing us to analyze a number of different situations separately. Subjects were provided with mobile phones which prompted them several times a day to fill out a questionnaire. In addition, they had the possibility to fill out a questionnaire whenever they wanted.

So far eight experienced hearing aid wearers with mean age 72,8 years (std: 5.6 years) participated in the study and filled out in total 1226 questionnaires. Subjects had a mild to moderate hearing loss with a mean pure tone average of 47.5 dB (std: 6.6 dB).

Preliminary EMA results suggest that the majority of subjects reported significantly better speech understanding with closed couplings. This is consistent with the result of the Göttinger sentence test in noise. Nevertheless, satisfaction ratings pooled over all questionnaires indicate that most subjects prefer open couplings to closed ones. However, satisfaction is situation dependent. Open couplings are preferred when listening to music or not actively listening to anything. If the background is quiet and not distracting from or adding anything to the target sound source, closed couplings are preferred.

66 Finding the sweet spot for phoneme in noise training

M. Serman, K. Kallisch, A. Mosebach, U. Giese
Sivantos GmbH, Erlangen, Germany

Reverse hierarchy theory proposes that rapid perception is based on high-level representations of the global, abstract categories of the perceived objects (Ahissar et al. 2009). Thus, in everyday communication and speech-in-noise understanding in particular, there is little time left for dwelling on fine spectro-temporal details of perceived speech. And yet, it is the fine details of speech sounds that are often ambiguous or unavailable to hearing impaired listeners and that would require repeated presentation. Phoneme training gives the listener the opportunity to concentrate on the detailed spectro-temporal representation, with repetition and feedback. However, the floor and ceiling effects of such training procedure diminish both the training benefit and the motivation of the listener. For each subject there is a narrow range of speech to noise ratios which will result in the largest training improvement. We shall call this the individual's training sweet spot. In the study reported here, an automatic procedure for determining listener's training sweet spot was developed and tested. The procedure, training results and the subjective experiences of normal hearing and hearing impaired subjects will be presented.

Ahissar, Merav, et al. "Reverse hierarchies and sensory learning." Philosophical Transactions of the Royal Society of London B: Biological Sciences 364.1515 (2009): 285-299.

67 Performance prediction of the binaural MVDR beamformer with partial noise estimation using a binaural speech intelligibility model

C. F. Hauth
Medizinische Physik and Cluster of Excellence Hearing4All, University of Oldenburg, Germany

N. Gößling, S. Doclo
University of Oldenburg, Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, Oldenburg, Germany

T. Brand
Medizinische Physik and Cluster of Excellence Hearing4All, University of Oldenburg, Germany

An objective evaluation of binaural noise reduction algorithms allows for directly comparing the performance of different algorithm realizations. In this study, a binaural speech intelligibility model (BSIM), which mimics the effective binaural processing of human listeners, is used to predict the performance of the binaural minimum-variance distortionless response beamformer with partial noise estimation (BMVDR-N), which

aims at preserving the speech component in a reference microphone and a scaled version of the noise component. The BMVDR-N beamformer is evaluated with respect to a predicted change in SRT depending on the parameter η, which controls a trade-off between noise reduction and binaural cue preservation of the noise component. The results show that BSIM benefits from the preserved binaural cues suggesting that the BMVDR-N beamformer can improve the spatial quality of a scene without affecting speech intelligibility.

68 Blind modelling of binaural unmasking for binaural speech intelligibility modelling at positive and negative SNRs

C. F. Hauth, **T. Brand**
Medizinische Physik and Cluster of Excellence Hearing4All, University of Oldenburg, Germany

The equalization cancellation (EC) model predicts the binaural masking level difference by equalizing interaural differences in level and time and increasing the signal-to-noise ratio (SNR) using destructive and constructive interferences. Here, a blind EC model is introduced that relies solely on the mixture of speech and noise, replacing the unrealistic requirement of the separated clean speech and noise signals in previous versions. The model uses two parallel EC paths, which either maximize or minimize the EC output level in each frequency band. If SNR is negative, minimization improves the SNR by removing the interferer component from the mixed signal. If SNR is positive, maximization improves the SNR by enhancing the target component. Either the minimizing or maximizing path in each frequency band is selected blindly based on an envelope frequency-selective amplitude modulation (AM) analysis. The requirement of considering positive SNRs is investigated using a binaural speech intelligibility experiment, where SRTs are obtained at positive SNRs. Results show a clear binaural release from masking for speech in noise at positive SNRs. The suggested AM-steered selection in the EC stage demonstrates that a simple signal driven process can be used to explain binaural unmasking of speech in humans.

69 A realistic test platform for near end listening enhancement

C. Chermaz, C. Valentini-Botinhao
The Centre for Speech Technology Research, The University of Edinburgh, United Kingdom

H. Schepker
Dept. Medical Physics and Acoustics and Cluster of Excellence Hearing4all, University of Oldenburg, Germany

S. King
The Centre for Speech Technology Research, The University of Edinburgh, United Kingdom

NELE (Near End Listening Enhancement) aims at improving the intelligibility of speech playback in noise. NELE algorithms are often evaluated in very controlled acoustic conditions, e.g. using synthetic speech shaped noise as a masker and not accounting for reverberation. While this is advantageous in terms of reproducibility, the benefit of NELE algorithms in real-world scenarios, e.g. for public announcements or telephone calls, may be overestimated. In order to create a more realistic test platform, two representative real-life scenarios were simulated: a large and crowded public space (the cafeteria) and a small domestic environment (the living room), which represent respectively a source of stationary and of fluctuating noise. Binaural impulse responses of real spaces [1] and live noise recordings were used for the simulations.

A listening test with N=24 normal hearing subjects was conducted. Intelligibility scores (in terms of correct keywords percentage) for unmodified speech were compared to those of a milestone study [2] on speech intelligibility in noise. Results indicate that higher SNRs are needed in order to achieve the same intelligibility levels when realistic noise is used, with differences of up to 8.6 dB. Preliminary results for a selection of NELE algorithms suggest that realistic noise proves to be more challenging also for modified speech, notwithstanding the type of modification.

This study exposes the gap between controlled lab conditions and the proposed real-world simulations, where the latter can provide a more meaningful prediction of the performance of NELE algorithms (and possibly other technologies) in real-life scenarios.

References

1. Kayser, Hendrik, et al. “Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses.” EURASIP Journal on Advances in Signal Processing, 2009: 6.

2. Cooke, Martin, et al. “Evaluating the intelligibility benefit of speech modifications in known noise conditions.” Speech Communication 55.4 (2013): 572-585.

70 Digits recognition in noise in school-aged population: effect of age, gender and number of spoken languages

J. Lagacé
Associate Professor, University of Ottawa, Canada

C. Giguère, L. Desormeau, A. Cameron, V. Vaillancourt
University of Ottawa, Canada

Objective – To evaluate the effect of age, gender and the number of spoken languages on the speech recognition in noise performance measured with the French version of the Canadian Digit Triplet Test.

Background – The importance of evaluating speech perception in noise within regular audiological evaluations of adult and children populations has been advocated for some years. Audiologists report the scarcity of valid speech in noise tests for not including this type of measures in their regular evaluation routine. A Canadian English and French version of the Digit Triplet Test (CDTT) has been developed by the University of Ottawa and Toronto (Ellaham et al. 2016, Canadian Acoustics 44(3), 220-221) in an attempt to fill this gap. The test uses an adaptive procedure to find the speech recognition threshold defined as the signal-to-noise ratio at which 50% of digit triplets are correctly identified.

Methods – The speech recognition threshold of 48 normal hearing French speaking children was measured with the CDTT. Two lists of 24 digit triplets (e.g., 5-2-8) were presented in a 65-dBA masking noise. The listeners were asked to enter the digits heard on a keypad.

Results – Between groups comparisons will be presented, as well as the effect of gender and number of spoken language.

Conclusions – The CDTT is ideal for measuring speech in noise performance in children as it requires no practice and can be quickly administered.

71 On the articulation between acoustic and semantic uncertainty in speech perception: Investigating the interaction between sources of information in perceptual classification.

O. Crouzet
LLING, Laboratoire de Linguistique de Nantes, France

E. Gaudrain
CRNL, Centre de Recherches en Neurosciences de Lyon, France

Listeners processing speech signals have to deal with two main classes of entropy. Ladefoged & Broadbent (1957, see also Sjerps & McQueen, 2013) showed that modifications in the resonant frequencies of a carrier sentence would affect the interpretation of a given vowel. For example, in the sentence "Please say what this word is: bet", increasing the frequency of F1 on the whole sentence, excluding the final word, would lead listeners to identify this final word as "bit" rather than "bet". Listeners may also take linguistic information into account, among which lexical hypotheses based on word co-occurence probabilities and semantic relations (e.g. McClelland & Elman, 1986). A tension can then take place between these two sources of information and the uncertainty that is associated with each of them must be stabilized in order to reach a perceptual decision. It is not clear however, how these sources of uncertainty are pondered in perception.

We are currently setting-up an experiment in which we plan to investigate this issue by independently manipulating (1) semantic relationships between words and (2) acoustic relations between a contextual part and the final word in this sentence. For example, based on word pairs that contrast on their vowel target only and for which the vowels are close to each other in articulatory / acoustic space (e.g. french "balle" vs. "belle", pronounced /bal/ vs. /bEl/ – eng. "ball" vs. "beauty"), 3 types of sentences are generated: (1) a sentence that would semantically "prime" the word /bal/ ("Le joueur a dévié la", eng. "The player deflected the"), (2) a sentence that would favour the word /bEl/ ("Le prince a charmé la", eng. "The prince charmed the") and (3) a neutral and / or semantically incongruous sentence in both cases "Le journaliste a parlé de la", eng. "The journalist talked about the").

We will present listeners with fully ambiguous final words (acoustically located between e.g. /bal/ and /bEl/) in contexts where semantic influence varies (sentence-types 1/2/3) and is balanced with acoustic manipulations of formant frequencies favouring one word or the other. This will let us describe how these sources of information interact in vowel classification. We are currently selecting sentence and word materials using French models for word embeddings (Mikolov, 2013) in order to provide a set of materials that will fit our expectations. We would highly benefit from discussing these materials and their modes of selection during the conference.

72 Comparison of binaural MVDR-based beamforming algorithms using an external microphone

N. Gößling, **S. Doclo**
Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, University of Oldenburg, Germany

Besides reducing undesired sound sources, an important objective of a binaural noise reduction algorithm is to preserve the spatial impression of the acoustic scene for the hearing aid user, such that no mismatch between acoustic and visual information occurs and the binaural hearing advantage can be exploited. Although the binaural minimum variance distortionless response (MVDR) beamformer is able to preserve the binaural cues of the target speaker, it distorts the binaural cues of the background noise. Hence, several extensions have been proposed, aiming at preserving the binaural cues of the background noise [1]. Because the performance of noise reduction algorithms is partly limited by the physical design, requiring the microphones to be integrated into the hearing devices, the usage of one or more external microphones that are spatially separated from the head-mounted microphones has been recently explored [2-3]. It has been shown that an external microphone enables to improve both the noise reduction performance as well as the binaural cue preservation performance.

In this contribution, we present a comparison of several binaural MVDR-based beamforming approaches that make use of an external microphone. First, the external microphone is merely used to provide an estimate of acoustic variables (e.g., relative transfer functions) that are required for the binaural beamforming algorithms. Second, the external microphone is used in conjunction with the head-mounted microphones as an additional input signal. Using realistic recordings of a moving target speaker in a reverberant room the performance of the considered binaural beamforming approaches is compared in terms of noise reduction performance (using speech-intelligibility-weighted SNR) and binaural cue preservation (using reliable ITD and ILD cues from an auditory model).

[1] D. Marquardt, S. Doclo (2018). “Interaural Coherence Preservation in Binaural Hearing Aids using Partial Noise Estimation and Spectral Postfiltering,” IEEE/ACM Trans. Audio, Speech and Language Processing, 26(7):1257-1270.

[2] J. Szurley, A. Bertrand, B. Van Dijk, M. Moonen (2016). “Binaural noise cue preservation in a binaural noise reduction system with a remote microphone signal,” IEEE/ACM Trans. Audio, Speech and Language Processing, 24(5):952–966.

[3] N. Gößling, S. Doclo (2018). “RTF-Based Binaural MVDR Beamformer Exploiting an External Microphone in a Diffuse Noise Field,” in Proc. ITG Conference on Speech Communication, Oldenburg, Germany, Oct. 2018, pp. 106-110.

73 Subjective evaluation of signal-dependent partial noise estimation algorithms for binaural hearing aids

J. Klug, N. Gößling, **S. Doclo**
Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, University of Oldenburg, Germany

Speech understanding is still a major challenge for many hearing aid users in complex acoustic scenes with background noise and interfering speakers. Besides suppressing undesired sound sources, an important objective of a binaural noise reduction algorithm is to preserve the spatial impression of the acoustic scene for the hearing aid user. Although the binaural minimum variance distortionless response (BMVDR) beamformer results in a good noise reduction performance and preserves the binaural cues of the target speaker, it distorts the binaural cues of the background noise, such that the target speaker and the background noise are perceived as coming from the same direction. Aiming at also preserving the binaural cues of the background noise, the binaural MVDR beamformer with partial noise estimation (BMVDR-N) was proposed, where a parameter allows to trade off noise reduction and binaural cue preservation of the background noise. In [1] a method was proposed to determine this trade-off parameter based on psychoacoustically-motivated boundaries for the desired interaural coherence of the background noise component.

In this contribution, we present a signal-dependent method to determine the trade-off parameter for the BMVDR-N beamformer based on the coherence between the noisy input signals and the output signals of the BMVDR beamformer [2]. We compare the performance of the BMVDR-N beamformer (both trade-off parameter methods) with the BMVDR beamformer for a realistic acoustic scenario with a target speaker, diffuse babble noise and an interfering speaker. Speech intelligibility was evaluated using the Oldenburg sentence test and spatial quality was evaluated using a MUSHRA test with N=11 normal-hearing subjects. In the presence of only diffuse babble noise, the signal-dependent trade-off parameter for the BMVDR-N beamformer significantly improves spatial quality compared to the BMVDR beamformer, while achieving the same speech intelligibility. Compared to the method in [1], the presented method achieves a significantly better performance, both in terms of speech intelligibility and spatial quality. When in addition an interfering speaker is present, the BMVDR-N beamformer (both trade-off parameter methods) significantly improves spatial quality without decreasing speech intelligibility compared to the BMVDR beamformer.

[1] D. Marquardt, S. Doclo (2018). “Interaural Coherence Preservation in Binaural Hearing Aids using Partial Noise Estimation and Spectral Postfiltering,” IEEE/ACM Trans. Audio, Speech and Language Processing, 26(7):1257-1270.

[2] J. Klug, D. Marquardt, N. Gößling, S. Doclo (2018). “Evaluation of Signal-Dependent Partial Noise Estimation Algorithms for Binaural Hearing Aids,” in Proc. ITG Conference on Speech Communication, Oldenburg, Germany, Oct. 2018, pp. 236-240.

74 The effect of acoustic and semantic cues on speech recognition in noise

K. Meemann, R. Smiljanic
The University of Texas at Austin, USA

Previous work has shown that listeners perform better in speech-in-noise tasks when the target speech has been produced clearly (e.g., Pichora-Fuller, Goy, & Van Lieshout, 2010) and when the speech signal contains sentence-level contextual information (e.g., Bradlow & Alexander, 2007; Smiljanic & Sladen, 2013). This benefit from clear speech modifications and semantic contextual cues has been shown to be modulated by several factors, such as the type and level of masking noise (e.g., Calandruccio et al., 2010; Payton et al., 1994) and listeners' experience with the target language (e.g., Mayo, Florentine, & Buus, 1997). While most research has examined the effect of some of these factors, only few studies have directly compared the intelligibility benefit of semantic and acoustic cues and their interaction with different types and levels of noise maskers.

The first goal of the current study was to explore to what extent listeners benefit from acoustic-phonetic and semantic intelligibility-enhancing cues. The second goal was to explore how these acoustic and semantic cues interact with energetic and informational masking at different signal-to-noise ratios (SNR). In two experiments, native English listeners heard meaningful noise-adapted (NAS) and clear speech (CS) English sentences, mixed with either speech-shaped noise (SSN), two-talker (2T), or six-talker (6T) babble, and presented at -5dB or -7dB SNR. In experiment 1, listeners heard sentences in which the final word was predicted by the preceding words (high-predictability sentences). In experiment 2, a different group of listeners heard sentences in which the final word could not be predicted from the preceding words (low-predictability sentences).

Results from both experiments showed that listeners benefitted significantly from CS and NAS for all masker types. Intelligibility gain from NAS compared to speech in quiet was significantly larger than the benefit from CS compared to conversational speech, indicating that the acoustic cues from NAS may overall be more accessible. For both experiments, the two speaking style modifications increased intelligibility most in SSN and least in 2T babble. This shows that speaking style adaptations improve word recognition most under energetic masking (SSN) and are less beneficial in listening conditions with less energetic masking (2T babble) that resulted from larger spectro-temporal dips. Results also revealed that the intelligibility benefit from NAS and CS was greater for high-predictability than for low-predictability sentences in all masker types. This suggests that listeners may be better at utilizing acoustic cues for speech recognition in noise when semantic cues are available.

76 The effect of stimulus choice on an EEG-based objective measure of speech intelligibility

E. Verschueren, J. Vanthornhout, T. Francart
ExpORL, Dept. Neurosciences, KU Leuven, Belgium

Recently an objective measure of speech intelligibility, based on brain responses, has been developed using structured Matrix sentences as a stimulus. We investigated whether this method also works with more natural running speech as a stimulus, as this would be beneficial for clinical application and required for neuro-steered auditory prostheses. We hypothesized that because the syntactic structure of natural speech is less controlled and more linguistic top-down processing is involved, the outcome measure could be different using a natural story compared to the Matrix sentences.

We recorded the electroencephalogram (EEG) in 19 normal-hearing participants while they listened to two types of stimuli: Matrix sentences, 5-word sentences containing a proper name, verb, numeral, adjective and object with 10 options per word category presented randomly, and a natural story. Each stimulus was presented at different levels of speech understanding by adding speech weighted noise. To investigate the brain responses we analyzed neural tracking of the speech envelope because the speech envelope is known to be essential for speech understanding and can be reconstructed from the EEG in response to running speech. The speech envelope was reconstructed from the EEG in both the delta and the theta band with the use of a linear decoder and then correlated with the real speech envelope. We also conducted a test-retest analysis to asses the reliability of our objective measure.

For both stimulus types and filter bands the correlation between the speech envelope and the reconstructed envelope increased with increasing speech understanding. In addition, correlations were higher for the story compared to the Matrix sentences. These results indicate that neural envelope tracking is affected by the stimulus choice and it can be enhanced by the use of more natural speech and speech understanding. These findings suggest that the choice of the stimulus has to be considered based on the intended purpose of the measurement.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 637424 to Tom Francart). Further support came from KU Leuven Special Research Fund under grant OT/14/119. Research of Jonas Vanthornhout (1S10416N) and Eline Verschueren (1S86118N) is funded by a PhD grant of the Research Foundation Flanders (FWO).

77 Do you understand the teacher? Case study: the effect of background noise and vocal effort on speech intelligibility in a classroom of a primary school

B. De Herdt, M. Schellens
Department Environmental Research, Provincial Institute of Hygiene, Antwerp, Belgium

To obtain a good understanding in classrooms, a suitable reverberation time (RT) and a low background noise level (BGN) are required, especially for children with communication disorders. An objective method to evaluate speech intelligibility is by determining the Speech Transfer Index (STI).

The present study investigated the effect of changing the BGN due to different classroom conditions and the effect of raising the level of vocal effort on the STI for two listening positions and one source position. The STI was determined and evaluated in accordance with the international standards IEC 60268-16. The RT was measured according to the European Standard ISO 3382-2:2008 and evaluated by the Belgian standard NBN S 01-400-2.

A MLS signal was played through the artificial mouth-directional sound source that was placed where the teacher was located (at a standing height of 1,60 m). There were two measuring positions: (1) located in the center of the classroom at a distance of 2 m from the source and (2) located at the back of the classroom at a distance of 4 m from the source. Both microphone positions were set on a sitting height of 1,40 m. Separate measurements of the impulse response (N= 5) and the BGN were conducted. These were combined when calculating the speech intelligibility in Dirac for different scenarios with different signal-to-noise ratios (SNRs).

The STI was evaluated as ‘good’ when there was a low level of BGN (position 1 = 0,67; position 2 = 0,60). During a regular lesson, the ambient noise level was induced by babbling between the children, moving of chairs, the use of electrical sharpeners and noise coming from the adjacent rooms. The STI was considered ‘poor’ for position 1 (STI = 0,32) and ‘bad’ for position 2 (STI = 0,29). When the teacher raised the voice, the STI became ‘fair’ for position 1 (STI = 0,48), but stayed ‘poor’ for position 2 (STI=0,45). The RT was evaluated ‘suitable’, thus it was indicated that the level of BGN was causing the low STI values. In fact, the child with the communication disorder was sitting at an inadequate position.

This study showed that although the teacher was raising the voice, it had only little effect on the STI. We can assume that the teacher is at risk of developing vocal problems. All children will benefit of lowering the BGN and optimizing the room in terms of acoustic insulation and absorption.

78 The role of non-linear processing in the prediction of speech intelligibility of hearing impaired listeners

H. Relano-Iborra, J. Zaar, T. Dau
Hearing Systems, Dept. Electrical Engineering,Technical University of Denmark, Denmark

A speech intelligibility model is presented, based on the computational auditory signal processing and perception model (CASP; Jepsen et al., 2008). CASP employs outer- and middle-ear filtering, a non-linear auditory filterbank (DRNL, López- Poveda & Meddis, 2001), adaptation loops, and a modulation filterbank, and has previously been shown to successfully account for psychoacoustic data of normal-hearing (NH) and hearing-impaired (HI) listeners in conditions of, e.g., spectral masking, amplitude-modulation detection, and forward masking (Jepsen et al., 2008; Jepsen and Dau, 2011).

This study shows the predictive power of the speech-based extension of CASP for NH listeners in conditions of additive noise, phase jitter, spectral subtraction and ideal binary mask processing. Furthermore, the model was adapted to individual hearing loss profiles in order to study the role of different processing stages, namely the auditory filter and inner hair cell transduction, in predicting accurate speech reception thresholds of HI listeners in conditions of speech in noise. The proposed model framework sheds light into the importance of non-linear processes in the auditory system for speech understanding, and how hearing losses that may linearize auditory processing affect HI speech intelligibility.

79 Characterizing the role of hearing loss in comodulation masking release

P. A. Mesiano, J. Zaar, B. Kowalewski, T. Dau
Hearing Systems, Dept. Electrical Engineering, Technical University of Denmark, Denmark

The detection of pure tones embedded in noise can be facilitated if amplitude modulations are imposed on the noise masker by multiplying it by low-frequency noise (i.e., modulated noise). This phenomenon is known as comodulation masking release (CMR). It can be quantified by measuring the reduction in the masked threshold observed when the masker is modulated, compared to a masker with the same spectrum and level but with an unmodulated envelope (i.e., unmodulated noise).

CMR in normal-hearing (NH) listeners has been observed through several experimental paradigms. Nevertheless, the perceptual mechanisms responsible for it are still unclear. In the studies investigating CMR in listeners with sensori-neural hearing loss (SNHL), an overall reduction in the amount of masking release was observed. Loss of sensitivity and reduced frequency selectivity were found to be linked to the reduction in CMR. However, the deficits in auditory processing directly responsible for the reduced CMR have not been precisely identified.

The aim of this study is to extend the investigation of CMR in relation to SNHL including other aspects of hearing. CMR was measured in a group of listeners with sloping, mild-to-severe SNHL. Large differences in the CMR across participants were observed, some showing normal CMR, others exhibiting reduced or absent CMR. The amount of CMR was related to several aspects of auditory processing, assessed by means of behavioral experiments. These included measurements of absolute thresholds, estimates of auditory filter width, cochlear compression and sensitivity to temporal fine structure. A statistical analysis of the results indicated a significant effect of frequency selectivity and sensitivity on the amount of CMR in HI listeners, in line with previous studies. No significant effect of the estimated cochlear compression ratio or sensitivity to temporal fine structure was observed. Additionally, simulations have been conducted with a computational model of auditory perception that has been shown to account for several aspects of hearing impairment. The model failed to predict the experimental results, even when the individual elevation of absolute thresholds, increased auditory filter bandwidth and loss of cochlear compression were included in the model front end. This suggests that other deficits in the auditory system, such as the processing of the temporal envelope of the stimulus, might be related to the CMR in listeners with SNHL.

80 Task dialogue between normal-hearing and hearing-impaired talkers in quiet and noise

A. J. Sørensen, E. N. MacDonald
Hearing Systems, Dept. of Electrical Engineering, Technical University of Denmark, Denmark

T. Lunner
Eriksholm Research Centre and Hearing Systems, Dept. of Electrical Engineering, Technical University of Denmark, Denmark

Maintaining an interactive conversation requires more resources than just understanding speech. Previous studies of the timing of turn taking in conversations suggest that interlocutors have to predict the end of each other's turn to sustain the rapid interaction that makes up normal, fluid conversation. Thus, while the presence of noise and hearing loss can make understanding speech more difficult, it should also make it more difficult to maintain the same fluid turn-taking dynamics in conversation.

In the present study, we recorded conversations between 12 pairs of native Danish young normal-hearing (NH) and older hearing-impaired (HI) listeners with mild presbyacusis in quiet and three levels of multitalker babble. The conversations involved solving the Diapix task, a spot-the-difference task between two almost identical pictures. Overall, the time the pairs took to complete the task increased with increasing noise level, suggesting that communication efficiency was impaired by the noise. Floor transfer offsets (FTOs), the intervals from when one talker stops and the other starts, were measured and the distributions were compared across conditions. With increasing background noise, both the median FTO and the standard deviation of the distribution increased. In addition, talkers held their turns significantly longer in increasing noise levels, which allows more time for speech planning and understanding. The HI's speech rates were constant over the four conditions, whereas the NH decreased their speech rates to match that of the HI in the loudest noises.

81 Speech & background levels in a realistic sound environment

N. Mansour, M. Marschall
Hearing Systems, Dept. of Electrical Engineering, Technical University of Denmark, Denmark

A. Westermann
Widex A/S, Denmark

T. May, T. Dau
Hearing Systems, Dept. of Electrical Engineering, Technical University of Denmark, Denmark

The use of realistic yet controlled sound scenarios for the evaluation of hearing-aid algorithms in a virtual sound environment (VSE) has the potential to positively impact the auditory quality of life of many hearing-impaired (HI) users. To do this in an ecologically valid way, these critical sound scenarios (CSS) need to be selected based on acoustic scenes that hearing-aid users experience as important through their difficulty and occurrence.

This study aims at selecting a set of appropriate CSSs based on results from literature and ecological momentary assessment (EMA) data, acquiring them in a real scene using a spherical microphone array, and reproducing them in an acoustically and perceptually valid way inside a VSE. A speech intelligibility task is implemented to obtain sound reception thresholds (SRT) for normal-hearing (NH) and hearing-impaired (HI) subjects and compare them to SRTs obtained with artificial background noise. In addition, a method for measuring in situ realistic speech levels between normal-hearing subjects is developed, and used to derive NH and HI speech intelligibility performance.

82 Relationships between envelope-following responses and speech intelligibility in noise: early markers of impaired hearing?

M. Garrett
Medizinische Physik and Cluster of Excellence Hearing4all, Oldenburg University, Germany

V. Vasilkov
Hearing Technology @ WAVES, Dept of Information Technology, Ghent University, Belgium

S. Uppenkamp
Medizinische Physik and Cluster of Excellence Hearing4all, Oldenburg University, Germany

S. Verhulst
Hearing Technology @ WAVES, Dept of Information Technology, Ghent University, Belgium

Despite having normal audiometric thresholds, many people have difficulties understanding speech in challenging listening environments. The origin of these suprathreshold hearing deficits are not well understood, even though human temporal bone studies provided evidence that a substantial amount of synapses and cochlear nerve terminals innervating the inner hair cells (IHC) can deteriorate due to noise overexposure or aging before audiometric hearing loss occurs (i.e., cochlear synaptopathy). This hearing deficit results in a reduced neural information transfer along the ascending auditory pathway and is believed to play an important role in our ability to understand speech in noise. To quantify synaptopathy in humans, non-invasive but indirect electrophysiological measures of peripheral hearing such as the envelope following response (EFR) have been proposed. EFRs are sensitive to synaptopathy in animal models, but their diagnostic sensitivity in humans as well as their relationship to speech intelligibility are still unclear.

This study aims to clarify how the EFR relates to different aspects of speech encoding in listeners with normal or impaired hearing. We consider young normal-hearing (yNH), elderly normal-hearing (oNH) and elderly hearing-impaired (oHI) participants as groups reflecting different degrees of age-related synaptopathy (yNH vs oNH) or outer-hair-cell deficits (oNH vs oHI). We focus on broadband speech in noise (OLSA) as well as low-pass (< 1.5 kHz) and high-pass (> 1.65 kHz) filtered speech stimuli to study how different cochlear frequency regions (and associated coding mechanisms) contribute to the EFR-vs-speech intelligibility relation in the different subgroups. The stimuli for the EFR were 4-kHz centered amplitude-modulated tones, narrow-band noise and sharp-envelope modulated tones. The results of this study shed light on the relationship between speech intelligibility and the underlying neural coding mechanisms and its impairments; an important first step towards better understanding and characterizing suprathreshold hearing deficits.

83 The effect of impaired speaker's voice and noise on children's spoken language processing

I. Schiller, D. Morsomme
Faculté de Psychologie, Logopédie et Sciences de l'Éducation, Université de Liège, BE

M. Kob
Erich-Thienhaus-Institute, University of Music, Detmold, DE

A. Remacle
Faculté de Psychologie, Logopédie et Sciences de l'Éducation, Université de Liège, BE and Fonds National de la Recherche Scientifique, Brussels, BE

Past studies indicate that listening to either impaired voice or against background noise may compromise children's ability to process spoken language. However, the interaction of both factors remains largely unknown.

The aim of this study was to investigate single and combined effects of impaired speaker's voice and noise on spoken language processing in children (aged 5-6).

First-grade primary school children (n = 53) individually performed two listening tasks: A Minimal-Pair Discrimination task assessing speech perception and a Sentence-Picture Matching Task assessing listening comprehension. Speech stimuli were presented in four conditions: (C1) normal voice and no noise, (C2) imitated impaired voice and no noise, (C3) normal voice and speech-shaped noise, and (C4) imitated impaired voice and speech-shaped noise. Task score per condition was calculated as measure of performance.

Irrespective of task, children performed significantly lower when stimuli were presented in a combination of impaired voice and noise (C4) as compared to any other condition. The presence of only one adverse factor (C2 or C3) lowered performance in the speech perception task but not the listening comprehension task.

Results suggest that when processing speech, young school-aged children are highly vulnerable to the combined effect of impaired speaker's voice and noise. This could be due to increased auditory masking and reduced cognitive capacity available for linguistic processing. With only a single adverse factor present, children seem able to still use semantic or syntactic context cues for correct interpretation. However, performance drops when such cues are unavailable. Favorable listening conditions may be crucial for children's processing of spoken language and positive learning outcomes. Particularly in the educational context, where listening is affected by voice quality and noise, measures should be taken to enhance the transmission of the speech signal and reduce noise.