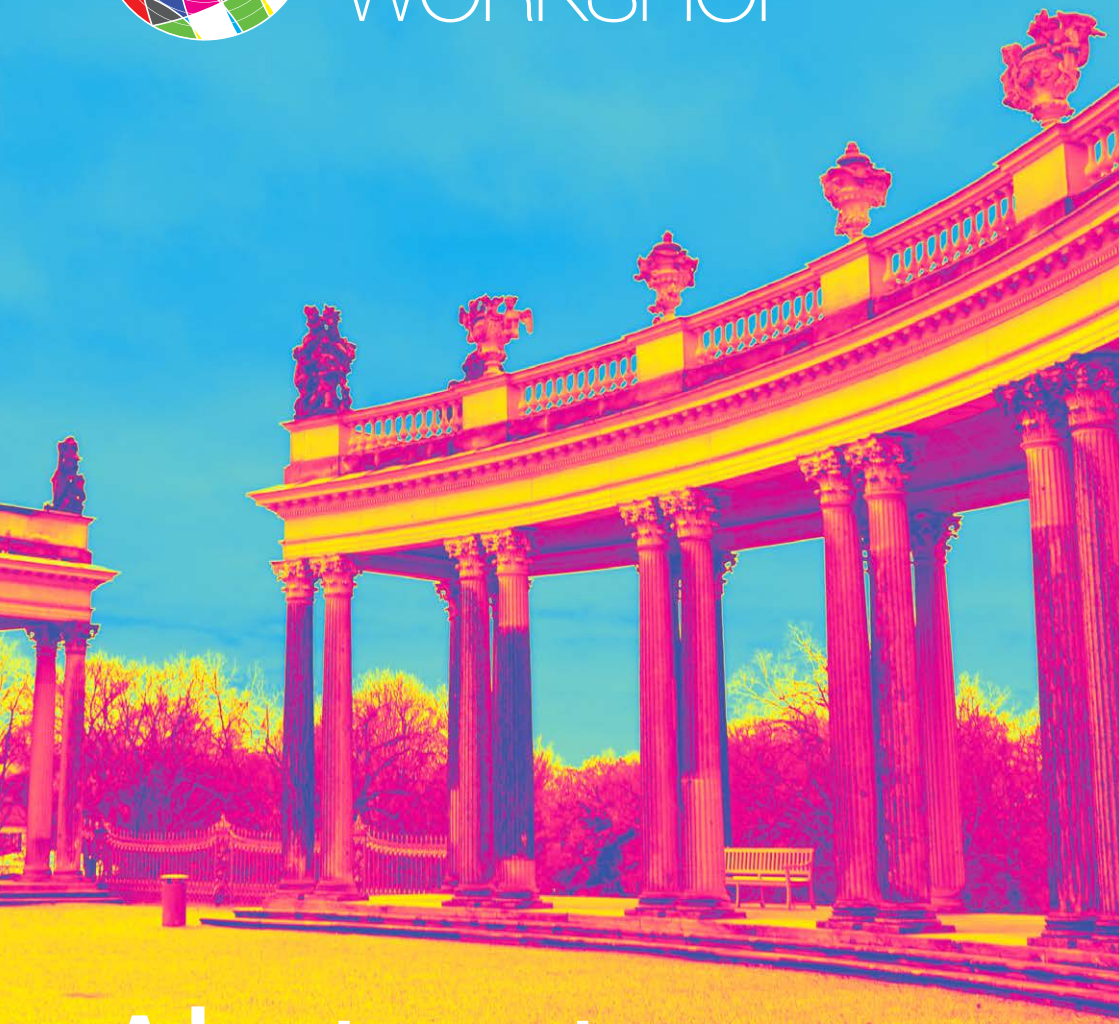
SPEECH IN NOISE WORKSHOP



Abstracts

Postdam, Germany | 11-12 January 2024

The 15th Speech in Noise Workshop was chaired by ***Outi Tuomainen****, University of Potsdam, Germany, with the help of the organisation committee:*

- *Laurianne Cabrera*
- *Etienne Gaudrain*
- *Antje Heinrich*
- *Chris James*
- *Damir Kovačić*

Coordinator: ***Thomas Koelewijn****, University Medical Center Groningen, Netherlands.*

https://2024.speech-in-noise.eu/
Contact information: info@speech-in-noise.eu

Date of publication: 22 December 2023.
DOI this version: 10.5281/zenodo.10423201.
DOI all versions: 10.5281/zenodo.10206866.

Cover illustration based on a photo by Kathleen Schneider (Creative Commons Attribution 4.0 International CC-BY).

The Speech in Noise Workshop is generously supported by:

oticon
life-changing technology

GN Advanced Science

WSAudiology

sonova
HEAR THE WORLD

Cochlear ®

Universität Potsdam

umcg

university of groningen

cnrs

Programme

Thursday 11 January 2024

09:00-09:45 Registration & Coffee

09:45-10:00 *Welcome*

10:00-10:25 Investigating speech processing in paediatric and adult CI users using combined fNIRS/EEG measurements
Kurt Steinmetzger
Tinnitus Center, Charité – Universitätsmedizin Berlin, Berlin, Germany

10:25-10:50 Neural correlates of stream segregation from childhood to adulthood
Elena Benocci, Claude Alain, **Axelle Calcus**
Université libre de Bruxelles, Belgium

10:50-11:15 What about speech?
Hannah J. Stewart, Erin K . Cash, Lisa L. Hunter, Jonathan E. Peelle, Jennifer Vannest, David R. Moore
Lancaster University, UK | Cincinnati Children's Hospital, USA

11:15-11:45 Coffee, Picture

11:45-12:10 Exploring mechanisms behind dynamic multi-talker listening
Hartmut Meister
University of Cologne, Germany

12:10-12:35 Speech in noise in the n200 study in Linköping Sweden
Henrik Danielsson, Erik Marsja
Linköping University, Sweden

12:35-14:05 Lunch

14:05-15:05 **Keynote —** Speech technology in everyday situations – is multimodality the answer?
Naomi Harte
Trinity College Dublin, Ireland

15:05-15:35 Coffee and Poster setup

15:35-18:00 Poster session 1

18:00-19:00 *Transit (30 min by Tram 96)*

19:00-22:00 Dinner at Kades Restaurant (Große Weinmeisterstraße 43B)

Friday 12 January 2024

09:00-11:30 Poster session 2

11:30-11:55 Coffee, and poster removal

11:55-12:20 Perception of the self voice and other voices during speech motor control
Abbie Bradshaw
University of Cambridge, UK

12:20-13:20 Lunch

13:20-13:45 **Colin Cherry Award 2023 —** Use of eye-tracking and pupillometry to assess speech-on-speech masking in a visual world paradigm
Khaled Abdel Latif, Thomas Koelewijn, Deniz Başkent, Hartmut Meister
Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital of Cologne

13:45-14:10 Pi-SPIN: Paraphrase to improve Speech Perception in Noise
Anupama Chingacham, Vera Demberg, Dietrich Klakow
Saarland Informatics Campus, Saarland University, Germany

14:10-14:35 How the types of sound reflections influence speech intelligibility in rooms
Nicola Prodi
Department of Engineering, University of Ferrara, Italy

14:35-15:00 Entraining alpha oscillations to facilitate auditory working memory: A TMS-EEG study
Kate Slade, **Jessica L. Pepper,** Elise J. Oosterhuis, Bjorn Herrmann, Ingrid S. Johnsrude, Helen E. Nuttall
Department of Psychology, Lancaster University, Lancaster, UK

15:00-15:20 *Business meeting: Colin Cherry Award 2024, next SPIN meeting and closing remarks*

15:20-15:50 Coffee and Goodbye

Talks

Thursday 11 January 2024, 10:00—10:25

Investigating speech processing in paediatric and adult CI users using combined fNIRS/EEG measurements

Kurt Steinmetzger
Tinnitus Center, Charité – Universitätsmedizin Berlin, Berlin, Germany

In case of severe hearing loss or congenital deafness, cochlear implants (CIs) represent the method of choice to restore hearing and enable language acquisition. Yet, little is known about how exactly the cortical processing of speech differs between acoustic and electrical hearing. In a first study, we thus tested unilateral adult CI users with preserved normal hearing in the other ear by separately presenting both ears with vowel sequences, while simultaneously recording functional near-infrared spectroscopy (fNIRS) and EEG data. Results showed smaller and delayed auditory cortex activity when the CI ear was stimulated. In a second study, paediatric CI users were tested during and after the first year of CI use and compared to an age-matched normal-hearing control group. In response to vowel sequences as well as to running speech, cortical activity was again markedly smaller in CI-based hearing. Despite trends in this direction, activity levels did not increase significantly with more CI experience. However, the less experienced CI group showed an abnormal shift of activity to the right hemisphere in response to running speech that was not observed in the other two groups. Overall, these data hence showed that, except from an initial adaptation phase, activity patterns were qualitatively similar but attenuated in electric hearing.

Thursday 11 January 2024, 10:25—10:50

Neural correlates of stream segregation from childhood to adulthood

Elena Benocci[1], Claude Alain[2], **Axelle Calcus**[1]
1. Université libre de Bruxelles, Belgium | 2. University of Toronto, Canada

In noisy backgrounds, listeners parse the concurrent auditory streams (“stream segregation”), and selectively focus on one stream as it unfolds over time (“selective attention”), performing the auditory scene analysis. Stream segregation is thought to remain immature in 12 year-olds with normal hearing. Here, we sought to investigate developmental changes in the neural signature of auditory stream segregation from childhood to adulthood. Children (n = 17), adolescents (n = 12) and young adults (n = 20) were presented with sequences of sounds consisting in coherent auditory figures imposed on stochastic backgrounds (“tone clouds”). These figure-ground sequences have been shown to elicit distinct EEG responses, including the object-related negativity (ORN) and P400, which reflect the processing of concurrent au-

ditory objects. Participants were also presented with a consonant identification task in three conditions: in quiet, in the presence of one interfering talker, and in the presence of speech-shaped-noise. Results indicate a progressive improvement in the behavioural figure-ground segregation. Amplitude of both ORN and P400 decreased with age; and was not significantly correlated with the behavioural performance. There was no clear relationship between ORN/P400 and speech perception in noise. Yet, our results suggest a protracted development of stream segregation from childhood to adulthood. This will be discussed in relation with the literature on auditory scene analysis and the development of the central auditory pathways.

Thursday 11 January 2024, 10:50—11:15

What about speech?

Hannah J. Stewart[1,2], Erin K . Cash[2], Lisa L. Hunter[2,3], Jonathan E. Peelle[4], Jennifer Vannest[3], David R. Moore[2,3,5]

1. Lancaster University, UK | 2. Cincinnati Children's Hospital, USA | 3. University of Cincinnati, USA | 4. Northeastern University, USA | 5. University of Manchester, UK

Despite passing standard audiological testing, 10-20% of children experience listening difficulties (LiD), particularly in noisy conditions. These children are often recommended for further auditory processing disorder (APD) assessment. In this study we examined magnetic resonance imaging magnetic resonance imaging (MRI) data from a large longitudinal study assessing whether LiD is due to sensory, cognitive or speech processing problems. We analyzed 81 children aged 6-12 years: 39 were typically developing; and 42 were classified as having LiD from caregiver reports of everyday listening difficulties (the ECLiPS).

We used resting state MRI where we can explore how different brain areas work together while the child is at rest, i.e. not performing an explicit task. We used a data driven analysis to examine the connectivity of three brain networks: Speech, Sound, and Visual. We found an extensively deficient Speech network in children with LiD compared to TD children. However, they had a very similar Sound network and no significant differences in their Visual network.

The extensive deficits in the Speech network connectivity were in higher level, speech processing brain areas rather than in the primary sensory processing brain areas. The strength of Speech network connectivity significantly related to the children's listening skills (ECLiPS, CCC-2), auditory processing skills (SCAN-3:C) and cognition (a composite of attention and memory). But not their speech-in-noise skills (LiSN-S). Our analysis also showed a maturational increase in connectivity throughout the Speech network for children with LiD. In summary, our study shows that children identified as having LiD through a caregiver report showed extensive changes in brain network function. We conclude that listening difficulties in children are mediated by speech-specific mechanisms.

Exploring mechanisms behind dynamic multi-talker listening

Hartmut Meister
University of Cologne, Germany

In communication situations of daily life multiple persons often speak at the same time. Such conditions, coined by the term “Cocktail Party Problem” (Cherry, 1953), place high demands on both the auditory system as well as cognitive abilities. Multi-talker listening requires the segregation of competing speech streams in order to focus attention on the message of interest (Shinn-Cunningham & Best, 2008). Stream segregation relies on different acoustic features, such as voice, spatial and intensity cues.

However, realistic scenarios typically involve conversational turn-taking, that is, the talker of interest can change dynamically. This has implications for auditory attention, as multiple sources need to be monitored and the focus of attention has to be switched when the target changes. Compared to static conditions with only one talker of interest, such dynamic multi-talker situations typically come at a “cost”- reflected in decreased speech recognition or increased reaction times, demonstrating the corresponding cognitive load (e.g., Brungart & Simpson, 2007; Lin & Carlile, 2015; Meister et al., 2020).

In a series of studies, our aim was to shed light on different mechanisms of dynamic versus static multi-talker listening (Meister et al., 2020; Wächtler et al. 2021; Wächtler et al., 2022). The studies are based on a paradigm with three spatially separated talkers presenting matrix sentences simultaneously. The talker of interest was dynamically changed with different switching probabilities, putting different strain on attentional demands. Costs of dynamic listening were calculated in relation to static conditions and different error types, such as confusion between the three talkers or omissions, were determined to get a more detailed insight into the different effects. The talk will present several analyses addressing the mechanism of dynamic multi-talker listening with a focus on age and hearing loss.

Funding: Deutsche Forschungsgemeinschaft (ME2751/3-1).

References:

Brungart, D.S., Simpson, B.D. (2007). Cocktail party listening in a dynamic multitalker environment. Percept. Psychophys. Jan 69 (1), 79e91.

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. 25, 975e979.

Lin, G., Carlile, S. (2015). Costs of switching auditory spatial attention in following conversational turn-taking. Front. Neurosci. Apr. 20 9, 124.

Meister, H., Wenzel, F., Gehlen, A. K., Kessler, J., & Walger, M. (2020). Static and dynamic cocktail party listening in younger and older adults. Hearing research, 395, 108020.

Shinn-Cunningham, B.G., Best, V. (2008). Selective attention in normal and impaired hearing. Trends Amplif. Dec 12 (4), 283e299.

Wächtler, M., Kessler, J., Walger, M., & Meister, H. (2021). Costs of dynamic cocktail party listening: Investigating the effects of cognitive abilities and hearing impairment. JASA EL, 1(7), 075201.

Wächtler, M., Kessler, J., Walger, M., & Meister, H. (2022). Revealing Perceptional and Cognitive Mechanisms in Static and Dynamic Cocktail Party Listening by Means of Error Analyses. Trends in hearing, 26, 23312165221111676.

Speech in noise in the n200 study in Linköping Sweden

Henrik Danielsson, Erik Marsja
Linköping University, Sweden

Background: The n200 project in Linköping Sweden was initiated to investigate the relationship between hearing, cognition and speech in noise. The initial 200 participants with hearing loss and hearing aids have been complemented by 200 participants without hearing loss. In addition, a smaller group of participants with hearing loss but without hearing aids was also found and tested when the group without hearing loss was recruited. The present study investigates the predictors of speech in noise in the 3 different groups for two different speech in noise tests, that is the Hearing In Noise Test (HINT) and the Hagerman sentences test. The Reading span task was used as an indicator of cognition, while better ear PTA was used as indicator of hearing ability.

Method: The two larger groups had comparable mean age (61 years), while the smaller group was older (69 years). Predictably, the hearing aid users had largest hearing impairment (37 dB better ear PTA), followed by the smaller group (28 dB) and the group without hearing impairment (10 dB). Correlational analyses were followed up by regression analyses to investigate the predictors of HINT and Hagerman in separate analyses.

Results and discussion: The pattern of predictors was different between the different groups and the two speech in noise tests. Reading span was a significant predictor for the Hagerman sentences but not the HINT test for all groups. Age and PTA were significant predictors for the hearing aid users and for the group without hearing loss, regardless of speech in noise test. For the smaller group, age but not PTA were significant predictors. It should be noted that the amount of explained variance was generally smaller for the HINT test and smaller for the groups without hearing aids. Results will be discussed in relation to the difference between the speech in noise tests and differences between the groups.

Keynote lecture

Speech technology in everyday situations – is multimodality the answer?

Naomi Harte
Trinity College Dublin, Ireland

Speech technology, such as automatic speech recognition (ASR), has seen impressive increases in performance in the past decade. Still however, ASR architectures demonstrate poor robustness when faced with the noisy scenarios that are mundane to us humans and don't unduly disrupt our ability to continue our conversations. Speech is not something we just hear though, it's also something we see. In a conversation, we seamlessly signal and monitor a multitude of visual cues including eye gaze, facial expression, hand gestures and head nods. In parallel, we interpret linguistic information and prosody. In a noisy room, we pay attention to lip movements. In this talk, I examine how we can integrate multimodality when developing more robust speech technology. In particular, I will focus on my group's recent and ongoing work in audio-visual speech recognition and conversational analysis. I'll also discuss how by truly understanding the nature of speech interaction, and by working in multidisciplinary teams, that we can build neural architectures better suited to this challenging task.

Perception of the self voice and other voices during speech motor control

Abbie Bradshaw
University of Cambridge, UK

Processing of self-generated speech auditory feedback is known to play a critical role in speech motor control. This has been demonstrated using the altered auditory feedback paradigm. Specifically, speakers exposed to a predictable sustained perturbation of real-time speech auditory feedback (e.g. a change in the first or second formant) gradually start to adapt to this perturbation; for example, by shifting their produced formant frequencies in an opposite direction to the perturbation. Less is known however about the impact of simultaneous perception of other voices on such speech motor adaptation, and whether adaptation is robust across different speaking contexts. In particular, despite speech typically being a social act, few previous studies have examined speech adaptation in contexts involving a social element. In this talk, I will firstly present some of my work looking at speech motor control during synchronous speech; the act of speaking in time with another speaker. Through a series of studies this research has shown that (1) synchronous speech induces vocal convergence (that is, causes a participant's voice to become more similar acoustically to their synchronisation partner), (2) synchronous speech affects speech motor adaptation to simultaneous formant perturbations, and (3) the effect of synchronous speech on speech motor adaptation depends on whether convergence aligns or conflicts with such adaptation, and not simply on auditory masking. In the second part of my talk, I will present my ongoing work investigating speech motor control and reports of perceptual experience of the self-voice under conditions

of speaking with degraded auditory feedback. This degradation of speech feedback is implemented using real-time noise vocoding, a technique that allows one to vary the level of spectral detail present in the speech signal. This provides an (incomplete) simulation of speech perception in individuals with cochlear implants. This work will investigate whether speech motor compensation for formant perturbations is intact when speaking with noise-vocoded speech feedback, as well as the effects of both expectedness and clarity of speech feedback on explicit reports of perceptual experience of the self-voice. I will discuss implications for the relationship between implicit motor control and conscious perception for self-produced speech, and for speech motor control in individuals with cochlear implants.

Friday 12 January 2024, 13:20—13:45

Colin Cherry Award 2023

Use of eye-tracking and pupillometry to assess speech-on-speech masking in a visual world paradigm

Khaled Abdel Latif[1], Thomas Koelewijn[2], Deniz Başkent[2], Hartmut Meister[1]
1. Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital of Cologne | 2. Department of Otorhinolaryngology, University Medical Center Groningen, the Netherlands

Everyday communication frequently includes situations with several persons speaking simultaneously. Such speech-on-speech listening is usually challenging, as understanding the person of interest requires the segregation of competing speech streams as a prerequisite to focus attention on the target talker. The features for segregation, like voice, spatial and intensity cues, are not always readily available.

In the pursuit to understand the intricacies of speech-on-speech segregation at a fine-grained temporal level, we developed a Visual World Paradigm (VWP, Tanenhaus et al., 1995) based on matrix sentences (The Oldenburg Sentence Test, Wagener et al., 1999). The VWP visually presents elements of the matrix sentences parallel to the spoken stimulus and is based on the finding that gaze fixations and speech processing are closely linked in time. Various Target-to-Masker Ratios (TMRs) were employed as cues to aid discriminating the target sentence from the masker sentence.

Our study, involving young normal-hearing individuals, unveiled that both gaze fixations and pupil dilation effectively reflected the segregation of competing sentences, yielding valuable insights into the processing of elements within the target sentence. Our data analysis included three parameters: the peak and slope of eye-gaze fixations to gauge speech processing, and the pupil dilation over time to assess the related cognitive load. The impact of varying TMRs

clearly became evident, with a significant decrease in the peak and the slope of fixations and a significant increase in pupil dilation in more challenging TMR scenarios. Notably, this held true even when the corresponding speech recognition was virtually perfect.

Funding: Deutsche Forschungsgemeinschaft (ME2751/6-1)

References:

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. Science (New York, N.Y.), 268(5217), 1632–1634. doi:10.1126/science.7777863

Wagener K., Brand T., Kollmeier B. (1999). Entwicklung und Evaluation eines Satztests in deutscher Sprache II: optimierung des Oldenburger satztests (Development and evaluation of a sentence test in German language II: optimization of the Oldenburg sentence test). Z. Audiol. 38 44–56.

Friday 12 January 2024, 13:45—14:10

Pi-SPIN: Paraphrase to improve Speech Perception in Noise

Anupama Chingacham, Vera Demberg, Dietrich Klakow
Saarland Informatics Campus, Saarland University, Germany

Considering the increasingly wide application of spoken dialog systems (SDS) in real-world noisy environments such as navigation and medical assistance, synthesizing noise-robust and better intelligible utterances has become a pressing priority. Nevertheless, current SDS are less adaptive to the listening difficulties of their interlocutors, in contrast to human speakers. Much research in the past has focused on modulating acoustic features like imitating the Lombard Speech to synthesize noise-robust speech.

In this presentation, I will first focus on our proposed strategy — replace a sentence with its better-intelligible paraphrase — to improve speech perception in noise. A new dataset called Paraphrases-in-Noise (PiN) was created by collecting the human perception data of sentential paraphrases in noisy environments. Our experimental results demonstrate that the choice of linguistic forms to represent a message introduces a significant difference in intelligibility among sentential paraphrases, in noise. In a highly noisy environment like babble noise at SNR -5 dB, replacing utterances with sentential paraphrases that have better acoustic cues resulted in an overall intelligibility gain of 33%. In the second part of the talk, I will present two novel approaches that we designed to synthesize noise-robust speech — (1) an intelligibility-aware paraphrase ranking model, and (2) a paraphrase generation model that optimizes for better intelligibility. To encourage further explorations on the mitigation of human mishearing in noise, we released the PiN dataset.

Acknowledgments: This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 232722074 – SFB 1102.

How the types of sound reflections influence speech intelligibility in rooms

Nicola Prodi
Department of Engineering, University of Ferrara, Italy

Depending on the room boundaries, the sound reflections following the direct sound in a room can be either specular or at least partly diffusive, the latter also named scattered. While specular reflections are correlated by nature, scattered ones are usually uncorrelated. Still limited research has been directed to the understanding of how our auditory system incorporates the early sound energy from either type of reflection and the results are not entirely consistent. The amount of early sound energy is highly relevant in the context of speech intelligibility of a target source in rooms due to the time limit of the useful energy integration. In addition, part of the binaural information coded in the target impulse response can be used to estimate the binaural unmasking. For this scope the interaural phase difference between target and masker is used in E-C based models together with the masker correlation. In similar modelling E-C frameworks the correlation of the target signal is almost disregarded: this is realistic as far as short source-receiver distance is considered or if the environment is very damped. Actually, in a real context, target correlation may vary, and is modulated either by the type of reflection, by reverberation, by distance or by a mixture of all the factors. Unfortunately, the accurate control of target correlation alone while keeping other speech qualities untouched is not trivial. In recent years much effort was made to disentangle the role of target correlation from other known monaural and binaural factors influencing speech intelligibility. In one experiment based on simulations the whole impulse response consisted of either specular or scattered reflections while monaural and binaural indicators were kept fixed. The noise was localized and it was either energetic or informational. The obtained spatial release from masking (SRM) differed for the two types of reflections, specular and diffuse, and the former ensured better performance. The E-C modelling approach helped to isolate the contributions of better-ear and binaural unmasking. In another experiment based on real measures the target correlation was further investigated by comparing symmetric scattered early reflections (that, although diffuse, keep an high left-right correlation) against unsymmetric ones that provide very low left-right correlation. The interferer's correlation was modulated too. Results helped in better understanding the role of target correlation and provide grounds for better acoustical design of rooms for speech.

Entraining alpha oscillations to facilitate auditory working memory: A TMS-EEG study

Kate Slade[1], **Jessica L. Pepper**[2], Elise J. Oosterhuis[2], Bjorn Herrmann[3,4], Ingrid S. Johnsrude[5], Helen E. Nuttall[2]

1. Lancaster Medical School, Lancaster University, Lancaster, UK | 2. Department of Psychology, Lancaster University, Lancaster, UK | 3. Rotman Research Institute, Baycrest Academy for Research and Education, Toronto, Canada | 4. Department of Psychology, University of Toronto, Toronto, Canada | 5. Brain and Mind Institute, University of Western Ontario, Ontario, Canada

Older adults often find listening to and remembering speech a challenging process, particularly in complex listening environments. Neural alpha oscillations may support working memory during speech perception. Specifically, alpha oscillations in parietal cortex may promote inhibition of distracting sounds, whereas alpha oscillations in temporal cortex may enhance attention to target sounds. Importantly, alpha activity during speech perception may be different for younger and older adults. We hypothesised that entraining alpha activity in these brain regions may facilitate speech perception. We investigated whether TMS delivered at an individualised alpha frequency (alpha-TMS) benefits auditory working memory, and how this may be affected by age.

Thirty younger adults (mean age = 20.9 years) and 15 older adults (mean age = 68.6) completed trials of an auditory working memory task. Participants attended to and recalled 9-digit sequences, whilst ignoring irrelevant sentences. Before the to-be-ignored sentences, participants received alpha-TMS. We investigated the effects of: (1) distractibility of irrelevant sentences (less vs. more distracting); (2) site of alpha-TMS (control vs. parietal vs. temporal); (3) age group (younger vs. older), on digital recall and alpha power. For alpha power, we also investigated differences across phases of the trial (attending vs. ignoring vs. recall).

Across all TMS conditions and age groups, digit recall was poorer in trials with more distracting to-be-ignored sentences [$F(1,39)=13.19$, $p=.001$]. Across all ages, sentences, and TMS sites, alpha power was highest during the digit attending phase and lowest during the recall phase [$F(2,80)=15.54$, $p<.001$]. There was also an interaction between trial phase and TMS site on alpha power [$F(2.7,108.5)=5.19$, $p=.003$]. Specifically, in the control and parietal TMS conditions, alpha power was highest during the attending phase, whereas, in the temporal TMS site condition, alpha power was highest during the ignoring phase. Alpha-TMS may modulate parietal and auditory alpha, which may influence independent inhibitory and attentional processes.

Posters

SESSION 1: Thursday 11 January 2024, 15:35-18:00

SESSION 2: Friday 12 January 2024, 09:00-11:30

P01 Effects of speaker adaptations in face-masked speech on working memory and voice perception

Cleopatra Christina Moshona
Engineering Acoustics, Technische Universität Berlin, Germany

Background: In everyday scenarios, interpersonal communication often occurs in challenging acoustic conditions. This is also the case when interlocutors converse with physical obstructions like face masks, which have been found to impair voice propagation and radiation. To ensure efficient message transmission, speakers modify their speaking style to overcome these obstacles and make themselves more understandable. This adjustment includes the use of Lombard speech. While such mechanisms can enhance speech intelligibility for listeners in noisy environments, they may not necessarily be advantageous in other cognitive contexts or situations characterized by minimal or no background noise. The alteration in voice quality associated with Lombard speech could affect how the speech signal is perceived, potentially making it more disruptive for specific cognitive tasks. Furthermore, the extent and appropriateness of such modifications might vary, depending on individual speaker characteristics.

Methods: In the present study we examined the effects of speaker adaptations in face-masked speech on working memory performance, voice quality perception and psychoacoustic-phonetic measures. For this purpose, 33 participants were presented with audio recordings of a native German speaker uttering matrix-type sentences with and without a face mask in conversational and Lombard speech style, contained in the BEMASK corpus. Listeners completed a self-paced cued-recall task in a quiet environment and subsequently rated perceived voice quality using unipolar semantic differential scales. Statistical analyses were carried out using GLMMs.

Results: Our findings align with previous research that indicates a decrease in recall ability while wearing a face mask, as opposed to not wearing a mask, but this effect was not statistically significant. However, adjusting one's speech style while wearing a mask impacted recall performance notably. Specifically, Lombard speech led to a significant reduction in recall, resulting in an average decrease of 5%. Psychoacoustic and phonetic measures of speech correlating with annoyance and increased vocal load were found to be higher in face-masked Lombard speech and participants perceived a significant change in voice quality between speech conditions.

Conclusion: In summary, our study replicated that face masks can slightly impact recall ability, but this effect was negligible. However, the adoption of Lombard speech, a common strategy for improving speech intelligibility, significantly reduced recall performance. These findings suggest that speech adaptations in the context of face masks may have unintended

consequences on cognitive tasks. They highlight the complexity of speech adjustments and the need for a nuanced understanding of their impact, especially in varying cognitive contexts and for individual speakers.

P02 Effects of context and auditory ability on speech perception in noise

Emily Tomasino, Patricia Bestelmeyer, Guillaume Thierry
Bangor University, United Kingdom

Being able to follow a conversation in a noisy environment is a critical skill. Whilst difficulty perceiving speech-in-noise (SIN) is common in clinical populations and even in healthy-hearing individuals, there is considerable interindividual variability in performance. Previous studies have shown that our ability to understand speech-in-noise (SIN) depends on both cognitive and low-level auditory skills, as well as the context in which information is perceived. Here, 155 participants took part in a battery of tests targeting the interplay between low-level auditory skills (such as temporal processing and sound grouping) and higher-order auditory processing skills (such as vocabulary and inhibition) when predicting SIN perception performance in healthy hearing individuals. We also investigated the role of context by presenting participants with a prime word in the clear, that was either related or unrelated to the target word presented with various levels of background noise (i.e., variable signal-to-noise ratios). Preliminary analyses with multiple linear regression revealed that the best predictors of SIN performance presented with context (related word pairs) were vocabulary proficiency and basic auditory skills (measured using the PROMS/the profile of music perception skills, Law & Zentner, 2014, doi:10.1371/journal.pone.0052508). The best predictor of performance for unrelated word pairs was sound grouping skill (measured with a figure-ground task, Holmes & Griffiths, 2019, doi:10.1038/s41598-019-53353-5). Further investigation will include a path analysis featuring all observed variables, this will give a more comprehensive view of the relationships between higher-order and lower-order processing skills and their role in SIN performance. A better understanding of SIN variability in healthy-hearing individuals could help design interventions and environments. For example, vocabulary, basic auditory skills (such as pitch perception), and auditory grouping skills could be targeted in training.

P03 Grandchild, speak clearly! Investigating algorithmic benefit for lisped speech understanding

Alexander Klingebiel[1,2], Cecil Wilson[1], Ulrich Hoppe[2], Marko Lugger[1], Maja Serman[1]
1. WS Audiology, Germany | 2. FAU Erlangen-Nürnberg, Germany

Elderly listeners often have trouble understanding mispronounced speech. Specifically, this may happen in communication with grandchildren, who may have a lisp or simply talk quickly and unclearly. Lisping is a popular term for different types of consonant mispronunciation and is a common stage in speech development. These communication problems are worse in noise, since the content of the noise can mask the already ambiguously pronounced consonants.

Here, we are interested in investigating the benefit of an adapted onset enhancement algorithm, as a method for improving speech understanding of lisped speech in noise. The algorithm was adapted to specifically enhance lisped speech onsets.

In order to investigate the effect of the algorithm, we recorded professional speech therapists from a speech therapy school in Nürnberg. The ability to produce different types of lisping is one of the speech therapist's learned assets. We asked them to read the fairy tale "Nordwind und Sonne" and minimal word pairs, in normal pronunciation and in three different types of lisping. From these recordings, we developed a study protocol, consisting of two specific tests.

Twenty normal hearing listeners participated in the study. Their subjective impression was assessed while listening to the fairy tale recordings. The objective performance was evaluated with the minimal word pairs test. In both tests the material was processed with and without the mentioned algorithm. The results of the study will be presented and discussed.

P04 Exploring attentive listening in noise through the just-follow conversation task

William M. Whitmer, David McShefferty
Hearing Sciences - Scottish Section, University of Nottingham

Karolina Smeds
ORCA Europe, WS Audiology

Being even a passive listener in a social gathering can be a challenge. There could be more than a story told by one person or a conversation amongst many that we may want to follow. In the just follow conversation task (JFC), a listener adjusts the signal level to where they can understand, with effort, the gist of what is being said in a background of noise. The JFC therefore provides insight into how we experience conversational listening by assigning an acoustic measure, a signal-to-noise ratio (SNR), to our self-perceived ability. Here, we use the JFC to probe if there is a difference between focused listening to dialogues vs. monologues, and how well we can distribute our attention across multiple dialogues or monologues. We further explore the psychometric properties of the JFC as a perceptual measure of attentive and aided listening.

Participants sat in the centre of a circular loudspeaker array and adjusted the level of one monologue, one dialogue, two monologues or two dialogues presented in the front hemifield to where they could just follow the speech. On each trial, participants made four adjustments. Signals were presented in an Ambisonics café background and uncorrelated same-spectrum noise, both at a fixed long-term average level of 67.3 dB(A). Bilateral hearing-aid users adjusted each speech type in each background both aided and unaided. Non-users repeated each condition to evaluate reliability.

Results show that just following a dialogue is generally perceived as more difficult than following a monologue, but the relationship is dependent on the type of noise and amplification. Individual JFC SNRs were correlated with speech subscale scores on the SSQ12 (Speech, Spatial and Qualities of Hearing Scale) as well as individual pure-tone threshold averages. No significant differences were found between unaided and aided conditions for hearing-aid users. JFC reliability was comparable to more objective speech understanding measures, but it may not be suitable to capture perceived conversational benefits for more subtle changes in hearing-aid processing.

P05 Unmasking attention: Investigating the competing acoustic and cognitive influences during spatial speech-on-speech listening

Georgie Maher, Sarah Knight, Sven Mattys
University of York, UK

Understanding speech-perception-in-noise (SpiN) requires modelling the interaction between bottom-up (acoustic) factors and top-down (cognitive) processes. However, the precise mechanisms by which cognition supports SpiN are unclear. In particular, the role of working memory (WM) remains debated. Some studies show that high-WM individuals have better SpiN. While this link is broadly established for older and/or hearing-impaired listeners, however, it is much less clear for young, normal-hearing adults. This may be partly due to: (1) the failure of existing studies to assess the multiple components of WM and account for other relevant abilities, such as attentional control; (2) lack of power to exploit the narrow range of individual variability in many WM tests; (3) differing degrees of acoustic difficulty in existing paradigms, usually implemented by manipulating energetic masking (EM; interference between speech and noise at the auditory periphery).

In this study, participants completed a selective listening task in which they were asked to transcribe the speech of one of two simultaneously-presented talkers. We filtered the speech into frequency bands that were either identical or non-overlapping between talkers (EM-present vs. EM-absent). We also manipulated perceived spatial distance between the talkers (collocated, i.e., diotic, vs. +/- 90^0 azimuth, i.e., dichotic). For EM-present stimuli, this resulted in maximal EM in the collocated condition and minimal EM in the dichotic condition, whereas spatial-attentional demands were maximal in the dichotic condition and minimal in the collocated condition. For EM-absent stimuli, only spatial-attentional demands varied across spatial distance. Participants also undertook a battery of cognitive tasks to assess three key compo-

nents of WM and attention: phonological loop, executive function and selective/divided attentional control. The study is currently being run online with a target of N=240. Results will be reported at the conference.

For EM-present stimuli, we expect that performance will be better in the dichotic than diotic condition due to spatial release from energetic masking. For EM-absent stimuli, however, we predict that performance will be better in the diotic than dichotic condition due to the cognitive cost of spatial attentional control in the dichotic condition. This hypothesised cognitive cost also leads us to expect that cognitive task scores will be more strongly linked to performance in dichotic than diotic conditions. When EM is severe, we expect that listeners will not be able to restore degraded speech via recruitment of cognitive resources, thus making the link between listening and cognition task scores weakest in the collocated, EM-present condition.

P06 A relationship between amplitude modulation neural processing, modulation masking and consonant-in-noise perception

Clémence Basire[1], Irene Lorenzini[2], Laurianne Cabrera[1]
1. Integrative Neuroscience and Cognition Center, CNRS-Univ de Paris, Paris, France | 2. Integrative Neuroscience and Cognition Center, CNRS-Univ de Paris, Paris, France I Laboratoire Ethologie Cognition Développement (LECD) Université Paris Nanterre France

Psychoacoustic research has highlighted the fundamental role of temporal modulations for speech perception in noisy environments. The present study seeks for a relationship between AM processing (using a behavioural task and an electroencephalography (EEG) measure) and speech perception in noise. The AM following response (AMFR, or envelope following response), an auditory potential, reflecting the brain activity following the modulation frequency of amplitude modulated tones, can be recorded at the scalp level with EEG. It was hypothesized that higher AM processing (higher magnitudes of AMFR and better abilities in AM perception behavioural task) would positively correlate with lower (better) speech-in-noise thresholds. Moreover, this relationship was expected to differ as a function of AM rates, as speech information is mainly conveyed by slow AM cues.

Fifty young adults with normal hearing (18-30 years) completed two experiments: 1) an EEG session measuring AMFR at two AM rates (8 vs 40 Hz using a pure tone centred at 1024 Hz sinusoidally modulated at 100%), and 2) a behavioural measure estimating consonant-identification thresholds in noise. Thirty-four adults of this group completed a third experiment: 3) a behavioural assessment of AM detection thresholds at 8 Hz. For the EEG experiment, Fourier transform (FFT) was performed on the averaged waveforms in each AM rate condition. The maximum magnitude value at 8 and 40 Hz was estimated individually and corrected by the EEG noise for each participant. For the speech-in-noise adaptive task, syllables of the form /aCa/ were presented. Six phonetic conditions were designed presenting a minimal change in place of articulation, voice or manner of articulation for fricatives or stop consonants. Syllables were presented in a XAB task to assess consonant identification thresholds within a steady speech-shaped noise. In the AM detection adaptive task, participants had to identify the modulated sound among two different sounds, which only one was modulated in ampli-

tude. Every time they succeeded the modulation depth of the AM sound decreased. In one condition of this task the carrier frequency of sound is a 500 Hz pure tone and in the other condition it is a narrow-band noise. This allow us to evaluated the masking effect (masking effect = score in pure tone condition - score in noise's one).

AMFRs were observed at each modulation rate. Thresholds for consonant-identification in noise were comprised between -11 and -19 dB SNR for all phonetic conditions averaged. Correlation between these three experiments (n=34) revealed that higher magnitude of AMFR only at slow AM rate (8 Hz) is positively correlated with lower (better) speech-in-noise threshold for stop consonants varying in place and for fricative consonants varying in voicing. Magnitude of AMFR at slow AM rate (8 Hz) is also correlated with masking effect. This study suggests that AM processing relates to some extent to consonant processing in noise.

P07 Examining the effect of voice training on speech-on-speech intelligibility and listening effort in cochlear implant users

Ada Bicer, Thomas Koelewijn, Deniz Başkent
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, The Netherlands

Understanding speech in multiple-talker situations can be challenging for people from all hearing backgrounds. However, achieving good speech-on-speech intelligibility is especially challenging when speech is degraded, such as in communication via a cochlear implant (CI). Additionally, understanding speech in multi-talker listening situations can be more effortful for CI users, compared to normal hearing listeners. For normal hearing listeners, listening to familiar or lab-trained voices can be more intelligible in speech-on-speech situations compared to listening to unfamiliar voices. In addition to these intelligibility benefits, voice training might also lead to a reduction in listening effort for normal hearing listeners in speech-on-speech situations. However, sometimes voice training might lead to a listening effort benefit that is not accompanied by an intelligibility benefit, as shown by previous studies with normal hearing listeners. Therefore, the question of whether voice training can lead to a benefit in improving speech intelligibility and/or reducing listening effort for CI users, remains open. The aim of this study is to examine if an explicit voice training can provide a benefit in either speech intelligibility or listening effort, or both, which might improve the quality of life of CI users.

In this study, an explicit voice training was implemented, which involved a talker identification task. During voice training, half of the participants listened to 3 female voices and the other half listened to 3 male voices. As a task, participants selected who was speaking by clicking on the talker's name, for each sentence that was presented. In total, 600 sentences were used during voice training, while one voice was presented more frequently (400 sentences) that was later used as the trained voice. Following the voice training, speech-on-speech intelligibility was measured by means of a Dutch version of the Coordinate Response Measure (CRM) test. Listening effort was measured during the CRM test using pupillometry. The CRM test involved a target sentence that consists of a call sign, a color, and a number (e.g. "Show the dog where

the blue five is"), uttered by the trained voice and an untrained voice, while an unintelligible speech masker was presented in the background, at 0 dB and +6 dB target-to-masker ratios. Results from this ongoing study will be presented.

P08 Neural correlates of the McGurk illusion in age-related hearing loss assessed by fMRI and EEG

Stephanie Rosemann[1], Mareike Daeglau[2], Stefan Debener[2], Christiane M. Thiel[1]
1. Biological Psychology, Department of Psychology, Faculty of Medicine and Health Sciences, Carl-von-Ossietzky Universität Oldenburg, Oldenburg, Germany I Cluster of Excellence "Hearing4all", Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany | 2. Neuropsychology, Department of Psychology, Faculty of Medicine and Health Sciences, Carl-von-Ossietzky Universität Oldenburg, Oldenburg, Germany I Cluster of Excellence "Hearing4all", Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

Previous research investigating cortical plasticity in age-related hearing loss provides evidence for cross-modal reorganization in the auditory cortex, additional recruitment of the frontal lobe, and increased coupling of the visual and auditory cortices for matching audiovisual input. These changes already begin to occur when hearing impairment is only mild to moderate. In addition, hard of hearing people are more prone to the McGurk illusion than those with normal hearing. Using functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), we here investigated the influence of mild to moderate hearing impairment on neural correlates of the McGurk illusion. Twenty-eight participants aged between 50 and 80 years participated in the study. They either had normal hearing abilities or showed a uniformly varying degree of mild to severe and symmetrical age-related hearing loss. None of them had any experience with hearing aids. The measurements taken included the McGurk illusion in the MRI (sparse-sampling approach) and in the EEG (64-channels), along with an assessment of hearing abilities by a pure tone audiometry and speech reception thresholds. In the McGurk task, we presented the syllables 'ba', 'da', and 'ga' in an auditory only, visual only, audiovisual congruent, and audiovisual incongruent conditions. The latter is supposed to initiate the McGurk illusion, typically achieved by presenting a visual 'ba' and an auditory 'ga', leading to the illusory percept of 'da'. Analysis of the behavioral data indicate a mean fusion response (i.e. that participants perceived the McGurk illusion) of 70% (range 0-100%). However, there was no significant correlation between the number of perceived McGurk illusions and hearing loss. Neural data relating hearing loss, McGurk illusion perception and cross-modal reorganization of the auditory cortex measured by fMRI and EEG are currently analyzed.

P09 Pupillometry reveals adaption to linguistic interference over time during speech-in-speech listening

Alex Mepham[1], Ronan McGarrigle[2], Sarah Knight[1], Lyndon Rakusen[1], Sven Mattys[1]
1. University of York, UK | 2. University of Bradford, UK

Listeners face various challenges when listening to speech in a background of competing talkers. Behavioural studies of adaptation during speech-in-speech listening show that performance can improve over the course of a test block, with faster improvement in unintelligible compared to intelligible maskers. However, it is unclear whether this pattern is reflected in changes in listening effort.

In this study, native British English listeners undertook a 50-trial speech recognition task in two masker conditions: intelligible maskers (English talkers) and unintelligible maskers (time-reversed English talkers). In Experiment 1 (N=40), participants first undertook an adaptive procedure to find the 50% SRT for each condition. Participants thus started the main task at ~50% correct in both conditions. In Experiment 2 (N=40), the same SNR was used for both conditions (-1.5 dB). In Experiment 3 (planned N=40), participants undertook an adaptive procedure such that they started the unintelligible masker condition at ~41% correct and the intelligible masker condition at ~66% correct. These values represent the mean performance levels from Experiment 2, but switched such that the intelligible masker was presented at the easier SNR and vice versa.

In Experiment 1, performance started around 50% in both conditions as planned. Improvement was faster in the intelligible than unintelligible condition, contrary to our expectations. Pupil dilation decreased over time at the same rate for both conditions. In Experiment 2, performance was higher in the unintelligible than the intelligible condition. Unlike Experiment 1, improvement was faster in the unintelligible than the intelligible condition. Pupil dilation mirrored that pattern, with a faster decrease in the unintelligible than intelligible condition. The data of Experiment 3 are being analysed.

These results suggest that the ease with which listeners learn to stream a target from a competing talker depends on both the intelligibility of the masker and relative SNRs. In Experiment 2, faster transcription improvement and a more pronounced decrease in effort for unintelligible than intelligible maskers suggest that the linguistic interference created by an intelligible masker leads to persistent cognitive demands associated with ignoring the masker's linguistic content. However, as indicated by Experiment 1, this pattern might arise from an initial performance advantage in the unintelligible masker condition (due to a more favourable SNR) at the start of the task. Results from Experiment 3, in which the SNRs are adjusted to ensure an initial performance advantage in the intelligible masker condition, will help to disambiguate these two interpretations.

P10 The irrelevant speech effect: A comparison between normal-hearing and hearing-impaired people

Nicolas F. Poncetti, Etienne Parizet
Laboratory of Vibration and Acoustics (LVA), Lyon, France

Open-plan offices are the most common office layout in tertiary sector. Despite a noise level below the regulations (<60 dBA), occupants of these offices complain about noise. Among the various noise sources, co-worker's voices and conversations seem to be the most annoying one, as employee surveys have shown. Moreover, not all employees working in open-plan offices are young people with normal hearing. They can be older (up to 60 yrs.) and have any level of hearing loss. So, the purpose of this study is to investigate the effects of mild hearing loss (onset of presbycusis) on performance in an open-plan office – particularly under the influence of the irrelevant speech (the Irrelevant Speech Effect). An analysis of the decrease in performance in the accomplishment of a cognitive task (a serial recall test) regarding the speech intelligibility level was performed with young normal-hearing participants under two hearing-conditions: with and without a hearing loss simulator, as well as with elderly hearing-impaired participants. During the task, the participants were exposed to five noise-conditions, with five different levels of intelligibility (ranging from STI 0.35 to 0.75), and silence. Afterwards, a subjective intelligibility measurement was performed to compare the signals' intelligibility level for each hearing-condition.

P11 Using verbal response time as a measure of listening effort for adaptive tests of speech in noise assessment in clinical settings

Chiara Visentin[1], Chiara Bonora[2], Nicola Brunelli[3], Andrea Migliorelli[3], Laura Negossi[3], Pietro Scimemi[2], Rosamaria Santarelli[2], Andrea Ciorba[3], Nicola Prodi[1]
1. Department of Engineering, University of Ferrara, Italy | 2. Department of Neuroscience, University of Padova, Italy | 3. Department of Neuroscience and Rehabilitation, University of Ferrara, Italy

Objective: The aim of this study was to investigate the use of verbal response time (RT) as a measure of listening effort within clinical settings. This was carried out in the context of speech-in-noise evaluations using staircase adaptive procedures.

Methods: A total of 58 participants were involved in the study, who were divided into three groups based on age and hearing thresholds: young participants with normal hearing (YL_NH, n=36, mean pure tone average PTA=3.5 dB nHL), older listeners with normal hearing (OL_NH, n=14, PTA=14.1 dB), and listeners wearing hearing aids (HA, n = 8, PTA=51.3 dB). The speech-in-noise test was presented through two loudspeakers placed in front of the listener and consisted of sequences of four disyllabic words. It was carried out under four listening conditions, created by combining two reverberation conditions (anechoic, reverberant with Tmid = 0.56 s) and two background noise types (stationary, fluctuating). An adaptive staircase procedure was used to determine the speech reception threshold for 80% correct word identification.

During the staircase procedure, the verbal RT for each sentence was recorded, defined as the time elapsing between the audio offset and the onset of the participant's verbal answer. The RT in a specific listening condition was calculated as the median value of the last 10 trials. Following each listening condition, participants provided a rating of their perceived listening effort using a visual analog scale.

Results: Statistical analyses of RTs revealed interactions between population, reverberation condition and background noise. In conditions with stationary noise, RTs were longer in reverberant compared to anechoic conditions. In all listening conditions, YL_NH had shorter RTs compared to HA. Additionally, RTs were significantly longer for OL_NH compared to YL_NH, but only in anechoic conditions and in the presence of fluctuating noise. Self-ratings of effort disclosed a significant main effect of the reverberation condition (lower perceived effort in anechoic conditions) and population (higher effort for HA in comparison to NL_NH). A significant positive correlation between RTs and self-rated effort was only found within the HA group.

Conclusions: Our findings provide an indication of the potential utility of incorporating verbal RT as a measure for assessing listening effort in clinical settings, particularly for speech-in-noise tests administered using a staircase adaptive procedure. At a high-performance level, RT is sensitive to changes in the auditory environment (background noise, reverberation) and disclose effects beyond intelligibility for the HA group. The absence of a correlation between RTs and self-rated effort, along with their different sensitivity, reinforce the argument that the two measures tap into separate cognitive dimensions. These results have practical implications for the clinical practice, specifically in defining new protocols for HA best-fitting procedures.

P12 Dividing attention bilingually: The benefits and costs of spatial separation between talkers in two different known languages

Emily Rice, Sarah Knight, Angela de Bruin, Sven Mattys
University of York, UK

Energetic masking (EM) refers to spectrotemporal overlap between a to-be-attended target and a competing masker, disrupting the target at the auditory periphery. Spatial separation between target and masker reduces EM, increasing target intelligibility. This benefit is known as spatial release from EM (SREM). When listeners attend to both sounds rather than one, spatial separation reduces EM, but increases cognitive load, as listeners must shift attention between two spatial locations. Nevertheless, spatial separation still appears to improve performance, demonstrating that SREM's benefits outweigh cognitive costs. However, existing research has focused on monolingual participants presented with one language. It is currently unclear how these effects operate when two different, but known, languages are spoken simultaneously, which is a regular occurrence for bilinguals. Furthermore, it remains an open question how the costs and benefits of spatial separation differentially affect the first language (L1) and second language (L2).

In the current studies, unbalanced Spanish-English bilinguals completed a selective listening (SL, N=98) and/or a divided listening (DL, N=80) task. Results are reported here for the 66 participants who completed both tasks. On each trial, participants heard one Spanish (L1) and

one English (L2) sentence simultaneously, either collocated (diotic) or dichotic (one sentence in each ear). In the SL task, participants were told in advance which talker to transcribe. In the DL task, they were not told which talker to transcribe until after stimulus presentation, so had to listen to both to perform successfully. Working memory was measured using a visual Letter Number Sequencing task.

In both tasks, performance was better for the L1 than the L2 sentences and in the dichotic than the collocated condition. There was no interaction between language and spatial separation. Thus, SREM was beneficial for both languages, and neither language benefited from SREM significantly more than the other.

However, in the L2 condition, the SREM benefit was smaller in the DL than the SL task, potentially because the increased cognitive demand associated with attending simultaneously to two spatially separated talkers had a greater impact on transcription of the less proficient language. Furthermore, working memory scores were positively correlated with listening accuracy in the DL, but not SL task, supporting an increased need for cognitive control when processing two languages simultaneously.

P13 Investigating the relationship between hearing, speech, language and cognition in typically developing children as a first step to diagnosing auditory processing disorder

Xuehan Zhou[1], Harvey Dillon[2], Kelly Burgoyne[1], Dani Tomlin[2], Alisha Gudkar[2], **Antje Heinrich**[1]
1. University of Manchester, UK | 2. University of Melbourne, Australia

A range of deficits can cause children difficulty when understanding speech in challenging listening environments, such as noisy classrooms. Children with listening difficulties are at risk of having poor long-term academic outcomes and social skills, especially when clinicians cannot detect or remediate their specific deficits. Deficits in auditory, speech, language, or cognition abilities may present in a similar manner. Currently, it is difficult to determine the cause of these difficulties. A systematic approach to differentiate between these causes in individual children has been devised (Dillon & Cameron, 2021, doi:10.1097/AUD.0000000000001069).

Children aged 6-12 years, enrolled in mainstream primary schools, go through a tri-level test battery; a combination of top-level speech perception in noise and reverberation ability, mid-level phoneme identification ability, and low-level acoustic resolution ability is applied (Dillon & Cameron, 2021). In conjunction with language and cognitive test scores, the combined approach allows for differentiation of the cause of the observed listening deficit.

The present study aims to investigate the relationship between hearing, speech, language and cognition as a first step to diagnosing auditory processing disorder (APD). Speech-sound identification ability in noise and reverberation, non-speech auditory processing abilities, language abilities and cognitive abilities will be used to predict the understanding of sentences in noise and reverberation. A more comprehensive understanding of the underlying factors contributing to listening difficulties is critical for the development of customised efficacious interventions to avoid and alleviate long-term consequences.

P14 The Reading Span Task as a means to measure the predictive ability of various aspects of working memory to speech in noise perception

Antje Heinrich
University of Manchester, UK

In addition to the ability to hear, accurate speech-in-noise perception also requires contributions from cognitive abilities. The cognitive ability that is examined most often in the context of speech-in-noise (SiN) perception is working memory (WM). According to Baddeley and Hitch (1974, doi:10.1016/S0079-7421(08)60452-1) two components are critical for WM: a storage component and a manipulation component. WM tasks differ in how they combine these two components. One WM task commonly used in the context of speech perception is the reading span task (RST). Why the RST works so well to predict speech-in-noise perception performance is not well understood. It is also not well understood whether it is the storage or the manipulation component that has the highest predictive value, or whether both components combine to maximise the task's predictive value. Finally, it remains to be understood whether the predictive ability of the RSTs' two components differs either for different groups of listeners or for different speech-in-noise tasks.

We assessed performance of the storage and manipulation components of an RST and related them to the perceptual accuracy of several SiN tasks in two groups of normal-hearing young listeners: English native speakers and English non-native speakers. This is the first experiment in a programme aimed to fully understand the predictive ability of the reading span task for various groups off listeners and tasks and to explore the use of the RST for audiological practice.

P15 Listening effort ratings for habitual and clear-Lombard speech in noise as predicted by a glimpse (release from energetic masking) measure

Esther Janse, Chen Shen
Radboud University Nijmegen, the Netherlands

Martin Cooke
University of the Basque Country, Spain

Earlier research developed the high-energy glimpse proportion metric (HEGP, Tang & Cooke, 2016, doi:10.21437/Interspeech.2016-14) to capture those speech-dominant spectro-temporal regions, or glimpses, that survive energetic masking when speech is presented in noise. Though the HEGP metric has been shown to correlate with intelligibility across speech types and signal-to-noise ratios, its connection to subjective listening effort ratings remains, to our knowledge, unexplored. Given the inverse relationship between intelligibility and subjective effort ratings (e.g., Simantiraki et al., 2023, doi:10.3389/fnins.2023.1235911), HEGP scores can be expected to predict subjective effort ratings.

Following up on our earlier acoustic analysis of speech from multiple talkers, we collected listening effort ratings for the same speech samples through an online experiment. Specifically, ratings were collected from 230 young adult normal-hearing raters for the speech of 48 talkers in two speaking styles (habitual and clear-Lombard), presented in speech-shaped noise at an SNR of -6. Raters rated their listening effort to understand the sentence content of the utterance on a scale from 1 to 7 (1 representing 'not effortful at all to understand' and 7 'extremely effortful to understand'). For all utterances, HEGP was calculated for presentation in speech-shaped noise at -6 SNR. Sentence-level acoustic measures (articulation rate, F0 median and range, and spectral balance) for these utterances were also available.

Based on the obtained ratings, we ask the following research questions. RQ1: Does HEGP predict listening effort ratings, and if so, does it do so differentially for habitual and clear-Lombard speech? RQ2: If HEGP predicts listening effort ratings, do sentence-level acoustic measures (i.e., articulation rate, pitch range, and spectral balance) explain additional variance in listening effort ratings?

To address these questions, we set up an initial linear-mixed effects model predicting listening effort ratings from speaking style and HEGP, and their interaction. In a second model, we added the four acoustic measures as predictors to the initial model to investigate whether inclusion of these further improved model fit. In both models, we included random intercepts for Talker, Rater, and Sentence, as well as by-talker random intercepts for speaking style to account for talker differences in the size of their speaking-style difference.

Concerning RQ1, results from the initial and second model showed that HEGP predicted listening effort rating equally strongly across speaking styles. Concerning RQ2, the second model proved to have a better model fit than the initial model, with all acoustic measures, except F0 median, predicting effort ratings. These results suggest that release-from masking metrics, complemented by acoustic measures that partly reflect talkers' clear-Lombard speech adjustments, explain subjective listening effort in noise.

P16 Speech recognition with different target and masker voices in a speech-on-speech masking task in normal hearing listeners and cochlear implant users

Verena Müller, Emeline Cordary, Pauline Burkhardt, Ruth Lang-Roth
University of Cologne, Department of Otorhinolaryngology, Head and Neck Surgery, Cochlear Implant Center, Faculty of Medicine, Germany

Objective: Speech recognition in a competing talker situation is extremely challenging. While normal hearing (NH) listeners can separate two competing talkers by their voices, this is hardly possible for cochlear implant (CI) users. This is partly due to the CI's limited processing of two important voice cues, the fundamental frequency (F0) and the formant frequencies (Fn). The aim of the study was to determine in how far speech recognition in a competing talker situation is influenced by the target's and the masker's voice and if speech recognition alters depending on which voice is the target and which is the masker.

Methods: 15 adult listeners with NH and 16 listeners with a CI participated in the study. The German Oldenburg sentence test, a matrix sentence test, was used as the test material. The sentences have the structure "name-verb-number-adjective-objective". The original male voice was manipulated regarding its F0 and its Fn, to create a female and a child-like voice. Always two sentences were superimposed, with the three voices acting as both, target and masker talkers. Additionally, all three voices were presented against modulated speech-shaped noise. Listeners had the task to repeat the sentence which began with the name "Stefan", a common German first name. The target-to-masker ratio (TMR) in dB at which listeners understood 50% was measured. Target and masker stimuli were both presented from the front.

Results: NH listeners' speech recognition was worst when the talkers with the same voice were superimposed, followed by the condition when voices were presented against the noise maskers. Speech recognition was best when competing talkers differed in voices. CI users' speech recognition was worst in the competing talker conditions and best when presented against the noise maskers. Within the competing talker conditions it seems that the child's voice was a more efficient masker compared to the male and the female voice.

Conclusions: For NH listeners results reveal that it is not the voice that matters, but whether there is a difference between the voices to influence speech recognition. This is different for CI users whose results showed that a voice which is higher in frequency might be a more efficient masker, and a better trackable target respectively, than a voice which is lower in frequency.

P17 The role of periodicity in speech-on-speech understanding in normal-hearing and hearing-impaired listeners

Paolo A. Mesiano[1], Hamish Innes-Brown[1], Tobias May[2], Johannes Zaar[1]
1. Eriksholm Research Centre, Snekkersten, Denmark | 2. Technical University of Denmark, Lyngby, Denmark

Background: Understanding speech in the presence of one or multiple competing talkers is a challenging auditory task that occurs often in daily life. While normal-hearing (NH) listeners can perform this task successfully, hearing-impaired (HI) individuals encounter severe difficulties in understanding speech in such auditory scenarios. The periodicity information of the competing speech signals, which is connected to the characteristics of their fundamental frequency, can provide useful auditory cues for target-speech understanding, but it is unclear how hearing deficits interfere with the access to such cues. This study investigated how the periodicity information in target and interfering speech contributes to speech intelligibility in young NH and older HI listeners.

Methods: 10 NH and 30 HI listeners participated in a two-competing-voices experiment. The HI group was divided into two subgroups: 15 listeners affected by high-frequency hearing loss (HI1) and 15 listeners affected by both low- and high-frequency hearing loss (HI2). In the experimental stimuli, the periodicity information of target and/or masker signals was either fully available (natural speech) or removed using noise vocoding (vocoded speech). The stimuli were played through two frontal loudspeakers (one for each competing signal). HI listeners were provided with linear-gain amplification.

Results: NH listeners performed best when natural speech was masked by natural speech. Vocoding the target or the masking speech separately did not affect intelligibility significantly, but vocoding both target and masker signals reduced the performance, producing the highest speech reception thresholds (SRTs) overall. Compared to NH listeners, HI listeners showed overall worse performances (with HI2 being worse than HI1 in all experimental conditions) and larger variability across listeners. For both HI1 and HI2 listeners, with natural target speech, vocoding the masker negatively affected speech intelligibility, but not significantly. In contrast, vocoding the target worsened speech intelligibility significantly, with the highest SRTs measured when target and masker signals were both vocoded.

Conclusions: The obtained findings suggest that (i) the severity of (low-frequency) hearing loss is a predictor of speech-on-speech understanding, (ii) NH listeners are challenged only when the periodicity information is removed from both target and masker signals, and (iii) HI listeners rely mostly on the periodicity information in the target speech, while the presence of periodicity information in the masker is useful to them only when no target periodicity information is available. Further research may be directed at exploring potential strategies for enhancing the relevant periodicity information to improve speech intelligibility for HI listeners.

P18 Cortical representations of function versus content words while listening to speech in natural soundscapes at different levels of simulated hearing loss: An fMRI study

Leo Michalke[1], Arkan Al-Zubaidi[1], Esther Ruigendijk[2], Jochem W. Rieger[1]
1. Applied Neurocognitive Psychology Lab and Cluster of Excellence Hearing4all, Oldenburg University, Oldenburg, Germany | 2. Department of Dutch and Cluster of Excellence Hearing4all, Oldenburg University, Oldenburg, Germany

Listening to naturalistic auditory stimuli elicits changes in brain activity which represent processing of speech from simple acoustic features to complex linguistic processes. Here, we used complex linguistic features, namely the distinction between function and content words from a recording of listening to natural speech, to examine the differences in stimulus processing between clear and degraded conditions.

We recorded fMRI data of 30 healthy, normal-hearing participants listening to the audio description of the movie "Forrest Gump" in German (Hanke et al., 2014, doi:10.1038/sdata.2014.3). The audio movie was presented three times in three different recording sessions in eight segments with different levels of simulated hearing loss (CS – clear stimulus, S2 – mild degradation, N4 – heavy degradation). The degraded stimuli were produced according to Bisgaard et al. (2010, doi:10.1177/1084713810379609). In each session, participants listened to the whole movie but with a randomized sequence of stimulus degradation levels.

We used speech annotations for the audio movie provided by Häusler and Hanke (2021, doi:10.12688/f1000research.27621.1) to categorize each spoken word into word classes – function words, content words, and rest (hard to classify, like interjections). These categories were then used as regressors to predict BOLD time courses. We added a word duration regressor,

and orthogonalized the other regressors. This allowed us to obtain word class specific activation estimates unbiased by average word length differences between classes. In addition, we did another analysis where pronouns were split from function words into a separate regressor.

Our results indicate distinctive spatial patterns of BOLD activity in response to function versus content words. Function words elicited significant frontal and temporal activations, whereas content words elicited significant parietal activations. Pronouns displayed the same patterns as the other function words, but with larger effect sizes. Word duration explained a lot of the activity in the temporal lobe presumably related to early auditory processing. Most activations remained across stimulus degradation levels, but with smaller effect sizes. The highest degradation level N4 showed some additional effects – with a slight increase in motor cortex activity.

Our study reveals distinct but overlapping cortical representations of content and function words when listening to continuous speech in natural soundscapes with different simulated hearing capabilities. It also points out the importance of accounting for differences in average word length between word classes.

Funding: This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2177/1—Project ID 390895286.

P19 The effect of motor resource suppression on speech perception in noise in younger and older listeners: An online study

Kate Slade[1], Alanna Beat[1], Jennifer Taylor[1], Christopher J. Plack[1,2], Helen E Nuttall[1]
1. Lancaster University, Lancaster, United Kingdom | 2. University of Manchester, Manchester, United Kingdom

Background: Older adults with hearing loss experience difficulty understanding speech in background noise. Speech motor resources may be recruited to assist challenging speech perception in younger normally hearing listeners, but the extent to which this occurs for older adult listeners is unclear. We investigated if speech motor resources are also recruited in older adults during speech perception. Specifically, we investigated if suppression of speech motor resources via sub-vocal rehearsal affects speech perception compared to non-speech motor suppression (jaw movement) and passive listening. Sub-vocal rehearsal may suppress motor resources by occupying them, hypothetically impairing speech discrimination.

Methods: Participants identified words in speech-shaped noise at signal-to-noise ratios (SNRs) from -16 to +16 dB in three listening conditions during which participants: (1) opened and closed their jaw (non-speech movement); (2) sub-vocally mimed 'the' (articulatory suppression); (3) produced no concurrent movement (passive listening). Data from 46 younger adults (M age = 20.17 years, SD = 1.61, 36 female) and 41 older adults (M age = 69 years, SD = 5.82, 21 female) were analysed.

Results: Linear mixed effects modelling investigated the impact of age, listening condition, and self-reported hearing ability on speech perception (d'). Results indicated that speech perception ability was significantly worse in older adults relative to younger adults across all

listening conditions. A significant interaction between age group and listening condition indicated that younger adults showed poorer performance during articulatory suppression compared to passive listening, but older adults performed equivalently across conditions.

Conclusions: The findings suggests that speech motor resources may be less available to support speech perception in older adults, providing important insights for auditory-motor integration for speech understanding and communication in ageing.

P20 Reconstruction of speech from a few channels using a single-speaker neural vocoder

Josef Schlittenlacher, Shiyi Xu
University College London, UK

Background: Speech has been shown to be intelligible when it is noise-vocoded in only three or four channels. Because of this, neural vocoders should in theory be able to generate high-quality speech from such small spectral representations when phonetic content is the only variable that changes during training but other attributes like speaker identity or speaking style are kept constant.

Method: In the present study, WaveGlow, a neural vocoder based on normalizing flows, was trained on Mel spectrograms with two, three, four, eight, or the typically used eighty channels on the LJ-Speech corpus. Twenty naive participants evaluated these sounds produced by WaveGlow and sounds processed by a noise vocoder with the same numbers of channels in an online experiment. They rated speech quality on a five-category category scale (excellent, good, fair, bad, poor) and reported the percentage of words recognized in a second run. Each condition was presented eight times in each run, the assignment of speech segments to conditions was randomized between participants and no speech segment was repeated within a participant.

Results: WaveGlow produced sounds with considerably higher sound quality and intelligibility than noise vocoders with the same number of channels in subjective and objective ratings. The participants rated the sound quality of the sounds produced by WaveGlow one to two categories higher than those produced by the noise vocoder, and reported to recognize 48 % of the words for two channels and 81 % for three channels for WaveGlow, but only 74 % for a noise vocoder with eight channels. The Short-Term Objective Intelligibility (STOI) metric showed a similar pattern than the mean opinion scores for sound quality.

Conclusions: Altogether, this shows that the neural vocoder successfully learned some general features of speech that are useful for naive listeners. However, the sounds produced by WaveGlow based on eight channels were only of “fair” quality despite near-perfect intelligibility and twice the training time as that for eighty channels, which may not make the reduced representation sufficient for use in text-to-speech systems.

P21 Evaluation of chosen DNN-based speech enhancement methods using Polish matrix speech intelligibility test

Szymon Drgas
Poznan University of Technology, Poland

Most of DNN-based speech enhancement algorithms are evaluated only using objective metrics (e.g. STOI or PESQ), or subjective quality assessment (as in deep noise suppression, DNS, challenge; Dubey et al., doi:10.48550/arXiv.2303.11510). Although previous studies showed the benefits that come from the DNN-based speech enhancement, it is not known how strong the effect coming from their use would be observed in standard speech audiometry tests.

In this work, the Polish sentence matrix test (Ozimek, et al., doi:10.3109/14992021003681030) was used to assess the chosen recent DNN-based speech enhancement methods. Publicly available pre-trained neural networks (Conv-TasNet, deep complex U-net, dual-path transformer network) were evaluated. The tests determined speech reception thresholds (SRTs), i.e. the SNRs of the speech mixed with babble noise, that after speech enhancement, give the intelligibility of 50%. The listening tests were completed by 20 participants with normal hearing. No fine-tuning of the tested neural networks to the specific conditions (language, speaker and noise) was done.

The algorithms were compared with respect to the measured speech reception thresholds and their computational and memory requirements. Additionally, a signal analysis was performed to compare the types of artifacts produced by the tested neural networks and objective speech intelligibility measures: STOI and HASPI.

P22 From 0 (dB) to 15 (dB) in 13 (years): Extended developmental trends in the spatial release of masking for highly informational maskers

Stuart Rosen[1], Shiran Koifman[1,2]
1. UCL Speech, Hearing & Phonetic Sciences, London, UK | 2. Carl von Ossietzky University of Oldenburg, Germany

There is growing evidence that spatial release of masking (SRM) develops over a much longer period through childhood and adolescence than had previously been suspected. Even so, children as young as 5 years old have shown at least some SRM. We have been investigating SRM using a modified version of the Coordinate Response Measure (CRM), the so-called Children's CRM (CCRM). Here all sentences, both targets and maskers, are in the form of commands 'Show the [animal] where the [colour] [number] is', with 6 animals, 6 colours and 8 numbers, all monosyllabic. Sentences are uttered by three male British English talkers. In each trial, a target sentence ('Show the dog …') from a randomly selected talker is presented simultaneously with two masker sentences, one from each of the other talkers, with no overlap in animal, colour or number. In the collocated condition, all talkers are placed to the front of

the listener in virtual space over headphones. In the spatially separated condition, the two masker sentences are placed symmetrically at ±45° to the participant's right and left. Listeners are required to click on an interface with 48 buttons to indicate the colour and number of the target sentence. Speech Reception Thresholds (SRTs, the SNR which results in ~50% correct identification of the colour and the number) are determined adaptively. SRM is calculated as the difference between the two conditions. Because all talkers are of the same sex (and with a similar F0), and masker sentences contain valid target words, this situation is one of high Informational Masking (IM), because there are few cues for a listener to use to attend to the correct target, especially in the collocated condition.

We have applied these tests to children aged 5-18, with two striking outcomes. First, in the collocated condition, there is very little change in SRT over age, ~2.6 dB on average over the 13-year age span of the children, with the mean SRT always at positive SNRs. Presumably, listeners are only able to use a loudness cue to focus on the target. Secondly, 5-year-olds displayed no SRM at all, with a clear non-zero value only occurring from age 7.5 years. The oldest children obtained ~15.5 dB release from masking with no indication of a plateau in performance even at age 18. It seems likely that this inability to use spatial separation to improve performance relates to immature executive functions, in particular, attention.

Acknowledgments: Thanks so much to Bethany Davison and Katie Wilkins who ran these studies as part of their MSc project in Speech & Language Sciences at UCL.

P23 Stable individual differences in audiovisual benefit across three levels of linguistic structure: Exploring the role of cognitive and perceptual abilities

Jacqueline von Seth, Máté Aller, Matthew H. Davis
MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom

Individuals differ substantially in their ability to use visual cues encoded in the speaker's facial movements to enhance speech perception. We recruited 114 normally-hearing participants to complete a three-part online experiment investigating the distribution, reliability, and predictors of individual audiovisual benefit for acoustically degraded speech. Rather than measuring changes in intelligibility due to adding visual speech, we measured the relative intelligibility of matched auditory-visual (AV) and auditory-only (AO) speech for materials at three levels of linguistic structure: meaningful sentences, monosyllabic words and isolated consonants. Participants performed speech identification tasks in AO, AV, and VO conditions in two sessions, set one week apart. Acoustic clarity was manipulated using a noise-vocoding procedure to create two levels of degradation. We matched report accuracy in high-clarity AO and low-clarity AV conditions by varying the mixing proportion of unintelligible 1-channel and intelligible 16-channel vocoded speech. In a third session, participants completed a battery of tests to assess their hearing, linguistic and cognitive skills.

We found that audiovisual benefit is stable (test-retest reliable) across experimental sessions using different item sets and significantly correlated across levels of linguistic structure and speakers. Multiple linear regression showed that individual differences in audiovisual benefit were explained by perceptual, rather than cognitive abilities. While auditory-only perception was predicted by verbal IQ, audiovisual benefit was predicted by better lipreading ability and relatively poorer hearing (estimated using the Digits-in-Noise test). These results represent the first step towards validating a short-form intelligibility-matched measure of individual audiovisual benefit and provide a foundation for further work assessing cognitive and neural (MEG) correlates of enhanced audiovisual speech benefit.

P24 Effects of sensorineural hearing loss on subcortical coding of speech in noise: A model with level- and fluctuation-driven efferent gain-control

Laurel H. Carney
University of Rochester, Rochester, New York USA

Daniel R. Guest
Universtiy of Rochester, Rochester, New York, USA

Afagh Farhadi
Purdue University, West Lafayette, Indiana USA

A recent model for subcortical neural coding of speech by neural fluctuations (NFs) is robust over the range of sound levels used in conversational speech and also in background noise. The responses of AN fibers tuned near spectral peaks (e.g., formants, or the broad high-frequency spectral peaks associated with fricatives) have small NFs, and AN fibers tuned in spectral valleys or on slopes have deep NFs. Thus, the spectral envelope is represented by the profile of NF depths along the tonotopic axis. The NF code is facilitated by cochlear tuning and saturating peripheral nonlinearities that result in capture (or dominance) of inner-hair-cell (IHC) and AN responses by energy near spectral peaks. The NF code is translated to a rate code at the level of the midbrain by fluctuation-sensitive neurons in the IC.

The operating points of the saturating peripheral nonlinearities depend upon cochlear sensitivity. Therefore, decreased cochlear gain due to sensorineural hearing loss (SNHL) reduces capture and decreases the contrast in the NF profile. As a result, SNHL degrades the representation of the spectrum at the level of the IC. We will illustrate subcortical speech responses for normal hearing and SNHL using a computation model for the AN that includes nonlinear cochlear tuning and saturating nonlinearities associated with IHC transduction and the IHC-AN synapse. We hypothesize that efferent cochlear gain-control pathways are critical for maintaining and enhancing NF contrast over a wide range of sound levels and of signal-to-noise ratios. We are exploring this hypothesis using a model with efferent pathways that convey both level-driven signals from the cochlear nucleus and fluctuation-driven signals from

the inferior colliculus to the medial olivocochlear cochlear (MOC) gain-control system. Key parameters of the efferent model that are currently under investigation are the bandwidths of the MOC-to-cochlea projections that are driven by energy vs. fluctuations.

Funding: Supported by NIH-DC-R01010813.

P25 An auditory loudness model with hearing loss

Lars Bramsløw
Eriksholm Research Centre, Oticon A/S

Auditory models have gained acceptance and wide use for modelling and understanding hearing and hearing loss. They can be used for qualitative analysis and interpretation, but certainly also for quantitative analysis and simulation of sound perception of any sound source. An interesting new application of auditory models with hearing loss are 'closed-loop' approaches, whereby deep learning is used to directly derive the appropriate hearing loss compensation by comparing a normal hearing and a hearing-impaired branch of the auditory model.

The 'AUDMOD' model presented here models masking patterns, specific loudness, and total loudness for any stimulus and hearing loss, as specified by a standard audiogram. It is a strictly perceptual / behavioural model, based on the Moore & Glasberg roex-filter model, combined with the Zwicker & Fastl loudness model. The auditory filter bandwidth depends on both signal level and hearing loss, both leading to increased upward spread of masking. The computational requirements are very low, and the simple model is differentiable, both making the model suitable for deep learning applications, including closed-loop hearing loss compensation. Furthermore, the model can be applied in a straightforward manner for analysis and interpretation of hearing aid functionality. In the future, AUDMOD will be added to the Auditory Modelling Toolbox, https://www.amtoolbox.org/.

The model output is compared to experimental data for classical psychoacoustics and to a branch of the Glasberg & Moore model. Furthermore, a few application examples are shown, related to hearing aid signal processing as well as speech perception, for both normal hearing and hearing loss.

P26 The relationship between hearing loss and autonomic nervous system activity during speech perception

Adriana A. Zekveld[1], Laura Keur-Huizinga[1], Nicole A. Huizinga[2], Niek J. Versfeld[1], Sjors R.B. van de Ven[2], Wieke A.J. van Dijk[1], Eco J.C. de Geus[2], Sophia E. Kramer[1]
1. Amsterdam UMC, location VUmc, The Netherlands | 2. VU University Amsterdam, The Netherlands

Previous studies indicated that the pupil dilation response to speech perception, associated with listening effort, is altered in individuals with hearing loss. Here, we aimed to assess the relationship between hearing loss and autonomic nervous system activity using cardiovascular and skin-conductance measures as well as pupillometry. We measured respiratory sinus arrhythmia (RSA) related changes in heart-rate variability, pre-ejection period of the left ventricle of the heart (PEP), skin conductance level (SCL) and the peak pupil dilation response during listening. RSA is sensitive to changes in parasympathetic activity, PEP and SCL are influenced by sympathetic activity, and the pupil response is influenced by both autonomic nervous system components.

Data of 125 participants were analysed (mean age 58 years, range 37-72 years). Pure-tone hearing thresholds (weighted pure-tone average of both ears at 500, 1000, 2000, and 4000 Hz, taking into account asymmetrical hearing loss) ranged from normal hearing to moderate levels of hearing loss. Participants performed adaptive speech reception threshold (SRT) tests targeting either 20%, 50% or 80% sentence intelligibility. Target sentences were pronounced by a female talker and masked with male interfering speech. Frequency-specific amplification of the stimuli was applied to increase audibility. Physiological measures (see above) and subjective ratings of effort, performance, difficulty and giving up were collected. We applied linear mixed effect models to test the effect of intelligibility (3 levels) and hearing loss on the dependent measures, controlling for age, sex, BMI, and tinnitus complaints.

More severe hearing loss was associated with worse SRTs and smaller peak pupil dilation. The peak pupil dilation and PEP reactivity were largest for 50% intelligibility as compared to both 20% and 80% intelligibility. No effect of intelligibility or hearing loss was observed on RSA or SCL. Intelligibility influenced all subjective ratings in the expected direction. Also, more severe hearing loss was associated with higher effort ratings, particularly in the 80% intelligibility condition.

The results replicated previous effects of intelligibility and hearing loss on the pupil dilation response. Also, changes in intelligibility level did increase sympathetic activity as reflected by PEP, specifically in the 50% intelligibility condition. RSA and skin conductance measures were not sensitive to hearing difficulties evoked by the current conditions in this sample. Speculatively, the associations between hearing loss and the pupil response might be related to more general effects of fatigue or motivation, as the subjective ratings confirmed relatively effortful listening, especially for individuals with more severe hearing loss.

P27 Development of speech perception in noise: Effect of auditory scene analysis and musical abilities

Elena M. B. Benocci, Axelle Calcus
Université Libre de Bruxelles

From bars to business meeting, we are accustomed to engaging in conversation within noisy environments. Yet children and adolescents encounter more challenges than adults when it comes to understanding speech in noise. Speech intelligibility in noise is likely influenced by auditory segregation, an aspect of auditory scene analysis that remains immature at least until late childhood. Notably, musical abilities also seem to play a role in the development of speech perception in noise. The aim of this study was to investigate the respective contribution of stream segregation and perceptual musical abilities on the development of speech intelligibility in noise. In the present study, we recruited children (n = 80), adolescents (n=40) and adults (n=80) non-musicians with varying levels of perceptual musical abilities. Participants performed a stochastic-figure ground discrimination task which evaluates auditory segregation and a consonant identification in noise task. Using structural equation modelling, we observed a developmental improvement in auditory segregation and speech intelligibility in noise. Moreover, the relationship between musical abilities and speech perception in noisy environments appeared to be mediated by auditory segregation. Additionally, our results suggest a developmental improvement of the mechanisms of auditory scene analysis and speech intelligibility in noise, from childhood to adulthood. These observations are in line with recent views stating that music may have emerged as a cultural creation relying upon preexisting adaptations for auditory scene analysis.

P28 Neural processing of degraded speech under divided attention: An fMRI – machine learning study

Han Wang, Patti Adank
Department of Speech, Hearing and Phonetic Sciences, University College London

Understanding spoken language often occurs under suboptimal listening conditions, such as processing degraded speech input or conversing in the presence of a competing talker. Neuroimaging studies suggest a role of frontal regions in compensating for degraded speech (Erb et al., 2013, doi:10.1523/JNEUROSCI.4596-12.2013) and that of cingulo-opercular attentional network in allocating resources between different tasks under distraction (Gennari et al., 2018, doi:10.1016/j.neuroimage.2018.06.035). However, no studies so far have revealed the combined effects of acoustic degradation and distraction on the speech processing network.

Using functional magnetic resonance imaging (fMRI) and machine learning (ML), we investigated the neural basis of processing degraded speech under divided attention. We examined brain responses of listeners (N=25) performing a sentence recognition task (4- and 8-band noise-vocoding) concurrently with a visuospatial task with two difficulty levels. Traditional general linear modelling (GLM) based fMRI-analysis revealed intelligibility-related responses

in the frontal and cingulate cortices and bilateral insulae but failed to detect neural correlates of visual-task difficulty. Using gradient tree boosting algorithms (Chen & Guestrin, 2016, doi:10.1145/2939672.2939785), we predicted task conditions from brain responses with high accuracy (60%) and significantly surpassing chance level (25%). Importantly, the algorithm further identified more elevated response in the ventral visual pathway (right inferior frontal gyrus) and right insula when the visual task imposed a high (compared to low) demand. Moreover, regions found sensitive to speech intelligibility including left supplementary motor and right middle frontal regions showed alleviated responses under a hard visual task, insinuating dynamic resource dispensing across the two tasks.

Our results unveiled the engagement of the frontal-temporal network in the processing of degraded speech under divided attention. These results suggest that ML can detect spatially complex and subtle non-linear neural activation patterns that are otherwise hidden by inferential statistics. The tree-based approach offers a robust and generalisable solution to predictions using fMRI data, whose sample size is often subject to external constraints.

P29 Predicting the effect of hearing loss on speech intelligibility using a physiologically inspired auditory model

Johannes Zaar
Eriksholm Research Centre, Snekkersten, Denmark

Laurel H. Carney
University of Rochester, Rochester, New York, USA

Background: Several speech-intelligibility (SI) prediction models have been developed for application to a large range of acoustical conditions. Most of the available models were designed and validated based on reference data for normal hearing (NH), for which individual differences in SI are typically small and for which simplistic linear simulations of auditory processing can provide sufficient predictive power. However, although many of these models have later been extended to incorporate aspects of hearing loss, it has remained challenging to accurately predict SI differences between listener groups with NH and hearing impairment (HI) and even more challenging to predict within-group individual differences in the HI population.

Methods: We recently introduced an SI prediction model (Zaar and Carney, 2022, Hear. Res. 426:108553) based on the recently proposed hypothesis that across-frequency fluctuation profiles in auditory-nerve (AN) responses are relevant for discrimination of complex sounds. A phenomenological model provided simulated AN responses for NH and individual HI listeners. The model is here evaluated using several data sets consisting of auditory profiling data and speech reception thresholds (SRTs), measured in a range of noise conditions, all collected in both NH and HI listeners. The model was calibrated using NH data obtained in a single noise condition; predictions were then obtained as a result of differences in the stimuli (different noise conditions) and by incorporating pure-tone thresholds and estimates of outer and inner hair cell (OHC and IHC) impairment into the auditory model (different HI listeners). Special attention was paid to the interpretation of pure-tone thresholds in terms of the underlying OHC and IHC contributions.

Results: The model accounted very well for SI across noise conditions in the NH group and accurately predicted the elevation of SRTs and the reduced masking release due to hearing loss. The measured and predicted SRTs for the HI listeners were strongly correlated for one data set and moderately correlated for another, smaller, data set. The model predictions for the HI listeners were strongly dependent on the interpretation of the pure-tone thresholds with respect to the underlying OHC and IHC contributions. However, individualization of OHC/IHC impairments based on loudness-scaling data did not increase predictive power.

Conclusions: The results indicate that the proposed model accounts well for effects of additive noise and hearing impairment on SI. The differential effects of OHC and IHC impairment in the model warrant further investigation – while individualization based on loudness-scaling estimates was not successful, other diagnostic measures may yield better results.

Funding: LHC was supported by NIH DC010813.

P30 Predicting supra-threshold speech reception deficits using the Audible Contrast Threshold test

Johannes Zaar[1], Peter Ihly[2], Takanori Nishiyama[3], Søren Laugesen[4], Sébastien Santurette[5], Chiemi Tanaka[6], Gary Jones[7], Marianna Vatti[5], Daisuke Suzuki[8], Tsubasa Kitama[3], Kaoru Ogawa[9], Jürgen Tchorz[2], Seiichi Shinden[8], Tim Jürgens[2]

1. Eriksholm Research Centre, Snekkersten, Denmark | 2. University of Applied Sciences Lübeck, Lübeck, Germany | 3. Keio University School of Medicine, Tokyo, Japan | 4. Interacoustics Research Unit, Kgs. Lyngby, Denmark | 5. Oticon A/S, Smørum, Denmark | 6. Oticon Japan, Kawasaki, Japan | 7. Demant A/S, Smørum, Denmark | 8. Saiseikai Utsunomiya Hospital, Utsunomiya, Japan | 9. OTO Clinic Tokyo, Tokyo, Japan

Background: The pure-tone audiogram is the main clinical diagnostic used for assessing hearing loss and provides the basis for the hearing-loss compensation applied in hearing aids. However, the audiogram does not necessarily reflect the hearing deficits that remain when audibility has been restored, for instance the crucial ability of individuals to understand speech in adverse conditions. These supra-threshold speech reception deficits can be measured using speech tests, but it has proven challenging to test speech reception directly in clinical settings due to limitations with respect to equipment, testing time, and standardized speech materials. A clinically viable test that is connected to supra-threshold speech reception deficits would thus represent a highly useful addition to the clinical assessment of an individual's hearing abilities.

Methods: The present study assessed to what extent the Audible Contrast Threshold (ACTTM) test, a novel quick-and-simple clinical spectro-temporal modulation detection test with built-in audibility compensation, can predict supra-threshold speech reception in noise. One hundred eight hearing-impaired participants, consisting of 81 native speakers of German and 27 native speakers of Japanese, participated in the study. The audiogram and ACT were obtained along with speech-reception thresholds (SRTs). SRTs were measured with the participants using hearing aids in a challenging setting with spatially distributed speech interferers. Four different hearing-aid settings were tested: amplification only, mild directionality and noise reduction (DIR+NR), medium DIR+NR, and strong DIR+NR.

Results: On the group level, SRTs were highest for the amplification-only setting and decreased with increasing levels of DIR+NR processing. The individual SRTs collected with amplification only were strongly correlated with ACT and – to a lesser extent – with the 4-frequency pure-tone average (PTA4). The predictive power of ACT and PTA4 was found to be complementary, as they both contributed significantly to predicting the amplification-only SRT in a two-predictor linear regression model. Furthermore, the two measures were also associated with the individual SRT benefit induced by the DIR+NR processing.

Conclusions: The results indicate that the ACT test yields a measure of spectro-temporal modulation (or audible contrast) sensitivity that is predictive of aided speech reception in a realistic environment. The ACT yielded better SRT predictions than the audiogram while also adding significantly to the predictive power of the audiogram. This suggests that the ACT indeed represents a clinical measure that predicts supra-threshold speech reception deficits, which may be used to complement the information obtained from the audiogram in the clinic. More research is needed to identify meaningful interventions for individuals with substantial supra-threshold speech reception deficits and to assess said deficits in different populations of listeners, including those with audiometrically normal hearing.

Funding: This project was funded by the William Demant Foundation (20-2461).

P31 Reduced speech understanding in walking-noise with and without hearing loss compensation

Lena Eipert, Kaja Strobel, Julia Warmuth, Ronny Hannemann
WS Audiology

Background: Hearing loss is linked to frailty, walking difficulties, and an increased risk of falling. Difficulties in walking, i.e., a reduced walking speed, decreased step length, as well as a higher variability in stride time, were already described for hearing impaired subjects. The interaction of this changed motion performance, i.e., it could also be considered as noise, on speech understanding remains unclear. Dedicated speech tests are usually performed with listeners seated in an acoustic laboratory with highly controlled conditions, although daily life often challenges listeners with additional tasks while perceiving speech. Previous investigations in virtual reality walking setups testing speech understanding in noise or speech background proofed the impact of physical load on cognitive resources and thus, speech perception in acoustic complex environments.

Rationale: Thus, we here investigated the effect of walking on the perception of monosyllabic words either with or without hearing loss compensation with hearing aids.

Methods: Speech perception thresholds, i.e., %-correct, for monosyllabic words (Freiburger speech test) were collected in a classical audiological laboratory setup with subjects seated in front of a speaker as well as in a dual task guided walk with subjects carrying the speaker for testing in a backpack. All listeners (n = 24) were hearing impaired and completed both tasks wearing hearing aids either without or with an amplification compensating for their individu-

al hearing loss. Hearing aids were equipped with integrated Inertial Measurement Unit (IMU) sensors. Data collections were App-based, and gait patterns were extracted from accelerometer data using the ear gait package (https://pypi.org/project/eargait/).

Results: For all listeners speech perception thresholds significantly declined during the dual task compared with the lab situation either with or without compensated hearing impairment. While speech perception thresholds improved for all listeners in both situations with compensated hearing loss, thresholds improved less for listeners during the dual task than during the laboratory testing. Further, gait patterns revealed longer stride times, decreased step length, and reduced gait velocity comparing walking with completing the dual task, i.e., walking and testing speech perception thresholds.

Conclusion: Motion activity such as walking appears to demand additional cognitive load while seeking to understand speech resulting in less beneficial hearing aid compensation compared to focused and undistracted listening. Further, besides speech perception thresholds, also gait patterns deteriorate to a state associated with walking difficulties. In conclusion, hearing aid compensation success not only depends on the hearing impairment of an individuum but also the cognitive load in addition to speech understanding.

P32 Effects of task-irrelevant whispered speech on short-term memory

Florian Kattner, Cosima M. A. Stokar von Neuforn, Patrizia F. Scholz
Health and Medical University, Potsdam, Germany

Lia Downing, Julia Föcker
University of Lincoln, Lincoln, UK

Irrelevant background speech is known to disrupt cognitive performance either by interfering with specific processes (e.g., serial rehearsal) or through attentional capture. Here, two experiments are presented to test the disruptive effect of whispered background speech in a serial recall task. According to an interference-by-process account, whispered speech should be less disruptive due to its reduced amplitude modulations and temporal fine structure compared to loud speech, thus providing weaker order cues to the auditory system and inducing less interference with seriation processes. However, due to the enhanced listening effort to process whispered speech (when presented in a comprehensible language), it could be argued that additional attentional resources will be demanded thus producing more disruption of serial recall. In Experiment 1, to-be-remembered letters were presented visually on the screen while either voiced or whispered to-be-ignored speech was presented via headphones. Half of the speech trials contained only a single German word that was played repeatedly (steady-state), whereas the other trials contained a full German sentence. Memory accuracy was significantly lower in the presence of sentential speech compared to steady-state words, indicating interference-by-process. More importantly, whispered speech produced more disruption compared to loud speech, indicating an additive effect of attentional capture potentially due to enhanced listening effort required to process semantic information. To test specifically whether the disruptive effect of whispered speech was due to enhanced listening effort, the same German speech materials were presented to a sample of English participants who did not

speak or understand German in Experiment 2. It was found that in this case whispered speech was less disruptive than loud speech, indicating that psychoacoustic speech properties may dominate when no listening effort is demanded as in case of an incomprehensible language.

P33 An experimental setup to measure cochlear implant output of ecological stimuli

Floris Rotteveel, Bert Maat, Deniz Başkent
University Medical Center Groningen, The Netherlands

Etienne Gaudrain
Lyon Neuroscience Research Center, CNRS UMR5292, Inserm U1028, Université Lyon 1, Lyon, France

Researchers interested in the exact stimulation presented by a cochlear implant (CI) require an experimental setup that can measure clinical CI processor output. This work describes such a setup and aims to provide all the information and code needed to recreate a similar one. Additionally, the setup allows precise adaptation of realistic personalized stimulation tables retrieved from ecological sound stimuli, to be used in experiments with a research interface.

Since the coding from acoustic signal to stimulation tables is highly complex in modern processors, it can be difficult for researchers to understand what signals are being presented to the CI recipient listening to complex stimuli such as speech. Computational models for relating sound input to processor output exist, such as BEPS+ by Advanced Bionics, and the Nucleus Matlab Toolbox (NMT) by Cochlear. However, these models can prove insufficient in several ways: (1) they are not updated as often as clinical processors, (2) apply simplifications to decrease computational time, (3) lack a brand-neutrality needed to compare between processing of different manufacturers, and (4) make it difficult to recreate inputs that incorporate the multiple directional microphones of modern processors. The experimental setup described here overcomes these difficulties by allowing a researcher to measure the output of a clinical CI processor with all its bells and whistles.

The setup is a chain of devices which starts and ends at a laptop running Python. Inside a sound treated room, a loudspeaker plays stimuli to the CI processor, which is set up in an ecological fashion by attaching it to the ear of a KEMAR dummy. An implant-in-a-box retrieves CI signals from the processor and, through a load board, transmits the voltage signals of each channel to two synchronized oscilloscopes.

An example experiment is used to illustrate the use of the system. Here, the voice pitch (F0) of recorded Dutch spoken syllables was adapted, and the availability of F0 pitch cues in the CI output was studied using the experimental setup.

The voltage output was processed to infer pulse timing and current information. From here, analyses were done to retrieve information related to pitch, such as spectral centroid, amplitude modulation, and salience. This output was compared to the outputs generated by BEPS+ and NMT.

P34 An interactive evaluation of gaze-directed beamforming in noisy conversations.

John F. Culling[1], Emilie D'Olne[2], NIamh Powell[1], Bryn D, Davies[1], Patrick Naylor[2]
1. Cardiff University, UK | 2. Imperial College London, UK

Gaze-directed beamforming has potential for future hearing aids by selecting sound from the direction of the user's gaze. Gaze control is intended to allow users to redirect the beam to the current speaker, but its overall effectiveness in continuous conversation with natural exchanges of the conversational floor has not been evaluated. We evaluated the effectiveness using a simulation of an 8-microphone array mounted on spectacles. In phase 1, participants watched 6 segments (165 seconds each) of a zoom call between two parties located at ±15˚. Interfering talkers were added in at 0˚, ±60˚, ±135˚ and 180˚. One condition simulated binaural hearing over headphones using head-related transfer functions. The other condition used an eye-tracker to select filters for each source and for each video frame from a look-up table. The table was based on the predicted directionality of the microphone array in diffuse noise for a minimum-variance distortionless response beamformer. The participant's gaze and the headphone output were recorded. Participants completed a short questionnaire about the conversation after listening to each segment. In phase 2, the intelligibility of each sentence was measured formally. Each individual participant's recording was presented to a new participant, sentence-by-sentence for transcription. An individually tailored video cut back and forth between the two talkers in accordance with the eye-tracking record of the corresponding phase-1 participant in order to provide identical lip-reading cues. In phase 1, the scores on the questionnaire were approximately doubled by use of the beamformer and, in phase 2, the number of words correctly transcribed also doubled.

P35 To appear or to disappear? Investigating change detection in hearing-impaired listeners

Michelle Kosminski[1,2], Maja Serman[2], Sascha Bilert[2], Ulrich Hoppe[1]
1. FAU Erlangen-Nürnberg, Germany | 2. WS Audiology, Germany

We live in a world where different, complex stimuli constantly compete for our attention. The ability to perceive the appearance or disappearance of a sound (i.e., a change) in a busy environment is crucial for us to feel safe, especially in unfamiliar environments, as we often hear a change before we see it.

Change detection in the auditory domain has so far been investigated mainly with normal-hearing listeners. Brungart et al. were among the few who investigated this ability with normal hearing, as well as unaided and aided hearing-impaired subjects. The task was to identify and localize an appearing or disappearing sound in an auditory scene consisting of multiple everyday sounds, presented in the front hemisphere. The authors show evidence that hearing-impaired listeners perform better unaided than when using their own hearing aids (HA).

We developed a similar experiment, in which we first reproduced Brungart et al.'s results with normal-hearing subjects. However, our experimental design was expanded to include sounds appearing and disappearing from the back hemisphere, thus creating a more realistic setup. This experimental design was then used in a study with 14 hearing-impaired listeners to investigate the influence of hearing loss and the impact of hearing aid processing on change detection abilities. The trials in the experiment consisted of four different everyday sounds played simultaneously, with one of these sounds either appearing or disappearing halfway through the short trial. The sounds in the scene were presented either from the front, back or both hemispheres. Subjects performed the experiment unaided and with two types of HA processing. The task was to identify and localize the appearing or disappearing sound. Additionally, subjects completed a questionnaire assessing their real-life hearing loss problems so as to explore the relevance of change detection in everyday life.

Consistent with previous findings, the subjects performed significantly better in the "appear" change trials compared to "disappear" change trials regarding both sound identification and localization. However, no significant differences were observed between unaided or aided performance. Significant differences were found between aided performance in the sound identification task with two types of HA processing. Additionally, a significant correlation between one of these HA processing and the subjective assessment of real-life hearing loss difficulties points towards the importance of change detection in daily lives of aided hearing-impaired listeners.

P36 Principal Components Analysis of amplitude envelopes from spectral channels: Comparison between music and speech.

Agnieszka Duniec, **Olivier Crouzet**, Elisabeth Delais-Roussarie
Laboratoire de Linguistique de Nantes - LLING / UMR6310, Nantes Université / CNRS, France

Keywords : perception, cochlear implants, natural signal statistics, efficient coding

Introduction: The efficient coding approach predicts that perceptual systems are optimally adapted to natural signal statistics. Sensory system would have evolved to encode environmental signals in order to represent the greatest amount of information at the lowest possible resource cost.

Previous studies applied Factor Analysis (FA) on amplitude modulations channels from natural speech signals in order to estimate optimal frequency boundaries between channels. While some authors argued that 4 channels would be sufficient to represent the main contrastive segmental information in natural clean speech, comparison of speech statistics with perceptual performance led to suggest that 6 to 7 frequency bands would be required to optimally represent vocoded speech.

However, research on music perception in cochlear implanted listeners sheds light on potential limits associated with this hypothesis. Indeed, performance observed on vocoded signal material in normal-hearing listeners as well as in cochlear implant users is systematically better for speech signals than for music. It is therefore crucial to compare statistical properties of

music and speech in order to reach a better understanding of the relation between characteristics of various auditory communication signals and their possible optimal coding in auditory perception.

We applied the same FA method on 2 different sets of data: (1) a database of free music recordings (Free Music Archive, https://github.com/mdeff/fma), (2) a free corpus of speech signals (Clarity Speech, doi:10.17866/rd.salford.16918180).

Method: Analyses were carried out using the Matlab environment and mirrored previous studies. Sample signals were passed through a gammatone filterbank (1/4th ERB bandwidth, approx. 100-120 channels depending on the higher-frequency limit) and their energy envelope was extracted. This amplitude modulation matrix was then run through FA, and Principal Components (PCs) were independently rotated. Channels that covary in amplitude envelope should be grouped as a single Principal Component. Methods for automatically determining the optimal number of Principal Components as well as to estimate frequency boundaries between these PCs were developed. As our aim was to compare speech and music, for which typical signal bandwidths differ, two higher-frequency limits were compared (8000 Hz vs. 16000 Hz).

Results: Focusing on a reduced number of PC combinations that would compare to previous conclusions on speech according to which 4 to 7 PCs would be optimal, we find that cumulative explained variance is similarly located between 35% and 50% for music and between 39% and 52% for speech. Our estimates of frequency boundaries identified do not match those of previous studies. Boundaries are not fixed and depend on the type of natural signals (speech vs. music): variation in (1) boundary location and (2) PC Rank/frequency relations. Perceptual studies are in preparation that will help assess the validity of these measures.

P37 Temporal processing of slow amplitude modulations and consonant in noise perception in children using cochlear implants

Elodie Zuccarelli[1,2], Charlotte Benoit[1,2], Léo Varnet[3,4,5], Christian Lorenzi[3,5], **Laurianne Cabrera**[6,4,7]

1. Universite Paris Cité, Paris, France | 2. Assistance Publique Hôpitaux de Paris, Paris, France | 3. Laboratoire des Systèmes perceptifs, Paris, France | 4. CNRS, Paris, France | 5. Ecole Normale Supérieure, Paris, France | 6. Integrative Neurosciences and Cognition Center, Paris, France | 7. Université Paris Cité, Paris, France

The aim of this study was to evaluate the development of auditory temporal processing and speech in noise abilities in children using cochlear implants (CIs) in order to identify potential predictors of cochlear implantation success in childhood.

We included 25 children aged 5 to 12 years, implanted before the age of 4 years. They performed several psychophysical tasks measuring (1) AM detection thresholds for a slow modulation rate (8 Hz), (2) AM masking, (3) consonant identification thresholds in a stationary speech-shaped noise. The AM tasks were directly delivered through one electrode of the implant using a research interface provided by Advance Bionics (direct stimulation). Moreover, children with CIs completed standardized assessments of receptive vocabulary, communica-

tion skills, non-verbal reasoning, non-verbal working memory and the level of parental education was registered. Their perceptive thresholds were compared to two groups of children with normal-hearing (NH), matched either in chronological age, or in hearing age.

The results showed no effect of age on 8Hz-AM detection thresholds and children using CIs who were electrically stimulated showed better AM detection thresholds than acoustically stimulated children with NH. However, children using CIs showed significantly worse thresholds in the masking condition than children with NH. Children with CIs showed poorer consonant identification thresholds in noise compared to children with NH. Finally, regression analyses indicated that the only significant predictor of consonant identification in noise was parental education. No relationship between AM detection thresholds and consonant perception in noise was observed.

In conclusion, the processing of masked AM in children using CIs is not similar to their NH peers of the same chronological or hearing age. Their ability to identify consonants in stationary noise was impaired but not significantly related to AM processing at 8 Hz. Further studies are required to assess the value of other psychoacoustic tests that can be used clinically to better predict language learning in children using CIs.

P38 The effects of age and hearing on turn following in the presence of informational vs energetic maskers

Alexina Whitley, Timothy Beechey, Lauren V. Hadley
University of Nottingham, UK

Background: Many of our conversations occur in non-ideal situations, from the hum of a car to the babble of a cocktail party. In conversation, listeners are required to switch their attention between multiple talkers, which places demands on both auditory and cognitive processes. Speech understanding in cocktail party situations appears to be particularly demanding for older hearing-impaired listeners. As such, this study examined the relation of age and hearing ability on performance in a speech in noise talker switching task.

Methods: Older adults (65-81 years) with a range of hearing abilities, and younger adults (21-30 years) with normal hearing, took part in an online speech recall task using the coordinate response measure (CRM) corpus. For each trial, two utterances were presented one after the other, analogous to a conversational turn switch. The first target sentence was presented in quiet, and the second target sentence was masked either by noise (steady-state speech shaped noise) or speech (another CRM sentence). The two target sentences were either spoken by the same voice or different voices.

Results: Relative to conditions in which the target talker remained the same between sentences, participants were less accurate when the target talker changed, particularly when the original target talker became a masker for sentence 2. Listeners with poorer speech-in-noise reception thresholds (assessed via a digit triplet test) were less accurate at recalling target information in the second sentence in both noise and speech masked trials, and made more masker confusions in speech masking trials (i.e., erroneously reporting masker information

instead of target information). An interaction between switch type (i.e., same or different talker) and speech-in-noise thresholds revealed that participants with poorer hearing in noise received less benefit from the presence of a consistent speaker. Additionally, worse ability to distinguish between talkers (lower accuracy on Bangor Voice Matching Test) was associated with decreased accuracy and increased masker confusions in speech masking trials.

Conclusions: Our findings replicate those reported by Lin and Carlile (2019, doi:10.1038/s41598-019-44560-1) regarding the cost of following a switch in talker, extending these earlier findings to older adults with a range of hearing abilities. Furthermore, we demonstrate that greater difficulty in speech recall after a turn change relates both to reduced audibility (i.e., speech reception threshold) and reduced ability to distinguish between competing talkers (i.e., voice perception). This provides evidence in support of anecdotal reports of difficulty following conversational turns by people with hearing impairment.

P39 The right-ear advantage in static and dynamic cocktail-party situations

Moritz Wächtler[1,2], Pascale Sandmann[3,2], Hartmut Meister[1,2]
1. Jean-Uhrmacher-Institute for Clinical ENT-Research, University of Cologne, Germany | 2. Faculty of Medicine and University Hospital Cologne, Department of Otorhinolaryngology, Head and Neck Surgery, University of Cologne, Cologne, Germany | 3. Department of Otolaryngology, Head and Neck Surgery, Carl-von-Ossietzky University of Oldenburg, Oldenburg, Germany

When speech stimuli are presented dichotically, an advantage of the right ear ("right-ear advantage", REA) can often be observed, which manifests itself in a better speech recognition performance compared to the left ear. Explanations for this effect often refer to a specialization of the left hemisphere for language in combination with either superior contralateral pathways (structural model) or rightward-shifts of attention induced by speech (attentional models). There is evidence that the REA is increased when cognitive load is high (Penner et al., 2009, Fumero et al. 2022). With this in mind, it is worth investigating how the REA behaves in static (constant target talker) versus dynamic cocktail-party situations (unpredictable target talker changes), as the latter are associated with higher cognitive load. Results from a previous study (Wächtler et al., 2020) indeed provided first evidence for a greater REA in dynamic relative to static situations, although ceiling effects in performance complicated interpretation of the results.

In the present study, a cocktail party situation was simulated with three competing talkers at different positions (-60, 0, and +60 degrees azimuth angle) using a matrix sentence test. The target talker was indicated using a keyword. In the static condition, the position of the target talker remained constant and was announced in advance, whereas in the dynamic situation it changed in an unpredictable manner after each trial. To avoid ceiling effects in performance, speech stimuli were either presented at low sound pressure levels or processed with a noise vocoder. Data from 16 young normal-hearing adults were included in the study.

We discuss to what extent the different cognitive demands of the two conditions (static/dynamic) influence the REA and whether there is an interaction with the type of signal degradation (low level/vocoder). Furthermore, we present the results of a detailed error analysis to investigate how far the results support structural and attentional models of the REA.

Funding: Deutsche Forschungsgemeinschaft (ME2751/3-2).

References:

Fumero MJ, Marrufo-Pérez MI, Eustaquio-Martín A, Lopez-Poveda EA (2022). Divided listening in the free field becomes asymmetric when acoustic cues are limited. Hearing research, 416, 108444. doi:10.1016/j.heares.2022.108444

Penner IK, Schläfli K, Opwis K, Hugdahl K (2009). The role of working memory in dichotic-listening studies of auditory laterality. J Clin Exp Neuropsychol, 31(8):959-66. doi:10.1080/13803390902766895.

Wächtler M, Wenzel F, Kessler J, Walger M, Meister H. (2020). What are some of the challenges in dynamic cocktail party listening?. Speech in Noise Workshop 2020 (SPIN), Toulouse, France. Zenodo. doi:10.5281/zenodo.8101983

P40 Binaural beamforming taking into account spatial release from masking

Johannes W. de Vries[1], Steven van de Par[2], Geert J. T. Leus[1], Richard Heusdens[1], Richard C. Hendriks[1]

1. Delft University of Technology, Delft, The Netherlands | 2. Carl von Ossietzky University, Oldenburg, Germany

Hearing impairment is a prevalent problem that comes with many daily life challenges. These challenges, mainly in the areas of speech intelligibility and sound localisation, can have consequences ranging from social isolation to physical danger. Even though current hearing aid technology can partly alleviate these issues, in practice many acoustic scenes are still challenging. One of the shortcomings of spatial filtering in hearing aids is that speech intelligibility is often not optimised for directly, as that is a subjective measure and more difficult to quantify. Instead, most developed beamformers focus on maximising the speech-to-noise-plus-interference ratio (SNIR), which is only one factor that influences intelligibility. The psychoacoustic effect known as spatial release from masking (SRM) is usually not considered, but can also be a dominant factor.

In this paper, a signal model is developed that explicitly takes SRM into account in the beamforming design. This is achieved by transforming the binaural intelligibility prediction model developed by Beutelmann and Brand (2006, JASA 120:331) to a signal processing framework. The main phase of this model, the equalisation–cancellation (EC) phase, can be represented as an internal beamformer that accounts for the spatial filtering of the auditory system. Internal masking noise is added to the hearing aid signals to model hearing thresholds, which is used to personalise the model. When concatenated with a typical beamformer signal model, the output of this extended model can be used to analyse existing beamformers and design new beamformers one step closer to how the auditory system perceives binaural sound.

It can mathematically be shown that the binaural minimum variance distortionless response (BMVDR) beamformer, which is known to be optimal in maximising the SNIR of the beamformer signals, is also an optimal solution for the extended, perceived model. This seems to suggest that SRM does not play a significant role in improving intelligibility after optimal beamforming is already performed. What is different compared to the classic case, however, is

that the solution is no longer unique; the solution space has a number of degrees of freedom dependent on the number of microphones in the hearing aids. These degrees of freedom can be used to preserve binaural cues of interferent sources, while still achieving the same perceived performance of the BMVDR beamformer. The proposed beamformer might in practice be sensitive to intelligibility model mismatch errors, and the practical performance needs to be studied in more detail.

P41 Perceptual learning of dysarthric speech requires phonological processing: A dual-task study

Patti Adank
Speech, Hearing and Phonetic Sciences, University College London (UCL), United Kingdom

Han Wang
Department of Speech, Hearing and Phonetic Sciences, University College London, London, United Kingdom

Taylor Hepworth, Stephanie A. Borrie
Department of Communicative Disorders and Deaf Education, Utah State University, Logan, United States

Rationale: Listeners can adapt to speech signals that can be initially difficult to understand, including artificially degraded signals such as noise-vocoded speech. We recently demonstrated (Wang et al., 2023) that listeners can perceptually adapt to noise-vocoded speech under divided attention (using a dual task design). Here, we evaluated the role of divided attention in perceptual learning of naturally (neurologically) degraded speech, i.e., dysarthric speech (Borrie & Lansford, 2021). We conducted an online between-subject experiment with four groups (N = 192). We examined the reliance of perceptual learning of dysarthric speech on selective attention to establish if perceptual adaptation to degraded speech qualifies as an automatic cognitive process.

Methods: Participants completed a speech recognition task in which they repeated forty sentences spoken by a male dysarthric speaker, in a between-group design. Participants completed a speech-only task or performed this task with a dual task aiming to recruit domain-specific (lexical or phonological), or domain-general (visual) processes. If perceptual learning of distorted speech qualifies as a largely automatic process, we expected no difference in rate or shape of adaptation across the four groups. However, if perceptual learning of speech requires domain-specific processes that matched the type of variation present in the speech signal, we expected a lower rate of adaption for the phonological group.

Results: We observed perceptual learning for all groups, except for the phonological group. Speech recognition improvement in the single speech, lexical, and visuomotor groups was around 10-11%, while improvement in the phonological group was not significant (5%).

Conclusions: Perceptual learning of dysarthric speech can occur under divided attention, as long as the dual task does not require phonological processes. Perceptual learning of speech is thus a largely automatic process, but engagement of domain-specific processes distorts learning.

References:

Borrie, S. A., & Lansford, K. L. (2021). A perceptual learning approach for dysarthria remediation: An updated review. Journal of Speech, Language, and Hearing Research, 64(8), 3060-3073, doi:10.1044/2021_JSLHR-21-00012.

Wang, H., Chen, R., Yan, Y., McGettigan, C., Rosen, S., & Adank, P. (2023). Perceptual Learning of Noise-Vocoded Speech Under Divided Attention. Trends in Hearing, 27, doi:10.1177/23312165231192297.

P42 A phoneme-scale evaluation of multichannel speech enhancement algorithms

Nasser-Eddine Monir, Paul Magron, Romain Serizel

Université de Lorraine, CNRS, Inria, Loria, Nancy, France

The impairment of auditory function resulting from hearing loss significantly undermines the capacity for speech comprehension, especially in scenarios where speech is mixed with various competing sounds. Several signal alterations, such as spread spectrum or masking phenomena, can make it challenging to differentiate between phoneme formants due to their overlapping frequencies. In this regard, speech enhancement appears as a promising solution to mitigate the adverse effects of ambient noise on the intelligibility and clarity of spoken language. Recent advances driven by deep learning empirically support the effectiveness of enhancement models in improving speech intelligibility in complex acoustic environments.

Such algorithms are typically evaluated for their ability to restore intelligibility and speech quality for individuals with normal hearing. Consequently, these assessment strategies may not always offer relevant insights for individuals suffering from hearing impairments. In particular, models are commonly evaluated at the utterance level, which aggregates errors across diverse phonemes. This results in potentially overlooking certain phonemic categories, which are of particular importance for individuals with impairments. Indeed, the influence of noise and its mitigation differs among phonemes, owing to their signal-level attributes which stem from their specific production model in the vocal tract. For impaired individuals, these differences are particularly salient due to reduced spectral and temporal resolutions, and the occurrence of masking effects.

To overcome this issue, in this study we perform a comprehensive assessment of speech enhancement algorithms at the phoneme level. We categorize phonemes according to their distinct articulatory models (e.g., plosives, fricatives, nasals) as it creates groups with similar signal characteristics. We consider four state-of-the-art multichannel speech enhancement models. Using publicly available datasets of clean speech and real-life noise, we simulate noisy mixtures in order to encompass various spatial conditions. In addition to the commonly-used utterance scale, we evaluate the models' performance in terms of distortion, artifacts and interference reduction at the proposed fine-grained phonemic scale. This outlines the algorithms' effectiveness in reducing noise interference according to specific phoneme-level features.

To summarize, this study tackles the challenges posed by hearing disorders, particularly the intricacies of phonemic decoding processes within the cochlea and the brain. It marks an initial step towards advancing speech enhancement models, as it enables the identification of specific speech components that require further emphasis.

P43 Rapid label-referent mapping with vocoded speech in young infants

Alan Langus
University of Potsdam, Germany

Mireia Marimon
Pompeu Fabra University, Spain

Amanda Saksida
Burlo Hospital, Trieste, Italy

The speech amplitude envelope plays a central role in speech perception and acquisition. The human auditory cortex tracks the speech amplitude envelope from birth. Envelope tracking correlates with speech perception in infants and speech comprehension in adults. The relative importance of the speech envelope for speech comprehension can be shown with vocoded speech – synthesized speech that simulates sound perception with a cochlear implant by dividing the speech signal into narrow frequency bands, then extracting their envelopes that are used to modulate noise in the same frequency band (i.e. channels). Adult listeners understand vocoded speech synthesized from as few as 3 channels and children from 4 or 8. While recent studies suggest that even young infants can discriminate vocoded speech during the first months of life, it remains unknown if young infants perceive vocoded speech as language or can acquire language from vocoded stimuli.

To answer this question 7- to 9-mo German-learning infants (N=36) participated in a label-referent mapping experiment. While infants listened to short trials (N=20) consisting of a familiarization and a test phase we measured their pupil size. In each trial, infants were first briefly familiarized with 2 object-label pairs and then presented with 4 test events: 2 Same trials (containing one of the familiarization object-label pairs) and 2 Switch trials (where the familiarization objects and labels were switched). The visual stimuli were 8 abstract Tetris-like objects. The auditory stimuli were 8 disyllabic nonce words uttered by 5 female German native speakers in Infant Directed Speech (4 speakers for familiarization and 1 for test events). The auditory stimuli were either natural speech or vocoded stimuli synthesized from 2, 4, 8, or 16 narrowband frequencies corresponding to the frequency bands of the human cochlea.

A cluster-based permutation test over the pupillary response at test revealed a significant interaction between Channel (Speech/2ch/4ch/8ch/16ch) and Trial Type (Same/Switch) in a window that started 1886 ms after test word onset and lasted for 789 ms (TSUM=64.42, P<.01). In this window we observed a significant difference between Same and Switch trials with natural speech and 16-channel vocoded speech, but not with vocoded speech composed of fewer channels. Our results show that infants can rapidly map labels to visual objects after only limited exposure to natural as well as vocoded speech. However, this ability to use speech that only contains envelope cues as labels to novel objects deteriorates quickly as the number of

frequency bands is reduced. This suggests that young infants do not perceive vocoded speech as readily as human adults and that stimulating only a few cochlear channels in infants may not provide sufficient auditory detail for perceiving spoken language.

P44 Exploring the effect of semantic context in dynamic cocktail-party listening

Moritz Wächtler, Hartmut Meister
Jean-Uhrmacher-Institute for Clinical ENT-Research, University of Cologne, Germany | Faculty of Medicine and University Hospital Cologne, Department of Otorhinolaryngology, Head and Neck Surgery, University of Cologne, Cologne, Germany

Cocktail-party situations are common in everyday conversation. They can either be static (target talker remains the same) or dynamic (target changes unpredictably). Due to the need to monitor multiple talkers and to switch attention from one talker to another, dynamic situations are associated with a higher cognitive load and thus a decrease in speech recognition performance relative to static situations, referred to as "costs" (see Lin & Carlile, 2015; Meister et al., 2020). However, the corresponding studies typically used matrix sentences which, due to their nonsensical nature, offer only little semantic context information compared to conversations from everyday life. As there is evidence that semantic context aids stream segregation and word recall (e.g., Meister, 2013), we assume that context effects can alleviate cognitive demand. Against the background that cognitive resources are limited, we hypothesize that higher context allows listeners to allocate more cognitive resources to the challenges of dynamic cocktail-party listening and therefore leads to lower costs.

Cocktail-party situations with three competing talkers located at different positions were simulated. Listeners were asked to repeat back words from the target talker, which had a higher voice than the two masker talkers. The listening situation was dynamic, meaning the target talker changed its position in unpredictable pseudo-random patterns. The three talkers either uttered matrix sentences such as "Simon orders forty wet bottles." that are assumed to have a low predictability (low context) or more meaningful everyday-life sentences like "The hen laid an egg." (high context).

The low and high context sentences were based on established speech tests for the German language. However, we did not use the original audio recordings of those tests but synthesized all sentences using a text-to-speech algorithm. In this way, we could improve comparability between high and low context materials by avoiding the talker differences present in the original recordings. In addition, this allowed us to extend the set of words used to form the matrix sentences, thus reducing the listener's chance of guessing words typically inherent in closed-

set material. Preliminary data from young normal-hearing listeners will be shown. The results will be analyzed using mathematical metrics for linguistic context effects and will be discussed against the background of cognitive speech recognition models.

Funding: Deutsche Forschungsgemeinschaft (ME2751/3-2).

References:

Lin & Carlile (2015). Costs of switching auditory spatial attention in following conversational turn-taking. doi:10.3389/fnins.2015.00124

Meister et al. (2013). Cognitive resources related to speech recognition with a competing talker in young and older listeners. doi:10.1016/j.neuroscience.2012.12.006

Meister et al. (2020). Static and dynamic cocktail party listening in younger and older adults. doi:10.1016/j.heares.2020.108020

P45 Co-speech and listening gestures during casual conversation in a noisy situation

Lubos Hladek, Yang Jiao, Bernhard U. Seeber
Audio Information Processing, Technical University of Munich, Germany

In face-to-face communication, people use gestures while they speak, but less is known about gestures during listening. In the current work, we test whether body movement behavior during conversation indicates the difficulties related to the increased noise level during the conversation. We developed a categorization system for conversational body movements in which we characterize movements on a physical level in terms of head, arm, leg and trunk movements. We conducted an experiment in which groups of three had casual face-to-face conversation while standing in a noisy audio-visual scene of an underground station created by the real-time Simulated Open Field Environment (rtSOFE). Full-body movements of one of the three participants were recorded using a motion capture system. Speech of the participants was reverberated and recorded. In the preliminary analysis, the motion capture data were labeled by one observer. We observed an increase in palm swipe gestures and complex hand gestures when the participant was speaking and an increase of contractive postures when they were not speaking. Nodding head movements were slightly increased during listening. Presence of noise (72 dB SPL) had slight influence on relative occurrences relative to No Noise condition. Body posture and head nods are typical backchannel cues during the conversation, which is in line with our statistical analysis. The present categorization approach of conversational body movements can complement existing linguistic tools to analyze personal communication from the audiological perspective.

Funding: This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 352015383 – SFB 1330, Project C5 and Bundesministerium für Bildung und Forschung (BMBF) (grant number 01 GQ 1004B).

P46 The imperfect invariance problem modifies cortical signals during listening in noise

Petra Kovács[1,2], István Winkler[1], Brigitta Tóth[1]
1. HUN-REN Research Centre for Natural Sciences, Institute of Cognitive Neuroscience and Psychology, Budapest, Hungary | 2. Budapest University of Technology and Economics, Budapest, Hungary

Imperfect invariances in speech pose a challenge to speech perception. When listening to speech in noise, finding regularities is hindered not only by noise but also the signal itself, as it varies due to coarticulation, underarticulation, and individual speaker characteristics. What is the minimal amount of invariance that is needed to perceive a coherent auditory object? To investigate the imperfect invariance problem in noise, we employed the stochastic figure-ground (SFG) task, which has been established as a suitable model for speech-in-noise listening. The SFG task uses random noise created from 50 ms long pure tones ("ground") and an embedded set of tones coherently fluctuating together over time ("figure"). We modified the number of tones available at different time points in the figure. We recorded the electroencephalogram (EEG) and analyzed event-related brain potentials (ERPs) while listeners tried to detect the figure segments.

Twenty-two healthy young adults listened to SFG stimuli of 3 s duration each. Half of the trials contained a figure, and half contained a random background only. Participants indicated whether there was a figure in each trial. All stimuli consisted of pure tones of 20 discrete frequencies, selected from a larger set. Figures contained a set of 10 potential repeating frequencies. We varied in three conditions how many of these frequencies were concurrently present in the stimulus: 10/10, 7/10, or 4/10 frequencies. While the overall number of repeating frequencies remained constant throughout the figure, the actual frequencies being present (permuted from the set of 10) changed at every 50 ms of the figure. This models the imperfect invariance problem: out of a set of possible components, only a subset is present in the stimulus at each time point, the subset varying throughout the whole stimulus.

We found that the number of repeating frequencies affects the object-related negativity (ORN) and P400 ERP responses to figure segments. Figures with 10/10 coherent frequencies elicited both ERP components with the largest amplitude, followed by figures at 7/10 coherent frequencies, and yet smaller amplitudes at 4/10 coherent frequencies. Further, we found no difference between ORN and P400 when comparing 4/10 coherent frequencies with the no-figure trials. These results suggest that a 40% invariance in frequency components is insufficient to detect an auditory object in noise (fluctuating energetic masking), but a 70% invariance is sufficient. We discuss these conclusions and their significance for speech-in-noise listening in the poster presentation.

P47 Spatial hearing training for young bilateral cochlear implant users: The BEARS approach

Bhavisha Parmar[1,2], Marina Salorio-Corbetto[3], Sandra Driver[4], Merle Mahon[5], Lorenzo Picinali[6], Helen Cullington[7], Pádraig Kitterick[8], Francis Early[9], Fleur Corbett[10], Dan Jiang[4], Deborah Vickers[3]

1. SOUND lab, University of Cambridge, Cambridge, UK | 2. UCL Ear Institute, London, UK | 3. SOUND lab, Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK | 4. St Thomas' Hearing Implant Centre, Guys and St Thomas' NHS Foundation Trust, London, UK | 5. Psychology and Language Sciences, Faculty of Brain Sciences, University College London, London, UK | 6. Dyson School of Design Engineering, Imperial College London, London, UK | 7. Auditory Implant Service, University of Southampton, Southampton, UK | 8. National Acoustics Laboratories (NAL) Macquarie University, Australia | 9. Department of Respiratory Medicine, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK | 10. Design Psychology Lab, Dyson School of Design Engineering, Imperial College London, London, UK

Background: Although sound localization and speech-in-noise perception are better for people with bilateral Cochlear Implants (CIs) compared to those with a unilateral implant, these skills remain far below those of normally-hearing children (Sarant et al., 2014; Sparreboom et al., 2015). A large body of research demonstrates that sound localisation can improve with training, underpinned by plasticity-driven changes in the auditory pathways for children and adults (Firszt et al., 2015; Yu et al., 2018). The use of audio-visual stimuli helps with task familiarisation, and the gamification approach contributes to improving engagement and attainment, which is crucial for children and teenagers. However, there is currently a lack of engaging, remote, multimodal training programmes for young people with bilateral CIs.

The BEARS approach: The BEARS project (NIHR201608) was established: (1) To develop the Both Ears (BEARS) training package, a set of virtual-reality games to train spatial hearing in young people with bilateral CIs through a series of PPIE workshops. (2) To develop the outcome measures needed to evaluate the effectiveness of the BEARS training package. (3) To conduct a large-scale confirmatory clinical trial to assess whether BEARS substantially improves hearing with two implants. (4) To understand the learning mechanism and process evaluation.

Results and conclusions: Over the last 2 years, the outcome measures and intervention have been developed and the clinical trial launched in June 2023. Here, we summarise the BEARS logic model, approach and next steps.

Funding: Supported by the National Institute for Health and Care Research (NIHR, grant number NIHR201608). MSC was supported by a travel bursary from Oticon Medical.

References:

Firszt J.B., Reeder R.M., Dwyer N.Y., Burton H., Holden L.K. (2015) Localization training results in individuals with unilateral severe to profound hearing loss. Hearing Research 319: 48–55. doi:10.1016/j.heares.2014.11.005.

Sarant J., Harris D., Bennet L., Bant S. (2014). Bilateral versus unilateral cochlear implants in children: A study of spoken language outcomes. Ear Hear. 35(4):396–409.

Sparreboom M.A., Langereis M.C., Snik F.M., Mylanus A.M. (2015). Long-term outcomes on spatial hearing, speech recognition and receptive vocabulary after sequential bilateral cochlear implantation in children. Research in Developm Disabil. 36:328–337.

Yu F., Li H., Zhou X., Tang X., Galvin J. J. III, Fu Q. J., Yuan W. (2018) Effects of training on lateralization for simulations of cochlear implants and single-sided deafness. Frontiers in Human Neuroscience 12: 287, doi:10.3389/fnhum.2018.00287.

P48 Developing and validating virtual-audio clinical tools for assessing spatial-listening skills for children with bilateral cochlear implants

Bhavisha Parmar[1,2], Marina Salorio-Corbetto[3,4], Jennifer Bizley[2], Stuart Rosen[5], Tim Green[5], Lorenzo Picinali[6], Ben Williges[7], Deborah Vickers[1]

1. SOUND lab, Department of Clinical Neurosciences, University of Cambridge, UK | 2. UCL Ear Institute, UK | 3. SOUND lab, Department of Clinical Neurosciences, University of Cambridge | 4. Cambridge University Hospitals, Emmeline Centre for Hearing Implants, UK | 5. Speech, Hearing and Phonetic Sciences, University College London, UK | 6. Audio Experience Design Group, Imperial College London, UK | 7. Sound Lab, Department of Clinical Neurosciences, University of Cambridge, UK

Background: Clinical tests for the assessment of spatial listening require multi-speaker arrays rarely available in clinical settings. A virtual-audio version of the Spatial Speech in Noise Test (SSiN; Bizley et al., 2015) leads to similar performance across spatial locations for loudspeaker arrays with normal-hearing listeners (Salorio-Corbetto et al., 2022). The aim of this work is to determine whether the virtual-audio versions of the SSiN and the Adaptive Sentence List (ASL; (MacLeod & Summerfield, 1990) using a spatial release from masking test configuration test yield comparable results than their loudspeaker versions for children with bilateral cochlear implants. Additionally, the efficacy of a centralisation app to identify the degree of balance between the ears was explored together with the findings from the virtual speech tests. The purpose of this work is to validate virtual assessments for use in a clinical trial with virtual reality spatial training games.

Method: A participatory-design approach was used to develop and finalise the virtual-audio implementations of the tests (Vickers et al., 2021). Ten children and young adults who wear bilateral cochlear implants and ten age-matched normal-hearing participants, will perform each test (SSiN and Spatial ASLs) in each implementation (virtual-audio or loudspeaker). The order of the tests and implementations were counterbalanced across participants. The participants also completed the centralisation task (i-balance app) using narrow-band noise and wide-band stimuli consisting of speech-shaped noise and a non-language specific speech-like stimulus (Holube et al., 2010). The interaural level differences for these stimuli were varied by the children using a visual/tactile interface so that the sound was perceived in the midline. Children were asked to show where they located or heard the sound relative to their head by colouring a drawing.

Results: So far, the virtual-audio applications were finalised. Eight participants with cochlear implants and six with normal hearing have completed the tests. Our outcomes will allow us to determine whether the virtual-audio versions of the tests have potential for clinical use,

provide the validation for use in the clinical trial and determine whether the results from the centralisation task used in the i-balance app are informative in terms of spatial hearing abilities for children with bilateral cochlear implants.

References:

Bizley, J.K., Elliott, N., Wood, K.C., Vickers, D.A., 2015. Simultaneous assessment of speech identification and spatial discrimination: A potential testing approach for bilateral cochlear implant users? Trends in Hearing, 19, p.2331216515619573. doi:10.1177/2331216515619573.

Holube, I., Fredelake, S., Vlaming, M., Kollmeier, B., 2010. Development and analysis of an International Speech Test Signal (ISTS). International Journal of Audiology, 49(12), p.891–903. doi:10.3109/14992027.2010.506889.

MacLeod, A., Summerfield, Q., 1990. A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. British Journal of Audiology, 24, p.29–43.

Salorio-Corbetto, M., Williges, B., Lamping, W., Picinali, L., Vickers, D.A., 2022. Evaluating Spatial Hearing using a Dual-Task Approach in a Virtual Acoustics Environment. Frontiers in Neuroscience, p.46.

Vickers, D.A., Salorio-Corbetto, M., Driver, S., Rocca, C., Levtov, Y., et al, 2021. Involving children and teenagers with bilateral cochlear implants in the design of the BEARS (Both EARS) virtual reality training suite improves personalization of the intervention. Frontiers in Digital Health, p.156.

P49 Speech-in-noise pupillometry data collected via a virtual reality headset

Tim Green
Speech Hearing & Phonetic Sciences, UCL, London, UK

Lorenzo Picinali, Tim Murray-Browne, Isaac Engel, Craig Henry
Dyson School of Design Engineering, Imperial College, London, UK

In an initial exploration of virtual reality (VR) for assessing speech recognition and listening effort in realistic environments, pupillometry data collected via a consumer Vive Pro-Eye headset were compared with data obtained via an Eyelink 1000 system, from the same normally-hearing participants presented with similar non-spatialized auditory stimuli. Vive testing used a custom platform based on Unity for video playback and MaxMSP for headphones-based audio rendering and overall control. For Eyelink measurements, head movements were minimized by a chin rest and participants fixated on a cross presented on a grey screen. No such constraints were present for Vive measurements which were conducted both using a 360° cafe video and with a uniformly grey visual field. Participants identified IEEE sentences in babble at two signal-to-noise ratios (SNRs). The lower SNR (-3 dB) produced around 65% word recognition, while performance for the higher SNR (+6 dB) was at ceiling. As expected, pupil dilation was greater for the lower SNR, indicating increased listening effort. Averaged across participants, pupil data were very similar across systems, and for the Vive, across the different visual backgrounds, thereby supporting the idea that VR systems have the potential to provide useful pupillometric listening effort data in realistic environments.

Ongoing experiments, conducted solely in VR and using spatialised sound, examine the effects of incorporating different aspects of realistic situations. In the aforementioned café video three computer monitors, each associated with a particular talker, are visible on tables. In one condition, as above, target sentences from a single talker are presented audio-only. In a second single-talker condition, a video of the talker reading the target sentence is also presented. In a third condition the target talker varies quasi-randomly across trials. A visual cue presented 2s prior to the target video allows the participant to orient to the correct screen. In each visual

condition, target sentences are presented against a background of café noise plus additional babble at two SNRs, differing by 9 dB. Comparison across conditions will be informative of the extent to which pupillometry data may be affected by factors such as dynamic changes in visual background and head movements of the type typically involved in multi-talker conversations.

Funding: Supported by the William Demant Foundation.

P50 Factor analysis of acoustic signals for the determination of optimal boundaries: Perspectives concerning cochlear implants and investigation of measurement variation

Olivier Crouzet, Agnieszka Duniec
Nantes Université / CNRS, France

Previous studies applied Factor Analysis on amplitude modulation from speech signals in order to estimate optimal frequency boundaries between channels. These may contribute to future improvements for specifying filterbank decomposition in cochlear implants. While some argued that 4 channels would be sufficient to represent the main segmental information, comparison of speech statistics with perceptual performance led to suggest that 6 to 7 frequency bands would be required to optimally represent vocoded speech. We applied the same approach on 2 different datasets: (a) free music recordings (Free Music Archive, https://github.com/mdeff/fma), (b) a free corpus of speech signals (Clarity Speech, doi:10.17866/rd.salford.16918180). An algorithm for the automatic computation of optimal boundary frequencies was also developed as results from the literature were based on visual judgements only.

As was expected, observed boundaries differ between speech and music: their distribution is organized differently though not homogeneously. For example, when selecting 7 modulation channels, size and direction differences vary between -3.8 and +10.7 semitones with the same determination method. Similar variation is observed either when maximal acoustic frequency is adapted for music or when it is kept constant for both conditions. Further, comparing our data on speech with results in the literature, estimated boundaries also differ. For example, determination of the optimal boundary frequencies for 4 channels gives rise to differences varying between +0.15 and +2.19 semitones. Such variation may relate to either the database content, the determination method (visual vs. automatic), or both.

Given such variability, it seemed crucial to investigate the level of variation that can be observed in conditions for which the type of signals and method are kept constant. Capitalizing on our development for the automatic estimation of boundary frequencies, we applied a procedure that is aimed at estimating variability in the measurements: concentrating on the music database, random portions of fixed duration for each music recording are extracted and the same Factor Analysis is applied. Automatic boundary determination between optimal channels is performed. For each random extract, a various sample of frequency estimates is

available along with information characterizing music and technical parameters (style, compression...). This approach and the final results will be presented. These may contribute to a better analysis of the importance of "efficient coding" for channel decomposition in cochlear implants by providing fine-grained data on variation in optimal frequency boundaries.

Acknowledgements: Agnieszka Duniec received PhD funding (2019–2023) from the RFI-Ouest Industries Créatives (RFI-OIC, Région Pays de la Loire) & Nantes Université.

P51 Interplay between working memory and speech recognition declines over time

Erik Marsja, Emil Holmer, Henrik Danielsson
Disability Research Division, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

Background: Age-related changes in auditory and cognitive functions are well-documented, with increased hearing thresholds (e.g., Wiley et al., 2008) and reduced working memory capacity (WMC; e.g., Wingfield et al., 1988) among older adults. Moreover, aging has been linked to poorer speech recognition in noise (e.g., Marsja et al., 2022), highlighting the multifaceted impact of age on auditory and cognitive domains. Our study examined the dynamic relationship between auditory and cognitive changes over time to shed light on the direction of influence between the two. To this aim, we employed change score modeling.

Methods: We analyzed data from 111 normally hearing individuals from the n200 study (https://2024.speech-in-noise.eu/proxy.php?id=81). At Time 1 (T1), their mean age was 61.2 years (SD = 8.00), and at Time 2 (T2), their mean age was 67.0 years (SD = 8.06). We used Latent Change Score modeling to explore the changes in WMC and speech recognition in noise. To measure speech recognition in noise, we used signal-to-noise ratios from the Hearing in Noise Test during speech-shaped noise. The reading span test was used as a measure for WMC.

Results and Conclusion: Preliminary results showed a decline in WMC, signified by the negative relationship between Reading Span at T1 and changes in Reading Span at T2. This negative relationship indicates that individuals with higher initial WMC experienced subsequent declines in their cognitive abilities. Furthermore, our analysis revealed a negative relationship between changes in speech recognition in noise at T2 and Reading Span at T1. This relationship suggests that individuals with higher initial WMC experienced less decline in their speech recognition in noise over time. Further research with additional time points may be needed to fully elucidate the complex relationship between cognitive and auditory changes over time.

References:

Marsja, E., Stenbäck, V., Moradi, S., Danielsson, H., & Rönnberg, J. (2022). Is Having Hearing Loss Fundamentally Different? Multigroup Structural Equation Modeling of the Effect of Cognitive Functioning on Speech Identification. Ear Hear, 43(5), 1437–1446.

Wiley, T. L., Chappell, R., Carmichael, L., Nondahl, D. M., & Cruickshanks, K. J. (2008). Changes in hearing thresholds over 10 years in older adults. J Am Acad Audiol, 19(4), 281–292.

Wingfield, A., Stine, E. A. L., Lahar, C. J., & Aberdeen, J. S. (1988). Does the capacity of working memory change with age? Exp. Aging Res., 14(2), 103–107.

P52 Affective valence affects speech intelligibility in noise

Alexandra E. Clausen[1], Florian Kattner[2], Wolfgang Ellermeier[1]
1. Technical University of Darmstadt, Darmstadt, Germany | 2. Health and Medical University, Potsdam, Germany

Some indication already exists that both prosodic and semantic cues to emotional valence in speech utterances facilitate speech intelligibility in noise (e.g., Dor et al., 2021; Dupuis & Pichora-Fuller, 2014). Of all emotions, fear in the voice in particular, can lead to higher accuracy in word recognition in a noisy condition (Dupuis & Pichora-Fuller, 2014). In a first experiment, we wanted to test a possible effect of negative and positive semantic meaning of words on speech intelligibility. Participants had to recognize negative, positive, and neutral words at the end of a structured sentence in a recording partially masked by speech-spectrum noise. All sentences were spoken by the same neutral voice in the intelligibility test. In a second experiment, we investigated, if a negative or positive association to a voice can also lead to enhanced intelligibility. We experimentally manipulated the valence of voices pronouncing semantically neutral words through evaluative conditioning. In the conditioning phase, participants heard words spoken by clearly discriminable voices, which were followed by either positive, neutral, or negative images. Participants were then asked to recognize words spoken by the previously conditioned voices in an intelligibility test like in the first experiment. Enhanced speech intelligibility was expected for words with cues to negative and positive valence in both experiments.

References:

Dor, Y. I., Algom, D., Shakuf, V., & Ben-David, B. M. (2022). Age-Related Changes in the Perception of Emotions in Speech: Assessing Thresholds of Prosody and Semantics Recognition in Noise for Young and Older Adults. Front. Neurosci., 16, 846117. doi:10.3389/fnins.2022.846117

Dupuis, K. & Pichora-Fuller, M. K. (2014). Intelligibility of Emotional Speech in Younger and Older Adults. Ear & Hearing, 35(6), 695–707. doi:10.1097/AUD.0000000000000082.

P53 Influence of semantic context information on rollover in aided hearing-impaired listeners

Lukas Jürgensen, Michal Fereczkowski, Tobias Neher
Institute of Clinical Research, University of Southern Denmark, Odense, DK | Research Unit for ORL – Head & Neck Surgery and Audiology, Odense University Hospital & University of Southern Denmark, Odense, DK

Background: At low presentation levels, a level increase typically improves audibility and thus speech intelligibility. At high presentation levels, a level increase can lead to poorer speech intelligibility. Termed rollover, this effect has been observed in listeners with normal and elevated audiometric thresholds. In a previous study, we found rollover at above-conversational levels in normal-hearing listeners in background noise when tested with context-free but not context-rich speech materials. We therefore concluded that semantic context information can mask rollover because of compensatory top-down mechanisms. However, other studies observed rollover at conversational levels in aided hearing-impaired listeners who were tested

with context-rich speech materials only. In view of these discrepant findings, the current study investigated the influence of semantic context information on rollover at conversational and above-conversational presentation levels in aided hearing-impaired listeners.

Methods: Listeners with mild-to-moderate sensorineural hearing losses participated in speech intelligibility measurements in stationary speech-shaped noise. To compensate for their elevated hearing thresholds, individual linear amplification was provided by means of a wearable hearing-aid simulator. The speech materials included context-free sentences from the Danish DAT corpus and context-rich sentences from the Danish HINT corpus. The speech signals were presented at three levels: 65, 75, and 85 dB SPL.

Results: Initial results from this ongoing study indicate rollover with both types of speech materials. Performance decreases seem to occur at moderate levels (65 vs. 75 dB SPL) and particularly higher levels (75 vs. 85 dB SPL). Overall, this would seem to suggest that hearing-impaired listeners are susceptible to distortions arising in the auditory system at conversational and above-conversational levels, regardless of the availability of semantic context information.

P54 Exploring the relationship between stream segregation and speech-in-noise performance in cochlear implant listeners.

Nicholas Haywood, Marina Salorio-Corbetto, Ben Williges, Deborah Vickers
Universtiy of Cambridge, UK

Although stream segregation is typically studied with relatively simplistic stimuli such as pure tones, the parameters that influence basic auditory object formation may be relevant to understanding speech perception in noise. Prelimarary findings (n=9) from this research suggest that cochlear implant (CI) listeners may experience pure-tone stream segregation at slower presentation rates (longer inter-stimulus intervals) than normal hearing (NH) listeners. We speculate that this may reflect increased adaption/habituation in the auditory cortex from CI stimulation. If present, increased patterns of cortical adaption may potentially impair speech-in-noise comprehension, or the ability to follow rapid speech. We are currently measuring speech-in-noise thresholds in our CI group, and will present correlations between stream segregation and speech-in-noise measures.

This research was designed primarily to address aspects of stream segregation in CI users which remain poorly understood. In NH, segregation increases as the frequency separation (ΔF) between alternating tones is increased and/or the inter-stimulus interval (ISI) between tones is decreased. However, while stream segregation in CI listeners appears to be influenced by ΔF, ISI has not been found to affect segregation.

The task required listeners to detect a temporal delay imposed on a single tone. Stimuli were arranged so that any obligatory stream segregation should impair performance, and three ISIs were tested (50, 100, and 200 ms). Preliminary CI results (n=9) show delay thresholds increased with ΔF, and increased further with the addition of a segregation-promoting precursor sequence. Both observations are indicative of stream segregation effects. The CI group

showed clear stream segregation effects at ISIs of 100 and 200 ms, but ceiling effects imposed on performance in the 50 ms ISI conditions. In contrast, a NH group (n=9) performed well and showed clear stream segregation effects when the ISI was 50 ms, but showed reduced stream segregation effects for the two longer ISIs. The pattern of results suggests stream segregation may persist at slower presentation rates for CI listeners. We will compare these measures of stream segregation to measures of speech-in-noise performance.

P55 Predictive sentence processing of L2 speech in noise: Differential effects for different types of linguistic cues

Rebecca Carroll[1], Angela Patarroyo[2], Holger Hopp[1]
1. TU Braunschweig, DE | 2. HU Berlin, DE

Incremental processing of speech leads listeners to build expectations and to make predictions about upcoming sentence information. This eye-tracking study explores the time course of L2 sentence processing and the (anticipatory) integration of different types of linguistic information in different types of noise. We investigate (a) whether L2 listeners continue to predict during sentence comprehension even when noisy speech causes phonetic unreliability or difficulties in information integration, and we examine (b) the degree to which different linguistic cues are affected by different types of noise in L2 listeners. This way, we aim to investigate which cues are particularly vulnerable in L2 predictive processing and why.

Using a visual world paradigm, we tested predictive processing among 72 German advanced L2 listeners of English in three linguistic conditions (discourse, morpho-syntax, lexical semantics), and across three acoustic conditions (quiet, stationary noise, multi-talker babble noise), which were presented to three subgroups (= noise groups). Noise can impact information integration in differential ways. Whereas stationary speech-shaped noise serves as an energetic masker, lowering the salience of bottom-up information (e.g. inflectional information, function words), multi-talker babble combines energetic and informational masking, thus adding cognitive load. Language proficiency is known to modulate effects of noise on speech perception and processing, raising the question which linguistic cues are particularly vulnerable in L2 predictive processing, and how different noise types may affect the processing of these cues.

Results from cluster-based permutation analyses and lmer suggest overall effects of noise group and of linguistic condition. As expected, the use of lexical semantic information remains predictive across noise groups. Listening to speech in stationary noise leads to general delays in prediction. We also observed some interactions of noise group x linguistic condition. Comparing prediction in default vs. marked structures, our L2 listeners presented with sentences in stationary noise (compared to quiet) did not show any predictive use of inflection, suggesting processing delays based on perceptual difficulties with low-salient information. Listeners presented with sentences in multi-talker babble, by contrast, failed to predict based on discourse-related cues, suggesting that the use of discourse cues in an L2 is subject to greater difficulties when the integration of information is compromised. In all, these findings suggest that linguistic and acoustic aspects interact in predictive processing in the L2.

P56 Time of exposure for a reliable pupil dilation response to unexpected sounds

Amanda Saksida, Niccolo Granieri, Eva Orzan
Institute for Maternal and Child Health - IRCCS "Burlo Garofolo" - Trieste, Italy

Introduction: Pupil dilation can serve as a measure of auditory attention and as an additional measure of hearing threshold. Studies in infants and adults show a difference in responses to speech and other sounds. It is unknown, however, how much exposure is needed to reliably observe this difference at a comfortable levels of intensity, how reliable is the measure of pupil diameter response (PDR) in individuals at various intensity levels, and whether we can observe systematic differences in the response to the specific type of deviant sounds.

Methods: We observed the PDR to tones and speech (ling-6-sounds) stimuli during passive listening at different intensities in two groups of young adults (N = 24, ME = 29 years, DS = 3.9, 11 females). An oddball paradigm with 20% of deviant sounds was used in both experiments. The time windows, in which the presence of a deviant sound elicited PDR compared to the standard sound across different intensity levels, were estimated by computing the cluster-based statistic using the permuted likelihood ratio tests. The averaged values of these time windows were used to model the group responses and predict individual performance.

Results: In both groups, the augmented PDR was associated with deviant sound stimuli. At the highest tested intensity level (70 dB, reported as comfortable by all participants), the analysis of 10 deviant and 10 standard trials (but not smaller amount of data) yielded reliable model predictions (tones: sensitivity = 0.83; sensitivity = 0.75, positive-predictive-value (PPV) = 0.77; speech: sensitivity = 0.83; sensitivity = 0.5, PPV = 0.63). Averaged raw data per participant yielded even higher PPVs (0.92 and 0.83). Further analysis revealed that in the tone experiment, only high frequency deviant tones (2 & 4 kHz) elicited significant change in PDR, whereas in the speech experiment, consonants (/ss/ and /sh/) but not vowels (/i/, /u/) elicited significant change in PDR.

Discussion: In this study, the minimal amount of exposure to tone and speech stimuli at the comfortable hearing level needed to fit a regression model and to reliably predict the performance in individual participants was measured. This represents the necessary step in creating the PDR-based adaptive procedure with which auditory attention can be measured. We also show that the PDR does not only depend on the type of the sound (speech, noise, tones) but also on the internal categories (e.g. vowels vs voiceless consonants).

P57 The impact of across-frequency coherence on speech intelligibility in young and older normal-hearing listeners

Helia Relaño-Iborra, Torsten Dau
Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Lyngby, Denmark

People with clinically normal hearing thresholds may still encounter challenges understanding speech in adverse conditions, yet identifying such suprathreshold deficits has proven challenging. Motivated by auditory-modeling studies proposing an analysis of across-frequency coherence in the backend of speech-intelligibility (SI) prediction frameworks, we explored a 'distortion-sensitivity' approach to characterize listeners' perception of across-frequency phase coherence as a potential suprathreshold factor influencing speech perception. While a loss of phase coherence across frequency bands, such as through phase jittering, has been shown to be detrimental for SI in young normal-hearing listeners, less is understood about the sensitivity of older individuals and those with listening challenges to such distortions.

We manipulated across-frequency coherence by independently adding a random phase component to each frequency channel of a speech-and-noise mixture presented at a signal-to-noise ratio of 5 dB. SI was measured based on the spread of the distribution of phase values. Additionally, we obtained the just-noticeable-differences (JNDs) for across-frequency phase jitter, as well as speech reception thresholds (SRTs) for undistorted speech in the presence of stationary, fluctuating and speech-like maskers. Data from 10 young and 10 older NH listeners revealed lower JNDs in the older listener cohort, suggesting a reduced sensitivity to across-frequency coherence. However, no changes in the perception of jittered speech was observed. The results from this experimental work will be translated into model parameters, enabling the evaluation of a suprathreshold distortion component within a computational framework for the predicting the measured SRTs.

P58 Speech enhancement in hearing aids using remote microphones

Vasudha Sathyapriyan[1,2], Michael S. Pedersen[1], Mike Brookes[3], Jan Østergaard[2], Patrick A. Naylor[3], Jesper Jensen[1,2]
1. Demant A/S, Copenhagen, Denmark | 2. Department of Electronic Systems, Aalborg University, Aalborg, Denmark | 3. Department of Electrical and Electronic Engineering, Imperial College London, London, UK

Hearing aids (HAs) help people who are hard of hearing to improve speech perception in scenarios such as, speech in background noise, when hearing from a distance or when communicating with speakers with soft voices. Recently, to improve the potential noise reduction performance in HAs, methods that include remote microphones (RMs) to sample a wider spatial field have been proposed in literature. The signal transmitted by the RM have commonly been incorporated as an extra channel by the HA beamformers. Moreover, they incorrectly assume that the RM signal is transmitted instantaneously to the HA unit, in contrast to real world applications. The remote and local HA microphones are located on separate devices and their signals need not be synchronously sampled. Moreover, there will be a time differ-

ence of arrival (TDOA) between the acoustic signal received by the local HA microphones and the wirelessly transmitted signal received by the HA unit from the RM, which varies depending on the distance between the target source and the microphones, the distance between the microphones and the wireless transmission protocol used. As the TDOA increases, the potential benefit of using the RM reduces, thereby rendering them nearly useless. We use the binary estimator selection (BES) strategy that considers the fact there exists a TDOA between the HA and RM signals. We use linear minimum mean-square error (MMSE) filters, to estimate the desired target signal, using the HA and RM signals independently, and use the BES strategy to select between the two estimates in each time-frequency tile based on their corresponding normalized mean-square errors (nMSEs). By doing so, the proposed BES method, picks the estimate with the lower nMSE in each time-frequency tile. The benefit provided by this strategy is to use the RM signal more, when the TDOA is low, and to use the HA signals more when the TDOA is large. Therefore, we use the benefit of both the HA and RM signals, particularly when there is a TDOA between the signals received at the HA unit.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369.

P59 Word comprehension with and without noise: Longitudinal evidence in cochlear implant users using event-related brain potentials

Anja Hahne, Christina Wegewitz, Niki K. Vavatzanidis
TU Dresden, University Clinic, Dresden, Germany

Adult patients who receive a cochlear implant (CI) do not immediately understand spoken language. Rather, it is a complex learning process to gain access to the speech system via the implant. This is particularly true for understanding in a noisy environment, which often does not reach a satisfactory level, even in the long term. Our understanding of this learning process and its neurophysiological basis is rather limited.

In this study, objective brain potential measurements were used to investigate the development of word processing over time. The study was designed as a longitudinal study with measurement points 3 days, 6 weeks, 3 months and 12 months after initial implant activation. 23 post-lingual CI users [mean (SD): 64 (12); median 66, range: 36-81] with severe to profound contralateral hearing loss participated. In addition, a matched typical hearing control group [mean (SD): 64 (12); median 66, range: 33-80] was tested. A picture-word matching paradigm was used, i.e. a picture was shown on a screen accompanied by a spoken word. This word either matched the picture (correct condition) or did not match the picture (incorrect condition). Half of the acoustic words were accompanied by a stationary noise (ICRA1; SNR 5dB). The EEG was recorded and evoked potentials were calculated offline.

For the no noise condition, we observed a significant N400 effect in the ERP (= difference between incorrect and correct condition) already 3 days after first fitting. However, the onset latency was largely delayed. Onset latency decreased systematically over the measurement points. After one year of implant use, the ERP effect of the CI group was similar to that of

the control group. In the noise condition, the first weak and late N400 effect was seen after 3 months of implant use, and even after one year the N400 effect was still later and reduced compared to the no noise condition.

Using objective methods, we were able to follow the process of learning to understand words in CI users during the first year of implant use. While comprehension of words without noise was observed within a few days after initial activation, comprehension of words in noise is much more challenging and not comparable to the control group even after one year of implant use.

Acknowledgements: We thank MEDEL for funding this study.

P60 Development of the Turkish Digits-in-Noise test

Soner Türüdü, Thomas Koelewijn, Deniz Başkent
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Netherlands

Background: The Digits-in-Noise (DIN) test is a closed-set, adaptive, simplified Speech-in-Noise (SIN) test that uses digit triplets in a staircase procedure to measure Speech Reception Threshold (SRT) in noise. DIN was originally developed in Dutch and was later adapted to other languages. This study aims to create new Turkish DIN materials and assess the test-retest reliability of the Turkish DIN test. Further, we investigate the effects of speaker variability and diotic and dichotic sound presentation methods as additional measures for reliability, as at least the latter was clearly shown to affect the DIN outcomes of previous materials.

Methods: We recorded Turkish digits (0-9) from two male and two female native speakers, from which the ten clearest instances for each digit were selected. The selected recordings were then high-pass filtered at 80 Hz to remove low-frequency noise, trimmed, and ramped for onset and offset. In conducting the Turkish DIN test, contrary to fixed triplets used in some DIN tests, our digit triplets are not predetermined but randomly selected from a pool for each digit, ensuring no three digits are identical in any given trial. Each triplet is consistently composed of recordings from the same speaker, with 150 ms silence between digits and 500 ms silence at both ends. For testing, these digit triplets are being presented in a fixed-noise condition, configured to have the same long-term average spectrum as the digits spoken by each speaker. The presentation level is set at 65 dB SPL. The test starts with a -16 dB starting Signal-to-Noise Ratio (SNR), with an initial 4 dB step size, which is adjusted to 2 dB after the first correct response. A 4x2x2 factorial design has been applied to assess the impact of speaker variability (2 male, 2 female) and sound presentation methods (diotic, dichotic). Block randomization has been used to distribute the eight test conditions across two sessions - an initial test and a retest after a 15-minute break - for 15 participants to evaluate the stimuli's effectiveness and the test's reliability through repeated measures.

Results and Discussion: The data collection is ongoing and is expected to be completed by the 15th Speech in Noise (SpiN) Workshop. Upon completion of the data analysis, this study aims to determine the most suitable speaker's recordings for refining the Turkish Digits-in-Noise test in future studies.

P61 A new piano training improves speech-on-speech perception in cochlear implant users

Eleanor E. Harding[1], Etienne Gaudrain[2], Robert Harris[3], Barbara Tillmann[4], Bert Maat[1], Rolien Free[1], Deniz Başkent[1]

1. University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology, Groningen, NL | 2. Lyon Neuroscience Research Center, CNRS UMR5292, Université Lyon, FR | 3. Prince Claus Conservatory, Hanze University of Applied Sciences, Groningen, NL | 4. Laboratory for Research on Learning and Development (LEAD), CNRS UMR5022, Université de Bourgogne

Understanding speech masked by a competing talker, namely, speech-on-speech perception, remains a difficult listening situation for users of cochlear implants (CIs). The performance shows large variability amongst CI users, implying that the CI device may provide sufficient speech cues for the task and as a result, training may help increase the proficiency level. Previous studies with musician populations indicated that music training, via transfer of learning, can provide an advantage for speech-on-speech. In the present study, we used music-based training, a dedicated piano method developed to be suitable for CI users— Guided Audiomotor Exploration (GAME). GAME training is based on finger movements on the keyboard and extemporized combinations of short musical structures. Emphasis on finger movements helps stimulate motor neural networks as part of audiomotor integration, reinforced by improvisation exercises. Further, social interactions with an instructor or in a group setting seem to provide an enjoyable training experience. Twenty-four CI participants were (pseudo-)randomly assigned to one of the three groups: the GAME training for 6 months, a control group with Minecraft lessons, and a control group of no training (only test effects). Minecraft lessons for the first control group were designed and delivered in a way similar to the piano lessons, involving social interaction with a teacher as well as the motor component, but lacking the complex audio feedback involved in music making. Participants were tested before and after training, in a speech-on-speech task based on the coordinate response measure. Listeners had to identify a color and number in a sentence presented simultaneously with another speech stream from the same talker, with the same or altered voice parameters. The results show that the GAME training had a significant positive effect on speech-on-speech perception compared to no-training. In contrast, the Minecraft training showed some small improvement, but this was not significantly different from the no-training group. These results indicate that this type of musical training could be beneficial for CI recipients.

SPEECH
IN NOISE

https://2024.speech-in-noise.eu