SPEECH IN NOISE WORKSHOP

Abstracts

Lancaster, UK | 9-10 January 2025

The 16th Speech in Noise Workshop was chaired by ***Helen Nuttall*** *and* ***Hannah Stewart****, Lancaster University, United-Kingdom, with the help of the organisation committee:*

- *Laurianne Cabrera*
- *Etienne Gaudrain*
- *Antje Heinrich*
- *Chris James*
- *Damir Kovačić*
- *Outi Tuomainen*
- *William Whitmer*

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

https://2025.speech-in-noise.eu/
Contact information: info@speech-in-noise.eu

Date of publication: 7 January 2025.
DOI this version: 10.5281/zenodo.14610784.
DOI all versions: 10.5281/zenodo.14451504.

Cover illustration based on a photo of Castle Hill, Lancaster, by Humphrey Bolton (Creative Commons Attribution-ShareAlike 2.0 Generic CC-BY-SA 2.0)..

The Speech in Noise Workshop is generously supported by:

Programme

Thursday 9 January 2025

08:00-09:00	*Tech check for speakers*
09:15-10:00	Registration, Poster setup & Coffee
10:00-10:10	*Welcome*
10:10-10:35	Phoneme encoding in the inferior colliculus of gerbils, with and without noise induced cochlear synaptopathy **Warren Bakay,** Ana Belén Lao-Rodríguez, Manuel S. Malmierca *Bakay Lab, University of Roehampton, London, UK | Cognitive and Auditory Neuroscience Laboratory (CANELab), Institute of Neuroscience of Castilla y León (INCYL), Salamanca, Spain*
10:35-11:00	**Colin Cherry Award 2024 —** Caught in the cue: Multimodal attentional bias in alcohol and nicotine users **Cosima M. A. Stokar von Neuforn, Patrizia F. Scholz,** Florian Kattner *Health and Medical University, Potsdam, Germany; Health and Medical University, Potsdam, Germany*
11:00-11:25	Speech-in-noise impairments in Alzheimer's disease dementia **Meher Lad** *University of Newcastle, UK*
11:25-11:45	Coffee, Photo, and Poster setup
11:45-12:10	Spatial hearing training for young bilateral cochlear implant users: the BEARS approach **Deborah Vickers,** Marina Salorio-Corbetto, Bhavisha Parmar, Dan Jiang, Helen Cullington, Sandra Driver, Lorenzo Picinali *University of Cambridge, UK*
12:10-12:35	Language acquisition with cochlear implants **Niki K. Vavatzanidis** *Ear Research Center Dresden, Technische Universität Dresden, Germany*
12:35-14:00	Lunch
14:00-15:00	**Keynote —** Towards a multimodal view on the neurobiology of language **Linda Drijvers** *Donders Institute, Radboud University, Nijmegen, Netherlands*
15:00-15:25	The role of individual brain rhythm differences in understanding speech in noise **Anne Keitel** *University of Dundee, UK*
15:25-17:30	Poster session 1 — with coffee
17:30-19:00	*Free time and make your way to Greaves Park (20 mins by Bus 1A, 4, 42 or 100)*
19:00-22:00	Dinner at Greaves Park (Bowerham Road, LA1 3AH)

Friday 10 January 2025

09:00-09:30 *Tech check for speakers*

09:30-11:30 Poster session 2 — with coffee

11:30-11:55 Conversation success and behaviours in multiparty interactions
Raluca Nicoras, Rosa-Linde Fischer, Lauren V. Hadley, Karolina Smeds, Graham Naylor
University of Nottingham, Glasgow, UK

11:55-12:20 Big data insights from hearing aid sound environment classification and smartphone-based self-reported hearing-aid experiences
Charlotte Vercammen, Stefan Launer
Sonova AG, Stäfa, Switzerland | University of Manchester, UK | University of Leuven, Belgium

12:20-13:20 Lunch

13:20-13:45 Association of musical training with speech neural coding and perception
Kelly Whiteford
University of Michigan, Ann Arbor, MI, USA

13:45-14:10 A primer on building wearable vibrotactile devices for improved speech in noise perception
Sam Chin, Gabrielle Ozrech, **Emmie Fitz-Gibbon**
MIT Media Lab, Cambridge, Massachusetts, USA; Brown University, Providence, Rhode Island, USA

14:10-14:35 Weighting of cues to segmental and suprasegmental categorization in quiet and informational masking
Adam Tierney, Ashley Symons
Birkbeck, University of London, UK

14:35-15:00 *Business meeting: Colin Cherry Award 2025, next SPIN meeting and closing remarks*

15:00-15:30 Coffee and Goodbye

15:30-17:00 Tour of Lancaster Auditory Neuroscience Network's research spaces (Whewell Building)

Talks

Thursday 9 January 2025, 10:10—10:35

Phoneme encoding in the inferior colliculus of gerbils, with and without noise induced cochlear synaptopathy

Warren Bakay[1,2], Ana Belén Lao-Rodríguez[2], Manuel S. Malmierca[2]
1. Bakay Lab, University of Roehampton, London, UK | 2. Cognitive and Auditory Neuroscience Laboratory (CANELab), Institute of Neuroscience of Castilla y León (INCYL), Salamanca, Spain

Several studies describe how peripheral processing of auditory information changes after synaptopathy, but only for basic acoustic sounds (e.g., transients, noise, tones). Studies rarely consider speech stimuli, even though this is necessary to ultimately relate speech intelligibility deficits to synaptopathy. We used computer-generated phonemes that enabled the differentiation of temporal fine structure (TFS) and temporal envelope (TENV) coding as stimuli to record responses from individual neurons in the inferior colliculus of gerbils, both controls and animals with noise-induced cochlear synaptopathy (CoSy). These stimuli, including maskers, are identical to those used in collaborative parallel studies in humans and other animal labs.

CoSy has been quantified using a variety of non-invasive measures, in parallel with human studies, including auditory brainstem responses (ABR), distortion product otoacoustic emissions (DPOAE), cortical auditory evoked potentials (CAEP), and validated by immunohistochemistry to identify the number of sensory cells and synapses in the cochlea. Awake, freely moving gerbils were exposed to an octave band of noise (2.6–5.2 kHz) at 100 dB SPL, for 2h, in an antiparallel chamber. During exposures, animals were unrestrained within small cells in a subdivided cage (1 animal/cell).

Here, we present findings categorizing the responses to different speech stimuli from individual neurons in the inferior colliculus of anaesthetized gerbils and relate it to the frequency response areas of each neuron. We further present how various masking conditions affect these responses and present the efficacy of the coding to represent the acoustic stimuli. Finally, we present how noise induced cochlear synaptopathy affects these coding mechanisms and their efficacy.

Funding: Work supported ERA-NET NEURON JTC 2020 (CoSySpeech Project) and the Consejería de Educación y Cultura de la Junta de Castilla y León (grants SA252P20).

Colin Cherry Award 2024

Caught in the cue: Multimodal attentional bias in alcohol and nicotine users

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

Substance users often exhibit an attentional bias towards substance-related cues, which is known to contribute to addiction development and maintenance. This bias can be demonstrated in a dot-probe paradigm, with faster responses to targets that are preceded by spatially congruent substance-related cues. The present study (N=99) investigates biases of auditory attention in alcohol and nicotine by presenting visual substance-related cues either spatially congruent or incongruent with an auditory target stimulus. Specifically, either a pure tone or a spoken sentence was presented to the left or right ear and participants were asked to localize the tone or to provide a true/false judgment to the sentence. Substance-related (alcohol- and nicotine-related images), affective, or neutral visual cues were presented either spatially congruent or incongruent with the auditory target. Based on self-reported alcohol consumption, participants were categorized as low-risk (n = 64) or high-risk (n = 35) drinkers, while for nicotine consumption, they were classified as non-smokers (n = 60), occasional smokers (n = 30), or regular smokers (n = 9). In the tone localization task alcohol-related cues elicited a stronger congruency effect in high-risk drinkers compared to low-risk drinkers. In the sentence judgment task, high-risk drinkers also provided more accurate responses when alcohol-related stimuli were presented congruently. For regular smokers, the opposite pattern was observed: their accuracy was notably higher in trials with incongruent smoking-related stimuli compared to congruent trials. Occasional nicotine consumers displayed a greater attentional bias than non-consumers in congruent trials. These findings are interpreted through the framework of the incentive-sensitization theory, suggesting that both visual and auditory cues may enhance craving responses in substance users. This study highlights the potential benefits of incorporating multimodal cues in interventions designed to reduce craving and modify attentional biases.

Speech-in-noise impairments in Alzheimer's disease dementia

Meher Lad
University of Newcastle, UK

I will explore how auditory cognitive measures, particularly memory for basic sound features, can enhance our understanding of speech-in-noise (SIN) perception and its relationship with cognitive impairment, including Alzheimer's disease (AD). I will discuss how non-verbal auditory tasks provide a valuable alternative to traditional verbal SIN tests, reducing linguistic and cultural biases. My work also examines the reproducibility of these auditory metrics in both in-person and online research settings.

I will present findings from studies involving cognitively healthy individuals, those with mild cognitive impairment, and those with AD dementia, where participants completed SIN tasks alongside auditory memory tests for sound features. Cognitive assessments, audiometry and neuroimaging were also incorporated to study associations between auditory abilities, cognitive function, and brain structure. By analysing multiple predictive models, I will demonstrate that auditory memory for non-verbal sound features is a significant predictor of SIN performance and cognitive status. These findings suggest that auditory cognition could serve as a unique marker for cognitive decline and reflect underlying neurodegenerative changes.

Additionally, I will discuss the implications of remote research methods for auditory testing. By comparing in-person and online approaches, I found that online testing is a reliable and accessible alternative, with reproducible results that align closely with in-person data. Importantly, I will highlight how online methods enable broader participation, though certain demographic factors influence engagement.

Overall, I will aim to provide evidence that auditory cognitive measures are tools that can be used for early detection and monitoring of cognitive decline in Alzheimer's disease dementia.

Thursday 9 January 2025, 11:45—12:10

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

Deborah Vickers[1], Marina Salorio-Corbetto[1], Bhavisha Parmar[1], Dan Jiang[2], Helen Cullington[3], Sandra Driver[2], Lorenzo Picinali[4]

1. University of Cambridge, UK | 2. Guys and St Thomas' Hospital, London, UK | 3. University of Southampton, UK | 4. Imperial College London, UK

Aims: The aims of this research were to develop, through a series of user engagement workshops, the Both Ears (BEARS) training suite, a set of virtual-reality games to train spatial hearing in children and young people (CYP) with bilateral cochlear implants (CIs) and to develop the outcome measures needed to evaluate the effectiveness of the BEARS training suite in a confirmatory clinical trial. Here we focus on the spatial speech-in-noise components of the research.

Methods: Forty CYP (8-16 year olds) with bilateral CIs have helped with the participatory design of the BEARS training suite. This followed an action research methodology with multiple focus groups providing input and feeding back on BEARS games, apps, equipment and assessments to go into the clinical trial. The randomised control trial to evaluate the effectiveness of the BEARS training suite is underway and 140 CYP have completed the baseline assessments of spatial speech-in-noise abilities. These were assessed using the Spatial Speech-in-Noise test (SSiN-VA) and the Adaptive Sentence Lists (Sp-ASLs) which were implemented on an iPad and presented over headphones.

Results: The participatory design phase resulted in an interactive suite of engaging games to train spatial hearing, using speech, music and localisation activities. Following participant feedback, an iPad version was created for individuals for whom using a head-mounted display was difficult due to head size or balance issues. Preliminary results of the BEARS clinical trial

can not be unblinded but initial results with the speech assessments indicate that CYP with bilateral CIs show some degree of spatial hearing at the group level but on average their abilities are lower than their typical hearing counterparts.

Interpretation: The participatory design phase indicated the importance of including patients as co-creators of the BEARS training suite because it has resulted in an intervention that is engaging and meaningful to CYP. The virtual spatial speech-in-noise assessments effectively measure spatial hearing abilities and show promise for implementation into clinical practice as a routine approach for measuring spatial hearing.

Conclusion: The BEARS training suite is made up of multiple virtual reality games to train spatial hearing using a head mounted display and virtual acoustics presented over headphones. They were designed using participatory design and the games appear to be engaging for CYP to use for rehabilitation. Virtual assessments of spatial hearing are showing promise as an effective way to evaluate hearing without the need of a speaker array in a booth.

Thursday 9 January 2025, 12:10—12:35

Language acquisition with cochlear implants

Niki K. Vavatzanidis
Ear Research Center Dresden, Technische Universität Dresden, Germany

Cochlear implants offer an effective technical solution for deaf individuals. Yet, the auditory input they provide remains more limited than typical hearing, particularly in the spectral domain. This talk will explore how these limitations impact language development in deaf toddlers and young children with bilateral implants during their sensitive phase of language acquisition.

I will present recent and prior electrophysiological findings from our studies, focusing on various stages of speech sound processing and, ultimately, word comprehension. Particular focus is placed on congenitally deaf children, whose exclusive reliance on cochlear implants results in consistently noisier auditory input throughout their language acquisition compared to typically hearing peers.

This research aims to illuminate if and how early periods of deafness during the critical phase of language acquisition, coupled with degraded auditory input, influence developmental language milestones and how these outcomes compare to those observed in typically hearing children.

Keynote lecture

Towards a multimodal view on the neurobiology of language

Linda Drijvers
Donders Institute, Radboud University, Nijmegen, Netherlands

Face-to-face communication involves auditory signals, such as speech, and visual signals, such as visual speech and hand gestures. Especially these visual signals can enhance speech comprehension in adverse listening conditions, such as in noise. However, despite the abundance of visual expressions in language, most models and theories on the neurobiology of language are based on characteristics of (clear) speech and text, and they rarely consider multimodal signals. In this talk, I will argue that we need a multimodal view on the neurobiology of language, and that these visual signals are often taken on board immediately by listeners in creating and shaping an interpretation of the linguistic input. This talk will center around the question of how we, as language users, integrate auditory and visual signals into a coherent message, how this is orchestrated in the brain, and how we do this in both clear and adverse listening conditions. I will focus on how this is achieved both within and between conversational partners, and will discuss what is so 'special' about face-to-face communication.

The role of individual brain rhythm differences in understanding speech in noise

Anne Keitel
University of Dundee, UK

Individuals differ vastly in how well they are able to understand speech in challenging listening situations, for example when there is background noise. One explanation for these comprehension differences could be underlying variations in intrinsic brain rhythm patterns. Rhythmic activity at different timescales is ubiquitous in the brain and emerging research suggests that individual differences in these brain rhythms might be able to explain some language-related skills.

In our large population study "Brain Waves for Hearing", we are systematically testing whether individual variations in peak frequencies and strength of brain rhythms are associated with speech comprehension skills, and how peripheral hearing abilities contribute to this relationship. This study will recruit N = 500 participants from diverse backgrounds to address the bias towards underpowered and unrepresentative WEIRD (Western, Educated, Industrialised, Rich, and Democratic) samples in cognitive research. Participants undergo a 10-min resting-state electroencephalography (EEG) recording. This is used to comprehensively model individual intrinsic brain rhythm peaks and amplitudes. Participants also complete a speech comprehension task, in which they verbally repeat sentences that are parametrically modulated with different background noise levels. We hypothesise that the individual amplitude of some intrinsic brain rhythms will be associated with comprehension of noisy speech. This hypoth-

esis is based on a preliminary analysis of a previous dataset, which suggested that a higher individual theta (4-8 Hz) amplitude during rest seemed to be associated with better speech comprehension in noise. I will present first results of our population study.

Apart from contributions to our basic understanding of neural mechanisms of speech comprehension, our results could be used to inform non-invasive brain stimulation approaches to improve speech comprehension, for example with beginning hearing loss.

Friday 10 January 2025, 11:30—11:55

Conversation success and behaviours in multiparty interactions

Raluca Nicoras[1], Rosa-Linde Fischer[2], Lauren V. Hadley[1], Karolina Smeds[1,3], Graham Naylor[1]
1. University of Nottingham, Glasgow, UK | 2. WS Audiology, Erlangen, Germany | 3. WS Audiology, Stockholm, Sweden

In this talk, I will focus on speech as a component of conversation success, presenting results from a series of studies. I will begin by exploring how speech in conversations differs from speech in speech-in-noise tests, highlighting interactive elements such as turn-taking coordination and conversational repairs. I will then discuss the factors contributing to conversation success as perceived by people with normal and impaired hearing, highlighting the roles of speech and the speaker's contributions to successful interactions.

Maintaining conversation success can be particularly challenging in noisy environments, especially for individuals with impaired hearing, who face additional barriers. I will present findings from a lab-based multiparty interaction experiment that investigated how background noise, hearing loss, and hearing aids influence self-reported perceptions of conversation success. Then, I will give examples of behaviours, including vocal activity and head movements, that are linked to interlocutors' ratings of conversation success.

Finally, I will summarise results from a third-party observer study showing how multimodal input shapes perceptions of conversation success. Observers watched a series of successful and unsuccessful conversations with and without audio input. Their assessments provide a unique perspective on how conversational cues impact perceptions of conversation success.

Big data insights from hearing aid sound environment classification and smartphone-based self-reported hearing-aid experiences

Charlotte Vercammen[1,2,3], Stefan Launer[1,4]
1. Sonova AG, Stäfa, Switzerland | 2. University of Manchester, UK | 3. University of Leuven, Belgium | 4. University of Queensland, Australia

Hearing aids continuously analyze and classify the acoustic environment when they are in use. In addition, they track how many hours per day they are switched on. Their data logging thus allows us to estimate typical acoustic days for hearing aid wearers. Most commercially available hearing aids also allow for wireless connection to Bluetooth-enabled devices, such as mobile phones. These ecosystems provide an effective means of collecting (near-)real-time, self-reported feedback from hearing aid wearers, e.g., guided by surveys in a mobile application.

In this presentation, we will presents findings from an exploratory, retrospective analysis of real-world hearing aid data, collected globally. In particular, we will discuss insights from hearing aid sound environment classification and wearing time data of over three million hearing aid wearers, as logged by the software used to fit the hearing aids. We will also discuss insights from over one hundred thousand self-reports, collected from hearing aid wearers through a mobile application as part of their hearing care. These self-reports described positive and negative experiences with hearing technology, matched to how the hearing aid classified the sound environment at the same time the feedback was provided.

Association of musical training with speech neural coding and perception

Kelly Whiteford
University of Michigan, Ann Arbor, MI, USA

Numerous studies have reported links between musical training and enhanced neural processing and improved perception of speech. Such findings suggest a role for experience-dependent plasticity in the early auditory system, which may have meaningful perceptual consequences. However, the robustness and generalizability of any musician advantage remains unclear for several reasons. First, sample sizes have often been small and the samples have represented extreme ends of the musical spectrum; second, the nature and magnitudes of the advantages have often been small or inconsistent; and third, methodological differences and varying analytical techniques have complicated direct comparisons between studies. This multi-site study examined the robustness of the musician advantage across the adult lifespan by replicating and extending eight key experiments involving both perception and neural coding across a large sample of listeners (n > 300) at six universities (Boston University, Carnegie Mellon University, Purdue University, University of Minnesota, University of Rochester, and University of Western Ontario). All participants were tested on all eight experiments in a lab-

oratory setting, including speech and non-speech informational masking, speech perception in noise and babble, and two physiological measures of fundamental-frequency encoding of speech sounds using electroencephalography. Participants completed additional measures to control for potential confounding factors, including a measure of musical aptitude, a cognitive assessment (Ravens Advanced Progressive Matrices), a measure of extended high-frequency hearing, and survey questions related to personality and socio-economic status. This talk will focus on a subset of the results related to speech neural coding and perception. All five previously published findings related to enhanced speech neural encoding in musicians failed to replicate. Of the three speech perception measures tested, only one remained significant after controlling for cognition and age. The results provide in-depth and highly statistically powered insights into the nature and robustness of the musician advantage across the adult lifespan.

Acknowledgements: Supported by NSF-BCS grant 1840818 and NIH grant R01 DC005216.

A primer on building wearable vibrotactile devices for improved speech in noise perception

Sam Chin
MIT Media Lab, Cambridge, Massachusetts, USA

Gabrielle Ozrech
San Jose State University, San Jose, California, USA

Emmie Fitz-Gibbon
Brown University, Providence, Rhode Island, USA

Recent research has demonstrated the efficacy of vibrotactile feedback in enhancing speech perception in noisy environments, benefiting both individuals with normal hearing and those with hearing impairments (Guilleminot & Reichenbach, 2022, doi:10.1073/pnas.2117000119; Rautu et al., 2023, doi:10.1038/s41598-023-43644-3). This approach shows particular promise for populations with age-related hearing loss and cochlear implant users, where additional sensory input can complement auditory processing. However, current vibrotactile solutions often rely on bulky motors or expensive commercial devices that require users to hold the haptic device,, limiting their practical application and accessibility. This approach also misses the opportunity for multi-tactor haptic configurations, which have been demonstrated to improve information throughput in other haptic applications (Tan et al., 2020, doi:10.1109/JPROC.2020.2992561).

We hope to lower the barrier-to-entry of additional exploration in this field by presenting cost-effective approaches to developing multi-channel haptic devices specifically designed for speech-in-noise applications. We outline methods for constructing affordable, wearable solutions that can be readily implemented by researchers in the speech perception field. We share our design and also provide some options for haptics that use commercial-off-the-shelf parts that can be purchased at large online retailers. Our suggested implementation leverages existing audio signals to drive the haptic feedback, eliminating the need for complex signal processing chains or additional sensors. The system architecture employs a multichannel design

that enables straightforward synchronization across tactors, crucial for maintaining temporal relationships in speech signals and spatial cues. We achieve this through direct audio-to-haptic conversion with latency below 10ms, ensuring that vibrotactile feedback remains perceptually aligned with acoustic input. This approach allows researchers to readily integrate the system with existing audio processing pipelines while maintaining the temporal precision necessary for speech enhancement applications. Through a systematic review of existing haptic design principles and best practices, we share guidelines for optimizing vibrotactile feedback in speech enhancement applications (Orzech et al. 2024, doi:10.1177/107118132412759).

Additionally, we present the results of a user survey exploring preferences for wearable haptic devices in hearing augmentation, specifically comparing head-worn and wrist-worn configurations designed for speech-in-noise perception and sound localization. Our study includes participants across different hearing profiles, from those with mild hearing loss to users of hearing aids and cochlear implants. We examine key factors including user comfort, signal perception, and integration compatibility with existing assistive hearing devices. Our results provide insight into user preferences and design considerations that may inform future development of haptic solutions for auditory rehabilitation.

Weighting of cues to segmental and suprasegmental categorization in quiet and informational masking

Adam Tierney
Birkbeck, University of London, UK

Ashley Symons
Royal Holloway, University of London, UK

Background: Speech is a highly redundant signal: any given feature is communicated across multiple acoustic cues. Initial consonant voicing, for example, is conveyed by voice onset time and the F0 of the subsequent vowel (among other cues), while word emphasis is conveyed by increases in pitch, amplitude, and duration. This redundancy makes speech robust, with comprehension possible even when the signal is distorted. However, information being spread across different cues poses a problem for listeners: how to make sure that the most important source of information is prioritized. Attention-to-dimension models suggest that selective attention is directed towards highly weighted cues and away from less useful cues. One way to test this hypothesis is to investigate whether relative cue weighting changes depending on whether target speech is presented in quiet or during informational masking. We tested this hypothesis by examining categorization of consonant voicing and word emphasis in quiet and in the presence of a single distracting talker.

Methods: Participants were presented with target speech to be categorized in one ear while either silence or a distracting talker was presented to the opposite ear. Stimuli were sampled from two-dimensional stimulus grids in which pitch and duration varied independently in the extent to which they conveyed whether a word was emphasized (Experiment 1) or whether a consonant was voiced or unvoiced (Experiment 2). Mixed-effects logistic regression was used to measure the extent to which pitch versus duration cues were weighted across individuals and conditions.

Results: Across both voicing and word emphasis categorization, individuals who tended to weight a single cue highly in quiet switched strategies under informational masking, moving towards integration across cues. Individuals who already used an integrative strategy did not, however, change their weighting because of distraction.

Conclusions: We find that perceptual strategies can, in some individuals, differ between quiet and distracting listening environments. Listeners seem to have difficulty solely relying upon a single primary cue during distraction, which supports the hypothesis that cue weighting draws upon selective attention. Moreover, these results suggest that perceptual strategies measured in quiet are not necessarily indicative of the ways in which listeners sample information across acoustic channels in more ecologically valid, complex listening situations.

Posters

SESSION 1: Thursday 9 January 2025, 15:25-17:30

SESSION 2: Friday 10 January 2025, 09:30-11:30

P01 Revisiting the speech intelligibility index with speech-in-noise perception

Jordi Boons, Gertjan Dingemanse, André Goedegebure
Department of Otorhinolaryngology, Head and Neck Surgery, Erasmus MC, Rotterdam, Netherlands

Background: In previous work, we found that the pure-tone average (PTA) had only a minimal effect on speech perception of the digits-in-noise (DIN) test when speech was presented well above pure-tone thresholds. However, a substantial association between PTA and DIN was found when the speech intelligibility index (SII) was used to account for audibility rather than level- and spectral adjustments during stimulation. This relationship may be driven by residual audibility effects that are not fully accounted for by the SII.

Rationale: Accurate audibility estimation is essential for investigating factors that influence speech understanding, such as cognition and suprathreshold processing. Adjusting for audibility ensures that differences in speech understanding can be attributed to factors of interest rather than to variations in hearing sensitivity.

Methods: Speech reception thresholds (SRTs) were measured with three DIN test variations. From 2011 to 2014, 3455 participants had a measurement with a fixed noise level of 55 dB SPL (DIN55), and from 2014 to 2016, 1513 participants were measured at 70 dB (DIN70). Of these, 711 also completed the DINvar test, which ensured audibility with spectrally adapted speech. We used the SII in two ways. SIISRT calculated the audibility of the speech-in-noise signal, while SIIspeech reflected audibility of the speech relative to hearing thresholds. Given the non-linear relationships between the SRT/SII and PTA, Generalized Additive Models (GAMs) were used to allow flexible modeling by fitting smooth functions.

Results: The relationship between SIISRT and PTA was stronger for DIN55 ($R_{adj}^2 = .16$, $F(5.09, 6.13) = 103.8$, $p < .001$) compared to DIN70 ($R_{adj}^2 = .11$, $F(4.41, 5.44) = 32.8$, $p < .001$). With DINvar, a weak relationship between SRT and PTA was found ($R_{adj}^2 = .01$, $F(2.57, 3.25) = 3.55$, $p = 0.01$). Among participants with valid DINvar measurements, of which PTAs are limited to 40 dB HL, the SIISRT and SRT of the DIN70 test were nearly perfectly correlated ($R_{adj}^2 = .97$, $F(4.1, 5.12) = 4342$, $p < .001$). We validated this with DIN55 for participants with a similar maximum PTA level relative to the sound level ($R_{adj}^2 = .94$, $F(5.42, 6.6) = 5900$, $p < .001$). After a linear correction with the SIISRT with the SIIspeech in these subsamples, only weak relationships with PTA were found (DIN55: $R_{adj}^2 = .07$, $F(2.43, 3.07) = 62.91$, $p < .001$ and DIN70: $R_{adj}^2 = .02$, $F(1, 13.64) = 13.64$, $p < .001$).

Conclusions: The SII may underestimate audibility effects in speech-in-noise perception. The relationship between SIISRT and PTA could be explained by the decrease in speech audibility relative to hearing thresholds at higher PTAs. The SII model should be adapted or extended to incorporate this audibility effect.

P03 Prosocial motivation and reward relates to speech perception in noise performance: Insight into the autistic listening experience

Ryan J. Oakeson, Jesca Herbert, Simon Roper, Hongyi Zhang, Stuart Rosen, Sophie Scott
University College London, UK

This study explored how prosocial motivation and social reward relate to speech perception in noise (SPiN) in a control group (N = 136). We investigated SPiN performance and subjective listening experiences across different masker conditions: 1-speaker, 2-speaker, and steady-state speech-shaped noise (SSN). We conducted a PCA on social trait questionnaires and created two dimensions of social reward differing only in prosocial reward. Results indicate that individuals who rated themselves higher in prosocial reward traits performed better in the condition that yielded the highest threshold out of the three conditions i.e., the 2-speaker condition. Age significantly influenced subjective ratings of listening effort across all conditions where older participants reported greater effort. These findings highlight prosocial motivation as a potential factor influencing SPiN abilities, particularly in complex listening scenarios, alongside the role of age in shaping subjective auditory experiences.

We also collected data in an autistic group (N = 27) using the same methods and procedures along with some additional measures. Findings in this sample showed SRTs in the 1-speaker condition were predicted by the individual's musical perception skills (MINI-PROMS) and SRTs in the SSN condition were predicted by the listener's working memory scores from a reading span task. Interestingly, working memory scores anticorrelated with ratings of prosocial reward whereas SSN SRTs correlated with a prosociality scale. In other words, higher prosocial ratings were associated with worse performance in SSN and worse working memory performance. These findings demonstrate a seemingly novel and distinct tradeoff in autism between prosocial orientation and cognitive ability when performing listening tasks amidst energetic noise.

P04 Assessment of listening effort for cochlear implant users with pupillometry and subjective ratings: A comparison of three speech-in-noise tests

H. Christiaan Stronks[1,2], Paula L. Jansen[1], Robin Van Deurzen[1], Jeroen J. Briaire[1], Johan H. M. Frijns[1,2,3]

1. Leiden University Medical Center, Netherlands | 2. Leiden Institute for Brain and Cognition, Netherlands | 3. Delft University of Technology, Netherlands

Background: Cochlear implant (CI) users experience listening as effortful, and especially in background noise. Listening effort can be measured with pupillometry, as based on the observation that the pupil dilates more when listening is more challenging. To assess listening effort in different noisy environments, a speech-in-noise test is needed where the magnitude of pupil dilation is dependent on the SNR. In contrast to typical hearing listeners (TH), however, PPDs have proven to be quite insensitive to SNR for CI users.

Objective: We have compared three speech corpuses to assess which one is most sensitive to measure the pupil response as a function of SNR for CI users. These were digits in noise (DIN), Matrix sentences, and LIST sentences.

Methods: Peak pupil dilation (PPD) was recorded for 18 CI users when listening to DIN triplets, Matrix sentences and LIST sentences at three different SNRs, namely at the speech recognition threshold where SI was 50% (0 dB re SRT), at + 6 dB re SRT, and in quiet. The DIN material consists of triplets drawn from a closed set of 10 digits; the Matrix sentences consist of 5 words with a fixed syntactic structure that provide no context; the LIST sentences are an open set of sentences that were developed especially for CI users.

Results: PPDs were significantly dependent on SNR for the DIN triplets and LIST sentences, but not for the Matrix sentences. For the DIN triplets, the PPD differed significantly between 0 dB re SRT and quiet, and between +6 dB re SRT and quiet. By contrast, for the LIST sentences the PPD was significantly different between 0 dB re SRT and quiet and between 0 dB re SRT and +6 dB re SRT.

Conclusions: The LIST sentences are useful for measuring listening effort with pupillometry in CI users around SRT, whereas DIN triplets are useful at more favorable SNRs. The Matrix sentences yielded PPDs that were unresponsive to changes in SNR and it proved to be a challenging listening test even in quiet.

Funding: This work was funded in part by the Dutch Research Council (NWO) within the Crossover Research program under the 'INTENSE' project (grant #17619) and co-funded by Advanced Bionics (Hannover, Germany).

Acknowledgements: We thank the study participants and Nicolas Furnon (Advanced Bionics) for technical support.

P05 Are you interested in making a monaural speech intelligibility model binaural?

Mathieu Lavandier
ENTPE, Ecole Centrale de Lyon, CNRS, LTDS, UMR5513, Lyon, France

James M. Kates, Kathryn H. Arehart
Department of Speech Language and Hearing Sciences, University of Colorado, Boulder, CO, USA

When a target speech source is spatially separated from competing sound sources, the target intelligibility is improved thanks to binaural hearing, compared to situations with co-located sources or monaural listening. This spatial release from masking (SRM) cannot be predicted by monaural speech intelligibility models. We are presenting here a binaural front-end that can be combined with monaural models to account for SRM. This front-end is based on the MBSTOI metric. It is implemented in the temporal domain, so that from the noisy speech signals at the two ears it produces a binaurally-enhanced monaural signal that can then be evaluated by monaural intelligibility models.

The front-end was tested here in combination with the monaural model HASPI (Hearing Aid Speech Perception Index) that allows intelligibility predictions for speech degraded by additive noise, reverberation, spectral changes, and nonlinear distortion. HASPI compares the degraded noisy speech signal to a clean speech reference, and accounts for hearing impairment by incorporating a model of the auditory periphery that can represent both impaired and normal hearing.

The model predictions, with or without the use of the binaural front-end, were compared to intelligibility scores from three datasets all involving sound reproduction with headphones, normal-hearing listeners, anechoic conditions, and a frontal (Danish) speech source. In dataset 1, a single stationary speech-shaped noise (SSN) was tested at ten azimuths around the listener at six signal-to-noise ratios (SNRs). In dataset 2, three azimuths and six SNRs were tested for an SSN or a non-stationary noise, with or without ideal binary mask (IBM) processing, which simulates noise reduction in hearing aids. In dataset 3, the competing sound was obtained by mixing in seven different proportions the ear signals resulting from three sources: an SSN simulated at 0-degree azimuth (co-located with the target speech), a diffuse noise coming from all directions, and an SSN simulated at 115-degree azimuth. This noise mixture was tested at eight SNRs.

Because HASPI was originally developed to predict percent correct for English sentences at positive SNRs, the model back-end was re-fitted to predict percent correct for Danish words measured at negative SNRs. This new fitting was done using only the co-located conditions of the three datasets, in which there was no SRM or binaural effects involved.

P06 Contribution of cognitive functions, musical abilities and auditory scene analysis in speech-in-noise perception

Elena Benocci, **Luna Prud'homme**, Axelle Calcus
Université Libre de Bruxelles, Belgium

10.17605/OSF.IO/G9SK7

From the latest trendy café to the poster hall at conferences, we evolve in noisy environments. Even accustomed to these experiences, not all individuals are equal when facing these challenges. Speech comprehension in noise is often thought to be influenced by sound stream segregation and selective attentional tracking, the two aspects of auditory scene analysis, but also by individual factors, such as cognitive, linguistic and/or musical abilities. The aim of this study was to examine the respective contribution of auditory scene analysis, perceptual musical abilities, memory, attention and vocabulary on speech intelligibility in noise. 80 adults participated in six tasks selected to assess their performance in those abilities. Structural equation modeling showed that perceptual musical abilities predicted auditory stream segregation and memory. These factors, in turn, predicted selective attentional tracking, which was a predictor of speech perception in noise. Performance in the attention and vocabulary tasks were not associated with the other variables. Our results highlight a complex relationship linking low-level factors to speech perception in noise via auditory scene analysis.

P07 Lyric intelligibility of musical segments for older individuals with hearing loss

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

The Cadenzateers
Universities of Leeds, Nottingham, Salford & Sheffield, UK

Background: Difficulty in understanding lyrics can be a major barrier to music enjoyment for people with hearing loss (Greasley et al., 2020, doi:10.1080/14992027.2020.1762126). One of the aims of the CADENZA project is to run machine-learning challenges to optimise technology to improve lyric understanding without sacrificing overall enjoyment of the music. To achieve this optimisation, developers need lyric-intelligibility metrics that are informed by the experiences of the target population. Currently, there are no data on the ability of older individuals with hearing loss to understand and recall lyrics.

Methods: Thirteen older participants with mostly mild-sloping hearing loss listened to and recalled 100 segments of popular music that varied in genre, duration and word count. These segments were used in a previous lyric intelligibility study with younger students without hearing loss. In each trial, participants heard a randomly chosen segment that was presented twice with a 5-s interstimulus interval over headphones at an A-weighted level of 65 dB plus individualised frequency-dependent gain (CAMFIT-nonlinear).

Results: The proportion of words heard correctly varied greatly across samples, from 0-100%, and varied as a function of genre. In comparison with the previous study using these stimuli with a different population, the results were reasonably well correlated, but there were several samples that were either easier or more difficult to recall for the older participants with HL. Individual intelligibility across samples was correlated with age; sample intelligibility across individuals and genres, however, was not correlated with word count or rate.

Conclusions: Improving the intelligibility of sung lyrics for those with hearing loss is a different challenge from spoken-speech enhancement. Not only does the enhancement need to be considered within the overall enjoyment of the music, but the variation in results — as seen in the current study — reflects variations in genre, orchestration and vocal quality in sung music.

P08 The impact of selective attention and musical training on the cortical speech tracking in the delta and theta frequency bands

Alina Schüller, Annika Mücke, Jasmin Riegel, Tobias Reichenbach
Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

10.1101/2024.08.01.606154

Understanding speech in noisy environments relies on the brain's ability to selectively attend to a target speaker, filtering out background distractions. Research suggests that musical training enhances this ability, likely due to improved auditory processing, but the underlying neural mechanisms remain unclear. Neural activity in the auditory cortex tracks the speech envelope, particularly within the delta (1–4 Hz) and theta (4–8 Hz) frequency bands, which are crucial for processing speech in noisy settings.

In this study, we examined the effects of selective attention and musical training on cortical speech tracking using magnetoencephalography (MEG) to record neural responses in 52 participants, both musicians and non-musicians. Participants focused on one of two competing speech streams, allowing us to measure how attention and musicianship influence neural tracking in the delta and theta bands.

Our analysis revealed that selective attention significantly modulates delta-band tracking, enhancing neural responses to the attended speech stream. Theta-band tracking, however, did not show significant modulation by attention, suggesting its primary role in lower-level acoustic processing, such as syllable parsing, regardless of focus. This aligns with findings that delta-band activity supports higher-level linguistic processing.

Interestingly, musical training did not significantly affect speech tracking in either the delta or theta bands. Both musicians and non-musicians exhibited similar neural responses, challenging assumptions that musicianship enhances all aspects of auditory processing. These findings suggest that while musicians benefit from enhanced auditory fitness, these advantages do not significantly affect cortical speech tracking. This is consistent with studies showing musical training enhances subcortical responses, with limited effects on cortical processing.

In conclusion, our study highlights the importance of delta-band activity in selective attention during speech perception in noise, while theta-band activity plays a more consistent role across both target and distractor speech streams. Furthermore, our findings challenge the view that musical training significantly alters cortical mechanisms of speech tracking, suggesting that its effects may be more limited to other auditory processing domains. These findings suggest that musical training's effects may be more limited to other auditory processing domains.

P09 The relationship between auditory processing, speech processing, language, and cognition in typically developing children and children with listening difficulties

Xuehan Zhou[1,2], Harvey Dillon[3,1], Dani Tomlin[2], Kelly Burgoyne[4], Helen Gurteen[5], Grace Nixon[2], Alisha Gudkar[2], Antje Heinrich[1]

1. Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, UK | 2. Department of Audiology and Speech Pathology, University of Melbourne, Australia | 3. Department of Linguistics, Macquarie University, Sydney, Australia | 4. Manchester Institute of Education, The University of Manchester, UK | 5. University of Queensland Centre for Hearing Research, University of Queensland, Brisbane, Australia

Background: Children with listening difficulties (LiD) often face challenges in understanding speech in noisy or reverberant environments, which has the potential to affect their academic performance in the long term. These difficulties may stem from deficits in hearing, auditory processing, language and/or cognitive ability, which complicates the clinical diagnosis. This study applied a structured tri-level listening test battery, ranging from non-speech auditory resolution, nonword identification in noise and reverberation, to sentence comprehension in noisy and reverberant classroom simulations. In addition to the tri-level test battery, language and cognitive assessments were included to explore the relationships between these abilities in typically developing children and those with LiD. The study aims to explore what affects speech perception in noisy and reverberant classrooms in typically developing children and those with LiD. The ultimate goal is to identify the underlying causes of LiD to support targeted clinical interventions.

Methods: Over 200 typically developing children from primary schools and 80 children with parent-reported LiD from clinics, all aged 6 to 12 years and with normal hearing, were recruited for the study. The clinical group included children experiencing a range of listening challenges, such as understanding speech in noisy environments. Participants completed a comprehensive assessment over 2–3 sessions. This assessment evaluated their abilities in a tri-level test battery (including non-speech auditory resolution, nonword identification and sentence comprehension in noise and reverberation), language tests (including auditory cloze tests and sentence recall) and cognitive tests (including memory, attention, and non-verbal intelligence).

Results: In both typically developing and clinical groups, non-speech auditory resolution abilities significantly predicted nonword identification in noise and reverberation. Nonword identification in noise and language abilities significantly predicted sentence comprehension in noise and reverberation. In the clinical group, language abilities were significantly associated with nonword perception in noise, a pattern not observed in the typically developing group. Furthermore, in the clinical group, better performances on the memory task associated significantly with higher sentence comprehension in noise scores.

Conclusion: Understanding children's listening in noise and real-life scenarios requires considering how auditory, speech, language, and cognitive abilities impact both task performance and everyday challenges. Compared to the typically developing group, the LiD group showed a significant association between language and memory abilities with nonword and sentence in noise tests performance, indicating a greater effect of cognitive and linguistic skills for the clinical population. These findings indicate that bottom-up auditory resolution skills and top-down cognitive skills, including language abilities and memory capacity, may affect speech understanding in noisy environments for LiD children. Incorporating multidisciplinary assessments into routine audiology evaluations can help to identify specific deficits, tailor interventions to individual needs, and enhance targeted strategies to improve outcomes for children with listening challenges.

P10 Development of speech recognition in noise in school-aged children: Bottom-Up and Top-Down Proficiency Factors

Theo Goverts[1,2], Koenraad S. Rhebergen[3], Alieke Breure[1,2], Marre Kaandorp[1,2], Cas Smits[4,2]
1. Department of Otolaryngology - Head and Neck Surgery, Amsterdam UMC, location vUMC, Netherlands | 2. Section Ear & Hearing, Amsterdam Public Health Research Institute, Netherlands | 3. Department of Otorhinolaryngology and Head & Neck Surgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, Netherlands | 4. Department of Otolaryngology - Head and Neck Surgery, Amsterdam UMC, location AMC, Netherlands

Background: Children require higher Signal-to-Noise Ratios (SNRs) than adults to achieve equal performance. These performances can be compared using a Proficiency Factor (P), defined as the quotient of the audibility (Speech Intelligibility Index, SII) that is apparently needed and the reference SII. P can be considered as the product of P- Bottom Up (BmUp) and P-Top Down (TpDn), which may both be developing in children. These Proficiency Factors can be used to predict the acoustic requirements (SII or Speech Transmission Index, STI) for children in for example classrooms.

Aims: The aim of this study was to investigate the development of the SRT in school-aged children for relatively "simple" DIN stimuli (Digit-In-Noise), and more complex sentences with keyword scoring (Sentences-In-Noise, SIN-kw). Furthermore, we examined/analyzed age effects of P-BmUp and P-TpDn for these tests.

Methods: The SRT of the DIN and SIN-kw were measured with an adaptive procedure at 50 and 85 % target points in 48 school-aged children (age 4-12) with normal hearing and typical development. Sixteen adults with normal hearing were included for reference data. The rela-

tions between outcome measures and age were explored. For each outcome measure, P was estimated as the SII divided by the reference SII. P-BmUp was based on the SRT at 50% DIN target point. P-TpDn was determined by comparing the total P to the P-BmUp.

Results: Analysis of variance showed a significant effect of age for all outcome measures, with better scores for older ages. DIN_85 performance of the children could be fully predicted by P-BmUp. For SIN-kw_50, an extra age-related P-TpDn was needed. The SIN-kw_85 results could be described using the SIN-kw_50-based P-TpDn and the DIN_50-based P-BmUp.

Discussion and Conclusion: Speech recognition develops with age in school-aged children. For DIN, the results can be described by the development of Bottom - Up Proficieny. SIN-kw can be predicted using an additional Top -Down Proficiency. These Proficiency Factors can be used to adjust acoustic requirements (SII or STI) for children. School-aged children need substantial higher SNR (i.e. SII or STI value) for effective speech reception in noise perception compared to older normal hearing listeners.

P11 Asymmetric pupil response to speech and music in toddlers with cochlear implants

Amanda Saksida[1], Sara Ghiselli[2], Eva Orzan[1]
1. Institute for Maternal and Child Health - IRCCS "Burlo Garofolo" - Trieste, Italy | 2. Ospedale di Piacenza, Italy

Left ear advantages for tonal stimuli and music, indicating prevalently right hemisphere processing, and conversely, right ear advantages for speech, indicating left hemisphere processing, have been found repeatedly, from birth on. The monaural capacity to attend to and process speech or music may be useful in understanding the asymmetric processing strategies also in bilateral cochlear implant (CI) users. Namely, regardless of the reported benefits of early binaural implantation, children with CIs experience overall poorer speech perception in noise, and poorer music perception. While toddlers with bilateral CIs exhibit increased attention to speech, indexed by increased pupil dilation, only in the absence of significant background noise, no study has explored the monaural capacities to attend to and process speech and music in CI users, nor adults nor children. The present study explores pupillary behavior of 12 early bilaterally implanted congenitally deaf toddlers (1.5-4-y-old) while they were passively listening to speech (rhymed verses) or instrumental music at a constant intensity level (60 dB SPL), in quiet or with a 10 or 0 dB SNR babble noise, in an ecological environment (Ambisonics semi-sphere), with only the left or only the right cochlear implants switched on. The analysis of variance for the linear mixed model fitter by REML shows significant interaction between the stimulus (music, speech) and listening ear (left, right), irrespective of the background noise level. The overall increase in pupil dilation for music in right ear listening, and for speech in left ear listening, may indicate increased listening effort when listening to the respective stimuli with the "non-advantageous" ear. These asymmetries in pupil response may in turn reflect the hemispheric asymmetries in processing speech and music in early bilaterally implanted children, indicating that their rehabilitation protocols could take advantage of processing differences between the two ears.

P12 A data-limit account of spatial and spectral release from masking

Sarah Knight[1], Yue Zheng[2], Georgie Maher[2], Ronan McGarrigle[3], Sven Mattys[2]
1. Newcastle University, Newcastle upon Tyne, UK | 2. University of York, UK | 3. University of Leeds, UK

Speech-on-speech listening involves selectively attending to a target talker while ignoring a simultaneous competing talker. Spatially separating the talkers improves performance, a phenomenon known as spatial release from masking (spatial RM). However, it is also possible to improve performance by spectrally separating the talkers, i.e., filtering them into non-overlapping frequency bands (spectral RM). In both cases, RM benefits derive from enhanced availability of the target signal.

The relative benefit of spatial vs. spectral RM is currently unknown. Furthermore, it is unclear how listeners' ability to exploit spatial vs. spectral cues is related to individual differences in cognitive abilities. It has been suggested that cognitive resources are of greater utility when less of the target signal is available, implying that the cognition/performance relationship should be strongest when spatial and/or spectral separation (i.e., RM) is limited or absent. However, the data-limit account (Norman & Bobrow, 1975, Cognitive Psychol 7(1):44) suggests that cognitive resources cease to be of use when the target is severely degraded, implying that the cognition/performance relationship should in fact be weakest when there is no RM.

In this study, participants (N=240) completed a selective listening task in which they transcribed the speech of one of two simultaneously-presented talkers. We filtered the speech into frequency bands such that the talkers were either spectrally overlapping or interleaved (spectral RM vs. no spectral RM). We also manipulated perceived spatial distance between talkers, presenting them either at +/-90° azimuth (dichotic) or collocated (diotic) (spatial RM vs. no spatial RM). We additionally administered a battery of cognitive tasks to assess three key components of WM/attention: phonological loop, executive function and selective/divided attentional control. Factor analysis was used to derive a single cognitive score for each participant.

Spectral RM was at least as effective as spatial RM in improving transcription performance, with the best performance observed when both types of RM were present. Cognitive scores were significantly positively correlated with both spatial and spectral RM benefits. Additionally, cognitive scores best predicted performance in the three RM conditions, with the weakest correlation observed in the condition with neither spatial nor spectral RM.

These results suggest that listeners can gain as much benefit from spectral as spatial cues during speech-on-speech listening. These RM benefits appear to be supported by cognitive processes, with larger RM benefits associated with better cognition. Finally, cognitive abilities were least predictive of performance when no RM was present, supporting a data-limit account.

P13 Investigating the second-language effect and hearing acuity on Digits-in-Noise test outcomes

Soner Türüdü, **Thomas Koelewijn**, Deniz Başkent
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, Netherlands | Research School of Behavioral and Cognitive Neuroscience, Graduate School of Medical Sciences, University of Groningen, Netherlands

Background: The Digits-in-Noise (DIN) test is an adaptive speech-in-noise assessment that measures speech reception thresholds (SRTs) using digit triplets. While DIN tests are available in multiple languages, studies investigating the effect of second-language proficiency and its combined effect with hearing loss on DIN test outcomes are limited. This study investigates the effect of second-language proficiency on DIN outcomes by comparing Turkish and Dutch DIN tests in native Turkish speakers with varying Dutch language proficiency and different hearing levels.

Methods: In this ongoing study, so far nine native Turkish individuals between the ages of 18-65 yr were included. Participants' Dutch proficiency levels range from A0 to C1 according to the Common European Framework of Reference for Languages (CEFR). Participants underwent pure tone audiometry (0.25-8 kHz), performed the Turkish Child-friendly Coordinate Response Measure (T-CCRM) test, and DIN tests in both Turkish and Dutch. The DIN test implementations had two variables: (1) Sound presentation, which could be diotic or antiphasic; and (2) Speaker, including Dutch male, Turkish male, or Turkish female. For all six tests, noise level was fixed at 65 dB SPL and the starting SNR was -6 dB.

Results and Discussion: Preliminary data of participants with varying levels of hearing acuity indicate observable trends (with no statistical analyses yet). Turkish DIN SRTs, for both male and female speakers, seem slightly lower than Dutch DIN SRTs. Dutch language proficiency seems to positively affect the Dutch DIN test outcomes. Increased hearing loss seems to negatively affect DIN SRTs in both languages, aligning with expectations. Antiphasic sound presentation resulted in lower DIN SRTs compared to diotic. T-CCRM scores were generally within normal ranges but showed decreased performance at an SNR of –6 dB, possibly due to the influence of participants with hearing loss. Data collection is ongoing, and these findings will be expanded and re-evaluated with additional participants by the 16th Speech in Noise (SpiN) Workshop.

P14 Consonant and word identification in noise in 5-to-10 year old children with normal hearing

Laurianne Cabrera[1], Charlotte Benoit[2,1]
1. Integrative Neurosciences and Cognition Center, CNRS-Univ Paris Cité, France | 2. Assistance Publique-Hôpitaux de Paris, France

Speech-in-noise perception abilities continue to develop until adolescence. The mechanisms underlying this rather long development are still not clear. The aim of this study was to evaluate the developmental trajectory of speech-in-noise perception in normal hearing children.

Three groups of normal-hearing children were included: 5-6 years old (N = 28), 7-8 years old (N = 20), 9-10 years old (N = 20); as well as an additional group of young adults (N = 20). All the participants completed two adaptive psychophysical tasks measuring their consonant and word identification thresholds in two types of noise conditions: a steady speech-shaped noise and a 2-talker babble noise.

Results revealed that speech identification thresholds of adults are significantly better than those of children regardless of their age in the presence of the 2-talker babble noise. In the presence of the speech-shaped noise, the consonants identification thresholds are similar to those of adults at the age of 7-8 years, and from the age of 5-6 years for words.

These results suggest different developmental trajectories for speech-in-noise perception abilities depending on the type of speech stimuli to process (syllables vs. words) and type of noise, with children showing similar thresholds to adults at different ages in each task. This could imply that the specific mechanisms underlying masking release continue to develop until late childhood.

P15 Exploring listening efficiency in a lexical decision task as a measure of hearing-aid outcomes at realistic signal-to-noise ratios

Laura Winther Balling
ORCA Labs, Lynge, Denmark

Frederic Marmel
ORCA Labs, Stockholm, Sweden

To better understand the difficulties of listeners with hearing loss and the performance of hearing aids (HA), there is a need to develop outcome measures that can uncover differences (between listeners or between HA signal processing schemes) at the kind of moderately positive SNRs that are prevalent in everyday situations. Standard speech tests are often not useful at these realistic SNRs because accuracy is near ceiling. Moreover, standard speech tests do not capture differences in how much effort HA wearers have to exert, a crucial factor for living with hearing loss.

Outcome measures that integrate accuracy and effort, a concept referred to as “listening efficiency”, hold the promise of being able to discriminate between listening conditions even if they show almost equal intelligibility. Listening efficiency has previously been implemented as the ratio of accuracy to response times, and more recently as differential evidence accumulation rates estimated in a cognitive model of decision making that fits accuracy and response times jointly. In this study, we aimed to explore the concept of listening efficiency and its ability to reveal effects of HA signal processing and SNR on the performance of HA wearers, using the auditory lexical decision task (LDT). This is a well-established task in experimental psychology, and because it measures word recognition, it taps into many of the same cognitive processes as the speech tests more typically used in audiology. Participants are presented

with a mixture of words and nonwords and are asked to respond as quickly and as accurately as possible whether each is a real word. Response times may be taken to reflect their underlying capabilities and the effort they exert.

Nineteen experienced HA wearers completed an LDT presented over a fixed 60-dB HINT noise at an SNR of either 5 or 10 dB, with the test HAs set to an omnidirectional or a directional program. Participants also completed a HINT test and questionnaires about listening effort. LDT response times and accuracy, as well as the ratio of accuracy to response times, were analyzed with linear mixed models. The SNR and HA-program manipulations only had small effects, and analyzing the ratios did not reveal larger effects than analyzing the accuracy or the response times alone. Additional work to derive listening efficiency from evidence accumulation rates, and to assess whether it leads to enhanced sensitivity to the SNR and HA program manipulations, is ongoing and will be presented at the conference, along with reflections on the merits of the task.

P16 Development of a questionnaire to measure social participation in noisy environments for people aged 60+ with hearing loss

Rebecca L. Bright, Emma Holmes
University College London, UK

Background: Several studies have shown an association between age-related hearing loss and social isolation, which may have implications for cognitive health for older adults. Questionnaires have long been used to measure older adults' social activity and social isolation across several disciplines. These scales continue to be adapted to capture changing lifestyle and societal factors, such as increased use of digital communication and a broader range of social activities available. However, changes in social participation over time may reflect the cognitive and acoustic demands of particular social environments, and items used in current questionnaires are unlikely to be sensitive to differences in these demands across social settings. In the current work, we aimed to develop a new questionnaire to understand how the auditory-cognitive demands of an environment (e.g., the presence and nature of background noise) affect social participation for people aged 60+ with hearing loss.

Methods: First, we reviewed existing tools that measure social participation and used this to generate a long list of social activities. We then developed a questionnaire to measure the frequency, ease, and satisfaction of involvement in these social activities. We refined the questionnaire based on a focus group of healthcare workers. We piloted the questionnaire in interviews with adults aged 60+ who have hearing loss to determine the comprehensiveness of the activities included and the usability of the questionnaire.

Results: The focus group suggested adding a free text entry option to the questionnaire to enable participants to describe the nuanced elements that influence the ease or satisfaction of each activity. Based on their responses, we also made some practical adjustments to improve usability of the tool. People with hearing loss who completed the interviews indicated that the activities included in the questionnaire cover the range of activities they participate in. Preliminary data from the questionnaire responses imply dissociations between the frequency of

participating in social activities, compared with the ease and satisfaction of involvement. In free-text responses, participants shared insights into the environmental and individual factors that determine the ease or satisfaction of each social activity. We plan to further analyse the data based on the auditory-cognitive demands of each environment or situation.

Conclusions: Our preliminary testing demonstrated the usability of the new questionnaire for adults aged 60+ who have hearing loss. Future work will test the questionnaire in a larger group of participants and measure changes in social activities across time.

P17 The third Clarity Enhancement Challenge

Jon Barker[1], John Culling[2], Trevor Cox[3], Michael Akeroyd[4], Graham Naylor[4], Simone Graetzer[3], Jennifer Firth[4], Jianyuan sun[1]

1. University of Sheffield, UK | 2. Cardiff University, UK | 3. University of Salford, UK | 4. University of Nottingham, UK

The Clarity Enhancement Challenges (CECs) seek to facilitate development of novel processing techniques for improving the intelligibility of speech in noise for hearing-aid users through a series of signal-processing challenges. Each challenge provides entrants with a set of stimuli for development and testing of their algorithms. Evaluations are conducted with similar but unseen material. Algorithms are permitted to use unlimited processing resources, but must be causal in the sense that output at time t must be independent of input at t+5 ms (i.e., in use, the algorithm could cause a lag of no more than 5 ms). The performances of the algorithms are assessed using objective measures of speech intelligibility and subjective measures conducted with a panel of hearing-impaired listeners. In CEC3, three different tasks were prepared, which each increased aspects of realism when compared with CEC2. For Task 1, the synthetised sixth-order ambisonic room impulse responses of CEC2 were replaced with real sixth-order ambisonic impulse responses, recorded in rooms using the em64 from MH Acoustics. As in CEC2, virtual ambisonic sources were processed to generate six hearing-aid input signals, three on each side of a head that rotates towards the target source. These impulse responses were used to generate virtual target sources within the room with up to 3 interfering sounds. For Task 2, the hearing aid input signals were directly recorded using microphones in hearing-aid shells during a listening task with target and interfering sounds presented from loudspeakers. Concurrent head orientation data was recorded using infrared motion tracking. For Task 3, real, often mobile interfering sounds were directly recorded in 6th-order ambisonics using the em64, 1) at various roadside locations, 2) at a railway station and 3) in a drinks party. Ambisonic impulse responses were also recorded in order to add target voices to these scenes and the hearing-aid microphone signals were again generated for a moving head. The performances of the algorithms were assessed using algorithmic estimates of speech intelligibility and objective measures conducted with a panel of hearing-impaired listeners. The enhanced signals from the challenge entrants showed that speech intelligibility as measured using the hearing-aid speech perception index (Kates & Arehart, 2021, doi:10.1016/j.specom.2020.05.001) could be substantially improved in each of the tasks. Further evaluation of the output-signal intelligibility for hearing-impaired listeners is ongoing and will be reported at the meeting.

P18 Listening effort is reduced with rapid adaptation to noise-vocoded speech under full and divided attention: evidence from pupil dilation and subjective rating

Rongru Chen, Emma Holmes, Patti Adank
University College London, UK

Background: Rapid adaptation has been robustly shown in various kinds of degraded speech, such as fast, accented, or noise-vocoded speech. While perceptual adaptation to speech is associated with enhanced task performance, how it affects the cognitive resources recruited to process degraded speech signals remains unclear. The current study employed pupillometry and the subjective rating method to investigate how listening effort changes during rapid adaptation to noise-vocoded speech. Furthermore, the effect of divided attention on speech adaptation and listening effort was also examined.

Methods: Sixty participants [median age 21, IQR (19 – 23), 28% men] were recruited in a between-subject design (single-task vs. dual-task condition). All participants listened to 72 six-channel noise-vocoded IEEE sentences presented in four blocks. Each sentence contained two consecutive pure tones ranging from 100-300 Hz with a 15% pitch difference. In the single-task condition, participants focused solely on recognizing noise-vocoded sentences (full attention). In the dual-task condition, participants both recognized speech and judged the direction of pitch shifts (divided attention). Pupillary responses during stimulus presentation were recorded using an EyeLink 1000 Plus eye tracker. After each block, participants rated their self-perceived listening effort on a seven-point Likert scale.

Results and Conclusions: Both pupillary response measurements and subjective ratings demonstrated that rapid adaptation to noise-vocoded speech significantly reduces listening effort. Our investigation of the relationships between attention, adaptation, and effort revealed two key findings. First, perceptual learning may be largely automatic, as divided attention did not impair adaptation as indicated by speech task performance. Second, while divided attention did not affect adaptation rates, it influenced listening effort dynamics across trials. Specifically, divided attention led to a more pronounced reduction in listening effort over time, as evidenced by greater decreases in pupil dilation across trials. Overall, these results suggest that listeners with divided attention can achieve the same level of adaptation as those with undivided attention while simultaneously showing greater efficiency in managing listening effort during speech processing.

P19 Binaural temporal fine structure sensitivity for children with developmental dyslexia

Sheila Flanagan, Brian Moore, Angela Wilson, Fiona Gabrielczyk, Kanad Mandke, Usha Goswami
University of Cambridge, UK

Background: Speech-in-noise perception is known to mature over the first 10 – 12 years of life. In this age range, children with language and/or reading difficulties have been reported to experience poor speech-in-noise perception compared with controls. However, the underlying aetiology for this finding is debated. Binaural Temporal Fine Structure sensitivity (bTFSs) is known to be beneficial for attending to sound sources in challenging environments. For young normal-hearing adults (YHNA), the upper frequency limit of bTFSs is known to be around 1400 Hz. Research has found the upper frequency limit of bTFSs to be significantly lower (worse) for typically developing children than for YHNA, with age being a significant predictor of the upper limit. If poor bTFSs contributes to impaired speech-in-noise perception in dyslexia (DYS), poorer bTFSs would be expected in DYS. In contrast, the Temporal Sampling (TS) theory of developmental dyslexia predicts that the perception of bTFS of speech may be preserved in children with dyslexia. By TS theory, reduced sensitivity to low-frequency envelope modulations is the core auditory impairment regarding DYS.

Methods: Binaural TFS sensitivity was measured using the Temporal Fine Structure-Adaptive Frequency (TFS-AF) test with 88 children aged 7-9.5 years (30 age-matched [CA], 20 male and 58 DYS, 31 male). Using a 2-up 1-down paradigm, the highest frequency at which interaural phase differences (IPD) of 30 degrees and 180 degrees could be distinguished from an IPD of 0 degrees was assessed.

Results: An LME model revealed no effect of group ($F[1,44] = 0.18$, $p = .68$), a significant effect of phase, with 30degrees lower than 180 degrees ($F[1,44] = 214.83$, $p<.001$), and no phase by group interaction ($F[1,44] = 0.04$, $p = .84$).

Conclusion: These results suggest that development of bTFSs is similar for DYS and CA children. Hence, the protracted developmental pattern of bTFSs was supported, with the upper frequency limit of bTFSs in children compared to YNHA being significantly lower ($p < .001$) for both levels of phase difference tested (30 and 180 degrees). A smaller frequency range of bTFSs may limit the benefit gained from spectral release from masking contributing to the known speech-in-noise deficit found in children when compared with adults. However, bTFSs was not found to be additionally impaired in DYS, supporting TS theory.

Funding: This work was supported by Fondation Botnar grant No. 6064, to U. Goswami.

P20 Convolutional neural networks improve decoding of selective attention to speech in cochlear implants users

Constantin Jehn[1], Adrian Kossmann[2,3], Niki Katerina Vavatzanidis[2,3], Anja Hahne[2,3], Tobias Reichenbach[1]

1. Friedrich-Alexander Universität Erlangen-Nürnberg, Germany | 2. Ear Research Center Dresden (ERCD), University Hospital Carl Gustav Carus, Dresden, Germany | 3. Technische Universität Dresden, Germany

Cochlear implants (CIs) are neural prostheses that use artificial electrical stimulation of the cochlea to restore hearing in severely hearing-impaired individuals. While modern CIs enable a majority of users to achieve good speech understanding in quiet environments, background noise and competing speech streams pose significant challenges. Auditory attention decoding (AAD) seeks to decode the attention of a listener in a multi-talker situation from electroencephalography (EEG) data. AAD may be used in the development of a neuro-steered CI, which aims to help CI users in challenging listening situations by amplifying the target speaker and attenuating background sounds. A variety of methods for AAD in normal-hearing individuals have been developed and evaluated over the past years, with deep neural networks (DNNs) proving superior to linear models in terms of decoding performance. However, although the feasibility of AAD in CI users has been demonstrated by several studies, the advantages of DNNs remain to be proven for CI users. Here we demonstrate how the implementation of a convolutional neural network (CNN) improves the decoding of selective attention to speech in CI users. First, we collected a substantial selective attention dataset from 25 bilateral CI users (15 female 10 male, median age 56 years ± 11.1), where stimuli were presented in a free field environment and EEG was measured simultaneously. Second, we implemented a CNN as well as a linear backward model for AAD. The CNN emerged as the superior method, as measured by the achieved decoding accuracy on all studied decision windows ranging from 1s to 60s. In conjunction with a learnable Support-Vector-Machine for speaker classification, the CNN achieved a maximal decoding accuracy of 74% (± 11%) on the population level and thereby significantly outperformed the linear backward model. These findings underscore the potential of DNNs with adaptable speaker classification as promising candidates for neuro-steered CIs, translating advancements made in AAD for normal-hearing individuals to benefit CI users.

P21 No evidence of musical training influencing the cortical contribution to the speech-frequency-following response and its modulation through selective attention

Jasmin Riegel, Tobias Reichenbach, **Alina Schüller**
Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany

10.1523/ENEURO.0127-24.2024

Previous research has shown that musical training can lead to differences in certain neural responses to auditory stimuli. In particular, musicians were found to exhibit stronger subcortical responses to speech sounds than non-musicians. Among these responses is the frequency-following response at the fundamental frequency of a speech signal (speech-FFR). However, the speech-FFR exhibits not only a subcortical, but also a cortical component. It remains unclear whether the cortical component of the speech-FFR may be influenced by musical experience as well. Moreover, we recently showed that the cortical component is modulated by selective attention (Schüller et al., J. Neurosci. 43:7429, 2023), but whether this attentional modulation is subject to musical training has not yet been studied.

In this study, we acquired magnetoencephalography (MEG) recordings to investigate the cortical contribution to the speech-FFR in 52 participants with varying levels of musical expertise. Participants were presented with two audiobooks from different speakers, with instructions to selectively attend to one while disregarding the other. We analyzed the data by computing temporal response functions (TRFs) to examine the source-reconstructed activity in the auditory cortex, specifically focusing on two acoustic features related to the fundamental frequency of speech, which characterize the speech-FFR.

Our findings revealed significant differences in attentional modulation for both acoustic features, consistent with our prior research. However, we observed no modulation of the responses based on musical training. Both musicians and non-musicians exhibited similar cortical contributions to the speech-FFR. Furthermore, the attentional modulation observed was not influenced by musical training. Our results suggest that the subcortical and cortical contribution to the speech-FFR play at least partly different roles in speech processing.

P22 Perceptual and sensorimotor learning with noise vocoded speech

Susan Black, Matt Davis, **Abbie Bradshaw**
MRC Cognition and Brain Sciences Unit, University of Cambridge, UK

Unfortunately the presentation of this poster has been cancelled.

Cochlear implants (CIs) are a device that can provide auditory input for individuals who are deaf, giving them access to spoken language. The vast majority of research has focused on how individuals with CIs perceive speech produced by others; however, CIs additionally provide access to auditory input from one's own voice during speech production, known to be critical for ensuring accurate and intelligible speech. This experiment aimed to investigate whether and how typical hearing participants can learn to use CI simulated speech input for perceptual learning and sensorimotor control of speech during production. CI simulation was achieved via noise vocoding, a technique that degrades the spectral detail in a speech signal in a manner similar to a CI.

The experiment was made up of two phases. First, a perceptual learning phase tested recognition of noise vocoded sentences before and after a training task which was manipulated between groups; either a perception training task where participants listened to noise vocoded sentences while reading matching text, or a production training task where participants read aloud sentences whilst hearing their own voice noise-vocoded in real-time. In the second phase of the experiment, all participants read aloud sentences while hearing their voice noise vocoded in real-time; after a baseline phase, an additional real-time manipulation of the formants of their voice (spectral properties of speech which determine the vowel sound) was made. The direction of this formant perturbation was manipulated between groups; assignment to these groups was counterbalanced relative to assignment to training conditions in phase 1. When speaking with non-vocoded feedback, speakers typically show corrective adjustments to their production of speech sounds in order to compensate for such formant perturbations; this is known as speech motor adaptation. This process ensures that our speech productions remain on target.

Preliminary results with N = 20 (target N = 30) indicate that both the perception and production training tasks in the first phase resulted in significant improvements in recognition of noise vocoded sentences; training type however did not affect the magnitude of this improvement. For the second phase, despite trends in the right direction, speech motor adaptation in our incomplete sample was not significant at the group level in response to the formant perturbations. Once the full sample has been collected, we will also compare adaptation between the two training type groups, to test if greater experience of speaking with noise vocoded feedback results in improved adaptation.

P23 Examining the effect of prior knowledge on speech processing in cochlear implant users: A visual world paradigm study

Khaled Abdel-Latif[1,2], Thomas Koelewijn[3], Deniz Başkent[3], Hartmut Meister[1,2]
1. Faculty of Medicine and University Hospital Cologne, Department of Otorhinolaryngology, Head and Neck Surgery, University of Cologne, Germany | 2. Jean Uhrmacher Institute for Clinical ENT-Research, University of Cologne, Germany | 3. Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Netherlands

Cochlear implants (CIs) help to restore hearing in individuals with severe to profound hearing loss. While many CI users can understand speech well in quiet environments, it becomes challenging when multiple people speak simultaneously. Previous studies in normal-hearing listeners have shown that knowing in advance who is going to speak and where to listen can improve speech understanding (Kitterick et al., 2010, JASA 127:2498–2508, doi:10.1121/1.3327507) and that prior knowledge of a speaker's location can reduce cognitive load during speech processing (Koelewijn et al., 2015, Hear Res 323:81–90, doi:10.1016/j.heares.2015.02.004). Still, it is unknown to what extent CI users can benefit from such information and how the presence or absence of prior knowledge impacts cognitive load during speech processing.

The aim of this study is to examine whether CI users benefit from information about the spatial position and the voice of a target talker when presented against a competing talker.

To understand the intricacies of speech-on-speech masking at a fine-grained temporal level, we use the Visual World Paradigm (VWP; Tanenhaus et al., 1995, Science 268:1632–1634, doi:10.1126/science.7777863; Abdel-Latif et al., under review), which is based on the finding that gaze fixations and speech processing are closely linked in time. We employ the VWP using matrix sentences from the Oldenburg Sentence Test (OLSA; Wagener et al., 1999, Z Audiol 38:44–56). Following Meister et al. (2020, JASA 147:EL19, doi:10.1121/10.0000499), two competing OLSA sentences are presented simultaneously, with the target sentence indicated by the keyword "Stephen." Participants are instructed to focus their gaze on icons representing the target sentence and to verbally recall the sentence after a retention period. Two different conditions are considered, one with and one without a priori information about the target talker's voice or the spatial position.

Preliminary results in terms of gaze fixation (as a proxy for attention), pupil size (as a proxy for cognitive load), and speech intelligibility are reported and discussed.

Funding: This project is funded by the Deutsche Forschungsgemeinschaft (ME2751/6-1).

P24 Effect of ambisonic order and loudspeaker positioning in spatial release from masking

John Culling, Sam Dighton, Ottopea Akrofi
Cardiff University, UK

Ambisonics is a method for reproducing a sound field that has been recorded at a particular location using multiple directional microphones. Most ambisonic systems are first-order using four directional microphones in a tetrahedral arrangement that are ultimately reproduced over at least four loudspeakers or within a head-mounted display using four virtual loudspeakers. A number of systems have been developed that extend the spherical-harmonic order by using more microphones and loudspeakers. Higher-order ambisonics should increase the spatial resolution of the reproduced sound, but the perceptual benefits of this development have not been extensively explored. In a series of experiments, spatial release from masking (SRM) was measured using a circular ambisonic reproduction system for a range of different ambisonic orders and for different loudspeaker selections. One experiment used ambisonic panning, while the second used impulse responses recorded form a 6th-order ambisonic microphone. These experiments found a progressive increase in SRM with increasing ambisonic order for speech presented against a single speech-shaped-noise interferer. SRM increased progressively up to at least 4th order and continued to asymptotically improve at higher ambisonic orders towards the level achieved with point sources. The second experiment also established that this effect was robust for different acoustic environments. Since a beneficial effect of higher ambisonic was confirmed using different methodologies and in both anechoic and reverberant environments, the spherical order of encoding format appears to form a common bottleneck in the necessary spatial information. Further experiments examining the effect of loudspeaker location for first-order ambisonics and the effect of ambisonic order on speech reception for ambisonic recordings in complex listening situations are on ongoing and will be reported at the meeting.

P25 The impact of effortful listening on the comprehension of intelligible broadcast dialogue

William D. Curry, Ben G. Shirley, Trevor J. Cox
University of Salford, UK

Martin Walsh
Xperi inc., San Jose, California, USA

Background: Difficulty understanding dialogue is one of the most frequent causes of complaints for media broadcast companies. Despite most productions presenting dialogue at speech-to-noise ratios (SNRs) in line with guidance and standards, reported subtitle use has increased among listeners with and without a hearing impairment. Background music and effects in broadcast audio can introduce significant listening effort even when objective intelligibility is high. Effortful listening can result in reduced recall and disrupt language processing. The impact of extended periods of effortful listening on the comprehension of broadcast content and how this affects viewer engagement is not known.

Aims: To investigate the influence of background noise levels on listening effort and dialogue comprehension in broadcast content.

Methods: Thirty-one adult (18-36 years) participants were presented with fifteen, two-minute clips from a radio broadcast in the presence of background noise. The degree of acoustic challenge was modulated by varying the speech-to-noise loudness ratio within the range commonly found in broadcast media (4LU, 10LU, 16LU). Following each clip, self-reported listening effort was recorded through ratings of five dimensions based on the NASA-TLX (Audibility, Mental Demand, Effort, Frustration, Engagement). Participants then completed a recognition task to assess their comprehension of the spoken content by probing for different levels of memory representation (surface form, propositional, situation model). Response times for the recognition task were recorded as a behavioural measure of listening effort. Electrophysiological (EEG) recordings were also taken throughout stimulus presentation and recognition task although this presentation covers only the participant response measures.

Results: Analysis of participant responses showed an increase in reported listening difficulty corresponding to reduced response accuracy at higher background noise levels. Analysis of the recognition task revealed a significant effect of loudness difference conditions on memory of the surface form and at the propositional level (between 16-10LU and 10-4LU respectively). A significant reduction in discrimination of contextually incorrect information was also observed. Results indicate that different information processing components were reduced differentially as acoustic challenge increased.

Conclusions: Extended periods of effortful listening had a negative impact on comprehension and memory of spoken media material. SNRs within the range commonly found in broadcast media can result in significant listening effort. A reduction in recognition for the surface form and an increased reliance on propositional information was observed between the two most favourable SNRs, with a decrease in propositional recognition at the lowest SNR. This points to a greater reliance on information summarisation as increasing acoustic and cognitive demand reduced capacity for encoding more precise details, highlighting the need to ensure that background levels are set to reflect the degree of comprehension necessary to understand and enjoy spoken media content, even when intelligibility is at ceiling level.

P26 Cognitive and auditory correlates of speech-in-noise perception in age-related hearing loss: An EEG investigation

Sankalpa Mahadev[1,2], Hari Prakash Palaniswamy[2], Rajashekar Bellur[2]
1. JSS Institute of Speech and Hearing, Mysuru, India | 2. Manipal College of Health Professions, MAHE, India

Unfortunately the presentation of this poster has been cancelled.

Introduction: Age-related hearing loss significantly impacts the ability to understand speech, especially in noisy environments. While changes in the peripheral auditory system contribute to this difficulty, increasing evidence indicates that central auditory processing and cognitive

functions also play critical roles in speech perception. This study investigates the complex relationship between age-related hearing loss, cognitive function changes, and difficulties in speech perception amidst noise in elderly individuals.

Method: The study used a prospective cross-sectional design with two groups of participants: 20 individuals aged 60 to 80 years with mild to moderately severe sensorineural hearing loss (Group 1) and 20 age- and education-matched controls with normal or near-normal hearing sensitivity (Group 2). Speech-in-noise perception was assessed using the Quick Speech-in-Noise (Quick SIN) test, which evaluated participants' performance across varying signal-to-noise ratios (SNRs) ranging from +20 to -10 dB. Cognitive functions were evaluated through computerized assessments, including attention network tests and N-Back tests, alongside simultaneous 32-channel EEG recordings.

Results and Discussion: The findings revealed significant correlations between central hearing abilities (measured as SNR Loss) and cognitive performance. A moderate to strong correlation was found between SNR Loss and cognitive scores, particularly in executive attention and working memory. A negative correlation was identified between EEG amplitude and hearing abilities, indicating that decreased neural responses were associated with increased central auditory loss. Furthermore, a positive correlation between central hearing abilities and reaction time suggested that individuals with more significant SNR loss needed more time to process information. The significant correlations between executive attention, working memory, and speech-in-noise perception indicate that cognitive processes related to task management (suppressing noise and focusing on speech) and information retention (retaining and recalling spoken information) are crucial for understanding speech in challenging auditory environments. In contrast, the lack of impact from alerting and orienting networks suggests that simply being aware or directing focus does not significantly improve one's ability to perceive speech in noisy conditions.

Conclusion: The results indicate that difficulties in speech perception in noisy environments arise from peripheral hearing loss, changes in central auditory processing, and cognitive decline, particularly in executive functions and working memory. This study provides valuable insights into the intricate relationships between auditory processing and cognitive functions in elderly individuals with hearing loss, emphasizing the need for comprehensive assessment and rehabilitation approaches that address both auditory and cognitive aspects of speech perception in noise.

P27 Reading with cochlear implants: the role of Cued Speech

Virginie Keukeleire, Brigitte Charlier, Cécile Colin
Université libre de Bruxelles, Brussels, Belgium

Background: Deaf adolescents have often been characterized as weaker readers compared to their typically-hearing peers. However, with earlier intervention and advances in hearing technology, a positive trend in reading comprehension has been noted among adolescents with cochlear implants, despite significant inter-individual variability in performance. However, cochlear implants do not transmit speech-related spectro-temporal cues properly. Therefore, cued speech, a lip-reading support enabling the development of accurate phonological

representations, might positively impact reading acquisition. This study aimed to examine reading comprehension abilities of deaf adolescents with implants, aged 10 to 15, who use or do not use cued speech at home.

Method: Fourteen cued speech users (CS group), fourteen non-cued speech users (nonCS group) and twenty-eight typically-hearing adolescents (TH group) performed two reading comprehension tasks (sentence and text) and a nonverbal IQ (a spatial short-term memory task and a nonverbal reasoning task).

Results: In sentence comprehension, there were no differences between CS and TH groups as regards accuracy and total response time. However, nonCS group had a worse performance and a longer response time than the two other groups, with specific difficulties regarding personal pronouns. As regards text comprehension, TH group had better performances than nonCS group in total score, principal idea, detail, prediction, context and explanation but the performances of CS group did not differ from that of the two other groups. For sequential information, CS and TH groups had a better performance than nonCS group.

Conclusions: CS adolescents aged 10 to 15 achieve comparable performance in reading comprehension for sentences and texts, compared with their TH peers. NonCS adolescents seem to demonstrate more weaknesses. Overall, the data provide a better understanding of the usefulness of cued speech on reading skills. Some additional individual analyses are nevertheless needed to disentangle the impact of other factors such as age and type of education.

P28 Association between characteristics of the envelope following response (EFR) and speech in noise processing

Mahmoud Abdellahi[1], **Yang Li**[1], Leny Vinceslas[1], Jesko Verhey[2], Ifat Yasin[1], Vit Drga[1]
1. Department of Computer Science, University College London, UK | 2. Department of Experimental Audiology, Otto von Guericke University Magdeburg, Germany

Speech understanding is known to become more difficult with increasing amounts of background noise even in the absence of hearing impairment. Some of these difficulties may be attributed to degraded neural representation of speech-related temporal cues. The envelope following response (EFR) is a far-field EEG response that can indicate how well a population of auditory nerve fibres phase-locks to the stimulus envelope, and thereby provide an indication of the robustness of temporal cue processing. The quality of the acquired response is affected by whether the eliciting stimulus is sinusoidally amplitude-modulated (the stimuli used most often to elicit and EFR; SAM-EFR) or rectangularly amplitude-modulated (RAM-EFR), which has also been used in recent studies. Robustness of the EFR (SAM or RAM) can be affected by background noise.

This study investigated characteristics of SAM-EFR or RAM-EFR in noise, and relevance to speech processing in noise. Characteristics of SAM-EFR and RAM-EFR were obtained for a signal of 110 Hz and carrier frequency of 2 kHz in the same participants for a range of modulators (80-200 Hz) and as a function of modulator cycles (2-10). Robustness was assessed by comparing the EFR responses in the absence or presence of noise. The EFR magnitude was

derived and compared across stimulus modulator frequency and number of cycles, as well as with, and without the presence of noise. Results of SAM-EFR and RAM-EFR are discussed with regards to temporal processing degradation in noise and relevance to speech envelope coding in noise.

P29 Multimodal communicative signals facilitating communicative success for hard-of-hearing individuals in noisy contexts

Anna Palmann[1], James Trujillo[1], **Linda Drijvers**[2], Floris Roelofsen[1]
1. University of Amsterdam, Netherlands | 2. Donders Institute, Radboud University, Nijmegen, Netherlands

10.5281/zenodo.14608615

Communication is inherently multimodal, requiring interlocutors to combine signals from the auditory with those from the visual modality. If, however, one of these modalities is impaired, as it is the case for people who are hard-of-hearing, compensation frequently occurs in the other modality. Thus, it is expected that people with hearing impairments differ from normal hearers in their use and processing of co-speech visual signals, i.e., gestures, head movements, or facial expressions. To date, it remains unclear to what extent people who are hard-of-hearing rely on visual information in communication with each other, for example by using sign-supported speech, lipreading or the enhancement of kinematic gesture and head features. In dyadic interactions, where speakers accommodate each other in order to facilitate mutual understanding, the production and comprehension of these signals are intertwined. In this research, we focus primarily on the production of multimodal communicative signals and indirectly also capture comprehension by investigating the role these signals play in communicative success in interactions of hard-of-hearing individuals in noisy contexts.

In a preliminary online survey, we collected data from hard-of-hearing individuals in the Netherlands. We inquired about their clinical history, their use of visual communication and hearing aids, communicative contexts that are perceived as challenging, and strategies for successful communication in such situations. Using a mixed-methods approach, we found, among other things, that although most participants regularly use hearing aids, the performance of these devices in filtering signals in background noise is still perceived as poor. Thus, communication difficulties in noisy contexts persist even among hearing aid users. To facilitate communicative success despite these challenges, participants reported a variety of strategies ranging from pre-planning the situational set-up to, most importantly, modifying acoustic and kinematic features during the conversation.

Building on these preliminary self-report findings, we will further investigate which multimodal communicative signals emerge in background noise and which of them are associated with communicative success in a naturalistic dialogue setting. For this, dyads of hard-of-hearing and control dyads of normal-hearing individuals will engage in free as well as in task-based dialogue while exposed to social and non-social background noise. In addition to analyzing acoustic and linguistic features of speech, we will use motion capture to analyze different aspects of gesture and head kinematics, facial expressions, and body posture. Using a data-driven machine learning approach, we will then assess which multimodal communicative

features play a role in successful communication, hypothesizing that communicative success and the production of multimodal communicative signals differ both between dyad groups and between type of background noise. With this research, we hope to provide insights into how to make communication easier and more successful for hard-of-hearing individuals, especially in noisy contexts.

P30 Can music training enhance/affect working memory and speech-in-noise perception in cochlear implant users? A randomized controlled study of EEG measures of improvement

Kathrin Mertel[1,2], Andrew Dimitrijevic[2], Michael H. Thaut[1]
1. Music and Health Research Collaboratory (MaHRC), University of Toronto, Canada | 2. Sunnybrook Cochlear Implant Program, Sunnybrook Hospital, Toronto, Canada

Background: A cochlear implant (CI) enables postlingually deafened people to understand speech but due to technical restrictions, users face great limitations in noisy conditions. Music training was shown to augment shared auditory and cognitive neural networks for processing speech and music and improves auditory-motor coupling which benefits speech perception in noisy listening conditions. These are promising prerequisites for multi-modal Neurologic Music Training (NMT) on speech-in-noise (SIN) perception in adult CI users. Furthermore, a better understanding of the neurophysiological correlates when performing working memory (WM) and SIN tasks after multi-modal music training with CI carriers may provide clinicians with a better understanding of optimal rehabilitation.

Methods: In the current study, 81 post-lingual deafened adult CI recipients will undergo a 4-week neurologic music therapy multi-modal training for 3 conditions (pitch, rhythm, and timbre). Pre- and Post-tests will analyze behavioral outcomes and apply a novel EEG approach recording alpha oscillation modulations to the sentence-final-word-identification-and-recall test (SWIR-EEG).

Outcome: The first behavioral results of short-term multi-modal music training on the enhancement of WM and SIN performance and their reflection in alpha oscillation modulations in prefrontal areas in post-lingual deafened adult CI recipients will be presented. Prospectly, outcomes could contribute to understanding the relationship between cognitive functioning and SIN besides the technical deficits of the CI. Targeted clinical application of music training for post-lingual deafened adult CI carriers to significantly improve SIN and positively impact the quality of life can be realized.

P31 Metacognitive awareness of lip-reading gains in young and older adults

Elena Giovanelli[1], Benedetta Desolda[1], Chiara Valzolgher[1], Elena Gessa[1], Tommaso Rosi[2], Francesco Pavani[1]
1. Center for Mind/Brain Sciences - CIMeC, University of Trento, Italy | 2. Level Up S.r.L., Trento, Italy

When listening to speech in noise, lip-reading facilitates communication. We investigated to what extent older and younger adults are aware of this benefit, and whether this metacognitive ability changes with age. Participants (N=40) completed a hearing-in-noise task while facing a human-like avatar in virtual reality. A panel of changing transparency was interposed between the listener and the avatar, to create pairs of conditions with different lip-reading accessibility in each trial. We measured performance, confidence and effort during audio-visual listening, and derived measures of real improvement (i.e., lip-reading gain) as well as metacognitive improvement (i.e., perceived gain in accuracy, confidence and effort) on a trial-by-trial basis. Results show that real improvements were comparable in the two age groups, who were also similarly aware of performance and confidence gains resulting from audio-visual lip-reading. Yet, older adults were less able to appreciate the listening effort reduction associated with higher lip visibility, particularly those with lower unisensory lip-reading abilities (as measured in a visual-only condition). Thus, while comparable awareness of audio-visual lip-reading gains exists in older and younger adults, the less efficient unisensory lip-reading skills of older adults may affect their appreciation of this effective speech processing behavior while listening in face-to-face communications. Since perceived effort concurs in the decision to implement (or not) adaptive behaviors, these findings have significance for the design of interventions aimed at ameliorating older adult's quality of face-to-face interactions.

P32 The impact of the semantic content of L1 and L2 maskers on L1 target transcription

Emily Rice[1], Sarah Knight[2], Angela de Bruin[1], María Muradás-Taylor[1], Sven Mattys[1]
1. University of York, UK | 2. Newcastle University, UK

Listening to a talker is more challenging when their language matches that of a competing (masker) talker (single-language context) than when target and masker speech are in different languages (dual-language context) (Brouwer et al., 2012, doi:10.1121/1.3675943). The Target-Masker Similarity Hypothesis attributes this to acoustical overlaps between languages, where greater similarity makes streams harder to separate. For bilinguals, when the target is the first language (L1), ignoring a second language (L2) masker may also be easier due to the L2's lower activation levels, causing less interference than an L1 masker. This raises questions about whether it is easier to ignore L2 speech primarily due to acoustical differences or also differences in activation levels.

The masker's semantic content might also influence interference (i.e., informational masking). Research in single-language contexts suggests that maskers undergo some semantic processing. For example, perception of a semantically meaningful target is more impaired by a meaningful (e.g., "Rice is often served in round bowls") than anomalous (e.g., "The great car

met the milk") masker. Most research compares interference caused by semantically meaningful vs. anomalous maskers; however, it is also possible to investigate whether the semantic relatedness of a meaningful masker to a target interferes with perception. If there is a semantic interference effect when the target is L1 and the masker is L1 or L2, it would suggest that maskers in any known language undergo semantic processing. However, if this semantic effect is absent with an L2 masker, it would suggest that bilinguals activate meaning only for L1 maskers due to the L1's higher activation levels.

We investigated this by presenting Spanish-English bilinguals (target N = 120, current N = 88) with pairs of simultaneous sentences. In each pair, the target sentence was Spanish (L1), while the masker was either Spanish (L1) or English (L2). The masker was either semantically related or unrelated to the target.

We hypothesise that transcription accuracy will be higher with an L2 than L1 masker, in line with the Target-Masker Similarity Hypothesis. If this effect results only from acoustical differences between the languages, then the difference between L1 and L2 masker conditions should not vary across semantic conditions. However, if different activation levels influence the masker language effect, then an effect of semantic relatedness will be apparent with an L1 masker, but not L2. This would indicate that L1, but not L2, background speech is semantically processed. Preliminary data from this study will be presented.

P33 Hearing at risk: Prevalence and impact of acquired hearing loss in adolescents

Ugo Benrubi, Axelle Calcus
Centre for Research in Cognition & Neurosciences, Université Libre de Bruxelles, Belgium

Background: Adolescents and young adults are increasingly exposed to high-intensity recreational noise, potentially leading to acquired hearing damage. Currently, no study has documented the prevalence of such acquired hearing loss (HL) among French-speaking adolescents in Belgium. However, studies conducted in similar industrialized regions of the world indicate that 17 to 19% of adolescents may experience slight/mild acquired HL. One major concern regarding slight/mild HL, as detected through pure-tone audiometry, is that it often goes unnoticed by the affected individuals. Research has shown that people who do not perceive any hearing issues are less likely to engage in protective behaviors that could safeguard their hearing. Moreover, even a slight degree of HL is significantly associated with poorer academic performance and speech recognition abilities.

Objective: This study aims to (i) estimate the prevalence of acquired hearing loss in a French-speaking Belgian adolescent population and (ii) explore its relationship with noise exposure, throughout development.

Methods: We conduct assessments in schools and community organizations, including pure-tone audiometry, speech perception in noise, Distortion Products of Otoacoustic Emission (DPOAE), and several questionnaires evaluating listening habits and noise exposure history since childhood.

Preliminary Results: Pilot data collected from a young adult sample (n=34, age = 21±2.2 years) revealed that 17.6% exhibited a mean pure-tone average (PTA) exceeding 15 dB HL. Additionally, our preliminary analysis showed that individuals who reported a lack of concern for their hearing health had significantly higher PTA values compared to those who actively cared about their hearing.

Discussion: Preliminary findings align with existing literature. We plan to expand our sample to approximately 500 adolescents to compare prevalence of acquired HL throughout development across the world, and look into possible causes of acquired losses. Results of this project could help shape public policies regarding noise exposure throughout development.

P34 Perception of prosody, emotional speech, and indirect speech in sensorineural hearing loss

Frederic Marmel[1], Haiping Huang[2], Elizabeth Davis[2], Eric Branda[3], Dina Lelic[4], Erin Picou[2]
1. ORCA Labs, Stockholm, Sweden | 2. Vanderbilt University Medical Center, Nashville, USA | 3. WS Audiology, Iselin, USA | 4. ORCA Labs, Lynge, Denmark

Most studies of the communication difficulties experienced by hearing aid (HA) wearers focus on speech understanding. Studying communication beyond speech understanding may shed light on hearing difficulties that are detrimental to quality of life but are currently overlooked. This ongoing study focuses on the perception of speech prosody; the aspect of communication that creates emotional tone and subtle grades of meaning. We are investigating whether HA wearers have difficulties perceiving several types of prosody: emotional prosody conveying basic emotions (anger, disgust, fear, happiness, neutral, sadness) or more nuanced emotions (adoration, amusement, anger, awe, confusion, contempt, desire, disappointment, distress, fear, interest, sadness), as well as prosody conveying intentions (sarcasm, teasing, white lie, sincere, questions and indirect requests). We are also assessing whether the degree of sensorineural hearing loss (SNHL) and the capability to understand prosody are related, and whether HAs are beneficial to prosody perception.

To evaluate prosody perception, we use speech stimuli extracted from four publicly available corpora: two corpora to evaluate the perception of basic and nuanced emotions, and two corpora to evaluate the perception of intentional prosody. The corpora are used in four separate closed-set tasks built on the same model: participants listen to speech stimuli one at a time and, after each stimulus, judge which emotion or intention has been expressed. Then, they rate their confidence in their judgment. The four tasks are presented in a pseudo-randomized counterbalanced order. Prior to completing the tasks, HA wearers are fitted with study HAs and given a 3-week acclimatization period. HA wearers complete testing twice, once without and once with the study HAs in a counterbalanced order. Age-matched participants with normal hearing (NH) complete the same tasks once, unaided.

Data collection is ongoing. Task performance and confidence ratings will be analyzed with linear mixed models including the hearing group (NH, HA unaided, HA aided) and the conveyed emotion (or intention) as fixed factors, and random intercepts for participants. Observing differences between NH participants and HA wearers (unaided) would confirm that SNHL can impair prosody perception for basic emotions independently from ageing, which has been previously reported by some —but not all— studies, and it could extend this finding to nu-

anced emotions and intentional prosody. The magnitude of the differences and the type(s) of prosody affected will inform our knowledge of how SNHL may impact quality of life, and the unaided vs. aided comparisons will inform us on whether (and to which extent) HAs can help.

P35 Neural signatures of stream segregation: from childhood to adulthood

Elena Benocci[1], Claude Alain[2], **Axelle Calcus**[1]
1. Université Libre de Bruxelles, Belgium | 2. Rotman Research Institute at Baycrest, Toronto, Canada

From classrooms to playgrounds, children's communication occurs in noisy environments. Although their peripheral auditory systems reach maturity approximately six months after birth, children experience persistent difficulties in perceiving speech in noise. When faced with noisy environments, listeners perform auditory scene analysis, allowing them to selectively focus on the relevant auditory target while ignoring interfering sounds. Stream segregation, a fundamental mechanism of scene analysis, involves organizing similar sound waves into a coherent stream, while distinguishing dissimilar acoustic components and attributing them to distinct sources. Two event-related potential components have been identified as 'neural signatures' of stream segregation: the Object-Related Negativity (ORN) and the P400.

Our study aims to examine (i) the maturation of neural correlates of stream segregation (ORN/P400) from childhood to adulthood and (ii) the development of the relationship between these neural correlates and speech perception in noise. Participants aged 8 to 23 years (children n=27, adolescents n=29, and adults n=27) were included in the study. ORN/P400 were recorded while participants performed an active stream segregation task. Participants also performed speech identification in noise tasks (behaviourally). Behavioral results indicate an improvement in both stream segregation and speech perception in noise from childhood to adulthood. Furthermore, stream segregation performance predicts speech perception in noise. Our neurophysiological findings suggest a reduction in amplitude of both ORN and P400 from childhood to adulthood, and P400 amplitude is a predictor of stream segregation performance.

Overall, our results suggest that the neural mechanisms underlying stream segregation follow a prolonged maturation trajectory, and support the progressive maturation of auditory scene analysis and speech perception in noise.

P36 The irrelevant speech effect in tonal languages for Mandarin and English native speakers

Yupeng Wu, Karolina Kluk, Rebecca E. Millman
University of Manchester, UK

Josef Schlittenlacher
University College London, UK

Background: Background sounds disrupt short-term memory, even if the target information is presented visually (Ellermeier and Zimmer, 2014, doi:10.1250/ast.35.10). Since speech has been found to be the most disruptive sound, this phenomenon was termed the Irrelevant Speech Effect (ISE). Previous studies have focused on various factors that contribute to the ISE, e.g., the type of sounds, speech intelligibility, or familiarity with the presented language. The present study focuses on the tonality of tonal languages. Compared with non-tonal languages, tonal languages contain more pitch changes and thus more changing state that causes the ISE. We hypothesize that, at least for native Mandarin speakers, tonal languages will cause a bigger ISE.

Method: 38 native Mandarin and 38 native English speakers were recruited and asked to memorize the order of a nine-digit sequence that was presented visually while the participants listened to background sounds. These sounds were either English, Mandarin, or Cantonese speech, or a continuous pink noise.

Results: Both Mandarin and English speakers confirmed that the least disruptive sound was continuous pink noise, with an average of 8.36 and 6.65 digits correctly recalled, respectively. This difference that Mandarin speakers could recall about one more digit than English speakers was found in each sound condition.English speakers experienced a larger disruptive effect in English (5.01) than in Cantonese (5.69) and in Mandarin (5.47). This result differs from previous studies that did not use tonal language sounds and found no effect of native language. Mandarin speakers experienced the highest ISE in Mandarin (6.51), followed by English (6.80) and Cantonese (6.94).

Conclusions: The results suggest that native language increases the ISE in tonal languages. The small ISE for Cantonese speech presented to Mandarin speakers suggests that the underlying cause is rather the native language than additional changing state due to tonality. A possible reason for the Mandarin speakers memorizing more digits than English speakers could be that Mandarin digits can be pronounced and thus rehearsed more quickly than English digits. We are presently investigating this further in a follow-up study where we present digits and Latin letters as to-be-memorized material.

We investigated whether the rehearsal language would affect the short-term memory and ISE. We found that both background sounds and to-be-memorized material significantly affect the correct rate, but only the background sounds significantly affect the size of ISE. This indicated that rehearsal language will only affect participants' short-term memory but will not affect ISE.

P37 Sound quality in DNN-based hearing-aid algorithms

Chuan Wen[1], Guy Torfs[2], Attila Fráter[1], Marjoleen Wouters[1], Matthias Inghels[1], Sarah Verhulst[1]
1. Hearing Technology @ WAVES, Department of Information Technology, Ghent University, Belgium | 2. IDLAB, Department of Information Technology, Ghent University, Belgium

Current hearing aids typically address outer-hair-cell (OHC) damage and associated hearing sensitivity loss, but do not consider age- or noise-exposure-related damage to auditory-nerve fibers (i.e., cochlear synaptopathy, CS). To compensate for individual and combined CS and OHC damage patterns, closed-loop systems which include biophysical models of (impaired) auditory signal processing can generate personalized sound processing algorithms. These closed-loop systems are most promising when the incorporated models are formulated as deep-neural-networks (DNN), such that the resulting sound processing algorithm can be obtained through backpropagation. One such method, CoNNear, is based on autoencoder models of auditory processing that comprise a modular and differentiable description of the cochlear mechanics, inner-hair-cell, and auditory-nerve fiber processing stages. The sound processors can be trained by minimizing biophysical auditory-processing differences between normal-hearing and hearing-impaired models, which can be embedded with AI-hardware.

However, these end-to-end systems come with a different kind of signal-processing artifacts than traditional sound processors. For example, the transposed convolutions included in CNN-based auditory processing modules can create tonal artifacts. The artifacts will propagate within the closed-loop framework to ultimately become overamplified and audible in the resulting hearing-aid algorithm.

To address this challenge, we propose a dilated CNN architecture that comprises a sequence of stacked memory blocks, which are most promising and artifact-free for closed-loop audio processing. To avoid using upsampling in the decoder, depthwise dilated 1-D convolutions are employed within each memory block to avoid artifacts while modeling the long-term dependencies of neural speech processing (e.g. cochlear impulse response durations and neural adaptation). We then employed the dCoNNear architecture to all auditory elements inside the closed-loop system as well as for the sound processors, and evaluated the sound quality as well as the compensation accuracy of the resulting algorithms. Our results show that dCoNNear cannot only accurately simulate all processing stages of non-DNN-based SOTA biophysical auditory processing system, but does so without introducing spurious and audible artifacts in the resulting sound processors. The predicted restoration accuracy for simulated auditory-nerve population responses shows that our algorithms can be used for both OHC and CS pathologies. The trained dCoNNear audio processors can process audio inputs of 3.2 ms in < 0. 3ms, which demonstrates its real-time capabilities. We conclude that the dCoNNear-based frameworks hold great promise for real-time and personalized hearing loss compensation strategies with high sound quality.

Acknowledgements: This work is supported by FWO Machine Hearing 2.0 and EIC Transition EarD-iTech101058278

P38 The association between socio-economic and lifestyle factors and auditory function in middle-aged adults.

Inez Sharp[1], Dialechti Tsimpida[2], Meher Lad[3], Helen Nuttall[1], Kate Slade[1]
1. Lancaster University, UK | 2. University of Southampton, UK | 3. Newcastle University, UK

Objectives: To investigate the associations between socioeconomic and lifestyle factors and measures of hearing ability; to better understand potential risk factors for hearing loss. This insight could help audiologists identify high-risk demographics, and ultimately contribute to addressing inequalities in hearing health.

Design: An online study which collected questionnaire data on participant demographics, lifestyle and socioeconomic status, including age, gender, ethnicity, region of residence, income, education, occupation, exercise frequency, height and weight, smoking status, and weekly alcohol consumption. Participants also self-reported their auditory function using the Speech, Spatial and Qualities of Hearing Scale 12 (SSQ-12 (Noble et al., 2013)) and completed an online digits-in-noise (DiN) task to assess speech perception ability. A sample of 274 45-65-year-olds (M age = 53.84, SD = 5.91) were recruited based on Office for National Statistics (ONS) income groups using Prolific.

Results: Two multiple regression models were conducted with the outcome variables of self-reported auditory function (SSQ-12) and behavioural speech perception ability (DiN). Being a regular smoker/tobacco consumer was significantly associated with worse self-reported auditory function (SSQ-12 scores) as compared to never-smokers. There were no significant predictors of speech perception ability (DiN task scores).

Conclusions: These data suggest that tobacco consumption may be associated with perceived hearing ability. It is possible that increased oxidative stress, induced by smoking, damages the inner ear, affecting hearing ability. However, we did not find an association between behavioural speech perception ability and tobacco consumption; which raises questions about the underlying mechanism. It is also possible that the measure of speech perception ability (DiN scores) was not sensitive enough to detect variations in hearing ability within our target sample. Tobacco consumption rates are higher among disadvantaged groups, which relates to previous findings that lower socioeconomic status correlates with hearing loss. Future research should explore the neurobiological mechanisms behind the effect of lifestyle and socioeconomic factors, such as smoking, on auditory function.

P39 Partial loudness of speech in 3-dimensional space: Role of signal-to-noise ratio and number of talkers

Lily Fisher, **Josef Schlittenlacher**
University College London, UK

Background: When computing the loudness of a target signal in the presence of other sounds, current computational models of loudness assume that the target signal and background noise are aggregated together on the basilar membrane without considering aspects of spatial hearing like the cocktail-party effect. In speech intelligibility, this effect has been shown to improve speech reception thresholds by up to 7 dB (Hawley et al., 2004, doi:10.1121/1.1639908). The present study investigates the partial loudness of speech in competing speech.

Method: The level difference necessary for equal loudness (LDEL) between the target in silence and the target in competing speech was studied for four signal-to-noise-ratios (SNRs; -6, -3, 0 and 3 dB). The target was always female speech at 45 degrees, the masker either a single male masker at the opposite side, or four simultaneous male speakers at 90 and 22.5 degrees on the same side, frontal and 45 degrees on the other side. Stimuli were recorded using an artificial head and presented via headphones. Twenty normal-hearing participants took part. LDELs were also calculated by a combination of the models of Glasberg and Moore (2005, 2005, J. Aud. Eng. Soc. 53:906) and Moore et al. (2018, doi:10.1121/1.5027246) that considers binaural hearing but not further processes of spatial hearing.

Results: The model predicted a decrease in partial loudness with increasing noise levels, reaching LDELs of -4 dB for the single-talker masker and -10 dB for the babble masker at an SNR of -6 dB. The experimental results followed this trend for the babble masker, reaching and LDEL of -7 dB at -6 dB SNR. For the single competing talker, all LDELs were close to 0 dB. A repeated-measures ANOVA confirmed these effects.

Conclusion: Overall, this study provided evidence for the importance of considering the effect of spatial separation in loudness perception. The surprising result of basically the same loudness in silence and that in the presence of a single competing talker that has 6 dB higher level suggests that partial loudness was probably judged during dips of the masker, and participants were able to fully focus on the target. The model predictions were reasonably close to the results for the babble noise.

Acknowledgments: This study was supported by the Royal Society, grant 232164.

P40 Speech-in-noise performance as an unhidden aspect of 'hidden hearing loss'

Stuart Rosen[1], Tim Schoof[1,2], Tim Green[1]
1. Speech, Hearing & Phonetic Sciences, University College London, UK | 2. Advanced Bionics, Valencia, CA, USA

10.5281/zenodo.14605521

Much interest surrounds the possible contribution of synaptopathy and/or neuropathy (SNpathy) to hearing difficulties that occur despite normal audiometric thresholds. We made extensive measurements in two groups of listeners with near-normal audiograms who were expected to differ greatly in the likelihood of SNpathy: 19 Y(oung) adults aged 18-25 with limited noise exposure and 23 M(iddle-aged) adults aged 44-61 with significant noise exposure. Speech reception thresholds (SRTs) were measured binaurally in speech-spectrum-shaped noise for two tasks (recognition of complex sentences and consonant identification in VCVs). To assess the use of temporal fine structure (TFS), target and masker were presented either diotically (S0N0) or with the masker in phase at the two ears and the target out of phase (SπN0). Stimuli were presented at 40 and 80dB SPL to assess claims that deficits due to SNpathy might be more prominent at high levels.

The main findings were: 1) Performance in the Y group was always better than that in the M group, with differences ranging from 0.3 to 3.5 dB across the 8 conditions. 2) SRTs worsened across age in the M group for some conditions, meaning that the younger M listeners could be performing similarly to the Y group. 3) Greater deficits were found for SπN0 than for S0N0 conditions, implicating some deficits in the M group for processing TFS. 4) Greater deficits were found for VCVs than for complex sentences. 4) The effect of level was small. 5) Group differences in SRTs were not related to a small difference in audiometric thresholds at frequencies ≤ 4 kHz, but mean thresholds at high frequencies (8 – 16 kHz) were a good predictor for VCV SRTs in all 4 conditions.

One complicating factor in interpretation is that a technical error meant that the masker only extended to 8 kHz whereas the speech stimuli were full bandwidth, so some part of the group difference may be due to the Y group having unmasked access to speech information ≥ 8 kHz where their thresholds were considerably better than the M group. In a separate study of Y listeners, we estimated the improvement in SRTs for 80 dB SPL, in phase VCVs, to be nearly 2 dB. Hence, group differences would have been smaller if listening had been properly restricted to speech information ≤ 8 kHz. Even taking account of this complication, it seems plausible that SNpathy could be a factor in these relatively small group differences, but that these effects differ little across level.

Acknowledgements: Funded by the RNID UK.

P41 Partial loudness of spatially separated speech in the presence of a competing talker

Emma Foottit, Josef Schlittenlacher
University College London, UK

Background: Loudness is typically reduced in the presence of other sounds. When both target and masker are presented diotically or from the same loudspeaker, partial loudness can be estimated well by considering energetic masking. For speech intelligibility, spatial separation of sounds improves intelligibility, even though the degree of spectral masking remains largely unchanged. This is known as the cocktail party effect. In the present study we investigated this effect on partial loudness by varying the spatial separation between a female target speaker and a competing male speaker.

Method: The target speech was positioned at 0° and 45°. In the respective reference conditions, the competing speech was positioned at the same angle and had the same sound pressure level as the target. Comparisons were made by twenty participants for spatial separations of 22.5°, 45°, and 90° using a 1-up/1-down two-alternatives forced-choice procedure. When the target was at 45°, the comparison stimulus was positioned towards or crossing the centre. Sounds were recorded with an artificial head and presented via headphones. After combining the target and masker, the signals to the left and right ear were switched randomly to reduce a-priori focus on the target. All stimuli had a duration of 2 seconds.

Results: For the 0° target, level differences required for equal loudness were close to 0 dB, implying that the target was perceived as equally loud when the masker coincided in direction and when it was positioned on the side. For the 45° target, the effect of spatial separation was small, less than 1.5 dB, but statistically significant, and increased with separation.

Conclusions: The level difference necessary for equal loudness (LDEL) was small compared to previous speech intelligibility studies that used similar separations and found differences in speech reception thresholds larger than 5 dB. The lack of any effect for the frontal reference was surprising. It may be that participants always knew where the target stimulus came from, which could have facilitated attention. Another explanation for the small, if any, effect in both conditions could be that participants based their judgment on segments when the target was loud and between syllables or words of the masker. This would suggest participants effectively judged partial loudness at a high signal-to-noise ratio, similar to loudness in silence, and thus is less dependent on the masker.

Acknowledgments: This study was supported by the Royal Society, grant 232164.

P42 Metacognition for hearing in noise and age-related hearing loss

Chiara Valzolgher[1], **Elena Giovanelli**[1], Elena Gessa[1], Tommaso Rosi[2], Giulia Pasin[3], Rebecca Ferrari[3], Matteo Pegoraro[4], Francesco Pontoni[4], Domenico Spinella[5], Giuseppina Tromballi[6], Giuseppe Nicolò Frau[6], **Francesco Pavani**[1]

1. Center for Mind/Brain Sciences – CIMeC, University of Trento, Rovereto, Italy | 2. LEVEL up, Laboratorio di Comunicazione Scientifica e Progettazione Didattica, Trento, Italy | 3. Department of Psychology and Cognitive Sciences (DiPSCo), University of Trento, Italy | 4. Pontoni udito e Tecnologia, Trento, Italy | 5. Studio UdiVer, Mori, Italy | 6. Unità Operativa di Otorinolaringoiatria Ospedale Santa Maria del Carmine Rovereto APSS Trento, Rovereto, Italy

Metacognition refers to the ability to understand one's cognitive skills and monitor them during a task, encompassing self-efficacy and locus of control more broadly. In auditory perception, this skill is particularly important in challenging listening situations (e.g., hearing in noise), as monitoring one's internal states and understanding one's abilities can facilitate task performance by informing the use of effective strategies. This skill becomes even more relevant with aging. Recent experimental studies have shown that metacognitive abilities related to listening in noise are preserved in older adults, in contrast to other domains like metamemory. In this study, we measured metacognitive skills related to listening in noise, examining both task-specific components (such as metacognitive monitoring) and more general aspects (such as self-efficacy, locus of control, and metacognitive knowledge) in three groups: 26 young adults with normal hearing, 26 older adults with normal hearing, and 27 older adults with mild hearing loss who do not use hearing aids. The primary objective was to explore whether metacognition for hearing in noise changes in individuals with age-related hearing loss. Secondly, we assessed the frequency with which these groups of participants reported using behavioral strategies to manage listening challenges in complex auditory situations. Our findings showed no significant differences in metacognitive abilities related to listening between young and older adults. Older adults with hearing impairment showed lower self-efficacy and a stronger external locus of control regarding hearing (i.e., a greater perception of having a hearing difficulty) compared to their normal-hearing peers. However, older adults with hearing impairment reported using certain behavioral strategies less frequently than their normal-hearing counterparts. These findings suggest that older adults maintain preserved metacognitive abilities in the auditory domain and good awareness of their hearing impairment when present. However, there appears to be a challenge for older adults with hearing loss in translating this awareness into practical everyday strategies. Therefore, our findings encourage future studies to test and design targeted support interventions for this population, with an emphasis on effectively measuring both behaviors and attitudes toward action.

P43 Speech-in-noise processing in Alzheimer's disease and primary progressive aphasia

Sophie A. Froud[1], Benjamin A. Levett[1], Lucy B. Core[1], Jessica Jiang[1], Charles R. Marshall[2], Jason D. Warren[1], Chris J.D. Hardy[1]

1. Dementia Research Centre, Queen Square Institute of Neurology, University College London, UK | 2. Centre for Preventive Neurology, Queen Mary University of London, UK

Background: In daily life, understanding spoken messages generally requires decoding of speech signals embedded in variably noisy acoustic backgrounds. This is a computationally demanding neural task that is likely to be vulnerable early in the course of neurodegenerative brain pathologies. However, how speech-in-noise processing is affected in different dementia syndromes, how it relates to peripheral hearing loss and the brain basis for impaired speech-in-noise perception remain unclear. Here we addressed these questions using a speech-in-noise test in a well-defined cohort of patients with Alzheimer's disease (AD) and language-led dementias (primary progressive aphasia (PPA) syndromes).

Methods: We administered an in-house Digit Triplets test in which participants listened binaurally to three digits presented in competition with speech-weighted white noise and reported the digits heard on each trial. Signal-to-noise ratio (SNR) was adjusted from trial to trial based on task performance following an adaptive staircase protocol, and mean SNR for speech comprehension was generated for each individual. 48 patients (15 with AD, 13 with logopenic variant (lv)PPA, 12 with semantic variant (sv)PPA, eight with nonfluent/agrammatic variant (nfv) PPA) and 22 cognitively-well, age-matched older individuals participated. Peripheral hearing function as indexed on pure tone audiometry and general cognitive functions were assessed using standard test procedures. Clinical diagnosis of dementia syndromes was supported on volumetric brain MRI and voxel-based morphometry (VBM) was used to identify structural neuroanatomical associations of speech-in-noise performance in the patient cohort.

Results: The AD, lvPPA and nfvPPA groups had significantly higher mean SNR (performed significantly worse) than cognitively-well older listeners on our speech-in-noise task, while the svPPA group did not significantly differ from healthy controls. SNR was significantly higher in the lvPPA group than in all other syndromic groups. Auditory cortical correlates of speech-in-noise performance were identified on VBM in the patient cohort.

Conclusion: AD and PPA syndromes are associated with separable profiles of speech-in-noise impairment, which appear to capture auditory brain dysfunction in these canonical dementia syndromes.

P44 Testing the moderating effect of interference control on the impact of fluid intelligence on SPiN

Ruijing Ning, **Erik Marsja**, Emil Holmer
Linköping University, Sweden

Cognitive factors have been found to be critical in supporting Speech perception in noise (SPiN) performance, as they shape how listeners process, filter, and prioritize auditory information amidst competing signals. Fluid intelligence—the capacity to reason and solve novel problems—and interference control—the ability to suppress irrelevant information—have been shown to contribute to SPiN outcomes. We hypothesized that interference control might moderate the role of fluid intelligence in SPiN, such that strong interference control would reduce the dependency on fluid intelligence during challenging listening conditions.

In this study, we assessed SPiN performance in 181 individuals with hearing impairment. Participants completed the Hagerman test under challenging conditions, where sentences were presented alongside a four-talker babble noise masker while they used hearing aids with linear amplification. We analyzed the signal-to-noise ratios (SNRs) required for participants to correctly repeat 80% of the words. Cognitive measures included fluid intelligence, assessed through Raven's Progressive Matrices, and interference control, measured as the sensitivity (A prime) in a Go/No-Go test. A linear regression model was used to examine the main effects of and interaction between fluid intelligence and interference control. Pure-tone average (PTA) was included as a control variable to account for hearing levels.

Our results indicate that both fluid intelligence and PTA significantly predicted SPiN performance, supporting the role of general cognitive ability and hearing acuity in challenging listening tasks. Contrary to our initial hypothesis, however, interference control did not show a main effect, nor did it interact with fluid intelligence to predict SPiN performance. Bayes factor analysis provided anecdotal evidence against including the interaction between fluid intelligence and interference control sensitivity in the predictive model, suggesting that this form of interference control does not moderate the relationship between fluid intelligence and SPiN under these specific high-cognitive-load conditions.

These findings indicate that while fluid intelligence is a valuable predictor of SPiN performance, interference control—at least as measured in our task—may not significantly influence SPiN outcomes in hearing-impaired hearing aid users under demanding noise conditions. We also found some evidence against the idea that interference control moderates the role of fluid intelligence in SPiN. Further research could explore alternative measures of interference control to clarify its role in SPiN and identify cognitive profiles that may require special attention in auditory support.

P45 The role of noise type and task demand in moderating the effect of hearing aid signal processing for speech in noise

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

Background: Hearing aids use various signal processing settings, such as linear and non-linear amplification and noise reduction algorithms, to enhance speech recognition in noise. Speech reception thresholds (SRT) indicate the minimum signal-to-noise ratio needed for speech recognition, tested at different task demands (50% and 80% accuracy). Noise types like speech-shaped noise (SSN) and four-talker babble present different challenges due to their distinct acoustic properties. Speech-state noise is continuous and predictable and has been suggested to be more straightforward to manage with linear amplification. At the same time, the dynamic four-talker babble benefits more from adaptive algorithms such as non-linear amplification and noise reduction. We aimed to examine whether the type of noise and task demand moderates signal processing settings.

Rationale: Acquiring knowledge of how different signal processing settings affect speech recognition in various noise environments, particularly cognitive load at different task demands, can be used to advance hearing aid technology and improve user experience.

Methods: Data from 215 individuals wearing hearing aids (mean age = 60.8 years, SD = 8.8) from the n200 database were used. The Hagerman task was used to assess speech recognition in noise performance. The main and interacting effects of hearing aid signal processing settings (linear amplification, non-linear amplification, and noise reduction), noise type (SSN vs. four-talker babble), and task demand (50% vs. 80% accuracy) on speech recognition performance were tested. SNRs were estimated for each signal processing setting at both 50% and 80% accuracy across the two noise conditions: four-talker babble and SSN. Linear mixed effects modeling was employed to analyze the data.

Results: Preliminary results suggest that the interaction between task demand and noise type influenced the effect of signal processing settings. Noise reduction was more effective in 4-talker noise compared to SSN noise. At higher task demands (80% SRT), noise reduction differed from nonlinear fast compression and linear amplification, which also showed differences from each other. At lower task demands (50% SRT), nonlinear and linear amplification showed slight differences, but noise reduction remained the most effective approach in both scenarios.

Conclusions: Noise reduction showed strong effectiveness in four-talker babble, especially under higher task demands (80% SRT). In contrast, at lower task demands (50% SRT), non-linear and linear amplification performed similarly in SSN. These results show the benefits of tailoring hearing aid settings to complex noise conditions, but that further research is needed to refine hearing aid technology for real-world listening environments.

P46 The strength of association between cognitive capacity and effective speech recognition in noise is moderated by noise and task conditions

Emil Holmer, Ruijing Ning, Erik Marsja
Department of Behavioural Sciences and Learning, Linköping University, Sweden

Background: Previous research indicates that cognition is an important predictor of speech recognition in noise in individuals with hearing loss. However, informational maskers, e.g., background babble, seem to tax cognitive resources to a greater extent than energetic maskers, i.e., non-speech noise, and relaxed expectations on task performance are associated with reduced cognitive dependency. In the present study, we investigated whether the prediction of cognitive capacity on speech recognition in noise is moderated by the interacting effect between type of background noise and task demands, in adult hearing aid users.

Rationale: We assume that effective speech recognition in noise is dependent on cognitive resources and propose that this is especially true when the tolerance for inaccurate task performance is low, and the speech signal is presented together with an informational masker. Thus, we hypothesized that the strength of the prediction of cognitive capacity on speech recognition in noise performance increases with the combined effect of informational background noise and high expectations on accurate task performance.

Methods: For this study, data from 215 adult hearing aid users (mean age = 60.8 years, SD = 8.8; mean PTA4 = 37.6, SD = 11.1) from the N200 database were used. To assess speech recognition in noise performance, the Hagerman task was used. Speech reception thresholds were estimated for 50 and 80 percent accuracy in two noise conditions: 4-talker babble and speech-shaped noise. To measure general cognitive capacity, the Raven's matrices test was used. Main and interacting effects of noise type (speech-shaped noise vs. 4-talker babble), task demand (50 vs. 80 percent accuracy), and general cognitive capacity (Raven's matrices) on signal-to-noise (SNR) ratio from the Hagerman task were tested in a linear mixed effects model. Age and PTA4 were included as covariates of no interest.

Results: As predicted, the strength of association between cognitive capacity and speech recognition in noise performance increased with the interacting effects between noise type and task demands. Greater cognitive capacity was associated with lower SNR's in all conditions, and this association was amplified in the 4-talker babble condition when task demand was high (80 % accuracy).

Conclusions: Effective speech recognition in noise is dependent on cognitive capacity, and the more difficult the listening situation becomes, the greater the dependency. Future research should identify factors that increase the cognitive burden of speech recognition in noise, and target how these can be modified to ease processing.

P47 Will I speak louder if I see you struggle understanding?

Elena Gessa[1], Chiara Valzolgher[1], Elena Giovanelli[1], Tommaso Rosi[2], Alessandro Farnè[3], Francesco Pavani[1]

1. Center for Mind/Brain Sciences – CIMeC, University of Trento, Italy | 2. Level Up s.r.l., Trento, Italy | 3. INSERM U534, Espace et Action, Bron, France

During face-to-face interactions, the quality of the conversation can be estimated on non-verbal cues expressed behaviorally by conversational partners. Here, we investigated whether a specific subset of these non-verbal cues (i.e., facial expressions of confusion, approaching movements towards the speaker) can prompt speakers to modify acoustic-phonetic characteristics of their speech to enhance clarity. We addressed this research question in the absence of acoustic feedback from the environment, without previous knowledge of the listener status and when no verbal feedback from the conversational partner is provided. In Experiment 1, we asked participants (N=22) to read short sentences to a listener seated in a separate room and visible through a head-mounted display. We informed participants that the listener hears their voice plus a background noise of different intensities, although this was not audible to them. The listener was, in fact, a confederate whose non-verbal cues were pre-recorded in silence with a 360 video-camera to convey three different levels of inferred listening effort (i.e., easy, medium and hard listening). We found that, with increasing non-verbal cues, participants reported that the listener experienced increased listening effort and reduced speech comprehension. Importantly, speakers also changed speech proportion (i.e., speech rate), voice intensity and fundamental frequency, which increased from easy to medium, and from medium to hard listening (as inferred from non-verbal visual cues). In Experiment 2 (N=12), we replicated the study but omitted from the cover story the information that the listener hears a background noise. Even in the absence of this contextual information of the listener's experience, the speakers modify acoustic-phonetic characteristics of their speech to enhance clarity as a function of non-verbal cues. Unlike Experiment 1, participants appeared to have prioritized clarity over loudness, because vocal intensity was the only parameter that did not change. Our experiments show that facial expressions of confusion and approaching movements towards the speaker can effectively elicit acoustic-phonetic adaptations, in the absence of acoustic feedback from the environment, without previous knowledge of the listener status and when no verbal feedback from the conversational partner is provided. We suggest that, during face-to-face interactions, non-verbal cues offer a significant advantage by providing the speaker with real-time feedback on the quality of the conversation. This instant feedback is not achievable through prior information about the listener's characteristics or the environment.

P48 The impact of noise on strategies for joint decision-making in triadic groups

Ingvi Örnólfsson[1], Axel Ahrens[1], Riccardo Fusaroli[2], Torsten Dau[1], Tobias May[1]
1. Hearing Systems Section - DTU Health Technology - Kgs. Lyngby, Denmark | 2. Linguistics, Cognitive Science and Semiotics - Aarhus University - Aarhus, Denmark

Verbal communication allows groups to make joint decisions by drawing on the knowledge and information contributed by each group member. Language-mediate joint decisions are common and important in many everyday life situations, and yet we know very little about how they are impacted by communicative barriers such as noise or hearing impairment.

To investigate this, we use a task paradigm in which triadic groups answer a series of binary geography questions (e.g., “Which country has the most inhabitants: A) Spain or B) Hungary?”). Group members first respond individually using a confidence scale ranging from -10 (fully confident in option A) to +10 (fully confident in option B). If there is disagreement, members are instructed to reach a joint decision through verbal communication in one of two levels of background noise. To quantify the effects of noise on joint decision-making, we employ a model-based approach in which asserts that the group forms its joint decisions by combining individual members’ prior decisions using one of three strategies. The first strategy, “group averaging”, equally weighs each member’s prior decision in forming the joint decision. The second strategy, “confidence slating”, incorporates only the most confident member’s input in the joint decision. The third strategy, “majority voting”, follows the preference of the majority. Confidence slating and majority voting can be considered suboptimal decision heuristics requiring less information sharing at the cost of not utilizing all available information. We hypothesize that these heuristics may be used when sharing information becomes more effortful and/or less reliable due to noise. The inclusion of these strategies is motivated by a previous study on collaborative decision-making in triads, where discussions in high noise levels led to more confidence slating but less majority voting. Using a model of the decision-making process with these strategies formally parametrized, we investigate how the different strategies are affected by background noise. Preliminary results suggest that confidence slating and majority voting are less prominent in high-level background noise, whereas group averaging is more common when discussions take place in quiet.

Joint decision-making tasks provide insights into how the utility of communication is limited by factors such as noise or hearing impairment. These tasks may eventually facilitate more ecologically valid assessments of hearing intervention benefits in real, face-to-face conversations.

P49 Development of a naturalistic conversation paradigm for use with wearable neuroimaging technologies

Liam H Collins-Jones, Nicholas Haywood, Marina Salorio-Corbetto, Deborah Vickers
University of Cambridge, UK

Background: Age-related hearing loss (ARHL) is a gradual loss in high-frequency hearing that begins to emerge in middle age, leading to difficulties in understanding speech-in-noise. Population-level studies have pointed towards a link between ARHL and dementia, and a lack of social interaction has been linked to an increased risk of developing dementia. However, no causal mechanisms have been definitively identified to explain these associations.

Established neuroimaging technologies restrict the range of environments in which neural data can be collected. Emerging advances in wearable neuroimaging technology, however, hold promise to enable researchers to study the neural basis of speech processing in ecologically valid contexts.

Here, we present a new paradigm for studying speech processing during a naturalistic conversation, mimicking everyday social interactions, whilst collecting data on neural activity with wearable high-density functional near-infrared spectroscopy (fNIRS).

Methods: The paradigm consists of two tasks. For both tasks, a participant and experimenter sit on opposite sides of a square (0.8m x 0.8m) table. The participant wears a high-density fNIRS cap. In task one, the participant and experimenter engage in a 3-minute conversation, assuming “speaker” and “receiver” roles which alternate every 15 seconds. Two other tables in the room are used to mimic a cafeteria set-up, each with a loudspeaker which plays 6-speaker babble noise or cafeteria noise to alter the acoustic soundscape. The task is run in quiet, with a loudspeaker on the farthest table presenting interferers, with a loudspeaker on the nearest table presenting interferers, and with both noise sources playing at the same time.

In task two, the experimenter and participant engage in an unstructured 6-minute conversation, conducted in each of the auditory scenarios used in task one. We intend to explore other approaches for modifying receiver (e.g. earplugs and black-out glasses) and speaker (e.g. complexity of language and speed) characteristics to determine how these affect communication.

Results: We will present results of behavioural scores in these scenarios and preliminary neuroimaging data collected during this paradigm with typically hearing participants.

Conclusions: Our naturalistic social communication paradigm will be used with wearable neuroimaging techniques in older adult cohorts with and without hearing loss to explore the neural basis of speech processing. We will further develop the paradigm for use with participants with hearing loss and/or cognitive impairment, contributing to the evidence base on whether there is a causal link between ARHL and cognitive impairment.

P50 The effect of semantic context in dynamic cocktail-party listening: Influences of age and hearing loss

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, Germany

Everyday communication typically entails situations in which multiple persons speak simultaneously ("cocktail-party"). Cocktail-party situations can be static (constant target talker) or dynamic (target changes unpredictably). Dynamic situations require listeners to monitor multiple talkers and to switch attention when the target changes, which is why they cause a higher cognitive load and thus lead to a decrease in speech recognition compared to static situations, which is referred to as "costs" (e.g. Lin & Carlile, 2015, doi:10.3389/fnins.2015.00124; Meister et al., 2020, doi:10.1016/j.heares.2020.108020). A common limitation of corresponding studies is, however, the use of matrix sentences that provide hardly any semantic context, as opposed to sentences from everyday life. Semantic context might help listeners to segregate auditory streams and recall words from memory (e.g., Meister et al., 2013, doi:10.1016/j.neuroscience.2012.12.006). We therefore hypothesize that high semantic context can reduce cognitive load and thus allows listeners to use a higher proportion of their cognitive capacity to overcome the challenges of dynamic cocktail-party listening, which in turn lowers costs. Moreover, data from Meister et al. (2013) suggests that this context benefit tends to be smaller in older than in younger listeners, at least when a divided attention task was used.

We simulated cocktail-party situations with three spatially separated competing talkers. Participants had to repeat back the words uttered by the target talker (who had to be identified by her voice). Verbal response times were recorded automatically to serve as a measure of cognitive load. In static situations the target talker's location remained constant and was announced prior to stimulus presentation, whereas dynamic conditions entailed unpredictable location changes of the target. Low context ("Lisa draws twenty heavy apples.") as well as high context ("The rocket flies into space.") sentences were used.

At the previous SPIN workshop we presented the results of a small sample of young normal hearing adults obtained with this setup (Wächtler & Meister, 2024, doi:10.5281/zenodo.10565092). To be able to study the effects of age and hearing loss, our current presentation will also include data from elderly people (> 60 years) with and without hearing loss.

We will examine the impact of semantic context on the costs of dynamic cocktail-party listening and how it is influenced by age and hearing loss. Moreover, verbal response times will be analyzed to shed some light on how semantic context could affect cocktail-party listening beyond recognition performance.

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

P51 Impact of distractor predictability during the perception of speech-in-noise

Alice E. Milne
Lancaster University, UK

Shuxia Cheng, Julia Wiktorowicz, Claudia Contadini-Wright
University College London, UK

The brain is highly sensitive to auditory regularities and leverages predictable information in the environment to process sensory input efficiently. Predictable sequences of tones have been found to be both less distracting and to require less cognitive effort than the same tones in a random order. Previous experiments have been restricted to auditory scenes comprising of only tone-pip sequences. In the current study, we seek to understand how tone-sequence predictability interacts with more complex auditory scenarios, specifically speech comprehension. Participants were presented with a spoken sentence embedded in Gaussian noise and were asked to detect key words in the sentence. Speech in noise thresholds were individually calibrated. In a third of the trials, they only heard the speech in noise (noise only). In the other two conditions, a distractor was also presented in one ear; the distractor consisted of a sequence of tone-pips that was either a highly predictable repeating pattern of five tones (noise + predictable) or the same tones in a random order (noise + random). We recorded pupil responses throughout. Pupillometry can be used as a proxy for cognitive effort, with larger pupil sizes associated with greater cognitive effort. Previous evidence showed that detection of a regularity induces a sharp decrease in pupil diameter, indicative of a reduced demand on cognitive resources. Consistent with those findings, we predicted that the noise-only condition would require the least effort and would be associated with the smallest pupil diameters, while the noise + random condition would be the most challenging and associated with the largest pupil sizes. Our results supported our prediction but dissipated over time, suggesting that listeners were able to rapidly adapt to the task, likely suppressing the information regardless of its predictability. We also explored differences in microsaccades and pupil dilation rates. This paradigm brings together two separate cognitive challenges: understanding speech in challenging conditions and processing of extraneous auditory information to determine the appropriate allocation of cognitive resources. It provides the foundation to further explore the interaction of these two processes in different populations.

P52 Association between self-reported difficulties in speech intelligibility in noise and Matrix speech-in-noise test performance

Leny Vinceslas[1], Vit Drga[1], Jesko Verhey[2], Ifat Yasin[1]
1. Department of Computer Science, University College London, UK | 2. Department of Experimental Audiology, Otto von Guericke University Magdeburg, Germany

Cumulative lifetime effects of noise exposure may or may not be related to reduced performance in speech in noise tests or elevated audiometric thresholds. In addition, some effects of noise-induced hearing impairment may be evident at the neural level, such as cochlear synaptopathy, but not necessarily be observable as raised hearing thresholds of the typical audiogram (often testing sound frequencies between 0.25 and 8 kHz). Individuals' experience of noise exposure is often difficult to measure, but a structured interview approach may prove useful in eliciting information regarding noise exposure across a range of noise exposure activities.

In the current study, the association between a participants' self-report of noise exposure and awareness of difficulties in following speech in ongoing noise were compared with their performance on the Matrix speech-in-noise test. The Matrix test is an adaptive speech-in-noise test using a larger speech corpus than other commonly used tests. All participants had hearing thresholds within normal limits, were assessed on their past history of noise exposure, and asked about their ease/difficulties following speech in ongoing noise.

For each participant, the Matrix test measures were fitted with psychometric intelligibility functions. The deduced intelligibility function parameters, i.e., speech reception threshold (SRT) and function slope, were compared with self-reported score of difficulties with speech intelligibility in noise and past noise exposure history. These resultant associations are discussed with reference to intra- and inter-participant variances in self-report of perceived difficulty and recall of noise experiences.

P53 Does puberty influence auditory processing?

Ellen Saliën, **Ugo Benrubi**, **Axelle Calcus**
Université Libre de Bruxelles, Belgium

Late Auditory Event-Related Potentials (LAERs) are widely used to investigate central auditory responses to sounds. Components of the LAERs reflect successive stages of sound processing, such as detection and discrimination, throughout the auditory pathway, and are influenced by (in)congruent visual information. From infancy to old age, LAERs show morphological and topographical changes. Their maturation is thought to follow a stepwise trajectory marked by distinct changes at the start and end of adolescence. Interestingly, the beginning of adolescence coincides with puberty onset, which triggers a cascade of hormonal changes that may lead to enhanced neural plasticity. Puberty could thus contribute to the maturation of sensory brain regions, including the central auditory cortex in the temporal brain region. To date, the effect of puberty on the development of the central auditory system has remained unexplored. Here, we hypothesize that pubertal maturation underlies the stepwise maturation observed

in LAERs. Our project aims to investigate the effect of puberty on the development of auditory and audiovisual speech perception. To date, we have tested 6 pre-pubertal children, 37 early-to-late-pubertal adolescents, and 18 post-pubertal adults. We plan to record LAERs from over 120 children, adolescents, and adults. Participants were presented with speech stimuli in an oddball paradigm across four conditions: audio only, audiovisual congruent, audiovisual incongruent, and visual only. Preliminary findings suggest that puberty drives certain changes in LAER morphology throughout adolescence, particularly in pre/early to mid/late pubertal adolescents. These results indicate that puberty may play a role in the functional maturation of the central auditory pathway, potentially contributing to the fine-tuning of adolescents' ability to process speech in ecologically challenging, noisy settings.

P54 The role of age-related changes in alpha activity during dual-task speech perception and balance

Jessica Pepper[1], Jason Braithwaite[1], Theodoros Bampouras[2], Helen Nuttall[1]
1. Lancaster University, UK | 2. Liverpool John Moores University, UK

Research suggests that older adults find it more difficult than younger adults to flexibly allocate attentional resources between two co-occurring multisensory tasks, such as perceiving speech-in-noise whilst maintaining balance. This attentional control may be reflected in oscillatory alpha activity, with increases in activity reflecting inhibition of different brain regions and decreases reflecting neural activation. However, there is limited research examining how alpha activity during dual-task conditions may change as a function of healthy ageing. This study aimed to investigate how younger and older adults reallocate attentional resources when perceiving speech-in-noise whilst maintaining balance, and how these age-related changes are reflected in alpha activity. Nineteen younger adults (18-35 years old) and sixteen older adults (60-80 years old) were asked to identify words in audiovisual sentences extracted from the Grid corpus. Participants completed this speech perception task with or without background noise, whilst standing in an easy balance position (feet side-by-side) or a difficult balance position (feet in tandem). Throughout the task, fronto-central and parieto-occipital alpha activity was recorded using EEG, to measure activation in brain regions associated with balance maintenance and audiovisual speech perception, respectively. Mixed ANOVAs revealed that all participants produced a weaker speech perception performance in noisy listening conditions. However, speech perception in the noisy listening condition was most accurate when participants stood in a challenging balance position, in contrast to our hypotheses. Whilst these behavioural effects were not reflected by fluctuations in parieto-occipital alpha power, decreases in fronto-central alpha power were greater in clear listening conditions compared to noisy listening conditions. Taken together, the results suggest that increasing cognitive load with a secondary multisensory task may not always be detrimental to balance maintenance in physically and cognitively fit older adults.

P55 Performance limitations of english-trained speech enhancement models on french phoneme categories

Imane Ghanem[1,2], **Nasser-eddine Monir**[2], Romain Serizel[2]
1. Bordeaux INP, Bordeaux, France | 2. Université de Lorraine, CNRS, Inria, Loria, Nancy, France

Unfortunately the presentation of this poster has been cancelled.

Background: Understanding speech in noisy environments is a challenge for many, especially individuals with hearing loss who rely on clear signals to comprehend spoken language. Speech enhancement (SE) technologies have become a key solution in hearing aids, aiming to reduce background noise and improve the intelligibility of speech. However, SE models trained on one language may not generalize effectively to others, as language-specific phonetic characteristics influence the model's performance.

Rationale: This study explores how SE models trained on English handle interference, artifacts, and distortions in both English and French, shedding light on potential limitations when applying language-specific SE models across different languages.

Method: We employed three SE models to enhance speech under varied noise conditions, using 1000 audio samples from LibriSpeech for English and FHarvard for French, a balanced dataset with phrases recorded by one male and one female native speaker for phonetic diversity. We also included the French LibriSpeech corpus, selected for its rich recordings, diverse speakers, and substantial recording hours.

Results: At the utterance level, English showed slightly higher initial interference but benefited from greater improvement through enhancement, with both languages performing similarly in terms of artifact control and overall distortion reduction. The amount of interference in the original signal, at the phoneme level, was similar for English and French across most vowels. However, English phonemes, particularly those with nasal, plosive and fricative characteristics, tend to exhibit a higher susceptibility to noise interference, while French phonemes show increased interference within specific consonantal categories, such as lateral and affricate sounds. Post-speech enhancement, French and English phonemes show similar levels of interference for sibilant, open-mid, and open categories, while French phonemes are notably more affected in plosive, approximant, and close-mid categories—with a particularly pronounced difference in affricate, lateral, and close sounds—whereas English phonemes exhibit higher residual interference in nasal and fricative sounds. Speech enhancement (SE) models trained on English generally reduce interference more effectively in English than in French across most phoneme categories, with minimal differences for sibilants and fricatives. The models introduce similar levels of artifacts for both languages, except for increased artifacts on French laterals and English affricates. Distortions are also comparable between languages, except for higher distortion levels in French laterals.

Conclusion: This study reveals language-specific limitations in English-trained SE models, as they reduce interference more effectively for English than French, with notable differences in phoneme category performance, highlighting the need for tailored SE solutions across languages.

P56 What you see is what you hear: Exposure to congruent social cues improves speech intelligibility under adverse listening conditions

Ruohan Guo, Fengjie Cheng, Bronwen Evans
Department of Speech, Hearing and Phonetic Sciences, University College London, UK

Background: Social information can affect speech processing. Studies have shown that manipulating listeners' expectations about accent regions or simply exposing them to culturally iconic objects (e.g., stuffed toys, [1]) can significantly bias phoneme categorization. These findings support exemplar-based models of speech perception, where non-acoustic information plays a crucial role in understanding and evaluating speech. Nevertheless, prior research has primarily examined social priming effects at the phoneme level, with findings yet to be replicated in British populations. In this study, we took a different approach and investigated how visual geographical cues depicting London or Glasgow influence the intelligibility of natural sentences produced in two accents: Standard Southern British English (SSBE) and Glaswegian English (GE).

Method: Fifty-three listeners aged 18-50 were recruited via Prolific and completed the experiment online. They transcribed 108 IEEE sentences at 3 signal-to-noise ratios (+3 dB, 0 dB, -3 dB), produced in SSBE and GE accents (2 male speakers per accent). In each accent & noise condition, one-third of the sentences were presented with congruent visual cues (to the accent region) and one-third with incongruent visual cues. The remaining sentences were paired with blank silhouettes giving no informative social information, acting as a baseline condition. Of the participants, 24 were SSBE speakers from southern England with minimal Scottish English exposure, while 29 were Scottish residents familiar with both accents.

Results: Transcription accuracy was operationalized as Token Sort Ratio and modeled using a linear mixed model in R. Unsurprisingly, performance decreased as the noise level increased ($\chi2(2) = 39.26$, $p < 0.001$). GE speakers achieved higher transcription scores than SSBE speakers overall ($\chi2(1) = 17.38$, $p < 0.001$; $z = 4.17$, $p < 0.001$): GE listeners performed equally well with SSBE & GE sentences at each noise level ($z = -0.61$, $p = 0.54$), while SSBE listeners performed more poorly with GE ($z = -5.12$, $p<0.001$). Notably, listeners from both accent backgrounds showed higher transcription accuracy when visual cues matched the accent region of the auditory stimuli compared to mismatching cues ($\chi2(2) = 3.94$, $p = 0.02$). Overall, the results suggest that socially meaningful cues can influence speech-in-noise recognition through a largely automatic process: exposure to visual cues activated corresponding regional concepts, which interacted with the processing of phonetic variation.

Reference: [1] Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. Linguistics, 48(4), 865–892.

P57 Cortical and subcortical timescale balance in speech recognition

Shota Murai, Takamitsu Watanabe
International Research Center for Neurointelligence (WPI-IRCN), University of Tokyo, Japan

Background: Speech perception relies on a wide range of neural time windows to integrate and segregate diverse neural inputs derived from auditory signals. Such a capacity to process temporal information is represented by the intrinsic neural timescale (INT), which quantifies neural dynamics in local brain regions. While previous studies have shown that multi-timescale processes in the language-related cortex underpin speech perception, how the processes in subcortical areas, particularly the inferior colliculus (IC) and medial geniculate body (MGB), affect speech perception, and whether mismatches of the timescales across cortical and subcortical areas disturb it, are unknown.

Rationale: Investigating the relationship between INT of language-related cortical areas and subcortical areas may clarify how these regions coordinate to support speech perception, especially in noisy environments.

Methods: To explore the relationship, we collected behavioural and functional magnetic resonance imaging (fMRI) data from 31 healthy right-handed participants, all of whom provided written informed consent in accordance with protocols approved by the University of Tokyo. We first recorded fMRI signals from participants during a sentence recognition task with varying levels of background noise (from −12 dB to 12 dB SNR); then, by applying hierarchical clustering analysis to noise-dependent brain activity patterns, we determined three primary cortical clusters: auditory-language, somatosensory, and fronto-parietal cortical networks. Along with this experiment, we calculated the INTs of these clusters and the bilateral IC and MGB (subcortical components) by examining the autocorrelation of resting-state fMRI signals recorded from the same participants. Speech recognition performance in the behavioural task was measured using the thresholds of a psychometric function.

Results: Partial correlation analysis indicated that the INT of the auditory-language cluster did not correlate with the MGB, but it did correlate with the other cortical clusters, the IC, and speech recognition performance. This finding underscores the auditory-language cluster's role as a hub that orchestrates information processing across the cortical and midbrain network. Furthermore, individuals with higher speech recognition performance had shorter INTs in the auditory-language cluster but longer INTs in the IC. In fact, individuals with a larger INT mismatch between the auditory-language cluster and the IC were likely to exhibit poor speech recognition performance.

Conclusions: This study emphasizes that a balance between cortical and subcortical INTs is essential for accurate speech processing, pointing to the auditory-language cluster as a hub facilitating efficient information transfer between cortical and subcortical areas.

P58 Can the short-time objective intelligibility metric predict recall accuracy of spoken sentences in face-masked speech?

Cleopatra Moshona
Engineering Acoustics, Technische Universität Berlin, Germany

Background: Previous research indicated that face-masked speech can impair memory accuracy in listeners, particularly in noisy environments. However, it is unclear whether this decline results from reduced intelligibility due to signal degradation, increased listening effort that diverts cognitive resources from memory encoding, or a combination of both. Dual-task paradigms are often used to explore these interactions, though they may not always be feasible in certain experimental designs.

Methods: The present study investigated how speech style adaptations affect listeners' working memory performance and retrospectively assessed the predictive value of the short-time objective intelligibility (STOI) metric. Eighty-two participants (48 female) listened to audio recordings of a female and male German native speaker uttering matrix-type sentences, with and without a face mask, in casual and Lombard speech. They completed a cued recall task, memorizing the last two words of 100 sentences, divided into 20 blocks, while exposed to task-irrelevant multi-talker babble with a speaker-adjusted, fixed SNR of +6 dB. A mixed binary logistic regression was calculated to predict recall accuracy using speech condition, speaker and listener sex as fixed factors and word position, serial sentence position, and block number as covariates. Random intercepts for participants and items and a random slope for serial sentence position by participant were included. A linear mixed-effects model was used to analyze the effects of sound condition and speaker on computed STOI scores and a point-biserial correlation evaluated the relationship between mean STOI and recall accuracy scores per sentence.

Results: The mixed binary model revealed significant effects on recall accuracy for all predictors except for speaker. Post-hoc comparisons indicated a 31.9% decrease in the likelihood of recalling keywords in masked casual speech versus masked Lombard speech, and a 27.4% decrease compared to unmasked Lombard speech. Recall odds were also 25.3% lower in masked casual speech relative to unmasked casual speech. Notably, female listeners demonstrated significantly higher recall accuracy odds, with a 48.1% increase compared to males. The linear mixed-effects model indicated significant effects of sound condition and speaker on STOI scores, with an interaction between these factors. Unmasked Lombard speech was more intelligible than masked casual speech, and the male speaker had higher scores than the female. Despite STOI scores following the same overall patterns as the experimental results, the point-biserial correlation revealed a weak positive correlation between mean STOI and recall scores per sentence, indicating that the outcome cannot be explained by intelligibility scores alone.

P59 Effect of selective attention training on speech in noise perception in older adults with near-normal auditory sensitivity

Sankalpa Mahadev[1], Srikar Vijayasarthy[2], Pruthvik S P[1]
1. JSS Institute of Speech and Hearing, Mysuru, India | 2. All India Institute of Speech and Hearing, Mysuru, India

Unfortunately the presentation of this poster has been cancelled.

Introduction: Speech perception in noisy environments involves bottom-up and top-down processing, with bottom-up integrating acoustic features and top-down involving higher cognitive skills like prediction and contextual thinking. Poor understanding may be due to decline in selective attention, and understanding how uncertainty and contextual limitations impact cognitive resources is crucial. Top-down training, in addition to devices enhancing signal to noise ratio, can also enhance speech perception in challenging listening situations.

Aim: The study aims to investigate if the training of selective attention leads to improvement in speech in noise perception in older adults with near normal hearing sensitivity. Further, to explore if training benefits are influenced by working memory and vocabulary in older adults with near normal hearing sensitivity.

Method: A quasi experimental study was conducted on 20 older adults with near-normal auditory sensitivity for improvement in speech in noise perception with & without selective attention training. The study included baseline, treatment/placebo, post-treatment, and follow-up assessments. Participants were assessed for speech perception in noise under different SNR conditions, Working memory and language proficiency & vocabulary level in baseline assessment. Experimental group underwent selective attention training using auditory & visual stroop tasks while the control group received placebo training for 6 days. Maintenance of training effects after one week was assessed during follow up assessment.

Results: The post-training results indicated that the experimental group showed significant improvement in SPIN scores at -2dB SNR with speech babble, indicating the effectiveness of selective attention training. However, the improvement in other SNR conditions was not statistically significant. The study also found that the benefits of the training were maintained one week after the intervention. Regression analysis revealed that pure tone average (PTA) and age were significant predictors of SPIN scores, particularly at challenging SNR levels.

Conclusion: Selective attention training using the auditory stroop task can significantly improve speech perception in noise, particularly in challenging conditions such as -2dB SNR with speech babble. These findings suggest that targeted cognitive training can enhance auditory processing in older adults, offering a potential intervention to improve communication in noisy environments. Further research is necessary to explore long-term benefits and the applicability of such interventions in real-world settings.

P60 Exploring infant responses to speech in noise with HD-DOT

Addison D. N. Billing[1], Robert J. Cooper[2], Deborah Vickers[1]
1. SOUND Lab, Department of Clinical Neurosciences, University of Cambridge, UK | 2. DOT-HUB, Department of Medical Physics and Biomedical Engineering, University College London, UK

Understanding how infants process speech in noisy environments is critical, as they frequently encounter complex communication settings from early infancy. This study measured cortical responses of typically hearing infants, aged 6 to 8 months, to varying levels of background noise using high-density diffuse optical tomography (HD-DOT).

We adapted nursery rhyme videos used in a well-established social behaviour paradigm to have different levels of intelligibility, for different stimuli. Infants watched audio-visual videos of women singing nursery rhymes with hand gestures, presented in three different competing babble noise (3 male and 3 female speakers) conditions: no background noise, +8 dB signal-to-noise ratio (SNR), and +4 dB SNR. Our intention was to create an objective psychometric performance function of speech engagement/perception as a function of SNR.

In the no background noise condition (high accessibility of speech), we observed significant bilateral activation in the Inferior Parietal Lobule (IPL), a region crucial for integrating audio-visual social cues. Additionally, the right dorsolateral prefrontal cortex (DLPFC) showed significant activation, suggesting that higher-order cognitive processing is engaged when speech cues are clear and undisturbed. Conversely, in the +8 dB SNR (medium speech access) and +4 dB SNR (low speech access) conditions, DLPFC activation was absent, indicating that this region's engagement may depend on the clarity of auditory signals.

The IPL continued to show bilateral responses across all noise levels, although the strength of the hemodynamic response was inversely related to the amount of background noise, suggesting that while the IPL is sensitive to auditory clarity, it can still respond to lower speech clarity input. Furthermore, the high-accessibility-of speech condition elicited significant activation in the left superior temporal gyrus (STG) and middle temporal gyrus (MTG), regions associated with language processing. This activation diminished with increased background noise, suggesting that these areas are likely more dependent on clear auditory input for optimal engagement.

The robust HD-DOT cortical responses highlight its potential as a tool for psychometric assessments in early childhood. Unlike EEG, HD-DOT is not susceptible to electrical interference, making it particularly suitable for evaluating infants using hearing aids or cochlear implants who are unable to provide reliable feedback. These findings pave the way for HD-DOT to be used in early audiological evaluations, offering insights into infants' auditory processing capabilities and aiding in the development of interventions during a critical period of speech and language acquisition and social development.

P61 Neuroplastic changes in the adolescent brain: A protocol to investigate the impact of puberty on speech-in-noise processing

Joseph F. Johnson[1], **Francesca Cavicchiolo**[1], Luna Prud'Homme[1], Antonin Rovai[1], Vincent Wens[1], Simon Dobri[2], Xavier De Tiège[1], Axelle Calcus[1]
1. Université Libre de Bruxelles, Brussels, Belgium | 2. Simon Fraser University, Burnaby, Canada

During adolescence, the human brain undergoes significant transformations that enhance cognitive, socioemotional, and perceptual abilities. This developmental stage is characterized by changes in brain networks that support these skills, including the complex tasks of detecting, following, and comprehending speech signals amidst environmental noise and competing voices. Puberty marks the onset of adolescence, potentially creating a sensitive period for neuroplasticity, where experience shapes brain development. Hormonal changes during this time correlate with observable physical changes, marking distinct stages of puberty.

Our study aims to investigate how these correlates of puberty relate to detectable functional and structural brain changes that could contribute to the fine-tuning of speech perception in noise at adolescence. Our protocol will include high-density electroencephalography (EEG) to assess the maturation of cortical tracking of speech in noise and functional connectivity. Additionally, we will analyze changes in grey and white matter using diffusion MRI with the Neurite Orientation Dispersion and Density Imaging (NODDI) modeling technique, which will help us discern the specific mechanisms of neuroplasticity involved. To investigate metabolic mechanisms supporting functional and structural neuroplasticity, we estimate GABA and glutamate neurotransmitter concentration using magnetic resonance spectroscopy in cortical regions involved in speech-in-noise processing. Speech perception in noise will be evaluated behaviorally by means of the Coordinate Response Measure paradigm (CRM); together with cognitive and linguistic factors known to contribute to its development. We will recruit a group of pre- and a group of early/mid pubertal adolescents. Across groups, participants will be age- and sex-matched, in order to isolate the effects of puberty, allowing us to compare hormonal influences and pubertal stages. Our objective is to delineate the trajectory of typically developing adolescent brains in regions related to cognitive and perceptual processes involved in speech perception in noise.

We will present adult pilot data utilizing these metrics to illustrate our protocol and support our modeling framework. This research seeks to clarify the interplay between puberty, neuroplasticity, and speech processing abilities in adolescents, paving the way for future research on the implications of these changes for learning and social interaction during this critical developmental period.

Funding: ERC grant n°101076968.

P62 The effect of hearing loss and hearing-aid use on cue-weighting strategies for phoneme categorisation in older adult listeners

Mengfan Wu[1], Yue Zheng[1], William Whitmer[2], Katrin Krumbholz[1]
1. Hearing Sciences, Mental Health and Clinical Neurosciences, School of Medicine, University of Nottingham, UK | 2. Hearing Sciences - Scottish Section, Mental Health and Clinical Neurosciences, School of Medicine, University of Nottingham, Glasgow, UK

Background: Hearing loss (HL) changes the way people hear sounds, such as speech, and this may change how listeners use different aspect of the acoustic information conveyed by these sounds. Previous studies have suggested that people with HL use different cue-weighting strategies for categorising phonemes. Some of these strategies may be altered by the use and/ experience with hearing aids (HAs).

Using the specific consonant-vowel continuum /da/ to /ga/, the current study aims to investigate how cue-weighting strategies are influenced by HA use, and how the strategies change in the presence of noise. In addition, we will also explore how individual cue-weighting strategies are related to individual differences in self-reported hearing outcomes.

Methods: Three groups of native English-speaking adults — older adults with bilateral HL that do or do not use HAs, as well as normal-hearing (NH) adults — will complete a phoneme categorisation task of a two-dimensional consonant-vowel continuum between /da/ and /ga/, varying either the consonant burst or the vowel formant transition, or both. Stimuli will be created from natural speech recordings by systematically varying the energy and duration of the consonant burst, and/or the frequency trajectories of the second and the third vowel formants. The task will be performed either in quiet or in speech-shaped noise. Participants with HL will be tested both with and without amplification prescribed according to the NAL-NL2 formula. Unaided self-reported hearing outcomes will be assessed by the short form of the Speech, Spatial and Qualities of Hearing scale (SSQ12).

Results: Phoneme categorisation judgements will be analysed in terms of the relative decisional weights of burst and formant cues, respectively. We expect that, without amplification, both HL groups will show a greater weighting of burst cues than the NH group, as HL has been shown to reduce sensitivity to formant cues. Moreover, as the burst should be more easily masked by noise, weighting of formant cues is expected to increase in noise for all listeners. Finally, amplification is expected to increase audibility, and thus weighting, of formant cues in the HL groups. This effect may be greater in the HA users than non-users, as reliance on burst cues should be less entrenched in HA users.

Acknowledgements: The study is funded by the Launchpad grant from the Medical Research Foundation. We would also like to thank Sven Mattys and the University of York for their support in stimuli recording.

P63 Bottom-up and top-down short-term audio-tactile training: Can training improve speech intelligibility and cortical speech-envelope tracking?

Brandon L. O'Hanlon[1], Lars Hausfeld[2], Lars Riecke[2], Barrie Usherwood[1], Christopher J. Plack[3,1], Helen E. Nuttall[1]

1. Lancaster University, UK | 2. Maastricht University, Netherlands | 3. The University of Manchester, UK

Neural speech-envelope tracking in the auditory cortex — wherein neural activations synchronise with fluctuations in the speech envelope — is enhanced with audio-visual and audio-tactile speech compared to audio-only speech. Whilst visual cues can also improve speech intelligibility, evidence has shown no benefit behaviourally with tactile cues. Understanding the potential benefits of tactile stimulation is crucial for developing new methods of aiding speech-in-noise perception in the absence of relevant visual cues. Audio-tactile training may improve intelligibility when tactile information is available, though the extent to which different training paradigms may be effective is unclear. Two audio-tactile training experiments were run, investigating both bottom-up and top-down driven training paradigms. In the first, 64 young participants (ages: 18-29; 21 males, 42 females, one non-binary) were tested over five EEG sessions. They received bottom-up tactile training through a speech-in-noise discrimination task with tactile stimulation that was either congruent with sentences heard (trained group) or incongruent (pseudo-trained group). At baseline, there was a significant increase in speech-envelope tracking accuracy with audio-tactile stimuli relative to audio-only, but no increase in speech intelligibility. After training, there was no benefit to congruent training for audio-tactile tracking or intelligibility. This suggests that speech intelligibility and tracking are not enhanced by bottom-up short-term audio-tactile training.

In a second experiment, we hypothesised that top-down audio-tactile training may provide the missing benefit to speech intelligibility and further enhancements to tracking accuracy with audio-tactile speech. Here, we present a preliminary pilot analysis of 30 young participants (ages: 19-30) who participated in a single-session electroencephalography experiment. Participants received either top-down audio-tactile training (discriminating between congruent and incongruent audio-tactile speech-in-noise) or bottom-up audio-only training (passively listening to audio-only speech-in-noise), with feedback after trials. After training, there was no significant increase in tactile benefit to speech intelligibility or tracking accuracy in either training group. So far, these preliminary findings suggest that a single-session top-down training paradigm is not sufficient for enhancing audio-tactile integration. Together, these experiments provide a lack of evidence for the benefit of short-term audio-tactile training to the perception of speech-in-noise, with discussion for how audio-tactile training may be improved in future research.

P64 How frequency-based processing may help clinicians improve sound quality perception in cochlear implant users

Louis Villejoubert
Sound Lab, Cambridge Hearing Group, Department of Clinical Neuroscience, University of Cambridge, UK

Lorenzo Picinali
Dyson School of Design Engineering, Imperial College London, London, UK

Kathleen Faulkner
Oticon, Ballerup, Denmark

Debi Vickers
Sound Lab, Cambridge Hearing Group, Department of Clinical Neurosciences, University of Cambridge, UK

The importance of sound quality in cochlear implants (CIs) has gained significant attention in recent years. While many CI users achieve near-normal speech recognition in quiet conditions, they frequently report poor sound quality. Although the impact of CI technology on sound quality degradation is well-documented, less is known about the contribution of specific "frequency-to-place factors" such as the frequency-to-place mismatch (FTPM) and electrode-neuron interface (ENI). Our study aims to develop innovative testing methods to examine the relationship between sound quality and these individual frequency-to-place factors.

The first method involves frequency manipulation of different vowels: /ɜ:/, /a:/, /u:/, and /i:/. Two types of manipulations were performed: in the first, the modification was restricted to the first formant, while in the second, it was applied to the entire vowel. Participants evaluated these shift configurations in terms of "overall preference" and recognition through paired comparisons using a preference rating scale. A Bradley-Terry-Luce model was used to determine the final ranking of the shifting configurations. These methods aim to reveal levels of adaptation to FTPM and ENI quality.

To further separate the effects of these factors, the second method assesses the quality of sound transmission across individual electrodes. This test uses a sound quality experiment based on chord comparisons. Participants were presented with pairs of chords and asked to rate the sound quality, comparing each chord with its inversion. Chord inversions varied by one note within the same harmonic framework, maintaining similar overall quality. These paired comparisons, conducted with various chord triads and their inversions, allow us to evaluate how individual electrodes may degrade sound quality in complex stimuli.

We tested 10 post-lingually deafened adult CI users with over 12 months of listening experience. Preliminary findings indicate good test-retest reliability for all three methods. Results from the vowel experiments show that some participants preferred a shifted configuration to enhance sound quality, though no clear link emerged between vowel and chord preference results. The chord comparisons revealed substantial differences in sound perception across electrodes, suggesting significant channel-to-channel variability, which appears to correlate with CI mapping settings.

These preliminary findings suggest that both testing approaches might provide valuable insights into frequency-to-place mapping and sound quality judgments in CI users. Further investigations are ongoing to examine the relationship with specific frequency-to-place factors. These methods show promise as clinical tools to guide CI setup adjustments, enhancing sound quality perception for CI users.

P65 The role of accent familiarity in speech recognition in noise by younger and older listeners

Kathleen McCarthy, Naomi Whittaker, Thomas Packer-Stucki
Queen Mary University of London, UK

Outi Tuomainen
University of Potsdam, Germany

Older adults display poorer speech processing abilities than younger adults in adverse listening conditions, such as background noise. These abilities can be further compromised by speaker characteristics, such as gender, age or accent, possibly driven by the added cognitive demands on the listener combined with age-related sensory and cognitive decline. In the current study we ask if accent familiarity boosts speech recognition in noise for younger and older listeners.

Participants were fifteen younger adults (YA, 18-30 years old) and nineteen older adults (OA, 65-80 years old). All were monolingual English speakers and had resided in East London, UK, since birth. All OAs passed the Montreal Cognitive Assessment (>26 out of 30). Sixteen of the OA had “normal hearing” (better-ear PTA <25 dB HL calculated across 0.25-4 kHz), and three had age-related mild hearing loss (25-35 dB HL in the better ear and typical sloping pattern at the higher frequencies). All YAs had normal hearing. Speech recognition was assessed using a speech-in-noise task, where participants listened to BKB sentences and were scored on the number of key words correctly repeated. We had 4 accent conditions: two regional and two second-language-accented (L2), with half of the accents being unfamiliar to the participants. Participants also completed the Speech Spatial Qualities of Hearing questionnaire and a language and accent experience background questionnaire.

As expected, YAs were more accurate than the OA in noise. We found a significant difference in accuracy between the regional and L2 accents in quiet — both the YA and OA showed greater accuracy for the regional accents than the L2 accents. No effect of familiarity was found in the quiet condition, likely because both regional accents were at ceiling. In noise, both groups had greatest accuracy for their familiar London accent, followed by the regional Dudley accent. This difference was significant for the YAs but not the OAs. Both groups had the lowest accuracy for the L2 accents and showed no advantage for their more familiar L2 accent. Taken together, these results suggest that regional accents might be easier to process than L2 accents in adverse listening conditions, and familiarity may boost speech recognition for YA but have limited impact for OAs. We are currently exploring the influence of listener-specific characteristics on the YAs’ and OAs’ speech recognition abilities e.g., accent experience, acoustic distance between the listener’s accent and stimuli.

P66 Relating amplitude modulation rate discrimination to speech-in-noise perception in cochlear implant users

Nick Haywood[1], Marina Salorio Corbetto[1], Brian C. J. Moore[1], Patrick Boyle[2], Josef Schlittenlacher[3], Debi Vickers[1]

1. University of Cambridge, UK | 2. Advanced Bionics, Hannover, Germany | 3. University College London, UK

For cochlear implant (CI) listeners speech information transmission is reliant upon the ability to process the amplitude-modulated (AM) envelope of speech sounds independently in different channels. This can be hindered for many reasons, not least due to, spread of electrical current or neural survival.

We employed a psychoacoustic task to explore AM processing. We recruited 12 typical hearing (TH) adults and 12 adult CI listeners. Modulated pure tones of two different rates (13 versus 40 Hz) were discriminated in a three-interval two-alternative forced choice task, where the modulation depth was adjusted adaptively to derive an AM discrimination threshold. Testing was conducted with and without speech envelope interferers on neighbouring channels. In the adaptive procedure the modulation depth of the interferer was adjusted in the opposite direction to the target. The modulation depth at threshold of the target AM stimulus is reported. Stimuli were delivered through headphones (HD600s). All front-end noise reduction features were de-activated on the sound processors of the CI listeners. Acoustic stimuli were presented at 5 centre frequencies, aimed to stimulate the same 5 electrodes in the CI users and equivalent frequencies for TH group, exploring AM processing across the frequency range up to 4 kHz, the important frequency range for speech.

Results were compared to speech-in-noise perception results using a children's coordinate response measure (CCRM) test in which a female speaker uttered a phrase, 'show the dog where the COLOUR NUMBER is' where the listener had to identify the colour and number which changed on a trial-to-trial basis. Six colours and nine monosyllabic numbers were included in the stimulus set. Speech was presented in 6-talker (3 male and 3 female speakers mixed) babble noise which was adaptively varied in level to obtain the speech reception threshold (SRT).

For all participants, AM discrimination was poorer in the presence of interferers than for absent interferers. For the CI listeners performance worsened with increasing centre frequency, an effect which may have been driven by a subset of individuals with particular high thresholds for higher centre frequencies. We speculate that the measure may differentiate between problems of neural survival and channel interaction which could cause these effects. Average (across all centre frequencies) AM thresholds in the presence of interferers correlates with CCRM SRTs for both groups.

P67 Investigating the effect of contact vs. non-contact sports on the brain's response to sound

Jessica R. Andrew[1,2,3], Mark Crawley[3], Andrew Thomas[3], Christopher J. Plack[1,4], Timothy J. Barry[2], Christopher J. Gaffney[2], Helen E. Nuttall[1]

1. Department of Psychology, Lancaster University, UK | 2. Lancaster Medical School, Lancaster University, UK | 3. Brainbox Ltd., Cardiff, UK | 4. Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, UK

Data indicates that sports-related concussions may cause damage to the auditory system or lead to difficulties perceiving speech in background noise. Understanding the effects of repetitive sub-concussive head impacts in contact sports remains limited, with no sensitive objective measure identified, or understanding of how sub-concussive impacts may affect auditory function. This study aims to investigate the impact of contact sports on the brain's auditory neural responses in young adult athletes. This research will use 48 participants, divided into two groups: athletes engaged in contact sports and those in non-contact sports. Each participant will undergo two electroencephalography (EEG) recordings: 1) subcortical and 2) cortical, under two conditions: in quiet and in noise, using a 170ms speech syllable /da/ for both conditions, with a 6-multi talker background babble only for the speech in noise condition. Preliminary results from a repeated measures mixed two-way ANOVA on a sample of 26 participants (13 per group) revealed a significant main effect of group ($F(1,23)=18.41$, $p<.001$) in the cortical N100 responses, with contact athletes showing a reduced N100 amplitudes ($M = -2.06\mu V$, $SD = 1.08$) compared to non-contact athletes ($M = -3.19\mu V$, $SD = 1.15$). There was also a significant main effect of condition, ($F(1,23)=19.43$, $p<.001$), where both groups showed reduced N100 amplitudes in noise ($M = -2.07\mu V$, $SD = 0.69$) compared to quiet ($M = -3.23\mu V$, $SD = 1.42$). There was also a significant interaction between group and condition ($F(1,23)=6.62$, $p=.017$). Pairwise comparisons revealed that the group difference in N100 amplitudes between contact and non-contact athletes was only significant in the quiet condition, but not in the noise condition. Subcortical responses (F0 amplitudes) showed no significant group or condition differences at this stage. These initial findings suggest that cortical auditory processing may be more vulnerable to repetitive sub-concussive head impacts than subcortical responses, especially in the presence of speech in noise. This research aims to offers potential insights due to understanding the role in auditory function following head impacts and potentially working towards an objective marker of auditory processing changes due to contact sports, with implications for understanding player safety and the effects of sub-concussive impacts on auditory function.

P68 Using a lateralisation task to predict asymmetries in spatial release from masking in children with bilateral cochlear implants

Marina Salorio-Corbetto[1], Bhavisha Parmar[1,2], Jennifer Bizley[2], Lorenzo Picinali[3], Debi Vickers[4]

1. Department of Clinical Neurosciences, University of Cambridge, UK | 2. Ear Institute, University College London, UK | 3. Audio Experience Design Group, Imperial College London, UK | 4. University of Cambridge, Department of Clinical Neurosciences, Herchel Smith Building, Cambridge

Background: Binaural hearing is crucial to everyday listening, and yet, the functional performance of children with bilateral cochlear implants is usually assessed for each ear separately. One reason for this may be space and equipment limitations affecting clinics. The BEARS project has developed clinical tools that require only a pair of headphones and a tablet to deliver spatial speech-in-noise testing. This toolkit includes, among others, a virtual-audio version of the Spatial Speech-in-Noise Test —the SSiN-VA — and a 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. These tests are being used in 11 UK-based hospitals to assess children involved in BEARS.

Rationale: SSiN-VA provides word-identification scores for each hemifield when the noise source is either on the same hemifield as the speech or on the opposite hemifield. An index of word identification is calculated for each hemifield, considering performance for each noise location. Asymmetries in these indexes across hemifields are thought to capture asymmetries in spatial release from masking. Differences in sound level across ears in the Balance app are thought to indicate lateralisation bias. We explored whether the average level difference across ears in the lateralisation task predicted the degree of asymmetry in the SSiN-VA task for one hundred and seven children who completed their baseline session.

Methods: 107 children with bilateral cochlear implants, aged 8-16 years, completed the Balance app task followed by the SSiN-VA task. Tests were presented using a table and headphones using spatialised stimuli. An index for each hemispace for the SSiN-VA and the average balance point across stimuli were calculated.

Results: A small significant correlation was found between the word identification index ratio, which characterises SSiN-VA asymmetry, and the Balance app average outcomes ($S = 156550$, $p = 0.016$, $\rho = 0.233$).

Conclusions: Although there seems to be a relationship between the test outcomes, further work is needed to optimise Balance to predict SSiN-VA at an individual level, as it would be required for clinical applications.

P69 Hearing loss compensation and simulation using a differentiable loudness model

Szymon Drgas
Poznan University of Technology, Poland

Lars Bramsløw
Erikholm Research Center, Denmark

Recent applications of auditory models with hearing loss are 'closed-loop' approaches. Deep learning is used to optimize the parameters of a neural network by comparing a normal hearing and a hearing-impaired branch of the auditory model.

Audmod (Bramsløw, 2024, doi:10.5281/zenodo.10473561) is a loudness perception model with hearing loss. The auditory filter bandwidth in the Audmod depends on both signal level and hearing loss, both leading to increased upward spread of masking. Therefore, the model takes into account complex dependencies that occur for multitone signals in loudness perception.

The aim of the work was to test the early application of the differentiable version of Audmod to hearing loss compensation (HLC) and simulation (HLS). In the HLC the input signal provided at the input of the Audmod with hearing loss is multiplied in the STFT domain by a gain matrix. The gain matrix is optimized, such that the specific loudness patterns computed by the model are as similar as possible to the loudness patterns computed by Audmod for normal hearing listeners for non-processed input. In the case of HLS, the signal at the input of normal hearing Audmod is multiplied by a gain matrix optimized to achieve similar specific loudness patterns as for hearing impaired listeners.

For HLC, the resulting signals were evaluated and compared to National Acoustics Laboratory revised, profound (NAL-RP) prescription using spectral measurements and the hearing-aid speech perception index (HASPI). For HLS, spectral measurements and informal listening have been used for preliminary evaluation.

P70 Predicting noisy speech intelligibility beyond acoustics: Including listeners' phonetic and language abilities.

Mark Huckvale, **Gaston Hilkhuysen**
University College London, UK

Scholars typically distinguish acoustic, auditory, phonetic and language stages in human speech perception. Intelligibility metrics however traditionally only account for changes in the acoustics of the noisy speech signal, although some now include information about the auditory sensitivity of hearing-impaired. Such metrics assume that cochlear impairment induces acoustic distortions, reducing the hearing abilities of people with a hearing loss. Unfortunately, such metrics predict the intelligibility of noisy speech for given listeners poorly. The current

study examines whether adding factors that account for the listeners' phonetic and language abilities can improve these predictions. The study is based on 13000 intelligibility judgments by 31 hearing-impaired listeners generated in the second Clarity Prediction Challenge.

To assess phonetic ability we start by computing the phone probabilities from the original and distorted speech audio. The correlation of these probabilities defines the degree of phonetic distortion, which is subsequently regressed on word-correct scores for each listener. The positions of these regression functions define a listener's phonetic ability.

We also computed word probabilities for the intelligibility sentences using trigram word probabilities calculated from the British National Corpus. The position of the regression functions of these word probabilities while predicting word-correct scores defines a listener's language ability.

A principal component analysis of auditory, phonetic and language abilities showed almost orthogonal auditory and language abilities, while phonetic abilities correlated intermediately with the other two abilities. A cross-validated regression analysis found that intelligibility predictions based on the short-time objective intelligibility measure (STOI) were improved when extended with the listeners' phonetic and language abilities. A regression model based on STOI, acoustic and phonetic abilities generalized best to unseen test data, giving a 9.8% relative reduction in prediction error

Outcomes indicate that phonetic and language abilities are promising predictors while estimating the speech intelligibility for specific listeners with hearing loss. It opens ways to account for intelligibility variation across speech corpora as well as differences between native and non-native listeners. Future research could focus on predictor interactions.

P71 The effect of internalized stereotypes on speech perception in older adults: Evidence from SRT tasks in quiet and noise

Vered Shakuf
Achva Academic College, Israel

One of the main concerns frequently raised by older adults is the difficulty they experience in hearing or participating in conversations. Indeed, a gradual decline in hearing ability occurs as a common and natural part of the aging process, associated with both physiological and cognitive factors. However, the literature on aging suggests that psychosocial factors may also contribute to age-related declines in performance. Specifically, internalized negative self-stereotypes about hearing loss can impact auditory performance, especially in tasks requiring speech comprehension or those presented in challenging conditions. This study examines the impact of these self-stereotypes on auditory performance in older adults.

Forty-one participants aged 60 and above were randomly assigned to one of two groups: a negative stereotype group (n=21) and a control group (n=20). All participants first completed a baseline Speech Recognition Test (SRT) in quiet to assess their speech recognition threshold, followed by a second SRT test in noisy conditions. For the noisy condition, the SRT test was presented with a broad-spectrum speech noise masker at 30 dB above each individual quiet

SRT threshold. In the intervention phase, the negative stereotype group read a fictional leaflet presenting negative stereotypes about age-related hearing loss, aiming to trigger stereotype threat. The control group read a neutral leaflet unrelated to age-based stereotypes. After reading, all participants completed a self-assessment survey to gauge their personal attitudes toward hearing loss. The SRT tests were then repeated in both quiet and noisy conditions to identify any changes in performance post-exposure.

Results revealed a significant decline in auditory performance following exposure to negative stereotypes in the negative stereotype group, while no significant change was observed in the control group. A moderate but significant correlation emerged between participants' negative self-perceptions regarding hearing and their decreased performance after stereotype exposure. These findings suggest that negative stereotypes about age-related hearing loss can adversely affect auditory performance, as measured by both quiet and noisy speech recognition tasks. The results of this study highlight the potential influence of psychosocial factors on auditory performance, which may inform future research and intervention strategies aimed at alleviating the impact of stereotype-related threats. This underscores the importance of addressing psychosocial influences in clinical practice, as stereotypes may distort test outcomes beyond physiological hearing capabilities.

P72 An automated digits-in-noise hearing test using automatic speech recognition and text-to-speech: A proof-of-concept study

Mohsen Fatehifar[1], Kevin Munro[1], Michael Stone[1], David Wong[2], Tim Cootes[1], Josef Schlittenlacher[3]

1. University of Manchester, UK | 2. Leeds Institute of Health Sciences, University of Leeds, UK | 3. Department of Speech, Hearing and Phonetic Sciences, University College London, UK

Aims: This study builds a framework for AI-powered speech-in-noise (SIN) tests. It specifically considers a digits-in-noise (DIN) test with Automatic Speech Recognition (ASR) and Text-To-Speech (TTS) and compares its performance against a benchmark test.

Methods: Three methods of DIN tests were compared: (1) An established DIN test, used as the Benchmark test; (2) Our implementation of the DIN test, called the Keyboard-based test, before adding TTS and ASR; (3) An AI-powered test, which was our implementation of DIN but extended to include TTS for synthesising the stimuli and ASR to transcribe the verbally repeated responses. Apart from stimulus generation and response capturing, its underlying code was identical to the keyboard-based test.

Ten hearing impaired and 21 normal hearing participants were recruited. The task in each test was to repeat three digits presented in babble noise to both ears in a sound-treated room. A 2-up/1-down approach was used to obtain the signal-to-noise ratio (SNR) corresponding to 71%-correct. A retest was carried out for both the Benchmark and AI-powered test. All tests were performed in a single session lasting no more than 90 minutes. The results (mean difference and 95%-limit of agreement — LoA) were compared using Bland-Altman analysis.

Results: The mean and 95%-LoA for test-retest reliability and comparisons between different methods were computed: the Benchmark test had a test-retest reliability with LoA of -0.4 ± 3.8 dB, and the AI-powered had a similar test-retest reliability of -0.9 ± 3.8 dB, indicating that the AI-powered test was as reliable as the Benchmark test.

Comparing the Benchmark test with the Keyboard-based method resulted in LoA of -0.7 ± 5.9 dB, and comparing the AI-powered test with the Benchmark test resulted in LoA of 0.0 ± 4.6 dB. The higher variability between the Keyboard-based test and the Benchmark test (5.9 dB) compared to the test-retest variability of the Benchmark test (3.8 dB) was likely due to differences in software implementations. The LoA for the Benchmark and AI-powered test (4.6 dB) was less than that of the Benchmark and Keyboard-based test (5.9 dB), indicating that the inclusion of TTS and ASR did not introduce additional variability and even improved the LoA. Which may be due to the elimination of typing errors and reduced distraction.

Conclusion: The developed framework works well: it adds little error compared to the test-retest reliability of the Benchmark test. The additional variance may arise at least in part from ASR errors. The results demonstrate proof-of-concept i.e., it may be possible to use ASR and TTS in a SIN test.

P73 Take-home cochlear implant processor with spectral and temporal enhanced processing (STEP): First results and real-world experiences

Damir Kovačić
Dept. of Physics, University of Split-Faculty of Science, Split, Croatia

Chris James
Cochlear France SAS, Toulouse, France

Cochlear implant users have varying capacities to use temporal and spectral cues. Sound coders used in CIs generally have fixed parameters such that the extent or availability of these cues cannot be individually tuned. Even when using experimental coders in the lab, live processing often has to be interrupted when parameters are changed, or only a limited set of parameters can be trialed. The possibility to hand over control of certain sound coding parameters to the user would fulfill some of the requirements of personalized medicine.

To address these limitations, we developed a take-home CI processor based on the University of Texas CCi-MOBILE platform. Real-time Spectral and Temporal Enhanced Processing (STEP, Kovačić & James, 2022, doi:10.1007/s10162-022-00854-2) was implemented in an Android smartphone application. In this first version, the CI user is able to switch between ACE, the standard advanced combination encoder, and STEP. In addition, the modulation bandwidth can be varied continuously, for example, to control temporal voice-pitch cues. We will report representative electrode-o-grams showing the effects of varying sound coding parameters. The first author will also share his real-world experiences listening through the device. Finally, we will demonstrate how the system allows CI users to control spectral parameters precisely by assigning specific frequencies to electrodes.

P74 Can we know audiovisual speech perception ability without measuring it?

Christian Sumner[1], Paula Stacey[1], Samuel Smith[2], Jens Roeser[1], Thom Baguley[1]
1. Nottingham Trent University, UK | 2. Eaton Peabody Lab, Harvard University, USA

Speech comprehension is often boosted by watching a talker's face. This is of critical importance for anyone who is hearing impaired, and the benefits that a hearing aid bring are likely to depend on whether or not the talker's face is visible. However, the influence of visual speech is not normally considered when treating hearing loss. This is partly because of large individual differences, which make it difficult to quantify the benefit without direct measurements. If we could develop efficient ways of predicting the boost provided by seeing a talker's face, this might make it practical to factor in when treating hearing loss. Here we ask the question: is it possible to predict audiovisual speech perception in an individual from measurements of the separate modalities?

We fitted a novel model of audiovisual integration to a normative dataset of ~200 individuals from a broad demographic, collected on-line, performing an open-set speech in noise task across three modalities: visual-only, auditory-only and audiovisual. Our model was able to fit the individual variation in audiovisual speech perception under the assumption that it depended only on auditory-only and visual-only speech perception (R~0.95), whereas unisensory performance varied widely across individuals. This suggests that the "audiovisual integration function" for speech is relatively consistent across a diverse population, with individual differences being attributable to differences in unisensory perception.

To test whether this model might have the predictive power to be of clinical use, we attempted to use this model to predict new data which was not used for fitting the model. We applied the model to a new set of data, collected in-person, where participants were either young (<35) or older (>50) and either had normal hearing or had a diagnosed hearing loss (most used hearing aids). Their visual-only and audio-only speech perception scores were used to predict audiovisual speech perception, using the model fit to the on-line data. The model was able to predict audiovisual performance quite data well. Exactly how well is left for the meeting, to provide intrigue. Overall, these data support the idea that we can predict audiovisual speech perception from perception of the individual senses.

P75 Interpreting the binaural speech intelligibility model as an internal beamformer

Johannes W. de Vries[1], Steven van de Par[2,3], Richard Heusdens[4,1], Richard C. Hendriks[1]
1. Delft University of Technology, Netherlands | 2. Carl von Ossietzky University of Oldenburg, Germany | 3. Cluster of Excellence Hearing4all, Germany | 4. Netherlands Defence Academy

Understanding speech is an integral part of daily life, and different solutions exist to aid people who are hard of hearing with their intelligibility issues. Because intelligibility measurements for these solutions can be expensive and time-consuming, models have been developed to predict intelligibility based on input sound signals. One such model is the binaural speech intelligibility model (BSIM) (Beutelmann & Brand, 2006, JASA 120:331), which, at its core, is a

concatenation of the equalisation–cancellation (EC) model and the monaural speech intelligibility index (SII). Although this model has been verified to predict binaural intelligibility well, it is mathematically somewhat convoluted. Consequently, the internal mathematical optimisation problem is solved numerically with an inefficient grid search. A later revision (Beutelmann et al., 2010, JASA 127:2479) improves on the computational complexity, but only approximates the original optimisation. In this work, the mathematical framework of the BSIM is reformulated, making it both more efficient and more straightforward to use in new contexts.

The main challenge of the BSIM is in finding optimal EC parameters. As the speech-related input signals are processed independently per frequency band, the model can first be transformed to the short-time frequency transform (STFT) domain. The attenuation and delay operations of EC on the input signals correspond to a complex scaling in this domain. Therefore, the entire EC phase can be interpreted as an 'internal beamformer' applied to the input signals, whose elements are determined by the attenuation and delay parameters. The output signal power of the EC phase can then be written in terms of the internal beamformer and the input cross power matrix, which is periodogram estimated from the input signal realisations. The effects of the artificial processing errors in the BSIM can be incorporated by element-wise multiplying the cross-power matrix with a matrix that depends on the error variances. In this description, the model's output SNR takes the form of a generalised Rayleigh quotient of the internal beamformer, such that the optimal EC parameters can be found efficiently through a generalised eigendecomposition.

Simulations of the original and novel implementations of the BSIM show that the internal beamformer framework results in similar intelligibility predictions but with a 10 to 20 times faster simulation time. Aside from the more efficient implementation, the mathematically more compact formulation allows the model to be used more easily in other applications, such as binaural beamformer design or machine learning approaches.

P76 The predictive ability of various aspects of memory to speech in noise perception

Antje Heinrich, Youn Ji Baek, Emanuele Perugia
University of Manchester, UK

In addition to the ability to hear, accurate speech-in-noise perception also requires contributions from cognitive abilities. Cognitive abilities examined most often in the context of speech-in-noise (SiN) perception tend to be related to memory, with working memory (WM) the most commonly investigated aspect of memory in this context. According to Baddeley and Hitch (1974, doi:10.1016/S0079-7421(08)60452-1) WM combines two critical components: a storage component and a manipulation component. It is not well understood whether it is the storage or the manipulation component that has the highest predictive value, or whether both components have to combine for maximum predictive value. Also, it is unclear whether the predictive ability of different WM components differs either for different groups of listeners or for different speech-in-noise tasks.

One WM task commonly used in the context of speech perception is the reading span task (RST). The RST exists in many different versions, possibly differing on the emphasis they put on the manipulation and storage component. Here, we are investigating to what extent the

setup of different RSTs favours either working-memory versus long-term memory, and which of these two components is more predictive of speech-in-noise perception in young and old English native speakers. Understanding the exact cognitive processes involved in various RSTs and their predictive value for speech-in-noise in various groups of listeners is the first step to use the RST in the clinic meaningfully as part of the hearing rehabilitation process.

P77 Creating a naturalistic classroom for lab-based auditory research

Brandon L. O'Hanlon[1], Liam Howard[1], Peter D. Tovee[1], Barrie Usherwood[1], Katharina Kaduk[1], Kiri Mealings[2], Steph Ainsworth[3], Hannah J. Stewart[1]
1. Lancaster University, UK | 2. Macquarie University, Australia | 3. Manchester Metropolitan University, UK

Paediatric auditory research typically takes place in controlled, lab-based research environments. Paradigm precision is high in these environments as confounding factors are easily controlled. However, this comes at a huge cost to ecological validity. For example, BKB-SiN sentences are often presented in a soundproof auditory booth to reduce miscellaneous background noise. Yet, in their everyday life, children neither learn nor play in a soundproof booth.

An alternative option is collecting data directly from schools to provide ecologically valid data. These school-based research environments, however, lack paradigm precision. For example, classroom environments are loud, with many distracting sensory elements (noisy peers, visual clutter). Peers may interact in a variety of different ways, and the tasks given to students by their teacher may be variable from day-to-day. These confounds are hard to control, despite providing greater application of research findings to a real-world setting.

One way in which a middle ground between laboratory and school-based research can be reached is through the creation of a naturalistic, lab-based environment. The L3 Framework provides a structured approach to designing such an environment. Using the L3 Framework, a naturalistic classroom setting has been constructed at Lancaster University that replicates the typical classroom that children may find themselves in at school. This classroom considers the utilisation of synchronised eye- and motion-tracking, audio monitoring, and fNIRS, all without compromise to the naturalistic setting. Here, we discuss this classroom design and its potential for furthering our understanding of the impact of children's hearing ability and technologies on learning.

SPEECH
IN NOISE

https://2025.speech-in-noise.eu

Lancaster University

university of groningen

umcg

cnrs

oticon life-changing technology

GN Advanced Science

WSAudiology

sonova HEAR THE WORLD

Cochlear®