SPEECH IN NOISE WORKSHOP

Abstracts

Paris, France | 12-13 January 2026

The 17th Speech in Noise Workshop was chaired by ***Laurianne Cabrera****, BabyLab, Integrative Neuroscience and Cognition Center, CNRS UMR 8002, Université Paris Cité, Paris, France, and* ***Léo Varnet****, Laboratoire des Systèmes Perceptifs, École Normale Supérieure, Université Paris Sciences et Lettre, Paris, France, with the help of the organisation committee:*

- *Etienne Gaudrain*
- *Antje Heinrich*
- *Chris James*
- *Damir Kovačić*
- *Helen Nuttall*
- *Hannah Stewart*
- *Outi Tuomainen*
- *William Whitmer*

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

https://2026.speech-in-noise.eu/
Contact information: info@speech-in-noise.eu

Date of publication: 8 January 2026.
DOI this version: 10.5281/zenodo.18187290.
DOI all versions: 10.5281/zenodo.18187289

Cover illustration based on the photo "Trocadero in the snow" by Carsten Sprotte (CC BY 2.0).

The Speech in Noise Workshop is generously supported by:

Programme

Monday 12 January 2026

09:00-10:00 *Registration & Coffee*

10:00-10:10 *Opening remarks*

10:10-10:35 How do visual mechanisms help auditory-only speech and speaker recognition?
Katharina von Kriegstein
Dresden University of Technology, Germany

10:35-11:00 How voice familiarity affects speech-in-speech perception
Emma Holmes
University College London, UK

11:00-11:25 Miscommunications in triadic conversations: Effects of hearing loss, hearing aids, and background noise
Eline Borch Petersen, Martha Reenberg Munck, Anja Kofoed Pedersen
ORCA Labs, Scientific Institute of WS Audiology, Lynge, Denmark

11:25-11:50 Coffee, Photo

11:50-12:15 The microscopic impact of noise on phoneme perception and some implications for the nature of phonetic cues
Léo Varnet
Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France

12:15-12:40 Adaptation to noise: A minireview of mechanisms and individual factors
Miriam I. Marrufo-Perez, Enrique A. Lopez-Poveda
University of Salamanca, Spain

12:40-14:00 Lunch

14:00-15:00 **Keynote —** Examination of speech coding in the human auditory nerve using intracranial recordings
Xavier Dubernard
Institut Otoneurochirurgical de Champagne Ardenne, CHU de Reims, France | Institut des Neurosciences de Montpellier, France

15:00-17:30 Poster session 1 — with coffee

17:30-19:00 *Free time*

19:00-22:00 Conference dinner at La Mère Catherine, 6 Place du Tertre, 75018 Paris

Tuesday 13 January 2026

09:00-10:00 Coffee

10:00-10:25 **Voice cue sensitivity and speech perception in speech maskers in children with hearing aids**
Laura Rachman, Pınar Ertürk, Gizem Babaoğlu, Basak Özkişi Yazgan, Gonca Sennaroğlu, Etienne Gaudrain, Deniz Başkent
University of Groningen / University Medical Center Groningen, Netherlands

10:25-10:50 **Cochlear implantation in children with single sided deafness**
Astrid van Wieringen
Experimental ORL, Dept Neurosciences, KU Leuven, Belgium

10:50-11:10 **Modelling the contributions of auditory, speech, language, and cognitive processes to speech-in-noise perception in school-aged children: A structural equation approach**
Xuehan Zhou, Harvey Dillon, Dani Tomlin, Kelly Burgoyne, Helen Gurteen, Grace Nixon, Alisha Gudkar, Antje Heinrich
The University of Manchester, UK

11:10-11:35 Coffee

11:35-12:00 **What does cognitive listening mean? Is it time for a rethink?**
Sven Mattys
University of York, UK

12:00-12:25 **Colin Cherry Award 2025 —** Age-related changes in alpha activity during dual-task speech perception and balance
Jessica L. Pepper, Theodoros M. Bampouras, Helen E. Nuttall
Lancaster University, Lancaster, UK

12:25-14:00 Lunch

14:00-16:30 Poster session 2 — with coffee

16:30-17:00 *Business meeting: Colin Cherry Award 2026, next SPIN meeting and closing remarks*

17:30-... Evening drinks at La Roue Libre, 80 Boulevard Richard-Lenoir, 75011 Paris

Talks

Monday 12 January 2026, 10:10—10:35

How do visual mechanisms help auditory-only speech and speaker recognition?

Katharina von Kriegstein
Dresden University of Technology, Germany

Understanding what is said and recognising the identity of the speaker are two important tasks that the human brain is faced with in auditory communication. For a long time, neuroscientific models of auditory communication have focused mostly on auditory language and voice-sensitive cerebral cortex regions to explain speech and voice-identity recognition. However, we now know that the brain uses even more complex processing strategies for recognising auditory communication signals, such as the recruitment of dedicated visual face areas. In my talk, I will present a short overview on our neuroscientific findings how visual face areas help processing auditory communication signals. I will then present our current work in which we translate our neuroscience findings to computational models.

Monday 12 January 2026, 10:35—11:00

How voice familiarity affects speech-in-speech perception

Emma Holmes
University College London, UK

People often face the challenge of understanding speech when competing speech is present—which involves a variety of cognitive processes, such as attention and prior knowledge. We have consistently found that familiarity with a person's voice improves the ability to understand speech-in-speech, using both naturally familiar (e.g., friends and partners) and lab-trained voices. In this talk, I will describe experiments in which we aimed to gain insights into the processes underlying the familiar-voice intelligibility benefit, and how people learn about new voices. These findings have implications for theories of speech perception, and potential applications for populations who typically find speech perception particularly challenging (e.g., older adults and individuals with hearing loss).

Miscommunications in triadic conversations: Effects of hearing loss, hearing aids, and background noise

Eline Borch Petersen
ORCA Labs, Scientific Institute of WS Audiology, Lynge, Denmark

Martha Reenberg Munck, Anja Kofoed Pedersen
WS Audiology, Lynge, Denmark

Speech understanding is difficult to assess in real-world communication. However, instances where interlocutors ask for repetitions, clarifications, or express difficulty hearing, can provide insights into the ongoing speech understanding. From conversations recorded from 25 groups of triads consisting of two normal-hearing (NH) and one hearing-impaired (HI) interlocutor, we identified and analyzed all miscommunication events to evaluate the effects of hearing impairment, background noise, and hearing-aid signal processing.

A subset of miscommunications are so-called other-initiated repairs (OIRs) where one interlocutor signals a communication breakdown, which is then jointly resolved by conversation partners. Verbal OIRs can vary in specificity, from the unspecific open requests (e.g. ‘What?’) to the very specific restricted offers (e.g. ‘Did you say blue?’). We hypothesize that in difficult communication situations, talkers will use open OIRs to a larger extent, as poorer speech understanding prohibits the formulation of more specific OIRs.

The results showed that interlocutors have more miscommunications when they suffer from impaired hearing, however increasing the level of background noise results in more miscommunications for all interlocutors. Open request OIRs were generally used more often than restricted and even more so in difficult communication situations. For the HI interlocutor, hearing-aid signal processing reduces the number of miscommunications and proportion of open OIRs used. Interestingly, the directional hearing-aid setting, which was experienced by the HI interlocutor at the high noise level, affected the number and type of OIRs produced by the NH conversation partners. To offer a potential explanation between the sound experienced by the HI interlocutor and the miscommunications made by the NH conversation partners, an analysis of the relation between miscommunications and conversational speech level will be presented.

The microscopic impact of noise on phoneme perception and some implications for the nature of phonetic cues

Léo Varnet
Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France

The effect of background noise on speech perception is a multifaceted phenomenon. Psycholinguists often attribute the reduced intelligibility in noise to energetic masking: weak elements of the speech signal are not audible anymore and cannot contribute to recognition. Yet, noise also introduces random fluctuations that disrupt perception in less obvious ways. Using a reverse correlation approach, we examined how trial-by-trial variations in the noise enveloppe influence phoneme categorization. Our results show that noise not only masks speech but also interacts with phonetic cues in systematic ways. This finding offers new insights into the hierarchical organization of phonetic cues and interindividual variability in speech perception.

Adaptation to noise: A minireview of mechanisms and individual factors

Miriam I. Marrufo-Perez, Enrique A. Lopez-Poveda
University of Salamanca, Spain

People with normal hearing recognize more syllables or words in noise when these speech tokens are delayed from the noise onset. This improvement in recognition is referred to as "noise adaptation". Here, we review our efforts to shed light on the possible mechanisms underlying noise adaptation and some individual factors that may affect it. We will show that users of cochlear implants show as much as adaptation to noise as do listeners with normal hearing. Since cochlear implants stimulate the auditory nerve directly and independently from the middle-ear muscle reflex or the medial olivocochlear reflex, this suggests that those peripheral reflexes are not necessary for noise adaptation to occur. We will also show that people adapt equally to highly fluctuating and stationary noises. This suggests that mechanisms other than neural dynamic range adaptation to the most common noise level are probably involved in noise adaptation. Lastly, we will show that impaired adaptation contributes to up to 10% of the speech-in-noise reception threshold loss shown by people with hearing loss. However, it does not contribute to speech-in-noise intelligibility difficulties related to age or to 'hidden' hearing loss.

Funding: Work supported by the Spanish MICIU, ref. PID2024-158683NB-I00.

Keynote lecture

Examination of speech coding in the human auditory nerve using intracranial recordings

Xavier Dubernard
Institut Otoneurochirurgical de Champagne Ardenne, CHU de Reims, France | Institut des Neurosciences de Montpellier, France

Early intracranial recordings of the auditory nerve, conducted by Møller et al. in the 1980s, explored how the nerve encodes simple sounds, such as clicks or tone bursts. We aimed to extend this work by investigating human auditory nerve responses to clicks, tone bursts, and phonemes, sounds crucial for speech perception. Our study focused on the relative importance of temporal fine structure and temporal envelope. Recordings of auditory nerve electrical activity were performed at the University Hospital of Reims on patients undergoing microvascular decompression surgery (NCT03552224) for trigeminal neuralgia or hemifacial spasm. This approach provided access to the human auditory nerve. Both normal-hearing and high-frequency hearing loss patients participated. Stimuli were delivered in a closed field (Etymotic ER1) and the auditory nerve activity was measured using a ball electrode (Ø1.6 mm, Inomed) connected to a Grass P511 amplifier. The generation and acquisition of the signals were entirely processed by a NI-PXI 4461 device controlled by a LabVIEW interface (National Instruments). In response to low-frequency tone bursts, the electrical signal recorded in normal hearing subjects clearly demonstrated phase-locked activity with the fine structure of the stimulus. The synchronization index calculated from a Fourier analysis of the nerve response was maximal at 700 Hz and declined progressively toward higher frequencies. Beyond 2 kHz, no phase-locked response could be observed. Surprisingly, a robust phase-locked response was observed in subjects in the 500-1000 Hz range even in cases of hearing loss up to 60-70 dB SPL. Phonemes (/o/, /u/, /i/, /y/ vowels and /du/, /bu/, /di/, /bi/ syllables) were delivered at 70 dB SPL in quiet or noisy environments. Our results also revealed a remarkable preservation of neural responses reflecting temporal fine structure in both normal and hearing-impaired patients. However, responses related to the temporal envelope were significantly reduced in those with hearing loss. This suggests that the basal region of the cochlea, which is often affected in hearing loss, plays a key role in processing the temporal envelope of sound. Intracranial recordings offer a promising way to study how the human auditory nerve encodes complex sounds. This technique has the potential to significantly improve our understanding of how hearing loss disrupts speech processing at the neural level.

Voice cue sensitivity and speech perception in speech maskers in children with hearing aids

Laura Rachman[1], Pınar Ertürk[2], Gizem Babaoğlu[2], Basak Özkişi Yazgan[2], Gonca Sennaroğlu[2], Etienne Gaudrain[3], Deniz Başkent[1]

1. University of Groningen / University Medical Center Groningen, Netherlands | 2. Department of Audiology, Health Sciences Institute, Hacettepe University, Ankara, Turkey | 3. Lyon Neuroscience Research Center, CNRS UMR5292, Inserm U1028, Université Lyon 1, Lyon, France

Listeners can make use of voice cues, such as fundamental frequency (F0) and vocal-tract length (VTL), to segregate speech signals from different speakers. Hearing loss and hearing aid use may lead to reduced access to relevant acoustic cues, and in children, perceptual cognitive mechanisms may still be developing, all affecting the perception of speech masked by competing speech. In this study, we assessed sensitivity to F0 and VTL voice cues and perception of speech in the presence of competing speech maskers differing in F0 and VTL in school-age children with hearing aids, taking into account developmental effects.

We measured just-noticeable differences (JNDs) of F0 and VTL cues and speech perception in single-talker speech maskers using the Turkish Child-friendly Coordinate Response Measure (CCRM-tr) in native Turkish speaking school-age children (aged 5-18) with hearing aids and with normal hearing. Stimuli were recorded by female speakers. Single-talker speech maskers were created by concatenating random chunks from the sentence database and manipulating F0 and VTL cues to produce voice differences between the target and masker speaker. The target and masker sentences were presented at different target-to-masker ratios.

Both groups of children show development effects of voice cue sensitivity and speech-on-speech perception as a function of age. The developmental patterns for F0 and VTL JNDs seem to differ. Children with hearing aids seem to catch up with age-typical development around teenage years for F0 JNDs, but not for VTL JNDs. For speech perception in speech maskers, both groups show a benefit of F0 and VTL differences between target and masker speech, but the VTL difference benefit is relatively smaller for the children with hearing aids. Our results show a large variability in children with hearing aids, with some children performing at the level of age-matched children with normal hearing, and others performing lower.

Given the development effects throughout childhood and teenage years, age-matched norm data is essential for the evaluation of speech perception in speech maskers in individual children with hearing aids with respect to their age, and for the adjustment of rehabilitation to fit their needs. The finding that many children with hearing aids perform at age-expected levels demonstrates that hearing aids can provide good compensation for hearing loss for voice cue sensitivity and for speech perception in speech maskers.

Cochlear implantation in children with single sided deafness

Astrid van Wieringen
Experimental ORL, Dept Neurosciences, KU Leuven, Belgium

Children with single-sided deafness (SSD) experience difficulties with localizing sounds and understanding speech in noisy environments and may also be at risk for speech-language delays. Over the past decade, 3 groups of children were followed up at regular intervals: children with SSD who had received a cochlear implant (CI) in the deaf ear at a very early age, children with SSD, and children with bilateral normal hearing. Longitudinal analyses show that the children with SSD who did not receive a CI were at risk for grammar skills, poorer speech perception in noise, sound localization, and verbal IQ. Children with SSD and CI were on par with normal hearing children regarding grammar and speech perception in noise. An overview of the main findings will be given. These confirm that children with prelingual SSD can benefit from a CI provided at an early age to support their development across multiple domains. The longitudinal data led to a federal policy change in health care in Belgium in 2024.

Acknowledgements: The studies were conducted by Anouk Sangen and Tine Arras between 2014 and 2024. The multi-center partners were Christian Desloovere, Jan Wouters (University Hospital Leuven, ExpORL, Dept Neurosciences, KU Leuven), An Boudewyns (University hospital Antwerp, University of Antwerpen), Ingeborg Dhooge (University hospital Gent, University Gent), Erwin Offeciers, Andrzej Zarowski (European Institute for ORL-HNS), and Birgit Philips (Cochlear Ltd Mechelen). The research was funded by the European Union [grant number FP7-60713], by VLAIO [grant number HBC.2020.2308], as well as the Research Foundation Flanders [grant number T002216N]. These grants were awarded to the KU Leuven. The cochlear implants and clinical support were funded by Cochlear Ltd throughout the study. All data have been published.

Modelling the contributions of auditory, speech, language, and cognitive processes to speech-in-noise perception in school-aged children: A structural equation approach

Xuehan Zhou[1], Harvey Dillon[2], Dani Tomlin[3], Kelly Burgoyne[1], Helen Gurteen[4], Grace Nixon[3], Alisha Gudkar[3], Antje Heinrich[1]
1. The University of Manchester, UK | 2. Macquarie University, Australia | 3. The University of Melbourne, Australia | 4. The University of Queensland, Australia

Understanding speech in noisy and reverberant environments, such as classrooms, requires the integration of auditory, language, and cognitive abilities. Many children struggle to listen in such settings. The vast majority of children with listening difficulties (LiD) are diagnosed with peripheral hearing loss. However, at least 5% of children referred to audiology services present with normal peripheral hearing. In these cases, the diagnosis can be challenging because the role of hearing, auditory processing, language, and cognitive deficits can differ between children. It is imperative to develop more sensitive and more consistent diagnostic tools for children with LiD because correct diagnosis is the first step to successful treatment. If LiD remains untreated, it can negatively affect long-term academic outcomes.

As a first step toward better diagnosis, we explored the relationships among auditory processing, phoneme identification, sentence understanding in noise, language, cognitive abilities, and reading performance in a representative sample of 221 school-aged children. Using a structural equation model enabled us to simultaneously consider how bottom-up auditory processing and top-down language and cognitive processing contribute to speech-in-noise outcomes.

Our structural equation model showed that bottom-up auditory resolution of nonspeech sounds supports phoneme identification in noise, which in turn facilitates sentence understanding in noisy and reverberant conditions. Sentence understanding in noise was strongly supported by top-down language abilities, while memory and intelligence contributed indirectly by enhancing language abilities. Reading performance in this cohort was primarily influenced by cognitive factors, particularly nonverbal intelligence. Notably, cognitive abilities had minimal direct impact on nonspeech auditory processing measures, supporting a conceptual separation between auditory processing and cognitive abilities.

This study is the first to simultaneously quantify the direct and indirect contributions of auditory, language, and cognitive abilities to children's speech-in-noise perception and reading performance. Our findings underscore the multifactorial nature of LiD and the clinical need for a multidisciplinary diagnostic approach for accurate diagnosis as a first step to effective remediation.

Tuesday 13 January 2026, 11:35—12:00

What does cognitive listening mean? Is it time for a rethink?

Sven Mattys
University of York, UK

Current SPiN models emphasize the contribution of cognitive resources to speech understanding in challenging conditions. However, these approaches (1) overly focus on moderately degraded signals, (2) do not adequately distinguish the mechanisms underlying L1 and L2 performance, (3) tend to underestimate the contribution of individual differences in perceptual and linguistic abilities, and (4) do not fully specify how resource engagement (i.e., effort) changes across the signal-quality continuum.

To address these shortcomings, we propose an integrative framework that characterizes listening as a continuum governed by three limiting factors: Data, cognitive (or mental) resources, and linguistic abilities. Building on Norman and Bobrow's (1975) distinction between data-limited and resource-limited processes, the Data-Resource-Language (DRL) framework extends this taxonomy to include a language-limited region. The DRL posits that the primary determinants of speech perception vary systematically with signal quality. When acoustic input is severely degraded (data-limited), comprehension is constrained by the capacity of our perceptual system, and engaging additional cognitive resources yields minimal benefit.

As signal quality improves to a moderate level (resource-limited), cognitive resources such as working memory and attention control become crucial for integrating incomplete or ambiguous speech cues. This is the region conventionally targeted by “cognitive listening” research. Under near-optimal conditions (language-limited), performance asymptotes are determined primarily by individual differences in linguistic knowledge (e.g., vocabulary, syntactic fluency, and discourse comprehension) rather than by perceptual or cognitive factors.

The framework also formalizes the documented nonlinear relationship between resource engagement and intelligibility, often depicted as an inverted U-shaped function, with maximal cognitive engagement occurring at moderate signal quality. Evidence from task-evoked pupil responses (TEPR) supports this claim, showing that effort and motivation peak when additional cognitive investment can meaningfully enhance performance.

The DRL framework offers testable predictions across listener populations. For hearing-impaired listeners, reduced access to auditory data shifts all three processing regions rightward on the low-to-high signal-quality continuum, increasing cognitive demand even in favorable conditions and explaining the more permanent state of operating within an effortful, resource-limited region for that population. For non-native listeners, limited linguistic knowledge shifts the language-limited region leftward, expanding the range in which both linguistic and cognitive factors interact to constrain performance.

By integrating perceptual, cognitive, and linguistic determinants of speech perception, the DRL framework provides a unified account of how listener-specific abilities and signal characteristics jointly shape performance and resource engagement. DRL offers new pathways for theory development, data re-analysis, and clinical intervention in speech and hearing sciences.

Tuesday 13 January 2026, 12:00—12:25

Colin Cherry Award 2025

Age-related changes in alpha activity during dual-task speech perception and balance

Jessica L. Pepper[1], Theodoros M. Bampouras[2], Helen E. Nuttall[1]

1. Lancaster University, Lancaster, UK | 2. Liverpool John Moores University, Liverpool, UK

Older adults find it more difficult than younger adults to allocate attentional resources between co-occurring multisensory tasks, such as perceiving speech-in-noise whilst maintaining balance. Attentional control may be reflected in oscillatory alpha activity, with increases in activity reflecting inhibition of different brain regions and decreases reflecting neural activation. This study investigated how younger and older adults reallocate attentional resources during dual-task speech perception and balance, and how these age-related changes are reflected in alpha activity.

Twenty-four younger adults (18-35 years old) and twenty-one older adults (60-80 years old) identified words in audiovisual sentences extracted from the Grid corpus. Participants completed this speech perception task with or without background noise, whilst standing in easy or difficult balance positions. 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 speech-in-noise performance was strongest during the challenging balance condition, in contrast to our hypotheses. Whilst these behavioural effects were not reflected in parieto-occipital alpha power, decreases in fronto-central alpha power were greater in clear listening conditions. Taken together, increasing cognitive load may not always be detrimental to balance in physically and cognitively fit older adults.

Posters

SESSION 1: Monday 12 January 2026, 15:00-17:30

SESSION 2: Tuesday 13 January 2026, 14:00-16:30

P01 Role of current spread and device coding in vocal-tract length perception in cochlear implants

Etienne Gaudrain[1], Sil van Zoest[2], Floris Rotteveel[2,1], Bert Maat[2], Deniz Başkent[2]
1. Lyon Neuroscience Research Center, Université Lyon 1, CNRS UMR5292, Inserm U1028, Lyon, France | 2. University Medical Center Groningen, Department of Otorhinolaryngology, Groningen, Netherlands

Besides phonetic information, speech contains indexical cues that are informative for, amongst other things, identifying who is talking. Vocal-tract length (VTL), which is related to the size of the speaker, is one of the important cues associated with speaker identity. Previous studies have shown that VTL perception is strongly impaired in most cochlear-implant (CI) users. Here we are exploring potential causes for this observation.

When speech is transmitted through a CI, it is first transformed into a series of electric pulses through a processing strategy. These processing strategies have been optimised for the transmission of phonetic information through the constraints of electrical stimulation. However, the strategies might not be optimal for other types of information such as VTL. Furthermore, the pulses generated by the strategy are then used to stimulate the auditory nerve by electrical stimulation through a rather conductive medium. This results in a large spread of excitation where the activated neurons are not necessarily located close to the stimulating electrode, thereby reducing the effective spectral resolution available to the listener. Consequently, both the processing strategy and the current spread at the electrode-neuron interface could play a role in reducing the information available to the listener and it can be difficult to disentangle the two.

To disentangle the contributions of these two mechanisms, we recorded the electrical outputs produced by a Cochlear Nexus 7 processor using the MP3000 coding strategy. These electrical outputs were then resonified using a vocoder that allowed us to control the amount of (simulated) current spread by adjusting the spectral width of the carrier. By comparing the VTL discrimination performance obtained in the original stimuli, to those obtained through vocoding with and without current spread, we could estimate how much the coding strategy alone and the electrode-neurone interface each contribute to the loss of sensitivity.

We collected VTL sensitivity from 16 normal-hearing adults across two variations of the experiment. Our results indicate that the MP3000 strategy preserves VTL information remarkably well, but that current spread quickly reduces the performance.

These results now need to be confirmed in actual CI users, where other measures of spectral resolution and current spread can be obtained using tests such as spectro-temporal ripple discrimination, as well as more objective measures of current flow based on eCAP measures.

P02 Oculomotor tracking of selective attention in speech-in-noise listening

Xena Liu, Maria Chait
Ear Institute, University College London, UK

Background: Successful interaction with complex auditory scenes in real life requires dynamic engagement of various cognitive processes, including arousal, attention, distractor suppression, and memory. Previous research has established that such information can be uncovered from our eyes, through orienting ocular responses like phasic Pupil Dilation (PD), Microsaccades (MS), and blinks.

Our recent studies using tone sequences have indicated that these oculomotor measures can provide instantaneous readouts of arousal and attention allocation in auditory processing. In the present experiment, we show that combining these measures provide rich information about time-to-time cognitive mechanisms in speech-in-noise processing.

Methods: We simultaneously recorded pupil and oculomotor movements (MS and blinks) with eye-tracking using a speech-in-noise working memory task (N=30). The task required subjects to focus attention on a target speaker (and report the words they produced) while ignoring temporally interleaved words from a distractor speaker, all embedded in multi-speaker babble.

Results: Distinct response patterns to behaviourally-relevant target and distractor words - reflecting fluctuations in instantaneous arousal and attentional engagement - were observed across pupil dilation (PD), microsaccades (MS), and blink time series. Time-locked additional analyses revealed robust response patterns to targets vs. distractors, as well as signatures of attentional states (e.g. attention lapse) associated with behavioural outcomes.

Each measure exhibited unique temporal characteristics, suggesting that they index different facets of attentive listening. Specifically, increased pupil responses (diameter and dilation rate) indicated heightened arousal prior to and following target words; while oculomotor inhibition was enhanced by the allocation of temporal attention, observed across both MS and blink dynamics, signalling a push-and-pull interaction between the oculomotor and auditory system, where visual sampling is reduced when attentional resources are deployed for the auditory modality.

Discussion: These results suggest that pupil and oculomotor dynamics can track time-to-time attentional allocation and arousal engagement processes in listening, even in difficult listening conditions. In addition, distinct patterns of different eye movements indicate potentially separate underlying neural pathways.

Understanding such underpinnings of oculomotor dynamics in hearing provide deeper insights into auditory attention and cross-modal interactions, as well as potential guidance for research into challenges in effortful listening, many of which possibly arising from breakdowns in distractor suppression in real-life complex auditory scenes.

P03 Language proficiency drives listening-related fatigue in those who communicate in their second language

Ronan McGarrigle, Jelena Havelka
University of Leeds, UK

Angela de Bruin
University of York, UK

Individuals who listen in their second language ('L2 listeners') are disproportionately impacted by adverse acoustic conditions; a phenomenon coined the 'non-native speech-in-noise disadvantage'. Evidence also suggests that L2 listening taxes cognitive resources to a greater extent than in L1 listeners. Very little, however, is known about the longer-term consequences of this sustained cognitive engagement. Listening-related fatigue refers to the daily life subjective experience of tiredness and exhaustion from effortful listening.

The current study recruited a large sample of L2-English listeners (N = 256) to explore linguistic and demographic factors that predict listening-related fatigue. Approximately half (n = 126) of this sample reported an Indo-European (IE) language (e.g., Italian) as their first language, whereas the other half (n = 130) reported a non-Indo-European (non-IE) language (e.g., Vietnamese) as their first language.

Non-IE L2 listeners reported higher overall listening-related fatigue than both IE L2 listeners and monolingual controls. Both self-rated and objective English language proficiency were found to be significant predictors of listening-related fatigue in L2 listeners. Additionally, language proficiency moderated the effect of L2 use on listening-related fatigue; individuals who reported increased L2 use were more likely to experience increased fatigue, but only if their language proficiency (both self-rated and objective) was low and/or moderate. Finally, exploratory mediation analysis suggested that the effect of L2 listener group (IE vs non-IE) on listening-related fatigue was largely attributable to lower language proficiency in the non-IE L2 group.

To conclude, language proficiency appears to be a strong determinant of listening-related fatigue in L2-English listeners and individuals whose L1 and L2 are more typologically distinct appear most susceptible to listening-related fatigue.

P04 Neural tracking of speech in real-time conversation: an EEG hyper-scanning study

Maria Perdiki, Paul Iverson
University College London, UK

We used EEG hyper-scanning to investigate speech processing of real-time continuous speech during conversation between non-native adults. In this study, seven dyads (n=14) of adult non-native English speakers completed a 'spot-the-difference' picture game in both their first language (L1) and English (L2). Participants were seated in separate rooms and conversed using stand-mounted microphones while listening to each other through EEG insert headphones. EEG was recorded simultaneously from both participants using 32-channel systems with bilateral mastoids. Conversations were recorded and speech amplitude envelopes of both takers were extracted and time-aligned to the EEG. We applied a multivariate temporal response function (mTRF) approach to estimate neural tracking at the acoustic level (e.g., speech envelope) for both perceived and produced speech in the speakers' first and second languages. Preliminary results demonstrated that we could obtain reliable auditory neural tracking using this paradigm. Further work will involve analysing the transcripts to derive lexical measures (e.g. lexical predictability and frequency) and correlating them with EEG in L1 and L2 conversations. We will extend these findings by testing additional real-world conditions, such as speech in noise.

P05 Audiovisual integration and gaze behaviour in degraded speech perception: An EEG and eye-tracking study

Xiyuan Li, Patti Adank
University College London, UK

Background: Speech is often perceived multimodally, with visual information from a speaker's face facilitating speech perception compared to auditory-only listening, especially in challenging listening conditions. Different gaze strategies, particularly where listeners fixate on the face, can affect speech perception outcomes. Yet, how visual cues influence degraded speech perception and how gaze patterns modulate the neural processing of speech in naturalistic settings remain poorly understood. Cortical tracking, which quantifies how closely the neural activity aligns with the temporal dynamics of speech (e.g., the speech envelope), provides an index of how effectively the brain encodes speech and enables studies using more ecologically valid stimuli. This study aimed to investigate whether audiovisual integration benefits occur when processing noise-vocoded naturalistic speech, and whether specific gaze patterns modulate neural encoding of speech envelope and audiovisual benefits.

Methods: Nineteen native British English adults were presented with naturalistic noise-vocoded stories under three presentation modalities (audiovisual, auditory-only, visual-only) at two degradation levels (4-band and 8-band) while electroencephalography (EEG) and eye-tracking data were recorded simultaneously. Cortical tracking analyses employed forward encoding models of the speech envelope, and audiovisual integration benefits were quantified using the additive model criterion [AV > (A + V)], which tests whether multisensory responses exceed the sum of uni-sensory responses. Two facial regions of interest (eyes and mouth) were defined for eye-tracking analysis. Gaze data were analysed using k-means clustering of the Eye-Mouth Index, which quantifies the relative fixation preference for mouth versus eyes, to identify distinct gaze patterns. Pearson's correlations were used to investigate the relationship between gaze behaviour and cortical tracking accuracy.

Results: While we did not find audiovisual integration benefits in cortical tracking for either degradation level, this study found a novel relationship between adaptive gaze strategies and neural speech tracking. Specifically, gaze analysis revealed distinct behavioural clusters, with participants categorised as either mouth-dominant or eye-dominant viewers based on their fixation preferences. Among the mouth-dominant viewers, reduced attention to the eyes relative to the mouth significantly correlated with enhanced neural tracking accuracy of the speech envelope in the more degraded condition. These findings provide novel evidence that adaptive, mouth-focused gaze strategies could functionally benefit neural processing of degraded speech and have important implications for understanding individual differences in multisensory speech perception.

P06 The inner ear's active process contributes to selective attention to speech in noise

Janna Steinebach, Tobias Reichenbach
Friedrich-Alexander-University Erlangen-Nuremberg, Germany

10.1101/2025.08.15.670505

Background: Understanding speech in noisy environments relies on the ability to focus on a target talker while filtering out competing input, a process typically associated with cortical auditory stream segregation. However, the auditory system is not strictly feedforward: descending efferent pathways, including the medial olivocochlear (MOC) bundle, project from higher auditory centers to the outer hair cells in the cochlea and regulate cochlear gain. This raises the possibility that selective attention can shape auditory processing at the earliest stages of the peripheral system. Otoacoustic emissions (OAEs), which originate from the cochlea's nonlinear mechanics, offer insight into these peripheral dynamics. While distortion product OAEs (DPOAEs) are usually elicited with pure tones, more complex, speech-like stimuli have rarely been used.

Methods: In this work, we employed speech-derived DPOAEs (speech-DPOAEs) to determine whether attention to speech modulates cochlear responses and whether such modulation depends on whether harmonic components are resolved or unresolved along the basilar membrane. The stimuli were generated from male and female speech by selecting harmonic overtones n and m (n < m) of the fundamental frequency to elicit cubic distortion products at 2n–m. Resolved harmonics are expected to produce distinct excitation peaks on the basilar membrane, whereas unresolved harmonics are assumed to overlap. The design allowed speech-DPOAEs from both voices to be measured simultaneously.

Results: Forty normal-hearing adults (18–31 years) were presented with two competing voices in one ear, while speech-DPOAEs were simultaneously evoked and recorded in the contralateral ear. Participants alternated their attention between the male voice, the female voice, or a visual distractor across repeated trials. We found that speech-DPOAEs corresponding to resolved harmonics were significantly reduced when the associated voice was attended relative to when it was ignored, indicating that attention reduces cochlear output for spectrally resolved speech elements. In contrast, no effect emerged for unresolved harmonics when both target and competing voices contained unresolved components in the same frequency range. Comparable modulation was also observed when auditory attention was contrasted with visual engagement, suggesting that intermodal attention can likewise influence cochlear activity.

Conclusion: The present results offer the first direct evidence that selective attention to speech in noise can act on the cochlear amplifier itself, consistent with efferent MOC control of outer hair cell function.

P07 Automated question and answering based speech-in-noise test using large language models

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

The aim of this study was to develop and evaluate a self-supervised conversational sentence-in-noise (SIN) test using machine learning and to compare its validity (detecting hearing loss) and reliability (consistency across two runs) with those of a standard clinical test. Two tests were used. The first was the new Conversational Hearing Assessment Test (CHAT), designed to evaluate hearing by measuring participants' ability to understand and respond to simple questions. In total, 500 statements and 500 corresponding questions were generated using ChatGPT and converted to speech with a text-to-speech (TTS) model. Each statement–question pair was mixed with background noise and presented to the participants. They were then asked to answer the question. Their spoken responses were transcribed with automatic speech recognition (ASR), and ChatGPT determined whether the answers were correct. The second test was the Adaptive Sentence List (ASL) test, a clinically established test, in which sentences were mixed with noise and participants verbally repeated the presented sentence. Their response was then scored by the researcher to determine whether they repeated the word correctly or not. Both tests measured speech reception thresholds using an adaptive up–down procedure. Reliability was assessed with Bland–Altman analyses, and limits of agreement between two runs of the same test were reported. Validity was evaluated using the area under the ROC curve (AUC and Youden index) to classify participants as normal hearing or hearing impaired. 40 native speakers participated, including 20 with normal hearing and 20 with hearing impairment.

For reliability, two runs of the ASL test had limits-of-agreement of ±3.1 dB, whereas CHAT had a limits-of-agreement of ±3.6 dB. In terms of validity, the ASL test reached an AUC of 0.93 and Youden index of 0.71 while CHAT had an AUC of 0.97 and Youden index of 0.89. These findings indicate that CHAT achieved validity and reliability comparable to the ASL test, demonstrating its potential as an ecologically valid, accurate and reliable screening tool for hearing loss. Moreover, because CHAT operates without the need of a human supervisor, it can be adapted for online delivery, allowing participants to complete the test remotely in their own homes.

P08 Do bone anchored hearing aids alleviate listening effort in patients with single-sided deafness? Evidence from pupil dilation, eye movements, and blink-rate analysis

Claudia Contadini-Wright[1], Thatchapat Anekpichayasit[1], Mark Chung[2], Nishchay Mehta[2,1], Maria Chait[1]
1. Ear Institute, University College London, UK | 2. Royal National ENT Hospital, London, UK

Single-Sided Deafness (SSD) poses unique challenges for individuals navigating acoustically complex environments. Although many can meet the everyday demands of listening by relying on their functioning ear, this often incurs a substantial cognitive cost— manifesting as mental fatigue, increased listening effort, and depletion of attentional resources. Bone-Anchored Hearing Aids (BAHA) provide a means of restoring audibility in unilateral hearing loss; however, their ability to alleviate cognitive load and listening-related fatigue, dimensions not typically captured by standard clinical assessments, remains uncertain.

To investigate this, we employed a spatially separated adaptation of the Coordinate Response Measure (CRM) task. Speech stimuli were presented to the deaf ear, with competing noise presented to the hearing ear. Participants (N = 16; ongoing) performed two listening conditions: High Load (BAHA off) and Low Load (BAHA on). An additional control group (N=30; normal hearing) was tested with the same set up and simulated SSD (through occluding one ear).

Listening effort and attentional allocation were assessed using three oculomotor measures: Pupil Dilation Response (PDR) as a marker of arousal and effort, Microsaccade Rate (MSR) as an indicator of moment-to-moment attentional allocation, and Eye Blink Rate (EBR) as a measure of general cognitive engagement. Results show that, despite improving behavioural performance on the task, there is little evidence that BAHA reduces arousal under high load, as reflected by PDR. However, BAHA-listening was associated with a higher MSR, suggesting reduced burden of allocation on attentional resources relative to the BAHA-off condition. EBR analyses revealed significant modulation after the sentence keywords, with rebound dynamics differing by load condition— indicating that BAHA may facilitate computation speed. These findings suggest that while BAHA may not fully alleviate all dimensions of cognitive effort, it may support more efficient attentional processing in SSD.

Importantly, these results demonstrate the need to integrate efficient and non-invasive physiological assessments into clinical research, enabling a more comprehensive understanding of the cognitive and perceptual impact of auditory interventions beyond simple, traditional measures of audibility.

P09 Alpha oscillations during effortful listening originate from different sources

Caroline Hamery[1], Guillaume Andéol[2], Sébastien Scannella[1]
1. Institut supérieur de l'aéronautique et de l'espace (ISAE-SUPAERO), Toulouse, France | 2. Institut de recherche biomédicale des armées (IRBA), Brétigny-sur-Orge, France, France

Listening effort, referring to the cognitive resources engaged when understanding speech in challenging conditions, can be assessed through behavioral and physiological measures. While the behavioral measures tend to be subjective, the use of physiological measures, such as electroencephalography (EEG), has shown mixed results. The multidimensionality of listening effort can thus be measured in different ways. Objective measures, such as EEG, offer the possibility to better understand the neural mechanisms underlying speech processing. In this study, we used EEG to examine how alpha oscillations are modulated during effortful listening and whether these changes originate from different neural sources. Alpha oscillations have been described in relation to listening in complex environments in contradictory ways. Some studies report increases in alpha power during effortful listening, while others describe a decrease in alpha power. One explanation for these inconsistencies seems to rely on the presence of multiple alpha generators in the brain. In this study, we investigated alpha oscillations across different difficulty levels in both speech-in-noise (SIN) and speech-in-speech (SIS) conditions.

Speech intelligibility, subjective listening effort and electrophysiological data were recorded during speech-in-speech and speech-in-noise performance of 30 participants. Spectral and time-frequency analyses were used to assess cortical responses to different levels of auditory difficulty. Because interpreting alpha oscillations in the time-frequency domain alone can be ambiguous, we extended the analysis by applying independent component analysis (ICA) to the EEG data. This approach allowed us to determine whether observed alpha dynamics reflected distinct underlying neural components. In addition, exploratory source localization was performed to identify the cortical origins of these components.

Results showed that, in the time-frequency domain, alpha power increased in the left temporo-parietal region during both SIN and SIS conditions, consistent with greater engagement in difficult listening. However, ICA revealed that these changes were associated with different independent components. Some components exhibited alpha synchronization, while others showed desynchronization. This pattern indicates that alpha oscillations during effortful listening are not generated by a single source but by multiple alpha generators with distinct functional roles. Alpha-band EEG activity may reflect greater neural inhibition, which could indicate a change in listening processes during the task. This modulation seems linked to task difficulty and may also relate to listening effort. By refining these results by extracting independent components related to alpha activity described in the literature, we confirmed that alpha power observed in the time-frequency domain is the result of different independent components, and that both enhancement and suppression of alpha can be observed during listening in SIN and SIS.

P10 A binaural model predicting psychometric functions for speech intelligibility in non-stationary noise and listeners with and without hearing loss

Anaïs Hiard[1], Thibault Vicente[2], Virginia Best[3], Jörg Buccholz[4], Mathieu Lavandier[1]
1. ENTPE, Ecole Centrale de Lyon, CNRS, LTDS, UMR5513, Vaulx-en-Velin, France | 2. Laboratoire d'Acoustique de l'Université du Mans (LAUM), UMR 6613, Institut d'Acoustique - Graduate School (IA-GS), CNRS, Le Mans Université, France | 3. Department of Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts 02215, USA | 4. Department of Linguistics-Audiology, Australian Hearing Hub, Macquarie University, New South Wales, 2109, Australia

The binaural speech intelligibility prediction model proposed aims to take into account some of the auditory mechanisms that influence speech intelligibility in noise (spatial release from masking, dip listening, hearing loss). This model considers better-ear listening (BE) and binaural unmasking (BU). The model predicts the psychometric function averaged across normal-hearing (NH) listeners and the individual psychometrics functions for hearing-impaired (HI) listeners, using an internal noise (IN) implementation that includes the individual audiogram of the listener and the overall level of the external stimuli. The model is developed on the basis of previous work which predicted only differences in speech reception thresholds (SRT). It shares a common structure with those models. The speech and masker signals at the listener's ears are filtered with a gammatone filterbank. In each frequency band, the advantages of binaural hearing (BE and BU) are estimated to compute an internal signal-to-noise ratio (SNR). To predict complete psychometric functions, a reference psychometric function is needed as input, whose characteristics relied on the target speech material considered and they are derived using experimental data measured in a given condition. This function is used to convert the internal SNR to percent correct. To consider dip listening, this scheme is calculated in time frames and averaged.

To validate the model, four datasets were used—each involving a different type of masker (stationary noise, modulated noise, 2 competing voices, 4 competing voices)—and four versions of the IN implementation were compared. The results showed that the model can capture well the difference in slopes highlighted in the literature: the slope of the psychometric functions in stationary noise are steeper than in modulated noise. The comparison of the different IN approaches indicates that it is necessary to set a ceiling value on the IN to get predictions that fit the data, otherwise the performances of the severely-impaired HI listeners are under-predicted. Compared to the models predicting only differences in SRT, the proposed model can capture the dependence of the perceptual effects (spatial release from masking, dip listening, hearing loss) on the SNR or intelligibility level considered.

P11 Audio-visual enhancement of speech perception across the lifespan: Evidence from behavioural and EEG data.

Mathilde de Kerangal[1], Mike Thornton[2], Enrico Varano[3], Tobias Reichenbach[2]
1. University College London, London, UK | 2. Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany | 3. Universität Zürich, Zürich, Switzerland

Hearing is essential for communication, but as people age, understanding speech in noisy environments becomes challenging. Age-related hearing loss contributes to this, but the alteration of brain mechanisms involved in the analysis of complex auditory scenes - such as attention, memory, or inhibition of irrelevant information - also plays an important role. Speech perception can be enhanced when the listener can both hear and see their interlocutor. Various types of visual signals can promote audio-visual (AV) enhancement, but it remains unclear whether age and cognition impact AV integration. The first aim of the current study is to investigate how age and cognitive factors affect AV integration when listeners are exposed to both natural and degraded visual speech signals. The second aim is to examine whether the neural processing of continuous speech is modulated by AV integration and ageing. Behavioural and EEG experiments (N=33, age range: 20-72) were used for these purposes respectively. Using the AV GRID corpus, two degraded versions of the videos were presented to the participants in addition to the unprocessed (natural) version. The target speaker was presented in a two-voice babble at an SNR corresponding to the speech reception threshold (SRT) of each participant. The subjects also completed several cognitive tasks targeting working memory (digit span and Corsi task), sustained attention (reaction time task), and inhibition (Stroop task).

The behavioural results demonstrated a gradual improvement of speech comprehension with the addition of details in the visual signals. SRTs correlated with age, working memory, and Stroop scores, but AV enhancements were unrelated to these factors, suggesting that differences between A and AV conditions were not driven by age or cognition. This indicates that AV integration for speech remains stable and robust across the lifespan.

The EEG data was used to reconstruct the target speech envelope of continuous speech. Focusing on the Delta band, reconstruction accuracies demonstrated a clear benefit of AV integration in the tracking of speech in all visual conditions. Additionally, AV reconstruction scores in the natural condition correlated positively with age, demonstrating a potential deficit of neural inhibition in older age. To conclude, AV benefit was observed behaviourally and in the EEG data for natural and degraded speech. The behavioural data showed that more visual details led to stronger enhancement regardless of age or cognitive abilities. EEG results showed improved neural speech tracking through AV integration but possible age-related deficits in neural inhibition.

P12 Adaptive infant speech networks under acoustic challenge revealed by HD-DOT

Addison Billing, Deborah Vickers
University of Cambridge, UK

Speech perception in noise (SPiN) is particularly critical during early development when infants and toddlers are acquiring language in environments that are often noisy, such as nurseries. Effective SPiN processing relies on coordinated neural networks supporting both top-down prediction and listening effort, as well as bottom-up auditory processing. Adult neuroimaging studies have identified involvement of several key brain regions in SPiN, including the left inferior frontal gyrus (IFG), the parietal lobe, sensorimotor areas, and temporal lobes. However, the mechanisms underlying SPiN in the immature infant brain remain unclear. Infants typically exhibit lower tolerance for background noise during language comprehension compared to adults, suggesting potential differences in neural processing strategies. To investigate this, we applied high-density diffuse optical tomography (HD-DOT) combined with Granger causality analysis to measure hemodynamic responses in 6- to 8-month-old infants exposed to audiovisual nursery rhymes under three conditions of speech clarity: clear (high clarity), and two degraded conditions at 8 dB and 4 dB signal-to-noise ratio. These conditions elicited distinct patterns of top-down versus bottom-up connectivity. Under the highest clarity condition, we observed strong top-down language prediction signals from the left IFG to the left inferior parietal lobule (IPL), which diminished as speech clarity decreased. Conversely, degraded speech conditions elicited enhanced connectivity between frontal and sensorimotor regions, consistent with increased listening effort. Right hemisphere engagement shifted between bottom-up and top-down connectivity patterns depending on the difficulty of the listening condition. These findings provide preliminary evidence that Granger causality analysis combined with HD-DOT is a viable method to investigate speech in noise processing in pre-verbal infants. Understanding these early neural mechanisms may inform interventions for infants at risk of speech and language difficulties in challenging acoustic environments.

P13 Reverberation impairs neural stream segregation of concurrent speech

Anna-Lena Krause[1], Sabina Wiley[1], **Lars Hausfeld**[1,2]
1. Department of Cognitive Neuroscience, Maastricht University, Netherlands | 2. Maastricht Brain Imaging Centre, Netherlands

Background: In everyday acoustic settings, listeners must often attend to one speaker among several concurrent talkers. The auditory system relies on cues such as spatial location and pitch to segregate the target from the distractor stream. Although previous studies have examined these mechanisms using brief, anechoic speech stimuli, the neural dynamics supporting speech segregation in reverberant and naturalistic environments remain unclear.

Methods: Participants (N = 18) listened to 30-second audiobook segments containing two concurrent speakers while high-density EEG was recorded. Speech mixtures were systematically manipulated across three acoustic dimensions: reverberation (high vs. low), spatial separation (co-located vs. separated), and pitch separation (small vs. large), using simulated binaural room impulse responses. Behavioral responses were recorded assessing intelligibility and subjective difficulty. Neural tracking was assessed using multivariate temporal response functions (mTRFs) based on direct, reverberant, and combined speech envelope models.

Results: Behaviorally, listeners' intelligibility significantly declined under high reverberation ($p < 10^{-8}$), while spatial and pitch separation exerted minimal effects. EEG prediction accuracy revealed significant main effects of both reverberation ($p = .019$; high < low) and model type ($p < 10^{-10}$; combined, direct > reverberant), indicating reduced neural tracking in more reverberant environments. Examination of the combined mTRF model showed early fronto-central responses (30–130 ms in low, 30–180 ms in high reverberation) for both target and distractor speech, suggesting prolonged early auditory processing under degraded conditions. A later, left-lateralized attentional enhancement (210–270 ms) emerged for target versus distractor speech in the low-reverberation condition but disappeared when reverberation was high. For reverberant speech, two target-selective processing stages (90–160 ms and 200–260 ms) were present in low but not in high reverberation, indicating increased difficulty and disrupted stream segregation.

Conclusions: These findings suggest that under low reverberation, the auditory system can separate target and interfering speech more effectively, possibly supported by reverberant cues that aid stream segregation of direct and reverberant speech. In contrast, strong reverberation appears to add to masking, limiting the auditory system's ability to dissociate target and distractor streams and reducing overall speech intelligibility.

P14 Progress toward developing a wearable microphone array and EEG recorder

Paul Iverson[1], Ana Campos[1], Maria Perdiki[1], Kathleen McCarthy[2], Bronwen Evans[1], Lin Wang[2]
1. University College London, UK | 2. Queen Mary University of London, UK

Assessment of real-world speech input has proven valuable for understanding child language acquisition, and has potential for understanding plasticity later in life (e.g., second language learning, hearing aids and cochlear implants, language disorders). That being said, real environments are noisy, making it difficult to conduct acoustical analyses on typical recordings with one or two microphones, and existing wearable recorders are mostly proprietary or non-programmable. We've recently developed a programmable microphone array recorder (4-16 microphones), with the intention of making it open source. The aim is to use microphone arrays and beamforming to better isolate speakers from background noise. We are also in the process of adding dry-electrode EEG in order to enable measures of attention, allowing us to assess what a listener is actively processing rather than measuring all ambiant sound. At present, our hardware and basic device-level programming for audio recording are in place, and we are beginning to trial the devices and develop post-recording processing pipelines. We've tested an early version of the device on adult native Japanese speakers in London, and have found that those who have more exposure to English in their daily life are better at understanding British-accented speech in noise in laboratory tests.

P15 Reflexive characteristics of the multimodal Lombard effect

Victoria Ivanova, Naomi Harte
Sigmedia Group, Trinity College Dublin, Ireland

Interaction is a multimodal process in which visual speech conveys a substantial amount of information. Speech and gesture are biomechanically linked, and evidence suggests that both are controlled by overlapping motor systems. Consequently, vocal speech phenomena are intrinsically entangled with gestures. Previous work investigating whether the Lombard effect extends to manual gestures has found that speakers produce gestures with more submovements and longer hold times when speaking in noise. However, such effects could reflect deliberate communicative strategies rather than a genuine Lombard response.

In fact, Lombard speech is characterized not only as a communicative adaptation but also as a reflex, showing rapid, low-level modulation through sensorimotor feedback pathways. It occurs even in non-interactive contexts, where speakers rely solely on auditory and somatosensory feedback in the absence of listener feedback. Additionally, it scales continuously with stimulus intensity rather than appearing as a binary effect. Therefore, to determine whether gestures are truly part of the Lombard effect, we must determine whether these same key characteristics apply.

We aim to determine whether gestures in noise show Lombard-like characteristics in a behavioural study with around 30 participants. Participants will complete a story-retelling task in a sound-treated studio under three conditions: quiet, 75 dB, and 85 dB multitalker babble, presented through speakers. They will watch short silent cartoon clips and then retell each story freely, without an interaction partner or explicit mention of gesturing.

Speech will be recorded via a lapel-mounted microphone to verify the presence and magnitude of the vocal Lombard response and enable automatic transcription. Gestural kinematics will be captured with wrist-mounted inertial measurement units (IMUs) and complemented by AI-based video motion tracking. From these signals we will extract indices of gesture velocity, submovements, hold duration, and spatial amplitude. Gesture occurrences will also be annotated by an expert coder assisted by gesture-detection software. This design tests whether gestures scale with noise intensity in the absence of communicative feedback, indicating a reflexive, Lombard-like modulation rather than a deliberate communicative adjustment.

If speech and gesture share partially common motor control mechanisms, we expect a general motor gain in noise, emerging both as increased vocal effort and as enhanced gesturing effort, reflected in changes in velocity, vertical amplitude, hold duration, or submovements. This might be further modulated in an interaction situation through a feedback loop, accounting for idiosyncratic differences. Our findings would have implications for models of speech–gesture integration. Currently, there is no agreement on how these two modalities are integrated and we hope to shed light on this question. Our presentation will feature illustrations from recordings and deeper insights into the data collection.

Funding: This study has emanated from research conducted with the financial support of Taighde Éireann - Research Ireland under Grant number 22/FFP-A/11059.

P17 Divergent cortical speech tracking at the cocktail party: Preserved in hearing aid users, impaired in cochlear implant users

Constantin Jehn, Jasmin Riegel, Tobias Reichenbach
Friedrich-Alexander Universität Erlangen, Germany

Anja Hahne, Niki Katerina Vavatzanidis
Ear Research Center Dresden (ERCD), University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

Background: Listeners with hearing impairment, including users of hearing aids (HAs) and cochlear implants (CIs), often struggle with speech comprehension in noisy backgrounds. Speech comprehension might be enhanced if the focus of attention was known and the corresponding speech signal could be enhanced. Recent research showed that the attentional focus can be recorded from EEG by considering the neural tracking of the speech envelope. However, this neural response has been predominantly studied in typical hearing subjects, leaving a critical gap in our understanding of how it functions in individuals with hearing loss.

Methods: We recorded EEG data while the participants attended to one of two competing talkers, a female and a male one, in a free-field acoustic environment. In contrast to prior studies, HA users as well as CI users used their personal, clinically-fitted devices. Cortical speech tracking was assessed through linear backward and forward models that related the EEG data to the speech envelope (1-8Hz). Moreover, we assessed speech comprehension using 3AFC multiple choice questions and subjective listening effort.

Results: Behaviorally, we found a significant increase in listening effort from TH to HA and further to CI users. Furthermore, comprehension scores were significantly decreased in CI users. Neurally, the backward model revealed that the envelope tracking of the HA and TH group did not differ significantly. As a result, the attention decoding accuracies in both groups were indistinguishable, reaching a maximal accuracy of 82% for 60s decision windows. In contrast, the CI group showed a profound deficit in attentional modulation, reaching a maximal decoding accuracy of 63% on a 60s decision window. The forward model further supported these findings. TRFs and encoding scores were largely similar between TH and HA users, whereas CI users' impaired cortical speech tracking was reflected in weaker attentional modulation of both encoding scores and temporal response functions.

Conclusion: This study reveals divergent neural envelope tracking between HA and CI users in a challenging multi-speaker environment. These patient-specific differences underscore the need to tailor future neurofeedback in hearing instruments to the unique neural processing of each user group.

P18 Electroencephalographic measurements of speech detection in noise in the first year of life

Clémence Basire, Laurianne Cabrera
INCC, CNRS, Université Paris Cité, France

Infants are known to have more difficulties detecting speech in the presence of a noisy background compared to adults. Most of the previous studies used behavioral methods to assess this capacity, whereas electroencephalography (EEG) may provide objective neural markers of this ability. Our study aims to assess infants' speech detection abilities in a noisy background by examining the auditory event related potentials (ERP) using a passive listening procedure.

We recruited a group of 25 three-month-old infants who were following-up at 10 months of age. We also recruited a group of 31 adults for comparisons. All listeners were awake throughout the EEG recordings and were presented with the syllables "ba" or "da" in a continuous stationary speech-shaped noise. The syllables were delivered via two speakers in front of the listeners and were randomly played at four different signal-to-noise ratios (SNR = +3, -4, -8, -15 dB). A total of 52 trials per SNR was played while measuring ERPs for about 20 min. The ERP amplitude after syllable onset was averaged over five frontal electrodes and in the time window 400:600 ms, corresponding to the P300-like ERP response.

A first analysis showed that for both infant groups and the adult group, only ERP amplitude obtained at the SNR -15 dB were not significantly different from the neural responses obtained during the noise inter-trial presentation. A linear-mixed model was then run on the mean [400-600ms] ERP amplitude and revealed a significant effect of Age ($p = .002$), SNR level ($p < .001$), syllables ($p <.001$) and interaction between Age and SNR ($p = .006$). Post-hoc analyses indicated that at 3 months of age, decreasing the SNR significantly decreased the ERP amplitude, while at 10 months of age the ERP amplitude differed significantly between the -15 and +3, -15 and -4, and -8 and +3 conditions, but not between any other pairs.

These findings showed in both infant groups an effect of SNR on the P300-like amplitude: the greater SNRs, the greater the ERP amplitude. This result is in line with adult studies showing that the P300 amplitude decreases as the difficulty of discriminating stimuli increases. Importantly, ERP response were significantly different from neural noise from SNR equal to -8 dB suggesting that as for the adults, the infants' brain is able to detect syllable at this SNR level. Our results are promising for objective clinical measures of speech in noise.

P19 Does familiarisation with dysarthric speech predict listeners' intelligibility ratings?

Ariadna Sanchez, Cassia Valentini-Botinhao, Simon King
Centre for Speech Technology Research, University of Edinburgh, UK

Speech disorders can affect people's ability to communicate with others effectively, as it limits a listener's ability to decode messages effectively. Particularly, we focus on a speech disorder type called dysarthria, where speech motor muscles weaken, resulting in imprecise articulation, slurred or slow speech, or changes in voice characteristics.

Previous work has seen a positive effect on intelligibility when listeners are previously exposed to a specific speaker with dysarthria, i.e. short-term perceptual adaptation. In these studies, listeners were presented with an explicit familiarisation phase prior to the transcription task for the same speaker. During the familiarisation phase, listeners listen or read-while-listen to stimuli. However, it is unclear how much exposure time is necessary for an improvement in the transcription task, as previous studies have proposed different lengths of exposure (Borrie et al., 2012). To the best of our knowledge, no study has controlled for different lengths of exposure in the familiarisation phase to study its effects in listener short-term perceptual adaptation for dysarthric speech. We hypothesise that exposing listeners to different amounts of stimuli (up to 5 minutes of recordings) prior to the transcription task will improve listeners' transcription accuracy rates, but that this improvement will not be linear, i.e. improvements will plateau as exposure time increases.

To investigate our hypothesis, we propose a between-subjects study where listeners are assigned to different exposure conditions. Conditions are divided by speaker (speaker A or B), and length of the familiarisation phase prior to testing (100 sentences, 50 sentences, 25 sentences, 0 sentences). After exposure, listeners complete a transcription task on unseen sentences from the same speaker as in the familiarisation phase. Speakers A and B are selected from the Speech Accessibility Project (Hasegawa et al., 2024), a dataset created to improve automatic speech recognition systems for those with a speech impairment. Speakers were selected based on their overall intelligibility, aimed to be between 40-60%, to avoid selecting speakers that might be too challenging to understand (and therefore we might see very minimal effects from exposure), or too intelligible (and therefore we would see ceiling effects). We will present our results for the aforementioned study, and discuss possible future directions for this research, such as replicating it with more challenging speakers (i.e. those with lower overall intelligibility).

References

Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric speech: A review of experimental studies. Journal of Speech, Language, and Hearing Research, 55(1), 290-305.

Hasegawa-Johnson, M., Zheng, X., Kim, H., Mendes, C., Dickinson, M., Hege, E., ... & MacDonald, B. (2024). Community-supported shared infrastructure in support of speech accessibility. Journal of Speech, Language, and Hearing Research, 67(11), 4162-4175.

P20 Listening effort during speech processing in quiet and in noise

Yue Zheng[1], Ethan Brims[1], Ronan McGarrigle[2], Sarah Knight[3], Sven Mattys[1]

1. University of York, UK | 2. University of Leeds, UK | 3. Newcastle University, UK

Both very quiet speech and speech in noise present challenges for the listener. Understanding quiet speech is demanding due to the intrinsic weakness of the signal, which, at very low levels, can be further interfered with by internal noise. In contrast, speech in noise involves energetic masking at all signal-to-noise ratios (SNRs) and might require additional cognitive resources to discriminate the target speech from background noise. Therefore, in this study, we asked whether understanding speech in noise is more effortful than understanding quiet speech.

In the current study, participants were asked to repeat back sentences played with or without a speech-shaped noise masker (Noise vs. Quiet) while having their task-evoked pupil response (TEPR) tracked as a measure of listening effort. Each participant first went through a staircase procedure to determine either the SNRs (Noise condition) or the dB SPL thresholds (Quiet condition) corresponding to 90% and 50% accuracy. Participants then completed the main listening task, in which they were again asked to repeat back sentences presented at these individually tailored intensity levels. The design, therefore, involved Listening Condition (Noise vs. Quiet) and Difficulty Level (90% vs. 50%). Subjective ratings of tiredness, effort, and accuracy were also collected for each of these four conditions.

The current findings are based on preliminary data from 14 participants (target N = 64). As expected from the staircase procedure, average performance was at approximately 90% and 50% accuracy in the respective Difficulty Levels across the two Listening Conditions, with mean SNRs at -2.68 dB and -6.18 dB for the Noise condition, and mean SPLs at 28.88 dB and 23.44 dB for the Quiet condition (90% vs. 50% condition, respectively). TEPRs were, overall, larger in the 50% than 90% condition. Listening Condition did not appear to significantly influence TEPRs. However, this 50%-vs-90% TEPR difference was substantially larger in the Noise than Quiet condition. Subjective measures correlated more strongly with performance than with the TEPR patterns.

These preliminary results suggest that understanding speech in noise is not more effortful than understanding quiet speech. The reduced 50%-vs-90% TEPR difference in the Quiet condition indicates that perceiving quiet speech might involve consistent cognitive demands regardless of intelligibility and performance levels. Quiet speech might therefore induce a steady state of attentional vigilance largely independent of signal quality and possibly linked to a constant level of internal noise. In contrast, the TEPR pattern in the Noise condition was more closely linked to signal availability (SNR) and performance, as observed in previous studies.

P21 The subjective and physiological signatures of listening effort

Johannes Wienen[1,2,3], Dorothea Wendt[1], Hamish Innes-Brown[1]

1. Eriksholm Research Centre, Snekkersten, Denmark | 2. Amsterdam UMC, Netherlands | 3. Vrije Universiteit Amsterdam, Netherlands

When listening situations become increasingly challenging due to background noise, individuals tend to disengage from listening. Based on previous research on physical and mental effort, we hypothesized that the decision to disengage is influenced by multiple factors including the speech-to-noise ratio, the potential reward of listening, current fatigue levels and recent errors in recall.

This study used an effort discounting paradigm in which participants were presented with offers indicating the task difficulty and potential reward on the upcoming trial. On each trial, participants chose whether they wanted to work (perform a speech-in-noise recall task to potentially earn credits) or if they wanted to rest. After each trial, participants provided a subjective fatigue rating and received feedback on their performance.

Different computational models were tested for their ability to explain participants' choices and fatigue ratings on a trial-by-trial basis. Additionally, pupillometry and fNIRS were used to measure pupil dilation and cortical blood oxygenation changes in response to the stimuli.

The data shows that participants chose to work less frequently in high-difficulty, low-reward conditions. These offers were also rejected more frequently in the second half of the experiment. Fatigue ratings increased slightly more after trials with recall errors compared to trials where the sentence was repeated correctly.

Preliminary results favor a computational model in which the subjective value of effort is characterized by the reward on offer, discounted by the associated task difficulty. The effect of task difficulty is scaled by an effort discounting factor that incorporates fatigue levels and performance feedback.

P22 Remembering spoken information: how noise impacts listeners' content and source memory

Isabel S. Schiller, Sabine J. Schlittmeier
Work and Engineering Psychology, RWTH Aachen University, Germany

Our ability to understand and remember information from spoken conversations in acoustically complex listening conditions is crucial for everyday communication. Oftentimes, we are required to recall both the content ("what was said?") and the source ("who said it?"). To systematically study how challenging listening conditions impact content and source memory, we adapted the Heard Text Recall (HTR) paradigm (Schlittmeier et al., 2023, doi:10.18154/RWTH-2023-05285). The HTR assesses comprehension and memory for spoken text based on short passages followed by open-ended content questions. In the present study, we added talker-related questions to this procedure, resulting in the HTR-TR version. This study employed the HTR-TR paradigm to explore how stationary background noise impacts listeners' ability to remember both conversational content and its corresponding talker. In an audio-only experiment, participants listened to short conversational passages from two spatially separated talkers (±60° azimuth) under quiet and pink noise (-3 dB SNR) conditions. We hypothesized that pink noise would impair content memory and, particularly, source recall, reflecting the increased cognitive demands of integrating both types of information. Results will be presented and discussed in terms of their relevance for studying speech-in-noise memory, with a focus on methodological enhancements and future applications of the HTR-TR paradigm in auditory cognition research.

P23 Prosody adaptively supports speech comprehension in noise

Tugba Lulaci[1], Pelle Söderström[2,1], Mikael Roll[1]
1. Lund University, Lund, Sweden | 2. National Acoustic Laboratories, Sydney, Australia

Speech unfolds rapidly and rarely happens under optimal conditions. Everyday listening requires the brain to extract meaning from incomplete and overlapping acoustic signals. Speech comprehension is thought to benefit from predictive mechanisms that integrate early cues to anticipate upcoming information. In Swedish, prosodic patterns (word accents: Accent 1 and Accent 2) on the word stem signal different suffix continuations, allowing listeners to generate expectations about the upcoming information before the word fully unfolds. This study investigated how background noise affects predictive use of such prosodic information.

Speech-shaped noise was used to mask the speech signal. Electroencephalography (EEG) was recorded while adult native Swedish listeners with normal hearing (n = 22, mean age = 22.5 years, SD =2.46, range = 18–28) heard sentences containing target words with a word accent followed by either an incorrect suffix (mismatch) or a correct suffix (match). Stimuli were presented in quiet and in two different signal-to-noise ratios (SNRs): 0 dB SNR and -5 dB SNR.

Behavioral results showed that predictive speech perception was affected by both listening condition and word accent. Response times increased in noise and mismatch trials. In the EEG results, mismatched stimuli elicited N400–P600 activity in quiet speech but these responses weakened when the speech signal was masked by noise. Crucially, the two prosodic patterns differed in their resistance to noise: accent 1 remained an effective predictive cue across conditions, whereas accent 2 lost its reliability as a cue when the masking increased. This difference suggests that prosodic cues vary in their acoustic informativeness and in how strongly they engage predictive mechanisms under adverse listening conditions.

These findings suggest that noise constrains—yet does not eliminate—the predictive use of prosody during speech comprehension. Prosodic cues continue to guide comprehension when the speech signal is degraded, though with less precision than in clear speech. This gradual weakening of predictive processes highlights how the brain adapts to uncertainty in natural listening and could offer a basis for future research on how such mechanisms operate in populations facing everyday hearing challenges.

P24 Role of sound onsets and offsets in cortical tracking of transient events during naturalistic speech listening

Adèle Simon[1], Maria Chait[1], Jennifer F. Linden[1,2]
1. Ear Institute, University College London, UK | 2. Department of Neuroscience, Physiology, and Pharmacology, University College London, UK

Background: The amplitude envelope of speech is crucial for comprehension, and cortical activity in the theta-delta bands has been shown to track this envelope. This tracking is thought to reflect responses to transient events in the speech signal, such as sound onsets (rapid increases in sound amplitude) and offsets (rapid decreases in sound amplitude). While onset tracking has been widely studied, offset tracking remains underexplored.

Methods: In this study, three independent datasets from different laboratories were analysed, consisting of continuous EEG recordings during audiobook listening in quiet from British (n=18) [1] and Danish (n=22) [2] participants or in noise from British (n=28) participants. An onset/offset model based on thalamic responses to sound transients in mice [3] was used to extract separately the onsets and offsets present in the broadband envelope of the continuous speech. Linear forward models using either speech onsets, offsets or both onset and offsets as regressors were used to assess separately how well each of the transient events are tracked during naturalistic speech listening. Additional analyses were performed on onsets and offsets extracted from a 16-channel gammatone spectrogram of the envelope to explore if onset and offset processing during speech listening relies on distinct spectral regions.

Results: Model performance, measured by the correlation between predicted and recorded EEG, was significantly above chance for all representations ($p<.001$), confirming that both onsets and offsets contribute to speech tracking. On the broadband envelope, Wilcoxon tests showed that onset-based models outperformed offset-only models ($p<0.001$, $d=0.69$), suggesting that onsets have a stronger effect. However, the combined model performed best overall, surpassing both offset-only ($p<0.001$, $d=0.85$) and onset-only ($p<0.001$, $d=0.56$) models. This finding highlights the complementary role of offsets in cortical speech tracking. Analyses of the envelope spectrogram are still being finalised, but preliminary results suggest that high-frequency channels may be particularly important for offset tracking.

Conclusion: Overall, our results suggest that sound onsets and offsets play a distinct role in cortical speech tracking both in quiet and in noise, with onsets exerting a stronger influence, but offsets providing significant complementary information.

References

[1] Etard O & Reichenbach T (2022). doi:10.5281/zenodo.7086208
[2] Simon A, Bech S, Loquet G & Østergaard J (2022). doi:10.5281/zenodo.7500806
[3] Anderson LA & Linden JF (2016). J Neurosci 36(6), 1977-1995.

P25 Influence of auditory and cognitive factors on speech in noise processing

Akshaya Jeyasri Kannan, Christian Sumner, Alexander Hardy, Jia Hoong Ong
Nottingham Trent University, Nottingham, UK

Speech-in-noise (SIN) perception—the ability to understand speech in challenging listening environments—is vital for effective communication. Even among individuals with normal hearing, SIN difficulties can impact daily life and social participation. Previous research has implicated both auditory and cognitive factors in SIN perception, yet their relative contributions remain debated. The present study examined these relationships in 75 adults with normal hearing using a comprehensive behavioural battery. Auditory measures included Frequency Difference Limen (FDL) and Gap Detection Threshold (GDT), while cognitive measures included Auditory Working Memory (AWM), Auditory Stroop, and the Sustained Attention to Response Task (SART). Spearman correlations showed that SIN performance correlated negatively with FDL ($\rho = -0.47$, $p < .001$) and positively with AWM ($\rho = 0.37$, $p = .001$), indicating that better frequency discrimination and stronger working memory are linked to improved SIN understanding. Stroop ($\rho = 0.26$, $p = .023$) and GDT ($\rho = 0.22$, $p = .059$) showed weaker or marginal associations. To assess the unique contributions of these variables, a multiple regression analysis was conducted. Both FDL ($p = 1.9e{-}05$) and AWM ($p = 0.006$) emerged as significant predictors of SIN performance, together explaining 36% of the variance ($R^2 = 0.36$, $p < .001$). In contrast, Stroop and GDT were not significant predictors when included in the model. These preliminary findings suggest that both auditory and cognitive mechanisms contribute to speech-in-noise perception, with frequency discrimination exerting the strongest influence and auditory working memory providing additional support. However, the study is ongoing, and further data collection—including additional tests like EEG recordings—is expected to provide deeper insights into the underlying mechanisms of SIN perception.

P26 Auditory processing of acoustic cues and speech recognition in noise in children with mild to moderate hearing loss

Charlotte Benoit, Inès Sonzogni, Laurianne Cabrera
Babylab, INCC, CNRS UMR 8002, Paris, France

Background: Speech-in-noise (SIN) perception is known to be affected by sensorineural hearing loss (SNHL). This is not only related to reduced audibility but also to impaired basic auditory mechanisms essential for speech processing that are affected by cochlear damage. The goal of the present study was to investigate the impact of SNHL in childhood on temporal and spectral processing, as well as SIN perception.

Methods: Preliminary results are presented with data collected among 12 children with mild-to-moderate SNHL and aged between 5 and 10 years. They were presented with four SIN conditions: consonant versus word identification in speech-shaped noise (SSN) versus 2-talker babble noise. In parallel, auditory temporal (amplitude and frequency modulations- AM/FM-fluctuating at 4 or 20 Hz) and spectral ripple density (tested at 0.5 or 2 ripple-per-octave) detection thresholds were measured, along with receptive vocabulary (using the Peabody Picture Vocabulary test) and nonverbal working memory measures (using the forward and backward versions of the Corsi test). Thresholds of children with SNHL were compared to those of children with normal hearing (NH) of the same age range.

Results: As expected, children with SNHL performed more poorly than NH peers in all SIN tasks. They also showed worse FM and spectral detection thresholds but preserved or even improved AM detection. Preliminary analyses showed that SIN thresholds did not correlate with temporal or spectral cue detection, neither nonverbal working memory. Only a weak relationship between consonant identification and receptive vocabulary was found.

Conclusions: As observed in previous adult studies testing the effects of acquired SNHL, our results in children with congenital SNHL showed that this condition affects FM and spectral processing, but that AM processing is relatively preserved. Children with SNHL displayed elevated thresholds in all SIN tasks used and a strong inter-individual variability. However, the analyses so far have not revealed any auditory or cognitive factor that emerge as dominant predictors of SIN in children with SNHL. To better capture inter-individual variability and fully assess the effects of age and development, data collection must continue.

P27 Nasal vowel perception with vocoded speech in Belgian French-speaking children: impacts of the number of channels and simulated insertion depth.

Sophie Fagniart
Language Sciences and Metrology Unit, University of Mons (UMONS), Mons, Belgium

Olivier Crouzet
Laboratoire de Linguistique de Nantes - UMR6310, Nantes Université / CNRS, Nantes, France

Background: Nasality is contrastive for vowels in french-language varieties: speakers differentiate vowel pairs like /E/ (as in tête', eng. _head_) vs. /E~/ (as in teinte', eng. hue, note that we use the SAMPA phonetic transcription system). However, it is usually advocated that such contrasts may involve two distinct relationships between the oral and the nasal member of the pair: a phonological one and a phonetic one. On a phonological account, the 3 nasal vowels {E~, A~, O~} contrast with the 3 oral vowels {E, A, O} respectively. However, considering phonetic grounds, the oro-pharyngeal configuration of the vocal tract in these 3 nasal vowels is much closer to the 3 oral vowels {a, O, o} respectively. Also, acoustic properties of nasal vowels are known to be complex as they involve a coupling between the oral and nasal cavities that gives rise to anti-resonances. These phenomena lead to various acoustic consequences that make the distinction between nasal vowels and their oral counterparts difficult to characterize in natural speech. The aim of the present study is to investigate how various parameters of speech vocoding may influence the perceptual identification of oral and nasal vowels by typical-hearing children and to provide perspectives concerning their recognition in the context of cochear implants.

Method: Vowel perception was investigated in 2-Alternative Forced-Choice tasks involving either an identification or a discrimination procedure among 8 children (all in their 8th year). Nasal vowels were associated with both their phonological oral counterpart or their phonetic oral counterpart. Speech sounds were artificially manipulated in order to systematically remove any duration or fundamental frequency differences between the stimuli while spectral content was preserved. Speech vocoding stimuli were generated using a Pulse-Spreading Harmonic Complex carrier with respectively 4, 12 or 22 channels and either a 0 or 3 mm simulated insertion depth. The original natural speech sequences were used as a control condition.

Results: With 4-channels, performance was similarly low for both insertion depths in the phonetic and the phonological contrast conditions. Though perceptual scores increased with the number of vocoded channels, they were significantly lower for phonetic pairs than for phonological pairs, and this difference was systematically more impactful when a 3 mm insertion depth was added.

Discussion: These results confirm that acoustic information associated with strict nasal resonance are strongly affected by various vocoding parameters and are particularly impacted by insertion depth.

P28 Differences in multimodal communicative behavior of hard-of-hearing and hearing individuals in social and non-social background noise

Anna Palmann[1], James Trujillo[1], Linda Drijvers[2], Floris Roelofsen[1]
1. University of Amsterdam, Netherlands | 2. Donders Institute for Brain Cognition and Behaviour, Radboud Universiteit, Nijmegen, Netherlands

Communication is inherently multimodal, requiring interlocutors to combine signals from the auditory and the visual domains to ensure mutual understanding, joint meaning-making, and ultimately communicative success. Hard-of-hearing individuals, who face difficulties in the auditory domain, might therefore rely more on visual communicative signals, such as gestures and facial expressions – especially in situations with loud background noise, which are perceived as particularly challenging. To date, it has not been empirically investigated how different forms of visual communication, such as sign-supported speech, gesturing, or lip reading, interact in dialogue situations with background noise for people who are hard of hearing. Furthermore, it is not well-established if the characteristics of their multimodal communicative behavior, i.e., the kinematic properties of their gestures and the acoustic properties of their speech, interact and differ from those of hearing individuals who engage in conversation in the same situation.

In the present study, we analyze video- and audio-recordings of conversations between hard-of-hearing dyads and between hearing dyads, who engaged in three rounds of conversation (free dialogue, joint decision-making task, director-matcher task) while being exposed to changing background noise (no noise, social noise, non-social noise) in the lab. We calculate the following kinematic features for each gesture instance per participant: maximum height, volume, sub-movements, rhythmicity, peak velocity, maximum distance from the body, and use of McNeillian space. For each speech instance per participant, we also calculate intensity and pitch. Using linear discriminant analysis (LDA), we investigate whether differences in these kinematic and acoustic features are big enough to reliably classify the hearing status (hard-of-hearing vs. hearing) of participants, as well as the type of background noise. Further, we investigate whether there is an interaction effect between the two.

A preliminary LDA based on kinematic features resulted in a successful classification between hard-of-hearing and hearing participants with an accuracy of 80%, indicating that their gesture behavior does indeed differ significantly. The role of speech and potential modulations of multimodal communication according to background noise type are still to be determined. Future analyses will also focus on whether kinematic and acoustic characteristics are predictive of communicative success, which we operationalize as a combination of self-report measures (questionnaires) and task measures (accuracy and reaction time).

With this research, we shed new light on the differences in communicative behavior between hearing and hard-of-hearing people in challenging listening conditions. By identifying patterns of particularly differentiating multimodal communicative signals, we ultimately aim to develop informed suggestions on how to create a more inclusive, pleasant, and welcoming communicative experience in situations with background noise.

P29 The influence of hearing loss on speech perception in music or noise

Adriana A. Zekveld[1,2], Laura Keur-Huizinga[1,2], Nicole A. Huizinga[3,2], Jorn H. Sangers[1,2], Niek J. Versfeld[1,2], Eco de Geus[3,2], Sophia E. Kramer[1,2]

1. Amsterdam UMC, Netherlands | 2. APH research institute, Amsterdam, Netherlands | 3. VU University, Amsterdam, Netherlands

In the standard Speech Reception Threshold (SRT) test, speech perception is typically assessed using sentences masked by stationary noise. In this study, we evaluated the ability of participants to perceive speech in the presence of ambisonics recordings of either background music or foodcourt noise, to increase the ecological validity of the test. We aimed to determine whether hearing loss would influence both the SRT and listening effort as measured by pupil dilation and subjective ratings. In addition, we applied an auditory and visual version of a Paced Serial Addition test to assess if any effects of hearing loss would extend to a less language-dependent task.

Ambisonics recordings of the masker signals were presented over 8 loudspeakers arranged in the horizontal plane around the listener, while the target sentences were presented from the front. A total of 129 participants (mean age: 58 years, range 37-73 years) completed the SRT task in either the background music or the foodcourt noise condition, targeting 80% correct intelligibility. Hearing acuity ranged from normal hearing to severely impaired (mean pure tone average (0.5 – 4kHz) = 27 dB HL, range -5 to 94 dB HL), and 55 participants performed the test with their own hearing aid(s).

The results showed that increasing hearing loss was associated with poorer SRTs and SRTs were better in the foodcourt masker than in the music masker condition. These results and the pupil response, subjective, and serial addition data will be presented and discussed.

P30 Beyond the audiogram: Using patient-reported outcome measure to uncover personalized hearing loss needs

Perrine Morvan
Institut Pasteur, Paris, France

Importance: Pure-tone audiometry and speech understanding tests are commonly used in clinical audiology but fail to capture communication challenges and daily-life difficulties experienced by individuals with hearing loss (HL), limiting the personalization of care.

Objective : To determine whether integrating Patient-Reported Outcome Measures (PROMs) into audiological practice can reveal specific functional and psychosocial needs not captured by traditional audiological tests.

Design: We conducted a population-based cross-sectional analysis on data collected between September 1, 2020, and December 31, 2022.

Setting: Multicentred study involving 700 audiology clinics across France.

Participants: A total of 91,297 adults aged 18 to 99 years with symmetrical HL were included based on the completeness of audiometric and PROMs data.

Exposures: Hearing loss severity, assessed using Pure-Tone Average (PTA) thresholds (500, 1000, 2000, and 4000 Hz) along with speech tests in quiet (Speech Reception Threshold, SRT) and in noise (Signal-to-Noise Ratio, SNR). PROMs were collected using the Client Oriented Scale of Improvement (COSI), where patients identified up to five listening situations for improvement with hearing aids.

Main Outcomes and Measures: The primary outcome was the association between specific PROMs and PTA. Regression models were used to identify significant PROMs-related predictors of hearing loss severity.

Results: The sample was gender-balanced (44,525 females; 46,762 males) with a mean age of 72.1 years (SD, 11.8). Most (81%) had mild to moderate HL (mean PTA 42.9 dB HL; SD, 11.3). Among 273,861 PROMs responses, individuals with slight to moderate HL revealed a wide range of clinical needs—such as increased social contact or avoiding embarrassment—that varied significantly by age and sex and were not explained by audiometric thresholds. In contrast, for those with severe to profound HL, needs were more homogeneous and primarily driven by PTA.

Conclusions and Relevance: PROMs identify specific, patient-reported needs that are not captured by conventional audiological measures—especially in slight to moderate HL where individual variability is high. Their integration into routine care provides a strong rationale for improving the management of milder HL and supports more personalized and effective rehabilitation strategies.

P31 Influence of noise setup on speech intelligibility during speech-in-noise testing for cochlear implant users

Anne van Alphen[1], H. Christiaan Stronks[1,2], Jeroen J. Briaire[1], Johan H.M. Frijns[1,2,3]
1. Department of Otorhinolaryngology and HNS, Leiden University Medical Center, Leiden, Netherlands | 2. Leiden Institute for Brain and Cognition, Leiden University, Leiden, Netherlands | 3. Department of Bioelectronics, EEMCS, Delft University of Technology, Delft, Netherlands

10.5281/zenodo.18153593

Background: Cochlear implant (CI) users have difficulty understanding speech in background noise. Directional microphones, also known as beamformers, are spatial filters that enhance speech intelligibility in noisy environments by attenuating sounds originating from the sides and rear. Beamformers can be monaural or binaural. Monaural beamformers generate a directional beam on only the cochlear implant. For binaural beamforming, a second device, such as a contralateral routing of signals (CROS) device, can be incorporated, by capturing sounds presented to the ear contralateral to CI ear (CROS-ear) and rerouting them to the CI, which subsequently enhances speech intelligibility in noise. Multiple research groups have investigated these beamformers for CIs with varying results. However, the variability in the test setups used complicates the comparison of different studies. The extent to which these different noise configurations influence the effectiveness of directional microphones is not known.

Rationale: Assess the influence of noise configuration on the effectiveness of beamformers in CI users.

Methods: 18 postlingually deaf CI users, monaurally implanted with an Advanced Bionics device, participated in this study. Speech recognition thresholds (SRTs) were assessed using the Dutch/Flemish Matrix test. The SRT was defined as the signal-to-noise ratio (SNR) where 50% of speech was recognized correctly. Four conditions were tested for CI users: CI-only, CI-CROS, CI with UltraZoom (monaural beamformer), and CI-CROS with StereoZoom (binaural beamformer). Four different noise setups were tested: a homogeneous noise field, 8-loudspeaker ring, 3-loudspeaker ring, and 1-loudspeaker setup. Speech was always presented from a loudspeaker positioned directly in front of the participant. Noise level was 60 dBA on average. The CROS-ear was plugged to minimize influence of residual hearing.

Results: No significant differences were observed between the various noise setups in the CI-only and CI-CROS condition. UltraZoom and StereoZoom significantly improved speech understanding compared to the CI and CI-CROS condition (all comparisons $p < 0.001$). Specifically, improvements were 5.5 and 5.8 dB SNR in 3-loudspeaker, 3.3 and 4.2 dB SNR in 8-loudspeaker, and 3.8 and 4.5 dB SNR in the homogeneous noise field, respectively. Activating UltraZoom resulted in significant differences in SRTs across all noise setups. Similarly, with StereoZoom enabled, SRTs differed significantly across noise conditions, with the exception of the comparison between the 8-loudspeaker and homogeneous noise fields.

Conclusion: In all noise configurations, activating a beamformer led to increased speech intelligibility in CI users. However, the type of noise setup and the direction from which noise originates used during speech-in-noise testing does influence speech understanding, which must be considered when comparing studies.

P32 Frequency and timing of head movement in 3- and 4-way conversation

John Culling[1], Lauren Hadley[2], William M. Whitmer[2,3]
1. School of Psychology, Cardiff University, UK | 2. Hearing Sciences - Scottish Section, University of Nottingham, Glasgow, UK | 3. School of Health & Wellbeing, University of Glasgow, Glasgow, UK

Three- and four-way conversations in background noise were analysed for the frequency and timing of head turns between other participants. In both situations, it was found that head turns between interlocutors were much more frequent when the individual was speaking, than when they were listening to exchanges between the others. In the four-way conversation, turns were more frequent with increasing sound level when listening, but were independent of sound level when talking. The overall frequency of head movements was also greater in the four-way conversations than in the three-way conversations, both when speaking and when listening. Moreover, the participants turned earlier, anticipating the exchange of floor more frequently than in the 3-way conversations. However, because these data come from different experiments, accounting for these differences is not straightforward. For instance, exchanges of conversational floor were roughly twice as frequent in the four-way conversation, which may explain the increase in head movement when listening, but not while speaking. The four-way conversations also manipulated hearing impairment and hearing-aid use. Hearing aids for the two hearing-impaired listeners in a conversation changed the behaviour of both groups with more frequent head movement for the impaired participants and less frequent movement for the unimpaired participants. Again, these effects were apparent both when a given participant was speaking and when listening to others.

P33 Identifying acoustic cues for English voiced stop consonants with reverse correlation: a pilot study

Simon J. Major
Worcester College, University of Oxford, Oxford, UK

Léo Varnet
Laboratoire des systèmes perceptifs, Département d'études cognitives, École Normale Supérieure, PSL University, CNRS, Paris, France

Maria Giavazzi
Laboratoire de science cognitives et psycholinguistique, École Normale Supérieure, PSL University, Paris, France

Auditory reverse correlation is an experimental paradigm that allows researchers to uncover the acoustic cues a listener relies on during an auditory categorization task. The method comprises introducing random fluctuations into a stimulus, relating these variations to participants' behavioral responses, and thereby identifying which acoustic features are most influential for a given perceptual decision.

In a previous study, we applied this approach to seven phoneme discrimination tasks to identify the acoustic cues underlying the perception of French plosive consonants (Carranante, Cany, Farri, Giavazzi, Varnet, 2024, doi:10.1038/s41598-024-77634-w). Six vowel-consonant-vowel stimuli (/aba/, /ada/, /aga/, /apa/, /ata/, /aka/) were presented in seven 1-interval 2-alternative tasks, allowing us to study the perception of two phonetic features: place of articulation and voicing.

We present here a pilot study aimed at extending this work to the perception of English voiced stop consonants. Two main changes were introduced to the paradigm of Carranante et al. First, the stimuli were productions of /aba/, /ada/, and /aga/ by a native English speaker, and all participants were native speakers of English. Second, a 1-interval 3-alternative task was implemented and the analysis was adapted to derive kernels from multiple-alternative experimental designs.

These preliminary results confirm the critical role of first and second formant transitions (both consonant-vowel and vowel-consonant transitions) for the perception of English place of articulation. In contrast, high-frequency release bursts appear to carry weaker perceptual weights on average.

P34 Modulation detection interference and its relationship to speech perception in noise at adolescence

Ugo Benrubi, Axelle Calcus
Université Libre de Bruxelles, Belgium

Understanding speech in noise (SPiN), particularly when the masker is also speech, is a complex ability that relies on multiple acoustic and cognitive cues, among which amplitude modulation (AM) plays a crucial role. Previous studies have demonstrated that sensitivity to AM is crucial for encoding speech envelopes and segregating competing sound sources. Beyond basic AM detection (AMD), modulation-detection interference (MDI) reflects how the perception of a target modulation is disrupted by concurrent modulations in other frequency channels. While AMD and MDI have been studied extensively, their relationship to SPiN remains poorly characterized. This study quantifies MDI in adults and examines how it relates to individual differences in SPiN intelligibility.

To date, twenty normal-hearing adults aged 18–29 years have participated in the study. Data collection is ongoing, additional adolescent participants is expected to be included by the time of the conference. Speech intelligibility thresholds (SRT_{50}) were measured using a French version of the Coordinate Response Measure, in which participants have to identify a target sentence (color–number combination) among two conditions: two competing female voices (speech-on-speech condition) and envelope-modulated speech-shaped noise (eSSN). Participants also completed a 3-interval alternative forced-choice (3IAFC) AM detection task. The target and masker were each carried by distinct pure tones at different carrier frequencies. The target carrier was modulated at 2.62 Hz, corresponding to the dominant modulation rate of the CRM target voice, while the masker carrier was modulated in six conditions: (1) no masker, (2) unmodulated masker, (3) 2.62 Hz (same rate as target), (4) 3.55 Hz (dominant rate of CRM maskers), (5) 16 Hz, and (6) 32 Hz.

We expect the magnitude of MDI to depend upon the distance between target and masker modulation rates. Specifically, AM detection thresholds are expected to be lowest for unmodulated or high-rate (16 & 32 Hz) maskers and highest when the masker shares the same modulation rate as the target (2.62 Hz), reflecting maximal interference between overlapping amplitude modulation filter banks. The magnitude of this interference is anticipated to reflect the listener's ability to segregate competing envelopes and the level of internal modulation noise. We also expect individual differences in MDI to relate to SPiN (SRT_{50}), particularly in the speech-on-speech condition, highlighting the importance of envelope-processing mechanisms for SPiN. Data collection in adults is ongoing, and these results will serve as a baseline for future developmental work investigating how MDI contributes to lasting SPiN difficulties in adolescents.

P35 Neural encoding of visually and acoustically derived speech features during audiovisual narrative listening

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

The presence of visual cues (e.g. lip movements) can enhance speech perception and the neural tracking of acoustic features, such as the amplitude envelope. It is unclear, however, how this neural enhancement relates to behavioural measures of audiovisual benefit. Here, we investigated the neural encoding of acoustically- and visually derived speech features, capturing amplitude, temporal, and spectral modulations in the speech signal during narrative listening. We collected concurrent MEG and eye-tracking data (n=35) alongside objective (3AFC comprehension) and subjective (intelligibility ratings) measures of story comprehension while participants listened to a speaker narrating a story (Varano et al., 2023, doi:10.1121/10.0019460) in auditory-only (AO), visual-only (VO) and audiovisual (AV) conditions. To assess changes in speech tracking due to changes in modality independent of changes in intelligibility, AO and AV conditions were matched in relative intelligibility using noise-vocoding. We also collected offline behavioural measures of lipreading ability and audiovisual benefit, speech reception thresholds (SRTs) and verbal IQ. Preliminary analyses found evidence for audiovisual benefit for visually-derived and spectral features, most prominently over occipital sensors, in line with a perceptual restoration of spectral dynamics via visual speech cues (Plass et al., 2020, doi:10.1073/pnas.2002887117). Replicating previous work (Aller et al., 2022, doi:10.1523/jneurosci.2476-21.2022; Bröhl et al., 2022, doi:10.1523/eneuro.0209-22.2022) we also found that tracking of speech features in temporal, but not occipital sensors during silent lipreading was associated with lipreading ability and audiovisual benefit. Ongoing work will extend these results by disentangling the relative contribution of modality-specific information through comparisons of neural encoding models including both auditory and visual speech features.

P36 Parietal alpha connectivity tracks listening effort across and beyond signal-to-noise-interference ratios in hearing aid users in realistic acoustic scenes

Payam Shahsavari Baboukani[1,2], Emina Alickovic[1], Jan Østergaard[2], Kasper Eskelund[3]
1. Eriksholm research center (part of Oticon), Copenhagen, Denmark | 2. Aalborg university, Aalborg, Denmark | 3. Oticon A/S, Smørum, Denmark

This study investigates how signal-to-noise-interference ratio (SNIR) affects auditory performance and neural responses associated with listening effort (LE), and whether other acoustic factors beyond SNIR contribute to this relationship. We collected a new dataset from individuals with mild-to-moderate hearing loss, all fitted with hearing aids (HAs). Participants listened to two competing audiobooks from front-facing loudspeakers, while 16-talker babble noise was presented from four background speakers to simulate realistic acoustic scenes. Six SNIR levels (5.47, –3.55, –2.13, –1.19, –0.64, and –0.27 dB) defined six conditions, each consisting of 12 one-minute trials. In each trial, participants attended to one audiobook while ignoring the competing audiobook and background noise, then completed comprehension questions and reported subjective LE.

Behavioral results showed a significant linear effect of SNIR on subjective LE ratings and a quadratic effect on comprehension questionnaire accuracy: perceived LE decreased steadily with improving SNIR, whereas comprehension performance plateaued at higher SNIR levels.

EEG analyses revealed a significant linear relationship between SNIR and local connectivity in the parietal alpha band. Parietal local connectivity correlated with subjective LE ratings, while spectral power in the same region and frequency band did not, indicating that local connectivity is a more sensitive neural marker of LE. While alpha power and connectivity may share partially overlapping neural origins, they appear to reflect distinct aspects of the neural processes underlying effortful listening.

Beyond SNIR effects, systematic differences emerged between the two talkers under identical acoustic conditions. One talker consistently elicited higher subjective and neural indices of LE, suggesting that talker-specific features—such as voice characteristics or linguistic content—can modulate cognitive demand independently of acoustic degradation.

Together, these findings highlight that LE is shaped not only by the external acoustic environment but also by intrinsic speech properties. Parietal alpha connectivity provides a robust neural correlate of LE, reflecting both graded SNIR effects as well as talker-specific effects, making it a promising tool to assess cognitive resources during complex, realistic listening scenarios in HA users.

P37 Examination of speech processing and speech retrieval in cochlear implant users and typical-hearing listeners

Khaled Abdellatif[1,2], Denise Gradtke[1,2], Thomas Koelewijn[3,4], Deniz Başkent[3,4], 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, Cologne, Germany | 3. Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Netherlands | 4. Research School of Behavioural and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, Netherlands

Taking part in verbal conversation involves multiple cognitive processes, as listeners must process and retain speech while preparing a response. These processes rely on the efficient allocation of perceptual and cognitive resources. In comparison to healthy hearing, sounds transmitted via cochlear implants (CI) are limited, especially in terms of spectro-temporal cues. CI users are therefore likely to work harder in verbal conversation than typical-hearing (TH) listeners due to the degraded signal quality, which may require additional cognitive compensation.

This ongoing study investigates the extent to which CI users and TH listeners can memorize and retrieve speech information, and how this affects speech processing and cognitive load. To this end, participants completed an auditory n-back task in which matrix-sentences consisting of three words—a numeral, an adjective, and an object (e.g., “seven red flowers”)—were presented in quiet. Participants were asked to repeat each sentence and indicate whether a specified word was presented in the current (zero-back) or the previous (one-back) sentence.

We show preliminary results on outcome measures, including speech intelligibility, retrieval accuracy, and reaction times. Listening effort and memory load—indexed by pupil dilation—and subjective ratings of task load were also assessed. In order to examine the influence of individual factors, several cognitive functions (processing speed, cognitive flexibility, focused attention, and working memory) and participant age were considered.

We hypothesize that increasing task difficulty—from the zero-back to the one-back task—will elevate working memory load, resulting in lower retrieval accuracy, longer reaction times, and greater pupil dilation, and that these outcomes will be associated with individual factors. Increasing task difficulty is expected to have a more detrimental effect on CI users compared with their TH peers.

P38 Evaluating phonetic distance metrics as predictors of speech intelligibility in noise

Maya Hola, Paul Iverson
University College London, United Kingdom

Measuring phonetic distance between speakers is essential for understanding speech intelligibility in noise, yet methods for quantifying these differences vary widely. Recent advances in AI-based speech representations offer new approaches to capture talker-listener similarity, but their effectiveness relative to traditional acoustic metrics remains unclear. We compared four types of phonetic distance metrics—legacy acoustic measure (ACCDIST), automated edit distances, embedding-based measures from self-supervised models, and novel logit-based measures—to determine which best predicts intelligibility within a homogeneous accent community. Forty standard southern British English (modern RP) speakers participated as both talkers and listeners in a sentence recognition task across four signal-to-noise ratios. Talker typicality—measured as average distance from a talker to all other talkers—strongly predicted intelligibility, with individual talker-listener distance providing additional explanatory power. Embedding-based metrics from an English phoneme recognizer optimally captured the phonetic variability relevant to intelligibility, with logit-based measures offering a computationally efficient alternative. These findings demonstrate that fine-grained phonetic variation within a single accent group systematically affects comprehension in adverse conditions, extending accent-distance theory to the idiolectal level and providing practical guidance for selecting distance metrics in speech intelligibility research.

P39 What are the links between impaired speech-in-noise intelligibility and hyperacusis?

Vincent Isnard
Université Paris Cité, Institut Pasteur, AP-HP, INSERM, Fondation Pour l'Audition, Institut de Recherche Biomédicale des Armées, Département Neurosciences et Sciences Cognitives, Unité Perception, IHU reConnect, Brétigny-sur-Orge, France

Hyperacusis, or the "discomfort for sounds that would be acceptable to most normally hearing people," is a complex hearing disorder whose pathophysiological mechanisms are often compared to those of tinnitus. However, some recent studies also suggest, albeit tentatively, a specific link with impaired speech-in-noise intelligibility. From a certain perspective, the discomfort associated with hyperacusis would interfere with speech processing, particularly in noisy environments. Nevertheless, hyperacusis and speech-in-noise intelligibility could share a common origin in participants overexposed to noise, manifesting as various forms of hidden hearing loss. In a recent study, we highlighted a strong link between speech-in-noise intelligibility, assessed by questionnaire and a logatome intelligibility task, and the degree of hyperacusis, assessed by questionnaire and visual analog scales (Fernandez & Isnard, 2025, doi:10.1016/j.heares.2025.109436). Hyperacusis participants also exhibited more hearing loss on the extended high-frequency audiogram. This new data supports the hypothesis that hyperacusis is linked to peripheral hearing loss, impairing the rapid temporal processing necessary for speech-in-noise intelligibility. Therefore, here, we propose to compare these data with the hypotheses of various studies highlighting such a link between speech-in-noise intelligibility and hyperacusis. Based on current knowledge, these analyses allow us to: (1) suggest making a clearer distinction between different types of hyperacusis, of peripheral or central origin; (2) support the recommendation for a more systematic joint assessment of speech-in-noise intelligibility and hyperacusis, including by rapid screening, to corroborate a diagnosis of hidden hearing loss and associated disorders.

P40 Effect of frequency-to-place factors on perception and adaptation in cochlear implant users

Louis Villejoubert[1], Lorenzo Picinali[2], Kathleen Faulkner[3], Deborah Vickers[1]
1. Sound Lab, Cambridge Hearing Group, Department of Clinical Neuroscience, University of Cambridge, United Kingdom | 2. Dyson School of Design Engineering, Imperial College London, United-Kingdom | 3. Oticon A/S, Smorum, Denmark

Sound quality in cochlear implants (CIs) remains a major challenge despite substantial progress in speech understanding. Many CI users achieve good speech recognition yet describe sounds as unnatural or distorted. A certain degree of this degradation may arise from “frequency-to-place factors” that affect how sound information is distributed along the cochlea and delivered to the auditory system - namely frequency-to-place mismatch (FTPM), neural responsiveness and current spread. Understanding these factors is crucial to optimise CI mapping and improve both sound quality and music perception. We developed two complementary psychoacoustic tasks to assess how these factors shape perception. (1) Vowel identification: Participants identified vowels generated from controlled combinations of the first and second formants, designed to stimulate electrodes independently. Each response included a confidence rating, enabling estimation of internal vowel maps and potential FTPM-related shifts. (2) Chord quality: Participants compared the sound quality of chords and their inversions using paired-comparison ratings, allowing assessment of electrode-specific degradation in complex harmonic sounds. Sixteen adult CI users (12 post-lingual, 4 pre-lingual) and normal-hearing controls were tested. One post-lingually deafened user was also followed longitudinally from switch-on to assess early adaptation. Individual X-ray and PECAP data provided physiological estimates of frequency-to-place alignment, current spread, and neural responsiveness. Preliminary results show high test–retest reliability. Several CI users preferred slightly shifted vowels, suggesting adaptation to FTPM. Chord comparisons revealed large electrode-specific variability, reflecting ENI differences. Together, these methods offer individualised insights into CI perception and hold promise for guiding clinical fitting to enhance naturalness and musical enjoyment.

P41 Reconstructing speech from EEG with Large Language Models

Louis Villejoubert, Deborah Vickers
Sound Lab, Cambridge Hearing Group, Department of Clinical Neuroscience, University of Cambridge, United Kingdom

Understanding how the brain encodes and reconstructs sound is central to hearing science and clinical audiology. Traditional EEG decoding methods, such as the envelope following response, have revealed important aspects of neural speech tracking but remain limited by their linear assumptions and restricted sound representations. These approaches capture only part of the complex, non-linear neural dynamics that support natural sound perception. Here, we introduce a new framework that applies a Large Language Model (LLM) architecture to decode EEG signals and reconstruct the corresponding sound. We adapted a pre-trained NeuroLM model - derived from GPT-2 and trained on large-scale EEG datasets -to predict neural representations of acoustic tokens. EEG signals from the SparrKULee dataset (105 normal-hearing participants, 64-channel recordings during audiobook and podcast listening) were converted into discrete “EEG tokens,” while the associated stimuli were tokenized using an EnCodec model at bitrates from 1.5 to 6 kbps. NeuroLM variants (254 M to1.7 B parameters) were fine-tuned to map EEG sequences to sound tokens, enabling continuous audio reconstruction from neural activity. Reconstructed signals were assessed using objective acoustic measures (mel-spectrogram correlation, PESQ, phase coherence) and subjective tests (intelligibility and MUSHRA ratings). The best-performing models produced intelligible and perceptually recognisable reconstructions, even for participants unseen during training. Beyond its methodological innovation, this work highlights the potential of LLM-based neural decoding as a clinical research tool. This approach could help evaluate and refine hearing technologies and investigate how individual factors - such as electrode placement or neural health - shape cortical representations of speech in cochlear implant (CI) users. It may also reveal how CI users compensate for degraded information or support EEG-based diagnostics, advancing understanding of auditory perception in both typical and impaired hearing.

P42 A differential digits-in-noise paradigm for assessing binaural and temporal masking release across age groups

Loes Desmedt[1,2], Nicolas Verhaert[1,2], Astrid van Wieringen[1], Jan Wouters[1], **Sam Denys**[1,2]
1. University of Leuven, department of Neurosciences, Research group Experimental Otorhinolaryngology (ExpORL), Leuven, Belgium | 2. University Hospitals of Leuven, Department of Otorhinolaryngology – Head and Neck Surgery, Leuven, Belgium

Binaural and temporal processing are fundamental to understanding speech in noisy environments, yet clinically accessible tools to assess these abilities—especially in children—remain limited. This study presents the development and validation of a differential Dutch-Flemish Digits-in-Noise (DIN) test paradigm aimed at quantifying binaural unmasking and modulation masking release effects using low-linguistic-load stimuli presented in noise. The test comprises multiple listening conditions that manipulate binaural phase information and masker modulation to isolate distinct auditory processing mechanisms.

Three studies were conducted: (1) a psychometric evaluation of the perceptual equivalence of the speech items in normal-hearing (NH) adults (N = 12); (2) the establishment of reference values and test-retest reliability in another NH adult cohort (N = 24); and (3) an evaluation in typically developing NH children aged 6–12 years (N = 34), comparing self-administered versus administrator-controlled test modes.

Results show that, owing to iterative perceptual equalization efforts, the test yields reliable speech reception thresholds (SRTs) across conditions, with consistent effects reflecting expected unmasking mechanisms. In children, SRTs improved with age, with difference measures (quantifying unmasking) being less sensitive to age-related variability. No significant performance differences were observed between administration modes.

These findings support the feasibility of the differential DIN paradigm for assessing binaural and temporal auditory processing in both research and clinical settings. By capturing key auditory unmasking mechanisms with minimal language demands, the test offers a promising tool for auditory profiling.

This study was funded by the Flemish Research Foundation (TBM-grant T000823N) and by the Clinical Research and Education Council of the University Hospital of Leuven (internal grant for academic studies).

P43 Speech-in-speech load triggers inattentional deafness

Franck Élisabeth[1], Guillaume Andéol[1], Véronique Chastres[1], Élodie Bayle[1], Isabelle Viaud-Delmon[2], **Clara Suied**[1]
1. Institut de Recherche Biomédicale des Armées, Brétigny sur Orge, France | 2. CNRS, IRCAM, Sorbonne Université, Ministère de la Culture, Sciences et Technologies de la Musique et du son, STMS, Paris, France

In multitalker scenes, tracking one speech stream while suppressing others taxes selective attention and working memory. We argue that this speech in speech load alone is sufficient to induce inattentional deafness—a failure to notice clearly audible, task irrelevant sounds—even when those sounds are well above threshold. Here, we developed a new paradigm, that can be run online, to study inattentional deafness in auditory-only conditions. We combined a multi speaker Coordinate Response Measure (CRM) task with a N-back task to keep listeners continuously engaged with target speech while competing sentences played to the opposite ear. On the final trial, an unexpected, non speech critical sound was presented at the same level as the sentences; awareness was probed only at the end of the experiment to capture inattentional deafness (one-shot paradigm). Listeners performed the speech task above chance across N levels, confirming engagement, yet still missed the critical sound on the majority of trials (on average 78% misses). Spatial attention modulated detection, consistent with the idea that resource allocation to the attended stream shapes what gets through awareness from elsewhere in the scene. Overall, the pattern demonstrates that speech competition creates sufficient cognitive load to generate inattentional deafness.

P44 AI-based real-time speaker separation: An alternative approach to beamforming.

Tom Berner[1], **Niki Katerina Vavatzanidis**[1], Tobias Reichenbach[2], Anja Hahne[1]
1. Ear Research Center Dresden, Technische Universität Dresden, Germany | 2. Chair of Sensory Neuroengineering, Friedrich-Alexander-University Erlangen-Nürnberg, Germany

Research Question: The so-called cocktail party problem describes the limited ability of hearing aid and cochlear implant users to follow a target speaker in noisy multi-talker environments. Classical approaches such as beamforming improve speech intelligibility using multiple microphones to extract and enhance the signal of a single speaker. The goal of this study was to develop an AI-based real-time speaker separation method that achieves good results using only one microphone.

Methods: A web-based system for blind source separation (BSS) was implemented, processing continuous audio data in 0.5-second windows. The separation is achieved using neural AI models (TDANet, TIGER) within a FastAPI/OpenVINO framework. The AI models, pre-trained on English data, were optimized for real-time application and applied to German audio data. Real-time processing was successfully implemented at a sampling rate of 16 kHz and a window length of 0.5 seconds. A WebAudio frontend with AudioWorklets handles streaming, playback, and visualization in real time - completely within the browser and without any special hardware requirements.

Results: The system achieved a latency below 300 ms on standard laptops (without GPU) and operated continuously without dropouts. Subjective reports from test participants indicated a clear separation of speaker voices and an improved intelligibility in overlapping speech. All signal processing was performed stably via the WebSocket stream, and visual feedback (waveform/STFT) enabled immediate assessment of separation performance in real time.

Conclusions: The combination of efficient model optimization and web-based architecture enables AI-based real-time speaker separation using only a single microphone. This concept shows potential for future intelligent hearing aids and cochlear implants that could perform AI-based separation directly on the user device.

P45 Intelligibility benefits persist for at least four weeks after a single session of voice identification training

Harriet J. Smith, Emma Holmes
Department of Speech Hearing and Phonetic Sciences, University College London (UCL), UK

Background: Listeners typically show a 10-15% improvement in speech-in-speech intelligibility when target speech is spoken by someone familiar, compared with someone unfamiliar—both for naturally familiar and for lab-trained voices. This improvement could have real-world benefits, if listeners were trained on voices they would later encounter in a noisy natural setting (e.g., public transport). However, this application depends on intelligibility improvements enduring over a prolonged period, which has not been tested systematically. This study examined whether, after one session of online training, participants showed a speech-in-speech intelligibility benefit for trained voices one day, one week, and four weeks after the initial training session.

Methods: 40 participants were trained to recognise three novel voices. During training, participants listened to each voice produce 166 natural, meaningful sentences (approximately 70 minutes in total). A speech-in-speech intelligibility test was completed immediately after training, one day later, seven days later, and 28 days later. On each trial of the intelligibility test, two voices simultaneously spoke different closed-set sentences, at two possible target-to-masker ratios (+3 and -6 dB TMR). Participants reported the words spoken by the target voice, which was either one of the three trained voices, or an unfamiliar voice. The exact sentences and unfamiliar voices differed across the four test sessions. All tasks were completed remotely online.

Results: Immediately post-training, participants showed an intelligibility benefit of ~13% for trained voices over unfamiliar voices. Critically, this familiar-voice benefit was significant at all timepoints, and the size of the benefit did not significantly change over time. Furthermore, these effects were present at both target-to-masker ratios. Overall intelligibility performance did not differ between the test sessions.

Discussion: Our results show that a single training session produces a substantial speech-in-speech intelligibility benefit for learned voices, which persists for at least four weeks without diminishing. Furthermore, this long-lasting benefit emerges after only 70 minutes of online training on three novel voices and is robust across varying levels of masking speech. These findings indicate that representations of new voices are formed rapidly yet remain stable over time. Our results also suggest that voice familiarisation training could be an effective practical tool for improving intelligibility, as a particular speaker's voice could be learned several weeks before it is encountered in the real world. In summary, this study demonstrates that brief voice training can produce long-term and practically meaningful improvements in speech-in-noise perception.

P46 A blind real-time capable binaural model for estimating subjectively perceived listening effort of normal-hearing and hearing-impaired listeners

Martin Berdau, Thomas Brand
Carl von Ossietzky University Oldenburg, Germany

Jan Rennies
Fraunhofer Institute for Digital Media Technology Oldenburg, Germany

Listening effort (LE) has become an established measure to assess communication situations that can provide more information for the perception of speech in situations, where speech intelligibility (SI) might be at ceiling. LE can be assessed using objective measurements as well as subjective ratings. LE and SI can be measured or assessed in hearing experiments and also predicted by models. Most models run offline on entire signals and may require clean source signals. Nevertheless, there can be cases where very frequent predictions might be required, such as in real-time applications. For example, a real-time model could be used in a hearing aid to automatically select the best algorithm at runtime to minimize induced LE. Usually, such applications have no source signals at hand, requiring them to work "blindly" on the ear signals only.

We propose a real-time, blind LE prediction model that uses block processing and accounts for binaural capabilities and hearing loss. The model consists of two stages: a binaural front-end and a monaural back-end. The front-end takes the ear signals as inputs and simulates hearing loss by adding threshold simulating noise. It then simulates spatial release of masking by simultaneously modeling binaural masking, using an Equalization Cancellation approach, and better ear listening. Thus, a binaurally enhanced single-channel signal is produced, which is then routed into the monaural back-end. The back-end uses a triphone classifier to make prediction on the subjectively perceived LE. The classifier outputs successive posterior probabilities for the set of known triphones. LE is then predicted by calculating a similarity measure over the course of consecutive posterior probabilities. Thus, lower similarity indicates lower LE, whereas greater similarity - potentially introduced by temporal smearing, e.g., through reverberation - indicates higher predicted LE.

The model was evaluated on data of subjectively perceived LE, measured using a novel real-time assessment method with normal-hearing listeners. It accurately predicted continuous, subjectively perceived LE ($R^2 = 0.86$) under conditions with varying signal-to-noise ratios (SNRs) and reverberation. Further evaluation of the model was conducted using stationary and modulated noise.

P47 Assessing Digits-in-Noise performance based on second-language proficiency and hearing acuity

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 measuring speech reception thresholds (SRTs) using digit triplets. While DIN tests are available in multiple languages, studies on the effects of second-language (L2) proficiency, particularly combined with hearing loss, on DIN outcomes are limited. This study compares Turkish (L1) and Dutch (L2) DIN test outcomes in native Turkish speakers of a wide age range (19-68 years, mean age: 31.43 ± 15.0 years) with varying Dutch proficiency and hearing levels, to investigate these effects.

Methods: Forty native Turkish individuals (self-reported gender: 22 female; 18 male) were included. Most had normal hearing (n=36), with four having mild to moderate hearing loss. Participants' Dutch proficiency ranged from A0 to C2 (CEFR). All 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: diotic or antiphasic; and (2) Speaker: Dutch male, Turkish male (S01), or Turkish female (S02). For all six DIN tests, speech level was adapted and starting SNR was -6 dB.

Results and discussion: Preliminary data from 40 participants indicate several observable trends. Consistent with previous studies with Dutch DIN, antiphasic sound presentations appear to result in lower (better) DIN SRTs compared to diotic across both Turkish speakers (S01 and S02). Observable differences also exist between the two Turkish speakers, with participants listening to S02 tending towards slightly better antiphasic DIN SRTs and for S01 towards slightly better diotic DIN SRTs. Participants' Turkish (L1) DIN SRTs seem better than Dutch (L2) DIN SRTs. Regarding the effect of L2, participants with higher CEFR levels tend to show better performance on the Dutch DIN (both diotic and antiphasic). An interesting trend observed is that higher Dutch proficiency also appears related to slightly better performance on some Turkish DIN tests (S01 diotic and antiphasic, S02 diotic). Increased hearing loss appears associated with poorer performance on the Turkish DIN tests. However, a clear relationship between degree of hearing loss and Dutch DIN performance seems less apparent. Turkish T-CCRM scores show a tendency to be poorer for participants with worse L1 Turkish DIN scores, but a similar relationship with L2 Dutch DIN scores is not evident. Data collection is still ongoing, focusing on recruiting particularly more participants with hearing loss. Updated results will be presented at the 16th Speech in Noise (SpiN) Workshop.

P48 Potential of deep neural networks for hearing loss compensation and noise reduction: In search of the best configuration

Szymon Drgas
Institute of Automatic Control and Robotics, Poznan University of Technology, Poland

Lars Bramsløw
Eriksholm Research Centre, Denmark

Recent applications of auditory models with hearing loss can be '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, thereby obtaining a new, unrestricted (and unknown) hearing loss compensation (HLC). The potential of such an approach has already been demonstrated by Leer et al, 2024 (doi:10.48550/arXiv.2403.10428), using a physiological auditory nerve model and Drgas & Bramsløw, SPIN2025 (doi:10.5281/zenodo.15433865), using a simpler loudness model.

Neural-based Noise Reduction (NR) can also be added, by adding noise only to the hearing-impaired branch. Different combinations of standard and neural HLC+NR can be combined into one (joint) or two (separate) deep neural networks (DNN). Likewise, different loss functions and DNN architecture can be applied, which are both critical for the result.

The aim of the work was to test the early application of a loudness model (AUDMOD: Bramsløw, SPIN2024, doi:10.5281/zenodo.10473561) for DNN-based HLC and noise reduction in different configurations: 1) Neural NR + Neural HLC, 2) Neural NR + NALNL2 HLC, 3) Joint HLC + NR, 4) No NR + Neural HLC, 5) NALNL2 HLC (standard reference). A selection of cost functions has been applied during training.

For HLC, the resulting signals were evaluated and compared to National Acoustics Laboratory nonlinear (NAL-NL2) prescription using spectral measurements and relevant objective metrics, such as STOI and the hearing-aid speech perception index (HASPI). Total loudness and loudness pattern comparisons are also presented.

P49 Neural encoding and behavioral discrimination of vowels in subjects with tinnitus

Pauline Devolder[1,2], Hannah Keppler[2,3], Sarah Verhulst[1]
1. Hearing Technology @ WAVES, Department of Information Technology, Ghent University, Zwijnaarde, Belgium | 2. Department of Rehabilitation Sciences, Ghent University, Ghent, Belgium | 3. Department of Otorhinolaryngology, Head and Neck Surgery, Ghent University Hospital, Ghent, Belgium

Background: Tinnitus has been associated with altered neural gain mechanisms in the central auditory system, potentially due to reduced peripheral input. Although neural overactivity and gain modulation are well-documented in animal models, their impact on speech-sound processing and perceptual outcomes in humans remains less clear. Investigating both the neural encoding and behavioral discrimination of speech-relevant sounds in listeners with tinnitus can provide insights into how central gain affects everyday auditory function.

Methods: The study included data from 120 participants evenly distributed across six groups: young normal-hearing individuals with and without tinnitus (18–35 years), older normal-hearing individuals with and without tinnitus (45–65 years), and older hearing-impaired individuals with and without tinnitus (45–65 years). The test protocol includes tonal audiometry up to 16 kHz, distortion product otoacoustic emissions, Flemish Matrix sentence test in quiet and in noise, a behavioral psychoacoustic vowel discrimination task (/o/-/u/) in quiet and in noise, frequency following response measurements using vowel /u/ and questionnaires on tinnitus and hyperacusis (TFI, THI, HQ).

Results: While audiometry and DPOAE's showed no significant differences between the groups, we observed that tinnitus-groups scored overall better on the behavioral vowel discrimination task compared to the control groups. Correspondingly, the young tinnitus group showed stronger neural encoding of the vowel envelope component in the FFR, while no significant group differences emerged for the temporal fine-structure response or in the older cohorts.

Conclusions: These results suggest that tinnitus may be associated with enhanced central encoding of slow temporal envelope cues and improved perceptual discrimination of vowels, consistent with compensatory central gain mechanisms.

Funding: Work supported by FWO G068621N AuDiMod.

P50 How does knowledge of who, what, and where influence speech-in-speech perception?

Elin Bonyadi, Harriet J. Smith, Emma Holmes
University College London, London, UK

Background: Speech-in-speech perception is challenging, but can be facilitated by prior knowledge. For example, advance cueing of the semantic context ("what") and spatial location ("where") of an upcoming target sentence each improves speech-in-speech intelligibility. Additionally, speech-in-speech intelligibility is better when the target sentence is spoken by a familiar rather than unfamiliar voice (known as the familiar-voice benefit), although explicit cueing of talker identity ("who") has not been tested. Here, we aimed to compare how advance cueing of talker identity, semantic context, and spatial location affects speech-in-speech perception.

Experiment 1: We recruited 24 participants without hearing loss, aged 18–45 years. Participants were first trained to identify the voices of three male talkers (for 166 sentences each). Then, in the speech-in-speech intelligibility task, participants heard two concurrent sentences (target and masker) amongst background noise on each trial, and repeated the target sentence aloud. The two concurrent sentences were spoken by different (male) talkers, had different topics, and were spatially separated. Before the sentences began, participants were visually cued to the talker identity (e.g., "John"), topic (e.g., "Animals"), or spatial location (e.g., "Left") of the upcoming target sentence, or saw an uninformative visual cue (baseline condition).

We found that semantic and spatial cues produced significantly better intelligibility than uninformative cues, with no significant difference in intelligibility between semantic and spatial cues. In contrast, talker-identity cues produced no significant benefit compared to uninformative cues.

Experiment 2: To test whether there is a talker-cue benefit to intelligibility under certain conditions, Experiment 2 focused on the talker- and uninformative-cue conditions. Twenty naïve participants were trained to identify three male talkers, but were only cued to one talker (trained for 166 trials) during the speech-in-speech intelligibility task.

We found a significant familiar-target intelligibility benefit for the trained voice compared to novel voices. However, participants gained no additional benefit from receiving a cue that signalled the identity of the upcoming target talker compared to an uninformative cue.

Conclusions: Our findings suggest that advance cues about semantic context and spatial location improve speech-in-speech intelligibility. Whereas, across two experiments, we found no evidence that talker-identity cues improve intelligibility, even when participants are sufficiently familiar with a voice to gain a familiar-voice benefit over novel voices.

P51 Profiling listening difficulties in children with language development concerns based on caregiver report

Hanne Falcone[1], Loes Desmedt[1,2], Hannah Keppler[3,4], Nicolas Verhaert[1,2], Astrid van Wieringen[1], Sam Denys[1,2]

1. KU Leuven, Department of Neurosciences, ExpORL, Leuven, Belgium | 2. University Hospitals of Leuven, Ear nose and throat disease, head and neck surgery, Leuven, Belgium | 3. Ghent University, Department of Rehabilitation Sciences - Audiology, Ghent, Belgium | 4. Ghent University Hospital, Department of Otorhinolaryngology, Head and Neck Surgery, Ghent, Belgium

Background: Listening difficulty (LiD) is the experience of struggling to understand speech, especially in noisy or complex listening environments, even when hearing thresholds are normal. Children with language deficits may experience LiD, resulting from interactions between auditory, cognitive, and language processing. The Evaluation of Children's Listening and Processing Skills (ECLiPS) is a validated caregiver-report measure designed to identify and understand listening and processing skills. Developing listening profiles using the ECLiPS may provide a better understanding of children's functional listening abilities and support more targeted diagnostic pathways.

Objectives: This study examines listening profiles of children with language development concerns (LDC) using the Flemish version of the ECLiPS (ECLiPS-FL). We hypothesize that these children primarily show weaknesses on the Language/Literacy/Laterality (L/L/L) scale, reflecting language-based roots of LiD. Additionally, we investigate associations between ECLiPS and standardized language tests, and whether language ability can be predicted from ECLiPS scores.

Methods: The study included 45 children aged 6-11 years (27 boys, 18 girls), referred for LDC, and 90 age-matched typically developing peers (57 boys, 33 girls). Children with hearing loss, intellectual disabilities, or limited language exposure were excluded. One caregiver for every child completed the ECLiPS-FL, providing composite scores on five subscales: Speech and Auditory Processing (SAP), Memory and Attention (M&A), Language/Literacy/Laterality (L/L/L), Pragmatic and Social Skills (PSS), and Environmental and Auditory Sensitivity (EAS). The children completed standardized language tests. For the current analyses, we focused on the Clinical Evaluation of Language Fundamentals – Fifth Edition (CELF-5), using the Core Language Score as a measure of overall language ability. Subsequent analyses examined group-level patterns and the association between ECLiPS and language scores.

Results: Compared to the reference data, children with LDC scored significantly lower across all ECLiPS subscales, with the lowest scores and the largest deviation from the normative group on the L/L/L scale. Despite this, the M&A scale was the strongest predictor of language performance on the CELF-5.

Conclusions: LiD are common in children with LDC, and mainly related to their language difficulties, as reported by caregivers. Quantitatively, scores on the M&A scale are most predictive of language ability as measured with the CELF-5. These findings point to a discrepancy: caregivers tend to notice overt language difficulties, while underlying mechanisms of LDC may primarily reflect cognitive factors related to memory and attention.

P52 Predictors of speech-in-noise understanding in a population of occupationally noise-exposed individuals

Guillaume Andéol
Institut de recherche biomédicale des armées, Brétigny-sur-Orge, France

In occupational environments, communication failures can increase the risk of accidents. Workers exposed to noisy environments often experience difficulties understanding speech in background noise. These difficulties may stem not only from the masking effects of noise but also from auditory damage resulting from chronic noise exposure. Current audiological assessments used in occupational health are insufficient for both evaluating speech-in-noise comprehension and monitoring auditory status over time. The objective of this study was to examine the relationships between various factors potentially influencing speech-in-noise understanding and to identify the most relevant predictors. Hearing thresholds at 12,500 Hz—a frequency beyond the range of conventional audiometry—showed a strong association with speech-in-noise performance. Routine monitoring of extended high-frequency hearing could therefore allow early identification of individuals at risk and implementation of preventive or rehabilitative measures before critical auditory decline occurs.

P53 Myelination and GABAergic maturation supporting speech-in-noise processing across pubertal development

Marta Puertollano[1,2], Joseph Francois Johnson[1,2], Francesca Cavicchiolo[1,2], Luna Prud'homme[1,2], Antonin Rovai[1,3], Simon Dobri[4], Xavier De Tiège[1,2,3], Axelle Calcus[1,2]

1. UNI – ULB Neuroscience Institute, H.U.B. – ULB University Hospital, Université libre de Bruxelles (ULB), Belgium | 2. CRCN – Center for Research in Cognition and Neuroscience, Université libre de Bruxelles (ULB), Belgium | 3. Laboratoire de Neuroanatomie et Neuroimagerie translationnelles, H.U.B., Université libre de Bruxelles (ULB), Belgium | 4. Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, Canada

Adolescence represents an extended period of neuroplasticity that promotes experience-driven adaptations across brain regions and networks supporting the cognitive and behavioural capacities required in dynamic social environments. During this developmental phase, adolescents increasingly engage with larger peer groups. However, their ability to parse concurrent streams and track a target voice in the presence of interfering speech keeps maturing well into adolescence. Such skill relies on a network of functionally and structurally connected regions including prefrontal cognitive control, parietal cross-modal association, and the fine-tuning of temporal auditory voice perception. Adolescence begins with puberty onset, which is associated with changes in sex-steroid hormones, some of which are known to contribute to the maturation of brain regions involved in speech-in-noise processing (SiN). As part of the SensationaHL pubertal development cohort collection, we aim to investigate the neuroplastic mechanisms and myelination trajectories linked to pubertal stages and hormonal markers, focusing on brain regions serving complex cognitive and auditory skills. We hypothesize that myelination trajectories will correlate with pubertal stage progression, extending past puberty-offset. We also predict puberty-associated increases in gamma-aminobutyric acid (GABA) concentrations, a key modulator of neuroplasticity, in SiN regions of interest: the left pars triangularis and left Heschl's gyrus. To address these aims, a cohort of 150 participants will be recruited and reassessed two years later, measuring pubertal stage transition surrounding puberty onset and offset. Pubertal hormone levels and stage assessments will be collected at the two timepoints, in addition to cognitive and auditory speech-perception tasks. Multimodal neuroimaging including single-voxel spectroscopy, diffusion imaging, and high-density electroencephalography is used at baseline and will be used again after two years to evaluate neurotransmitters concentrations, structural and functional connectivity and thus infer upon the mechanisms of neuroplasticity involved. Cross-sectional MRI preliminary results from approximately 70 participants (ages: 9-18 years; Early to Post-pubertal) collected to date will be presented in this poster. So far, preliminary findings suggest an increase in neurite density, decrease in orientation dispersion as well as maintained GABA concentration values within the targeted regions across age and pubertal status. These findings could allow to identify the neurobiological triggers of heightened auditory plasticity during adolescence and advance our understanding of how puberty shapes cognitive and auditory development in challenging environments.

P54 Development of auditory scene analysis and speech-in-noise perception at adolescence

Luna Prud'homme, Elena Benocci, Axelle Calcus
Université Libre de Bruxelles (ULB), Belgium

Understanding speech in background noise is a complex task, and children persistently perform worse than adults in such situations. When the competing noise is composed of interfering talkers (informational masking), this ability does not reach mature levels until late adolescence (around 16 years old). Speech perception in noise (SIN) is thought to be influenced by auditory scene analysis (ASA; the ability to segregate and track a sound stream while ignoring others), but also by cognitive abilities. The first aim of this pre-registered study is to investigate the development of SIN and ASA at adolescence. In adults, ASA has been shown to be a predictor of SIN. Yet the interplay between cognitive, auditory factors and speech perception in noise is likely different in development – especially given the protracted maturation of executive functions until late adolescence. Therefore, the second aim of this study is to determine whether and how the interplay between cognitive abilities, ASA and speech perception in noise develops from childhood to adulthood. Participants, aged 9 to 23 years old, were presented with different tasks to measure performance in SIN (energetic and informational masking), auditory scene analysis (stream segregation and attentive tracking), attention and inhibition, working memory and musical abilities. A power analysis indicated that a total sample of 200 participants was necessary to analyze the results using structural equation models. Data collection is ongoing. So far, n = 160 individuals participated in the study. Preliminary results suggest that both SIN and ASA improve until (late) adolescence. The relationship between ASA and speech in noise performance appears to also change during adolescence. The relationship between attentive tracking and speech perception in noise appears stronger in children and adolescents than adults. The opposite seems to be true for the relationship between stream segregation and speech perception in noise. Results will be discussed in line with developmental models of executive functions and their implications for the development of complex auditory processing at adolescence.

P56 Social motivation shapes listening effort in autism: evidence from an audiobook-in-noise paradigm

Ryan J Oakeson[1,2], Allison C Nugent[2], Sarah White[1], Sophie Scott[1]
1. University College London, UK | 2. National Institute of Mental Health, USA

Little is known regarding the listening experiences of autistic individuals during naturalistic speech-in-noise paradigms. Common traits that lead to a diagnosis of autism often lie within the domain of social communication challenges and altered reactivity to sensory stimuli—which together may result in autistic individuals struggling during social environments, particularly noisy ones. Autistic individuals have been shown to demonstrate variations in cognitive abilities such as theory of mind, working memory, and attention that have been believed to lead to the noted social communication challenges. However, according to social motivation theories, autistic individuals may lack a specific social motivation to engage in social interactions, which subsequently leads to an inexperience that results in weaker developed social-cognitive traits. Therefore, we hypothesized that motivation in the form of prosociality might be associated to the amount of effort an individual would exert during a social interaction in a noisy environment; specifically, more prosocial individuals might be exerting more effort. To address the various factors that affect listening performance and effort during noisy environments, we recruited a sample of autistic individuals (N = 27) from the Greater London Area with an age and IQ matched control group (N = 22). The methods implemented included behavioral tasks probing speech-perception-in-noise thresholds, working memory, musical perception skills, and social orientation via questionnaires. We then had participants listen to an audiobook-in-noise whilst an eye tracker recorded the evoked pupil responses during ideal and adverse listening conditions. Our results demonstrated several key findings. During the audiobook-in-noise, the autistic group (N = 13) demonstrated an increasing amount of effort and arousal throughout the duration of the audiobook in all conditions, whereas the control group (N = 11) only showed increasing arousal. Linear mixed models found that the key factors influencing mean evoked pupil size and variance in the autistic group were the speech condition, the occurrence of the condition, age, working memory, autistic traits, and prosociality. In the control group, age was the only factor having a significant effect on effort and arousal during the audiobook task. The findings here demonstrate that autistic individuals may be utilizing a dynamic combination of cognitive and social traits to help their engagement during a naturalistic speech-in-noise paradigm. Moreover, prosociality was strongly anticorrelated to working memory, thus there may be a complex tradeoff between social traits and cognition underlying listening effort in autism that studies should explore in the future.

P57 Examining the benefit of spatial, voice, and visual cues for speech recognition in a competing-talkers paradigm

Hartmut Meister[1], Denise Gradtke[1], Pascale Sandmann[2], Khaled Abdel-Latif[1]
1. Faculty of Medicine and University Hospital Cologne, Department of Otorhinolaryngology, Head and Neck Surgery, University of Cologne, Germany | 2. Department of Otolaryngology, Head and Neck Surgery, Carl-von-Ossietzky University of Oldenburg, Germany

In situations with competing talkers, three main mechanisms are beneficial for separating the different auditory streams and extracting the speech of the intended talker: Spatial cues based on interaural time and level differences help determine the position of the talkers, voice cues help distinguish between talkers and visual cues enable to identify the talker and extract speech features over and above those submitted by the auditory channel. However, individuals with hearing loss and particularly cochlear implant (CI) recipients may be limited in the use of the auditory cues, but may have enhanced abilities to exploit visual speech cues.

This ongoing study aims to investigate the relative contribution of these cues in groups of typical-hearing (TH) listeners and cochlear implant users in terms of speech recognition and cognitive load. To this end, a female target talker uttering matrix sentences (name-verb-number-adjective-object) is masked with two competing talkers, who can either have the same or a different voice and can either be spatially separated by ±30° from the target or colocated. In addition, these conditions are presented with and without additional visual speech cues, provided by a virtual character that enables the visualization of arbitrary speech materials. The target sentence always begins with the name “Stephen”, and the target-to-masker ratio is adaptively adjusted to achieve 75% correct word recognition. During stimulus presentation, gaze is tracked and pupil dilation is measured as a proxy of cognitive load.

It is hypothesized that the two study groups would exhibit different patterns: while the TH listeners primarily use auditory cues, CI recipients are expected to rely more heavily on visual cues and to be particularly limited regarding the benefit of voice cues. The present contribution outlines the study rationale and design and presents preliminary results for both TH listeners and CI recipients.

Funding: Supported by Deutsche Forschungsgemeinschaft (DFG), reference ME 2751/4-1

P58 An experimental paradigm for testing context-aware closed-loop speech enhancement systems

Jens Hjortkjær, Tobias Dorszewski
Technical University of Denmark

Context-aware hearing instruments leverage user-related signals, such as EEG, gaze, or first-person video, to infer which speech sources the listener wants to hear. This information can be used to steer speech separation algorithms, selectively enhancing target speech streams while suppressing background noise. Yet, experimental paradigms for evaluating such closed-loop systems are still missing. Here, we present a behavioral audio-visual speech test paradigm designed to quantify improvements in speech intelligibility achieved by closed-loop systems in competing conversation scenarios. The speech material comprises spoken numbers concatenated to form speech streams with naturalistic speech statistics, each paired with AI-generated talking faces. Listeners are presented with four speech streams grouped into two simultaneous dyadic conversations. Within each conversation, talkers alternate between number 'sentences' and backchannel yes/no responses, all presented in background noise. Participants are instructed to attend to specific conversations and detect repeated numbers and backchannel keywords. This test paradigm enables evaluation of closed-loop systems that selectively enhance the speech from attended conversational groups rather than isolated speakers. Here, the paradigm is used to demonstrate a gaze-steered enhancement system that integrates gaze direction across time to decode and selectively enhance the currently attended conversation. We show that the paradigm is sensitive to differences in keyword detection performance between conversation-based enhancement, speaker-based enhancement, and no enhancement.

P59 A method for measuring noise-induced voice adaptations and their effects on speech intelligibility and cognitive effort during conversations

Paolo A. Mesiano[1], Outi Tuomainen[2], Lorenz Fiedler[1], Johannes Zaar[1]
1. Eriksholm Research Centre, Snekkersten, Denmark | 2. University of Potsdam, Germany

During a conversation in presence of acoustic or auditory barriers (e.g., background noise or hearing deficits in one of the interlocutors), talkers adapt their speech production to enhance speech clarity, i.e., to produce "clear speech". Although the literature about clear speech is substantial, its effects on speech intelligibility and cognitive effort are underexplored, mainly because of the lack or inadequacy of the available testing methods. Crucially, standard speech intelligibility testing protocols utilize short target sentences, typically pre-recorded in quiet, and therefore do not allow assessing intelligibility of clear speech produced in background noise or intelligibility in conversational settings. This study explores a novel testing paradigm that, within the framework of a conversational scenario, aims to enable the investigation of voice adaptations in response to acoustic and auditory barriers and their effects on speech intelligibility and cognitive effort. A visual task was used to elicit a conversation between two participants, who were sitting in separate acoustic environments. The acoustic environments of the two participants were controlled separately, allowing the presentation of background noise as an acoustic barrier to none, one or both participants. In each trial of the test, triggered by the visual task, a measure of speech intelligibility was embedded in the conversation through the visual presentation of a written HINT sentence to one participant, who had to read it, and their interlocutor had to repeat it back. Speech intelligibility was scored from the repeated sentence as in standard speech intelligibility tests by a researcher who administered the experiment. During the conversation the participants alternated between talker and listener roles, allowing to separate cognitive effort into speaking effort and listening effort. During speaking turns, speaking effort was measured in terms of a set of voice features (intensity, fundamental frequency, speaking rate and spectral tilt) and pupil dilation. During listening turns, pupil dilation was used as a proxy measure of listening effort. This paradigm might offer a tool for (i) testing speech intelligibility in a more ecological conversational environment compared to standard speech intelligibility tests, (ii) measuring acoustic and linguistic changes in the voices of the participants in response to acoustic and auditory barriers that can be controlled separately for the two interlocutors, and (iii) measuring how these voice changes affect speech intelligibility and cognitive effort. The paradigm will be presented together with preliminary data from young normal-hearing participants.

P60 Evaluating the intelligibility of conversational speech for the CHiME-9 ECHI challenge

Robert Sutherland[1], Stefan Goetze[2], Jon P Barker[1]

1. School of Computer Science, University of Sheffield, Sheffield, United Kingdom | 2. South Westphalia University of Applied Sciences, Iserlohn, Germany

Improving the performance of assistive hearing devices, such as hearing aids, is crucial for supporting communication in noisy everyday environments. In this pursuit, a large dataset has been collected for an open challenge focusing on speech enhancement in conversations, "CHiME-9: Enhancing Conversations to address Hearing Impairment (ECHI)". The speech enhancement systems submitted to this challenge will be assessed using listening tests in March 2026. While there are standardised methods for speech quality assessment, research on assessing conversational speech intelligibility is more limited. This work highlights key challenges in designing such tests and proposes a novel evaluation scheme to assess speech intelligibility in conversations.

The CHiME-9 ECHI Challenge focuses on four-party conversations in noisy environments. The task is to remove background noise while preserving the speech of the three conversation partners, using audio recorded with hearing aids or Project Aria glasses. This dataset consists of 30 hours of audio from 49 sessions, with 194 unique speakers. Challenge participants will submit audio for the evaluation set, which consists of nine sessions.

To evaluate the effectiveness of these systems, intelligibility in the context of conversations must be defined, and then a natural task must be designed, with an emphasis on ecological validity. Preliminary pilots have identified several key challenges which make this kind of evaluation difficult. Segments must be designed appropriately, so they are semantically coherent and complete, and easy to parse, remember and repeat. Listeners must be provided with context so they can 'tune into' the sample in a one-shot listening paradigm. There must also be a consistent cue for the target, so that listeners know which speaker to focus on.

Consideration of these factors has led to the development of a new listening test methodology, which provides listeners with ecologically valid cues to encourage them to attend to a specific target speaker. To do this, listeners are provided with a clean speech sample of the target speaker, the prior context of the conversation, cues for when the target speaker is talking and training samples for each target speaker.

All tools developed for this project will be open-sourced, and the audio signals and intelligibility labels will be released to the research community after the conclusion of the CHiME-9 ECHI Challenge.

Acknowledgements: This work was supported by the UKRI AI Centre for Doctoral Training in Speech and Language Technologies (SLT) and their Applications funded by UK Research and Innovation [grant number EP/S023062/1]. This work was also supported by WS Audiology.

P61 The Diamonds Chamber: A new lab for assessing listening effort and speech reception in immersive audio-visual virtual reality

Chiara Visentin, Matteo Pellegatti, Nicola Prodi, Giovanni Turri, Marcello Bonfè
Department of Engineering, University of Ferrara, Italy

Giulia Tampone, Andrea Bertelli, Andrea Trevisani, Luca Tebaldi
e-Learning and Multimedia Services Office, University of Ferrara, Italy

Marianna Manuelli, Andrea Ciorba
Department of Neuroscience and Rehabilitation, University of Ferrara, Italy

Objective: The goal of this study is to bridge the gap between laboratory assessment and real-world hearing experiences by integrating standard speech reception tests with video information, virtual reality, and assessment of listening effort. An integrated audio-video test bench is designed, allowing tests to be conducted in lifelike and ecologically valid conditions.

Methods: The setting up of the audio-video test bench is divided into three phases. The first phase involves the design and development of a multi-channel audio system for spatial audio reproduction, with a high capacity of recreating sound fields that closely approximate target sound environments – whether real or virtual with high spatial accuracy. The system, installed in a large audiometric booth, consists of 41 loudspeakers arranged on an irregular array, enabling Higher-Order Ambisonics (HOA) playback. The second phase involves the development of a Unity-based software platform capable of simultaneously managing the playback of auditory stimuli, the control of visual scenarios—presented through a VR headset—and the collection of speech reception accuracy data. The third phase focuses on integrating into the software system the acquisition of behavioral and physiological measures related to listening effort, including verbal response time, pupillometry, skin conductance, and heart rate. Once the system has been fully developed and integrated, it will be evaluated in an initial clinical-audiological

application. In this phase, five typical real-life audio-visual scenarios will be reproduced, each representing different communication situations and background noise conditions. Examples include one-to-one conversations at short distances, such as those taking place in a café or a public park. The scenarios are created using 360° video recordings captured in the selected locations, combined with multichannel audio recordings. The speech test material used is the ITA Matrix sentence test, consisting of five-word sentences presented in an adaptive mode. Participants, either with normal hearing or with mild hearing loss will be asked to repeat each sentence they hear, allowing for the assessment of speech reception performance under realistic, ecologically valid conditions.

Expected results: It is expected that the proposed system will enable more accurate and ecologically valid assessment of speech reception and listening effort compared to traditional laboratory tests. The integration of behavioural and physiological measures is anticipated to provide a deeper understanding of real-world communication challenges. The outcomes will support the clinical adoption of immersive, multimodal testing protocols for hearing evaluation.

P62 Disentangling f0 and spectral envelope contributions to speech-in-noise comprehension via selective spectrotemporal modulation filtering

Michael Ernst
Fiebachlab for Cognitive Neuroscience/Peopleslab, Department of Psychology (FB05), Goethe University, Frankfurt am Main, Germany | Cooperative Brain Imaging Center (CoBIC), Frankfurt am Main, Germany

Léo Varnet
Laboratoire des Systèmes Perceptifs, UMR 8248, École Normale Supérieure, PSL, CNRS, Paris, France

The spectro-temporal modulation (STM) patterns present in speech carry information relevant for intelligibility, such as formants, pitch contour and syllable boundaries. The STM content of a signal can be represented and manipulated in the Modulation Power Spectrum (MPS) domain. Elliott & Theunissen (2009, PLoS Comput Biol, doi:10.1371/journal.pcbi.1000302) showed that certain regions of the MPS are critical for intelligibility. They found that speech-in-noise comprehension was significantly impaired when removing temporal modulations below 12 Hz or spectral modulations below 4 cycles/kHz. However, a limitation of this approach is that the MPS representation jointly encodes the spectral envelope and harmonic structure, which are thus both affected by STM filtering. When comprehension drops, it is therefore impossible to determine if the effect was purely due to the degradation of envelope information or due to the removal of fundamental frequency (f0) cues.

We address this limitation by applying the same filtering approach as Elliott & Theunissen, but separately processing the f0 contour and the spectral envelope. This makes it possible to independently transform these two components by including or excluding the f0 in the signal or by selectively reintroducing the f0 after filtering the spectral envelope. This approach thereby grants more specificity in the manipulations one can apply to a speech signal and should allow the disentanglement of the relative contributions of the acoustic components for comprehension.

Using the PyWorld package, our processing pipeline decomposes a speech signal into its f0 and spectral envelope, calculates the MPS from the latter, applies a lowpass filtering in the spectral or temporal modulation domain, and subsequently resynthesizes a new speech signal given the filtered MPS and the unchanged or filtered f0. Using the same corpus of 100 sentences as in the original study, we aim to test how filtering restricted to the spectral envelope of the signal may affect speech-in-noise comprehension. Stimuli will be randomly drawn, and their presentation balanced between filtering conditions and signal-to-noise ratios.

In this context, we expect speech comprehension to be higher compared to Elliot & Theunissen's study, given the preserved pitch contour (aiding in auditory unmasking and stream segregation). In particular, the minimal spectral-modulation cutoff for speech-in-noise comprehension should be lower than estimated in the original study and provide a better estimate of the required resolution of the spectral envelope itself, without interference from a parallel processing pathway for pitch.

P63 Precise cross-domain phenotyping as a link between auditory function and genetic causes

Shiran Koifman[1], Anna Warzybok-Oetjen[1], Sabine Hochmuth[2], Mareike Buhl[3], Crystel Bonnet[3], Paul Avan[3], Birger Kollmeier[1]

1. Medical Physics and Cluster of Excellence Hearing4all, Carl von Ossietzky Universität Oldenburg, Germany | 2. Department of Otolaryngology, Head and Neck Surgery, Carl von Ossietzky Universität Oldenburg, Germany | 3. Institut de l'Audition, Institut Pasteur, Paris, France

Precise auditory phenotyping is essential to establish a causal link between genetic variants and their functional consequences in untimely age-related hearing loss (uARHL), defined here as hearing loss occurring at least 20 years earlier than expected based on age-appropriate normative hearing thresholds and in the absence of identifiable non-genetic causes (e.g., noise exposure, ototoxic medication). Individuals with uARHL (≥40 years) and age-matched controls with normal hearing are characterised using a clinically applicable cross-domain test battery that includes, among others, pure-tone audiometry, suprathreshold psychoacoustic measures (speech intelligibility in noise, tone detection in noise, and categorical loudness scaling), as well as electrophysiological and vestibular measures. Parallel genotyping, performed in collaboration with the Institut de l'Audition in Paris, enables the first genotype–phenotype investigation in monogenetically caused uARHL.

Preliminary results reveal distinct auditory phenotypes in uARHL, including reduced masking release for speech in fluctuating noise, elevated tone-in-noise thresholds at 2 kHz, and steeper loudness growth functions compared to controls, with some effects interacting with age and/or pure-tone average. These findings emphasise the importance of suprathreshold processing measures for capturing functional deficits beyond audibility. By directly linking pathogenic variants to specific auditory performance profiles, this study provides first insights into the phenotypes of monogenetically caused uARHL and demonstrates the potential of integrated genotyping and phenotyping for precision diagnostics and profiling of age-related hearing loss.

P64 Electrophysiological correlates of speech perception in noise development at adolescence

Francesca Cavicchiolo[1,2], Marta Puertollano[1,2], Marlies Gillis[1,3,4,5], Joseph Francois Johnson[1,2], Luna Prud'homme[1,2], Vincent Wens[1,3], Xavier De Tiège[1,2,3], Axelle Calcus[1,2]

1. UNI – ULB Neuroscience Institute, Université libre de Bruxelles, Bruxelles, Belgium | 2. CRCN – Center for Research in Cognition and Neuroscience, Université libre de Bruxelles, Bruxelles, Belgium | 3. Laboratoire de Neuroanatomie et Neuroimagerie translationnelles, H.U.B. – ULB University Hospital, Bruxelles, Belgium | 4. Laboratory of Functional Anatomy, Université libre de Bruxelles, Bruxelles, Belgium | 5. ExpORL – Experimental Oto-rhino-laryngology, KU Leuven, Leuven, Belgium

Background: Adolescence is a period of heightened neural plasticity underlying the improvement of cognitive and perceptual skills required to succeed in complex social environments. Among these, the ability to perceive speech signals in the presence of noise or competing speakers is crucial in teenagers' daily-life interactions. The neurobiological mechanisms underlying speech perception in noise (SIN) protracted development remain poorly understood, but likely stem from changes at the endocrine and neural levels. As part of the SensationaHL pubertal development cohort collection, this study investigates the effects of puberty on the electrophysiological changes underlying auditory and cognitive development supporting SIN.

Methods: We will present electrophysiological data from a subset of participants (9–18 years old) enrolled in a two-year longitudinal cohort providing measures of pubertal stage, hormonal levels, cognition, SIN, and multimodal neuroimaging (HD-EEG and MRI). We will focus on HD-EEG measures of cortical tracking of speech (CTS) during passive listening to continuous speech in 3 conditions: quiet, energetic masking (EM; envelope-modulated speech-shaped noise) and informational masking (IM; two-talker babble).

Results: Preliminary findings from a subset of participants at T1 suggest developmental improvements in both SIN perception and CTS, particularly in IM. Moreover, better SIN performance correlates with higher CTS, suggesting a link between behavioural and neural measures. Finally, we are investigating the role of puberty in such neurobehavioral developmental trajectories during adolescence.

Discussion: Results will be interpreted in the context of adolescent neurodevelopment, with a focus on the potential role of puberty in triggering a second sensitive period for the maturation of complex auditory perception. Furthermore, such study holds potential for practical implications: the emphasis on speech-in-noise perception addresses a real-world challenge with potential effects on academic performance and social interactions during this sensitive developmental phase.

P65 Exploring auditory perception experiences in daily situations in autistic adults

Elena Silva
Tilburg University, Netherlands

Linda Drijvers
Radboud University Nijmegen, Netherlands

James Trujillo
University of Amsterdam, Netherlands

Autistic individuals often show differential sensory perception, including hypo- or hypersensitivities to sound. Previous research also suggests that autistic individuals often have difficulty processing intentional and affective cues in speech acoustics. However, it is unclear whether autistic individuals have difficulties with the perception of auditory information,and in particular, whether they have more difficulty with speech understanding than non-autistic individuals. Additionally, there is a growing recognition that many individuals identify as being autistic without having a formal diagnosis. Assessing the auditory perception experiences of both self-identifying and clinically diagnosed autistic individuals provides a more inclusive investigation, and helps to shed light on the comparability of these two groups.

We investigated self-reported auditory perception using the Speech, Spatial, and Qualities of Hearing Questionnaire (15-SSQ) among autistic (self-identifying, n=18; clinically diagnosed, n=45) and non-autistic adults (n=66). The study was conducted in the Netherlands, but the questionnaire and call for participation were in English and open to anyone regardless of country of residence.

Both clinically diagnosed and self-identifying individuals with autism reported significantly lower scores on the total 15-SSQ score and on the Speech subscale compared to non-autistic individuals, indicating challenges in overall quality of auditory perception, and speech comprehension. Clinically-diagnosed individuals also showed lower scores on the quality and spatial subscales compared to non-autistic individuals. We further found 1) that speech hearing is particularly challenging for many autistic individuals, and 2) full score and all scores for all subscales were statistically equivalent between the self-identifying and clinically diagnosed autistic individuals. Finally, autistic participants used the free-response portion of the survey to report several factors that they find to be particularly challenging about speech understanding in noisy environments.

Our findings first highlight the challenges faced by autistic individuals regarding auditory perception. Beyond hypo- or hyper-sensitivities, difficulty processing auditory information in general, and in particular speech, seems to commonly occur in autism, which could be contributing to social interaction difficulties. The finding that self-identifying and clinically diagnosed autistic individuals show similar patterns of hearing difficulties further emphasizes the need for more inclusive research practices that collect the experiences of all the individuals in the autistic community in the study of sensory perception in autism.

P66 Using real-time acoustic resynthesis and virtual reality to understand the needs of hearing-impaired listeners in conversational settings

Thomas Stolarski, Jon Barker
University of Sheffield, UK

Simone Graetzer
University of Salford, UK

Alastair Moore
Royal National Institute for Deaf People, Peterborough, UK

Evaluating the effectiveness of hearing aid algorithms in conversational settings is a challenging task. Setting up realistic scenarios involves substantial preparation and effort on the part of both researchers and participants. Additional challenges arise when trying to design environments that are both controllable and reproducible. One potential solution is to use virtual reality in combination with real recorded data.

This study uses a novel dataset, CHiME9-ECHI, that provides close-talk microphone and head-tracking data. Using this data, we are testing the feasibility of hearing aid algorithm evaluation in synthetic environments, where recorded ground-truth speech and head-tracking data from four conversational partners can be combined with spatial audio processing to replicate the audio as it would be received by the hearing aid microphones worn by one of the conversants.

With this synthetic environment, we can implement "virtual hearing aids" that can use the audio and sensor information to replicate the source enhancement strategies used by physical hearing aids, such as beamforming and other target-speaker enhancement techniques. A user can then wear a virtual reality headset, and the tracking data can be processed by the system to produce a simulated hearing aid experience in a conversational scenario.

The long-term goal of this system implementation is to allow for hearing aid algorithms to be assessed more easily and reproducibly. By providing a means to rapidly and iteratively adjust the parameters and algorithms employed, we can better identify which characteristics align hearing aids with the needs and preferences of their users, both generally and individually.

In the future, we expect that similar systems may allow for hearing aid manufacturers to better identify what customizations they may offer to their users, or to allow for a hearing aid fitting process where a user can virtually experience challenging noise environments and tune parameters in real time, creating a customized profile that could be uploaded to their hearing aids to create a listening experience better suited to their personal needs.

At the workshop, we hope to provide some early audio examples of conversations reprocessed using these techniques.

Funding: This project is supported by RNID.

P67 Development of an ecologically-valid speech intelligibility test using virtual acoustics

Fabian Hettler[1], Jonas Stohlmann[1], Florian Denk[2], Hendrik Husstedt[2], Jürgen Tchorz[1], Markus Kallinger[1], Tim Jürgens[1]

1. University of Applied Sciences Luebeck, Germany | 2. German Institute of Hearing Aids, Luebeck, Germany

Speech intelligibility (SI) testing plays a central role in audiology and hearing research, indicating how well speech can be understood in quiet or noise. Traditional SI tests rely on isolated sentences or single words presented in a highly controlled, but unnatural environment. While this approach ensures repeatability, it lacks ecological validity and often fails to reflect the listening abilities in everyday life. With the use of virtual acoustics and text-to-speech (TTS) synthesis, we designed and evaluated a SI test in a virtual acoustic everyday environment with conversational contents. The goal was to create a test design in a representative everyday situation, that is semantically coherent and acoustically authentic.

The virtual test environment simulated a busy cafeteria. Background noise was the recording of a real cafeteria with a spatial microphone array. Synthetic speech material generated by two different TTS providers (Google TTS and Acapela) was embedded into the ambient noise recordings and virtually presented from the frontal direction. The complete acoustic scene was presented in the anechoic chamber of the University of Applied Sciences Lübeck using a spherical 65-loudspeaker array with 7th order Ambisonics rendering. The ambient cafeteria noise was presented at a level of 65 dB SPL. Test target-stimuli consisted of TTS-generated dialogues containing both carrier phrases and keywords to be recognized, presented at fixed signal-to-noise ratios (SNRs) of −11 dB and −8 dB, corresponding to the SNR30% and SNR70% conditions, respectively. Twenty-four normal-hearing participants completed multiple test runs under both voice conditions. Speech intelligibility was determined by keyword recognition, and the results were used to derive psychometric functions for each condition. In addition to measuring SI, the participants gave subjective ratings of the voices' naturalness, the scenes' realism, and the perceived difficulty of each measurement by responding to six-step Likert scales.

The results showed that psychometric functions were successfully inferred from the data using both SNR conditions. Participants rated the SNR70% as significantly more realistic. While no significant differences in perceived difficulty were found between voice providers, Google voices were rated as more natural and exhibited smaller test-retest bias compared to Acapela. Acapela voices, however, yielded slightly steeper psychometric function slopes, suggesting greater measurement sensitivity near the speech reception threshold.

In conclusion, the study demonstrates that synthetic speech embedded in a simulated cafeteria environment enables reliable and ecologically valid SI measurements.

Funding: This study was funded by the German ministry of research, technology and space affairs.

P68 The time course of pupil responses for cochlear implant users and typical hearing listeners differs when listening to speech-in-noise

H. Christiaan Stronks[1], Jacolien Van Rij[2], Jeroen J. Briaire[1], Johan H. M. Frijns[1]
1. Leiden University Medical Center, Netherlands | 2. University of Groningen, Netherlands

Background: Pupillometry is commonly used to assess listening effort. For typical hearing (TH) listeners, the magnitude of the pupil response associates with the signal-to-noise ratio (SNR). However, recent studies have shown that the pupil response for CI users is less sensitive to SNR and depends much more on the type of speech material. Most of these studies assessed peak pupil dilation, which reflects the maximal amplitude of the pupil response during or just after sentence offset relative to baseline, but omits the temporal dynamics of the pupil response.

Objective: To assess the time course of the pupil response for CI users and TH listeners at different SNRs and speech stimuli.

Methods: Pupil responses were collected from 18 CI users and compared to 18 age-matched listeners using Matrix and LIST sentences administered at three SNRs: speech recognition threshold, where speech intelligibility (SI) was 50% (SRT); +6 dB re. SRT; and in quiet. Pupil waveforms were analyzed with generalized additive mixed modelling (GAMM).

Results: Preliminary analyses revealed that SNR effects were most prominent for TH listeners, where pupil dilation was significantly larger at more challenging SNRs across all SNR pairs for Matrix and LIST sentences. For CI users, SNR effects were smaller. For the LIST sentences, all SNR pairs differed significantly, but this was not the case for Matrix sentences. For both groups, SNR effects on pupil dilation were most pronounced seconds after sentence offset, but much less prominent during stimulus presentation. CI listeners showed significantly larger pupil dilations than TH listeners during sentence presentation and after sentence offset, in quiet and at +6 dB re. SRT. This effect was also most prominent after sentence offset. At 0 dB re. SRT, where speech intelligibility was identical for both groups, no differences were observed.

Conclusions: SNR effects for both groups were most pronounced after sentence offset, suggesting a more sustained listening effort under more challenging listening conditions. Similarly, differences between CI users and TH listeners were most prominent during the release of effort, which develops long after sentence offset and may not be captured by classical peak-picking outcome measures, notably the peak pupil dilation response.

Funding: 'INTENSE' project (grant #17619) within the Crossover Research Program of the Dutch Research Council (NWO), co-funded by Advanced Bionics (Hannover, Germany).

Acknowledgements: We thank the study participants for their dedication, and Paula L. Jansen, Robin van Deurzen, and Nicolas Furnon (Advanced Bionics) for technical support.

P69 Children listening to mispronunciations in continuous speech

Katharina Kaduk[1], Rebecca Holt[2], Liam Howard[1], **Hannah J. Stewart**[1]
1. Lancaster University, UK | 2. Macquarie University, Australia

Listeners need to track continuous speech and phoneme perception to communicate effectively. However, children's speech listening is usually assessed using sentence repetition tasks, primarily using one short sentence at a time (e.g. the BKB). These tasks do not measure the effort of mentally repairing and reconstructing the speech that was initially misheard.

To investigate how children perceive mispronunciations in continuous speech, we have created an event-related potential (ERP) paradigm using 10-minute narrated short stories. Here we present the results from 1) a behavioural study validating the perceptual distinctiveness of our mispronunciation manipulations; and 2) the pilot ERP study for one of the short stories.

In our stimuli validation study, we are currently testing 20 typically developing hearing children aged 6-11 years on a three-alternative forced-choice response paradigm (odd-one-out). Our stimuli consist of one syllable nouns (e.g., 'bag'). The odd-ones-out are mispronunciations of the trials' noun altered by systematically changing the initial consonants' place of articulation ('dag') or voicing ('pag'), while maintaining the manner, to yield non-words.

We predict that the children will successfully detect the odd-ones-out, with accuracy significantly higher than chance. We also predict that detection (as measured by RT and accuracy) will significantly differ for place compared to voicing articulation manipulations. Finally, we predict that detection will significantly improve with the children's age, receptive vocabulary and processing speed.

In our pilot ERP study, nouns in the middle and at the end of sentences have been randomly assigned as target words every few sentences throughout an Enid Blyton story. A third of these target words remain in their original form, a third are mispronounced as per the stimuli validation study (but with up to four syllables), and the final third are nonsense words. To maintain attention the children are asked comprehension questions at the end.

We predict that the hearing children will show a graded N400 response, with the largest amplitudes for nonsense words followed by mispronunciations and the original words. This would indicate that the nonsense and manipulated words were unpredictable, and that the children were listening to them. We also predict that the children will show reduced processing efficiency, due to fatigue, over the course of the story. As evidenced by smaller N400 amplitudes for the mispronounced and original words. However, we expect N400 amplitude to remain large for the nonsense words as they are perceptually distinct.

P70 Predicting speech in noise perception for cochlear implant listeners from a combination of measures of amplitude modulation rate discrimination and viability of the electrode-neural interface

Louis Riley
Imperial College NHS Healthcare Trust, UK

Nicholas Haywood, **Debi Vickers**
University of Cambridge, UK

Speech perception using a cochlear implant (CI) is reliant upon the listener's ability to process envelope cues in multiple independent channels.

In this research we use two measures to assess these abilities: (1) the Amplitude Modulation Discrimination test for Cochlear Implants (AMCI) and (2) the Panoramic Electrical Compound Action Potential Method (PECAP). Boths measures aim to estimate neural health and current spread. The AMCI is a perceptual measure that assesses AM rate discrimination abilities (13 versus 40 Hz) at the cortical/sub-cortical level in the presence or absence of a speech-shaped noise masker on adjacent channels. The PECAP measures eCAPs using a forward-masking artefact-reduction technique for every combination of masker and probe electrodes at the most comfortable level, a peripheral measure. The children's coordinate response measure (CCRM), an adaptive speech test using either a multi-talker or a speech-shaped noise masker was used to assess speech-in-noise perception.

A pilot study was conducted with ten adult cochlear implant listeners (Cochlear Nucleus device users) tested on five different target channels (19, 16, 13, 10, 7), to evaluate how the AMCI and PECAP measures relate to one another, to understand how they might relate to speech-in-noise perception and to determine the power calculation for a larger scale exploration of these effects.

The measures of current spread from the AMCI and PECAP seem to be related with a moderate-to-large effect size (r=0.60) but the measures of neural responsiveness had a smaller effect size (r=0.23) for the relationship, which was not significant for this pilot work. Different models were estimated for the prediction for speech-in-noise perception, with the best-fitting model including both the AMCI and PECAP measures as predictors (r=0.83; for both competing noise types).

The power analyses suggested that for some of the analyses the sample size was sufficient but to be able to fully understand the relationship between measures a larger sample of 26 participants are required.

P71 Do we talk differently to AI? Speech adaptation in conversations with ChatGPT

Jihoo Lee[1], Outi Tuomainen[2], **Jieun Song**[1]
1. Department of Digital Humanities & Computational Social Sciences, KAIST, Daejeon, South Korea | 2. Department of Linguistics, Potsdam University, Potsdam, Germany

With the introduction of voice-based Generative AI chatbots (e.g., OpenAI ChatGPT's Advanced Voice Mode, 2024; Grok AI's Voice Mode, 2025), people have begun communicating with AI systems through speech more frequently than before. Thus far, however, previous research has primarily focused on clear speech modifications towards AIs in controlled settings, using earlier, less-interactive systems. Furthermore, despite the advance in these technologies, there are still limitations in real-time conversations with AI such as delays in turn-taking. While people may find it effortful to communicate naturally with AI chatbots using speech, research on these issues remains largely lacking.

The present research aimed to investigate how people adapt their speech when they engage in spontaneous speech interactions with AI, compared to when they interact with a human interlocutor. To this end, subjects were asked to perform the spot-the-difference Diapix UK task with a human interlocutor and with ChatGPT-4o, where they conversed with each other to find 12 differences between two Diapix pictures. The subjects were native Korean speakers (age: 20~35 years), and their speech was audio-recorded using a headset microphone (DPA4488) during the task.

Their speech was analysed in terms of conversational timing and speaking style, by measuring i) Floor Transfer Offset; ii) the number of turns and backchannels, and iii) acoustic-phonetic characteristics (speech rate, f0, intensity). The preliminary results suggest that speakers exhibited almost no backchanneling behaviour when interacting with the AI and did not attempt to interrupt or take the floor from it. Turn-taking also occurred less frequently when talking with the AI than with a human interlocutor. In addition, the subjects tended to speak faster when talking with the AI. These findings suggest that, despite the advancements of voice-based chatbots as conversational partners, people still adopt a less interactive speaking style when conversing with them. Speakers may adopt this style in order to secure their turns, based on the assumption that AI would not be able to use turn-taking cues as humans do, which is essential for natural speech communication. It is also possible that speakers feel more anxious and exert more speaking effort when talking with AI, which needs to be further investigated. To accurately assess the naturalness of ChatGPT as a conversational partner, its speaking style and turn-taking behaviour will also be examined further. To date, we have collected data from seven subjects, and data collection is still underway to have a full understanding of the results.

P72 Neural speech tracking in older adults: The effects of speech rate and pause

Minhong Jeong[1], Jin Dou[2], Min Hee Kang[3], Jieun Song[1]
1. Korea Advanced Institute of Science and Technology, Daejeon, South Korea | 2. University of Rochester, New-York, USA | 3. KT Corporation, South Korea

Previous research has shown that older listeners have greater difficulty understanding fast speech than younger listeners. They may thus benefit from a reduced speech rate or additional pauses during fast speech, as this provides them with extra processing time. The present study aimed to examine how inserting pauses in linguistically appropriate positions, under varying speech rates, affects speech processing in normal-hearing older listeners. We examined this in terms of neural tracking of speech, specifically to understand how the underlying neural processing of speech relates to differences in perceived listening effort and speech comprehension performance.

To this end, electroencephalograms (EEG) were recorded from 18 native Korean older adults (mean age: 64.9), with normal hearing and cognitive abilities, while they listened to Korean sentences that varied in speech rate (artificially compressed vs. uncompressed) and pausing (with vs. without pauses). The rate for the fast condition was individually determined using an adaptive staircase procedure, to equalise speech comprehension accuracy across subjects. All naturally produced pauses were removed in the no-pause condition, whereas 500ms silences were inserted at prosodically appropriate positions in sentences used for the pause condition. After each sentence, subjects answered a comprehension question and rated their perceived listening effort. Neural tracking of the speech envelope was measured in EEG using multivariate Temporal Response Functions (mTRF). We also computed the predictability of each word in a sentence using a Korean large language model (Mi:dm 2.0) and examined how EEG responses varied as a function of the semantic predictability using mTRF.

The results found that speech rate did not affect speech comprehension accuracy in the main experiment, whereas subjective effort was higher when listening to fast speech than normal speech. In contrast, theta-band tracking was stronger at normal than fast rates. Pauses had no effect on any of the behavioural responses or neural tracking. When the degree of neural tracking was included in mixed-effects models predicting listening effort ratings, the results demonstrated that perceived effort increased as tracking increased, but only in the fast condition. In contrast, stronger tracking was significantly associated with poorer comprehension accuracy in the fast condition. These results suggest that speech tracking outside of the normal modulation rates was difficult for older listeners. However, speech tracking was enhanced when understanding fast speech was more difficult, likely because listeners expended greater listening effort. The semantic analysis is in progress to understand how lexical and semantic processing vary under these manipulations.

P73 Balancing visual and acoustic strategies in natural conversation: Effects of noise, hearing impairment, and talker position

Valeska Slomianka
Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark

Florence Mazzetti, Eline Borch Petersen
ORCA Labs, WS Audiology, Lynge, Denmark

During face-to-face communication, interlocutors' head and gaze orientation reflect a balance between two competing strategies: one that aims to maximize visual input by directing gaze toward the talker to support degraded acoustic signals through multimodal cues, and another that seeks to optimize acoustic input by orienting the head to improve signal-to-noise ratio. Previous research has shown that this balance is influenced by acoustic challenges such as background noise and hearing impairment, yet most studies have used passive listening tasks and fixed talker positions. This study aimed to enhance ecological validity by analyzing free conversations in a real-life environment, with varying seating arrangements and acoustic conditions.

We collected speech data from 12 triads consisting of one hearing-impaired and two normal-hearing participants who were familiar with one another. Each group engaged in 16 conversations in a canteen, half during noisy lunchtime and half during quieter hours. One participant—either the hearing-impaired individual or an age-matched interlocutor—wore eye-tracking glasses to record gaze and head movement. Seating arrangements were varied such that the eye-tracked participant faced either both interlocutors or one in front and one to the side.

We found that participants gazed less at talkers positioned next to them compared to those seated in front, both during listening and speaking. The offset between head and eye orientation during listening increased with more distal talker positions, higher background noise, and hearing impairment—suggesting a shift toward acoustic optimization. Hearing-impaired participants looked less at their interlocutors while speaking than normal-hearing participants, possibly reflecting increased cognitive load or reliance on other cues. Noise also increased the time spent gazing at the talker during listening, and in noisy conditions, hearing-impaired participants showed more stable gaze behavior with reduced visual scanning.

These findings highlight how acoustic and spatial challenges shape multimodal communication strategies, emphasizing the adaptive coordination of auditory and visual attention in conversational interaction. The reduced visual engagement with non-central talkers and the stabilization of gaze behavior in noise may have important implications for understanding compensatory mechanisms in hearing-impaired individuals and for informing rehabilitation approaches that consider both auditory and visual aspects of communication.

P74 Effect of children's age on performance on the virtual audio spatial speech-in-noise test for children with bilateral cochlear implants

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

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

Background: Binaural hearing is key to listening in real environments, and yet, the functional performance of children with bilateral cochlear implants is usually assessed for each ear separately. One reason for this may be the limited space and equipment available to clinics. The BEARS project has developed among other assessments, a virtual-audio version of the Spatial Speech-in-Noise Test, the SSiN-VA. This test potentially provides a measure of speech discrimination, spatial release from masking (SRM), listening effort, and relative sound-source localisation. For SSiN-VA to be adapted for clinical use, it is necessary to assess how children's age affects the ability of SSiN-VA to capture SRM and listening effort.

Methods: SSiN-VA provides word-identification scores for each hemifield, for trials where the noise source is close to the speech source, and for trials where the noise source is far from the speech source. It is possible to calculate an index of word identification for each hemifield, considering performance for each noise location. Asymmetries in these indexes across hemifields in SSiN-VA are thought to capture asymmetries in SRM across ears. To assess how age affects the relationship between SSiN-VA and SRM, we will assess whether the SSiN-VA speech discrimination index differences can be predicted by performance in the Spatial ASL task, which uses a traditional SRM paradigm, and whether predictions are affected by listener age. To assess whether SSiN-VA captures listening effort, we will assess whether SSiN-VA reaction time can be predicted by the outcome of the Vanderbilt Fatigue Scale, on the assumption that greater listening effort, leading to greater reaction times, will be associated with greater listening fatigue in daily life. The effect of age on these predictions will be assessed.

Results: Outcomes for 120 children from 14 UK cochlear-implant departments indicate that, overall, the group can perform the test, obtaining similar profiles as cochlear-implant users previously assessed using an array of loudspeakers. The SSiN-VA speech discrimination index difference is not correlated with the ASL SRM difference across ears for the group ($S=334318$, $p = 0.08$, $\rho = -0.16$). Analyses about the impact of age on this relationship and on the relationship between SSiN-VA reaction time and listening effort/fatigue will be available at the time of the conference.

Discussion and conclusion: Current assessments performed in the clinic do not generally capture spatial listening. However, this dimension is key to daily communication in real-world environment. The BEARS project has developed research tools to assess speech discrimination, SRM, listening effort, and relative localisation using simple equipment and no additional space requirements in clinic. However, it is necessary to ensure that these tools are accessible to a wide range of children beyond the study population. The effect of age on performance needs to be evaluated in order to guide adaptations of the current research tools for general clinical practice.

P75 Non-intrusive prediction of human speech recognition using mistuned binaural processing and posterior phoneme probabilities

Simon Weihe[1], Jan Rennies-Hochmuth[2,1], Thomas Brand[1]
1. University of Oldenburg, Dep. Medical Physics and Acoustics, Oldenburg, Germany | 2. Fraunhofer IDMT, Hearing, Speech and Audio Technology, Oldenburg, Germany

The prediction of speech intelligibility in real-time via a non-intrusive binaural model would be a convenient tool in research and applications for hearing aids. A front-end using Equalisation-Cancellation (EC) processing is one way to model binaural release from masking. Non-intrusiveness and real-time capability require the calculation of a binaurally speech enhanced signal, which can then be further analysed for speech intelligibility prediction. The modelling of inaccuracies in the human binaural processing is an additional requirement. To our knowledge, so far none of the published EC front-ends fulfil all of these requirements.

Human inaccuracies were previously modelled with Monte-Carlo simulations which, however, do not produce a defined signal and are not suitable for real-time processing. Therefore, we suggest replacing them with a deterministic mistuning of the interaural equalisation parameters.

This approach was evaluated with the standard and modified Speech Intelligibility Index (SII) and the Mean Temporal Distance (MTD) as respective back-ends, to compare the model predictions with previous studies and own measurements. Only the latter fulfils the requirement of non-intrusiveness and produces reliable predictions of speech reception thresholds (SRT) in reverberant rooms.

P76 A simple quantitative approach to assess the ebb and flow of group conversation

William M. Whitmer
Hearing Sciences - Scottish Section, University of Nottingham, UK | School of Health and Wellbeing, University of Glasgow, UK

Graham Naylor
Hearing Sciences - Scottish Section, University of Nottingham

Conversations, regardless of their modality, are a vital part of our everyday life. In recent years, we have tried to evaluate verbal conversation behaviour (e.g., timing, speaking style, movement) as a means to more realistically diagnose how well or poorly someone can converse with or without assistance. Withdrawal and dominance are often referenced as two conversational strategies used when conversation is difficult (e.g., in noise, with a hearing loss). We here look for evidence of both strategies in the behaviours of four-person conversations recorded in a lab. Tetrads composed of two older adults with and two older adults without hearing loss carried out topic-driven conversations under six conditions: in quiet and two levels of background babble with and without wearing hearing aids for the two participants with hearing loss. To assess the degree of interaction, we analysed vocal activity over the course of each conversation for contributions from each member; the binary outcome 'contribution' was calculated using a proportional threshold of speaking time in a given time window. There was no clear pattern across noise conditions of partial withdrawal or dominance for any given participant. This general lack of previously reported real-world strategies may be a symptom of reactivity in laboratory conversations. There was generally a higher degree of interaction (i.e., more polylogues) in conversations when the interlocutors with hearing loss were aided, but the effects of noise and aiding on the degree of interaction in the conversations were dependent on the time-window and contribution-threshold settings of the analysis. These results indicate that there may be inherent difficulties in reproducing in the lab the more extreme conversational behaviours reported in our everyday life. But with further work to better determine the analysis parameters, this approach can assess how different demands and assistance affects the degree of interaction, an essential facet of group conversation.

Funding: Work was supported by funding from the Medical Research Council [grant number MR/X003620/1].

P77 Exploring conversational entrainment in noise

Lena-Marie Huttner[1,2], Grace Carrier[1,3], Anna Josefine Munch Sørensen[1]
1. Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark | 2. Technical University of Denmark, Kongens Lyngby, Denmark | 3. University of Twente, Enschede, Netherlands

Over the course of an interaction interlocutors tend to behave increasingly like one another. This phenomenon, also referred to as entrainment, can be observed on multiple levels of conversational behavior, from movement to word choice to acoustic features of speech. Entrainment is believed to facilitate conversational flow and improve interpersonal rapport in social interaction. In speech, prosodic entrainment, including pitch (f0) and stress patterns, is often observed. Since f0 may serve as an important cue in turn-end prediction, f0 entrainment could potentially support interlocutors maintaining the flow of their conversation. However, entrainment is commonly studied in quiet settings. To the authors' knowledge, the effect of noise on f0 entrainment in conversation has not yet been studied. Some studies have found that people have reduced perception of f0 in noise. And since theoretical accounts suggest that entrainment relies on a close link between perception and production, we might expect that having reduced access to f0 would decrease entrainment. However, speakers are known to raise their f0 in noise, potentially as a compensatory strategy to enhance signal clarity. This could, in turn, preserve or even enhance entrainment despite the degraded perceptual environment. In this preliminary analysis, we take a first step toward investigating how background noise affects f0 entrainment in dyadic conversation. Whether such prosodic adjustments ultimately support conversational coordination remains an open question for future research. We collected data from 24 older (average of 63.2 years) native Danish speakers with age-adjusted normal hearing, grouped into 12 conversational pairs. The speakers in a dyad were of opposite gender and were asked to solve the Diapix task (a spot-the-difference task) together by conversing in quiet, as well as 60 dBA and 70 dBA multi-talker babble noise. Each condition was repeated twice, resulting in a total of six conversations per dyad. The conversations were recorded through head worn microphones. We extracted f0 from these conversations using the parselmouth library in Python. We will present preliminary results examining f0 entrainment between interlocutors in quiet as well as in different background noise levels. The results will offer initial insights into how background noise affects prosodic coordination with implications for future studies on conversational dynamics in adverse conditions. Future research could explore entrainment in noisy environments, particularly in relation to individuals with hearing loss, who often experience increased difficulty in perceiving prosodic cues and coordinating effectively during conversations in noise.

P78 Impact of low-frequency acoustic hearing on speech perception in noise and sound localization performance among various groups of cochlear implantees

Mariana Bandeira, Timo Stöver, Tobias Weißgerber
University Hospital Frankfurt, Goethe University, Frankfurt, Germany

The use of electric-acoustic stimulation (EAS) is an established treatment in patients with partial deafness and residual hearing in the low frequencies. Studies have demonstrated better speech perception in quiet and in noise and better sound quality compared to cochlear implant (CI) users with electrical stimulation only. The present study aimed to compare speech perception and sound localization abilities in EAS subjects with groups of normal hearing, bimodal hearing or electric stimulation (CI) only.

Freiburg monosyllabic scores in quiet were compared between ears with CI stimulation (n=1666) and ears with EAS stimulation (n=109). Speech reception thresholds (SRTs) in continuous noise using German Oldenburg matrix test (OLSA) were compared between ears with CI stimulation (n=880) and ears with EAS stimulation (n=85), also for binaural stimulation with either continuous or modulated noise. SRT scores in different spatial noise conditions and for different noise characteristics in free-field and under reverberation were evaluated for smaller subgroups of CI and EAS users. Additionally, mean error in sound localization was measured with an LED pointer method.

EAS users achieved significantly higher speech perception scores in quiet (5 percentage points) and significantly better SRTs in continuous noise (1.6 dB), while still more than 5 dB worse than normal hearing subjects. Under binaural test conditions, bilateral EAS users had better SRTs compared to bilateral CI users in continuous and modulated noise. Bimodal and bilateral EAS users showed a tendency of better SRTs in spatially separated speech in noise conditions (in free-field and reverberation) compared to bilateral CI users. There was a significant correlation between mean low-frequency residual hearing (PTA_{low}) and speech perception in modulated noise. EAS users with contralateral normal hearing showed localization abilities superior to all other CI or EAS groups with mean localization errors closest to normal hearing.

Even if the hearing performance among but also within various groups of cochlear implantees is very heterogeneous and varies greatly depending on the noise and spatial condition, EAS users show better speech perception than users of electric stimulation only in many listening conditions.

SPEECH
IN NOISE

https://2026.speech-in-noise.eu