SPEECH IN NOISE WORKSHOP

Abstracts

Virtual meeting | 20-21 January 2022

The 13th Speech in Noise Workshop was organised by selfless volunteers:

- *Antje Heinrich*
- *Rebecca Carroll*
- *William Whitmer*
- *Julien Pinquier*
- *Thomas Koelewijn*
- *Fanny Meunier*
- *Laura Rachman*
- *Tim Jürgens*
- *Christian Füllgrabe*
- *Anna Warzybok*
- *Gaston Hikhuysen*
- *Gloria Araiza Illan*
- *Marita Everhardt*
- *Gizem Babaoğlu*
- *Mustafa Yüdsel*
- *Soner Türüdü*

Coordinator: Etienne Gaudrain, Lyon Neuroscience Research Center, CNRS, Lyon, France.

Contact information: info@speech-in-noise.eu

The Speech in Noise Workshop is generously supported by:

Talks

Thursday 20 January 2022, 09:15—09:40

A method to convert between Speech Recognition Thresholds (SRT) and percentage-correct scores for speech-in-noise tests

Cas Smits
Amsterdam UMC, Dept. of Otolaryngology – Head and Neck Surgery, Amsterdam, NL

Different approaches have been taken in the development of speech-in-noise tests to quantify speech recognition abilities in noise. Speech-in-noise tests most often use either fixed-SNR procedures to measure the percentage of correctly recognized speech items at a fixed signal-to-noise ratio (SNR) or use adaptive procedures to measure the SNR corresponding to 50% correct (i.e., the speech recognition threshold, SRT). Both procedures have advantages and disadvantages. Unfortunately, a direct, simple, comparison of these measures is not possible yet.

It will be demonstrated that these measures can be converted by using a relatively easy analytical method, when the speech-in-noise test meets specific criteria. Formulae to convert between SRT and percentage-correct were derived from basic concepts that underlie standard speech recognition models. Information about the audiogram is not being used in the proposed method. The method was validated by comparing the direct conversion by these formulae with the conversion using the more elaborate SII model and a representative set of 60 audiograms ($r = 0.994$, respectively). Finally, the method was experimentally validated with the Afrikaans sentence-in-noise test ($r = 0. 866$).

The proposed formulae can be used when the speech-in-noise test uses steady-state masking noise that matches the spectrum of the speech. Because pure tone thresholds are not required for these calculations, the method is widely applicable.

Thursday 20 January 2022, 09:40—10:05

Text-to-speech and back — new ways in speech audiometry

Inga Holube[1], Saskia Ibelings[1], Jasper Ooster[2], Theresa Nuesse[1]
1. Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, Oldenburg, DE | 2. Communication Acoustics and Cluster of Excellence Hearing4all, Carl von Ossietzky University, Oldenburg, DE

Speech audiometry involves the use of speech recognition tests. In an open set format, sentences are presented via headphones or loudspeakers and the correct repetitions by the listener are noted by the examiner. Developing and administering speech tests is a complex and time-consuming process. Speech material must be recorded, adjusted, optimized, and evaluated with large numbers of normal-hearing listeners. New technologies like text-to-speech synthesis (TTS) and automatic speech recognition (ASR) make it possible to support this process. The quality of TTS is sufficient that it can replace recordings with natural speakers. The results of speech tests for normal-hearing listeners using synthesized matrix sentences or everyday sentences, are comparable to the results of tests with their natural speech version. Advances in TTS will enable easier development of larger speech corpora for existing speech tests or new speech material. Further, during the current pandemic, the need for online testing and maintaining greater distances between listeners and examiners in the testing room has increased. While matrix tests can be administered in a closed-set format with all response alternatives displayed on the screen, everyday sentences still require an examiner. An alternative is to record the responses and have them evaluated by ASR. This technique additionally allows calculation of verbal response time, i.e., the time between the end of the presentation and the beginning of the listener's response, as a measure of listening effort. This presentation will give examples of the application of both technologies.

Thursday 20 January 2022, 10:20—10:45

Ignoring people at a cocktail party: testing which acoustic characteristics of ignored speech influence attended speech perception

Hans Rutger Bosker
Max Planck Institute for Psycholinguistics, Nijmegen, NL

When having a conversation in a multitalker setting ('cocktail party' listening), listeners are required to selectively attend a to-be-attended talker and to ignore speech from other competing talkers. This talk discusses a research project that investigated which acoustic characteristics of the speech from competing talkers interferes with the perception of attended speech, and which do not. The surrounding acoustic context in which a given target speech sound is heard is known to directly impact perception. For instance, a vowel ambiguous between a short and a long vowel in Dutch is perceived differently depending on the preceding sentence's speech rate ('temporal context effect'). Similarly, the perception of a vowel ambiguous between /ɪ/ and /ɛ/ is influenced by the first formant characteristics of the preceding sentence ('spectral context effect'). Using these context effects, we tested whether listeners can successfully ignore the temporal and spectral characteristics of a competing talker when interpreting ambiguous speech sounds. Outcomes indicate that, while spectral context effects are strongly modulated by selective attention, temporal context effects are immune to selective attention. Thus, this project demonstrates that listeners are very successful at ignoring a competing talker's spectral voice properties, but have a hard time ignoring their speech rate.

Regular rhythmic primes do not (always) benefit speech-in-noise perception: Evidence for distinct outcomes for temporal attention depending on speech-in-noise manipulations

Anna Fiveash[1], Etienne Gaudrain[1,2], Nathalie Bedoin[3], Reyna L. Gordon[4], Barbara Tillmann[1]
1. Lyon Neuroscience Research Center, INSERM, CNRS, University of Lyon 1, France | 2. University of Groningen, University Medical Center Groningen, Netherlands | 3. Lyon Neuroscience Research Center, INSERM, CNRS, University of Lyon 1, University of Lyon 2, France | 4. Department of Otolaryngology – Head & Neck Surgery, Vanderbilt Kennedy Center, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Research has shown that regular music rhythm can boost grammaticality judgements in subsequently presented speech, even though speech is less regular. Regular rhythmic primes (~30s) presented before a set of naturally spoken sentences typically improve grammaticality judgements compared to sentences preceded by irregular primes or baseline conditions. This rhythmic priming paradigm is motivated by the Dynamic Attending Theory (DAT), which suggests that external rhythms entrain neural oscillations that persist over time and enhance subsequent processing. Enhanced speech processing is particularly valuable when the signal is degraded, as in noisy environments and for hearing-impaired listeners. To simulate this situation, in Experiment 1 (n = 31), participants (tested in the lab) detected grammatical errors in naturally spoken sentences presented in noise (-3 dB signal-to-noise ratio) that were preceded by regular and irregular rhythmic primes. Participants were more sensitive to grammatical errors after irregular compared to regular primes (measured by d'), suggesting that the noisy background reversed the typical regular prime benefit. To investigate this reversed priming effect, three additional experiments were run online. In Experiments 2 (n = 39) and 3 (n = 39), the speech envelope was either preserved (up to 20Hz with reduced temporal fine structure, TFS) or reduced (filtered at 5Hz with original TFS). We predicted that regular rhythmic primes would enhance processing when the speech envelope was preserved, but not when it was reduced. However, both experiments showed a small benefit of regular rhythms for grammatical error detection compared to irregular rhythms, and a stronger bias to respond grammatical after irregular compared to regular primes in Experiment 2. Experiment 4 (n = 40) investigated whether introducing a rhythmic modulation at the regular rhythm frequency (2 Hz) to the noise masker from Experiment 1 (maintaining -3 dB SNR) would re-instate the benefit of the regular prime. Regular and irregular primes did not differ in their influence on modulated noise processing, even though the direction of mean differences was similar to Experiment 1. These results suggest that rhythmic primes may influence speech processing differently depending on whether speech-shaped noise is applied to the speech (Exp. 1, 4), or whether the speech envelope and TFS are manipulated (Exp. 2, 3). Effects of rhythmic priming on speech-in-noise perception appears to be a promising line of enquiry to uncover the acoustic features important for rhythmic priming, and could provide insights into when speech processing in noisy environments is helped or hindered by regularity.

The impact of social observation during a speech-in-noise task: insights from cardiovascular and pupil responses of listeners with hearing loss

Bethany Plain[1,2], Hidde Pielage[1,2], Tanveer A. Bhuiyan[3], Thomas Lunner[2], Michael Richter[4], Gabrielle Saunders[5], Niek Versfeld[1], Adriana A. Zekveld[1], Sophia E. Kramer[1]
1. Amsterdam UMC, NL | 2. Eriksholm Research Centre, Snekkersten, DK | 3. Demant A/S, Smørum, DK | 4. Liverpool John Moores University, UK | 5. University of Manchester, UK

To communicate in difficult listening situations, individuals with hearing impairment must often invest substantial listening effort. In addition to the acoustic conditions, the degree of listening effort invested is affected by the social context of the situation. For some people, social context may increase the success importance of listening, thereby increasing their effort investment. However, for others, social context may lead to stress, particularly when there are communication breakdowns or misunderstandings. To manipulate social context in this study, we investigated the impact of social observation during a speech-in-noise task. We measured participants' pupil and cardiovascular responses. We anticipated that social observation would increase effort investment, reflecting in higher physiological reactivity and subjective effort ratings when observed, compared to when performing the task alone. Twenty-nine experienced hearing aid users were recruited (mean age = 65 years). Participants performed the Danish Hearing in Noise Test (HINT) alone and in the presence of two observers. HINT sentences were presented at two signal-to-noise ratios (SNRs), individually adapted to target 50 and 80% of sentences correct. Baseline pupil size, peak pupil dilation, mean pupil dilation and cardiovascular parameters (heart rate variability, pre-ejection period, blood pressure and heart rate) were measured. After each block, participants rated their subjective effort investment, stress, tendency to give up and their preference to change the situation to improve audibility. Social observation increased baseline pupil size and blood pressure measures, suggesting an increase in stress. Contrary to expectations, self-report, peak pupil dilation, mean pupil dilation and the other cardiovascular measures revealed no effect of observation. Performance was also unaffected by observation. Instead, performance, self-reported effort and stress, peak pupil dilation and mean pupil dilation were sensitive to the SNR, such that performance was higher at the easier SNR, whereas self-reported effort and stress, peak and mean pupil dilation were higher at the harder SNR. These results indicate an increase in effort investment at the harder SNR compared to the easier SNR. No cardiovascular measures were sensitive to SNR. Performing a speech-in-noise task while under observation resulted in an increase in stress, rather than an increase in effort investment. Interestingly, self-reported stress did not increase under observation, but instead varied as a function of SNR, perhaps revealing that the self-report and physiological parameters measured different constructs. This study emphasizes the need for the development of listening tests that better reflect real life communication scenarios, including social context.

Challenges and methods to design a child-appropriate speech-in-noise experiment in spatial auditory environments for young children

Janina Fels, Karin Loh
Institute for Hearing Technology and Acoustics, RWTH Aachen University, DE

The assessment of noise effects in controlled listening experiments for young children (aged three to six years) is a challenging topic. Tasks must be designed appropriately for this age group. Sound reproduction methods must consider children's smaller anthropometric head dimensions to deliver plausible spatial impressions of close-to-real life scenarios. This work focuses on the objective assessment of listening effort. For this, we developed a dual-task paradigm for three to six years old children. The primary task is a double-word recognition task. The secondary task is a memory task, which consists of recalling the correct order of five symbols. This paradigm was integrated into an experimental framework including plausible sound reproduction via headphones using individual headphone transfer function equalization and individualized head-related transfer functions (HRTFs), where the interaural time difference (ITD) is adjusted according to the head dimensions of the participants. We will present preliminary results from a pretest with adults, addressing the possible effects of different noise types (speech-shaped noise, irrelevant multi-talker babble, and relevant multi-talker babble). Further, we will discuss challenges and methods within the experiment's design process.

Keynote lecture

Perceiving and representing voice identity: effects of talker variability and listener familiarity

Carolyn McGettigan
University College London, UK

A talker can sound quite different from one occasion to the next, and this has implications for the ease with which listeners can recognise that person from their voice alone. However, until relatively recently, natural within-talker variability was almost entirely overlooked in studies of voice identity perception. In this talk, I will present our work that has addressed how within-talker variability impacts voice identity perception, and how performance is affected by familiarity (e.g. with the talkers themselves, their accent, the language being spoken). Our approaches have included different perceptual paradigms (sorting, recognition, discrimination), manipulations of voice acoustics (e.g. F0 and formant spacing), and interrogation of the neural responses to natural stimulation (using multivariate analysis of fMRI data). I hope to convince the audience of the importance of including naturalistic variability in our studies of (speech and) voice processing, and will conclude by presenting some key questions for future research in this area.

Prediction of the ASR quality by analysis of the sound environment

Sébastien Ferreira[1], Jérôme Farinas[2], Julien Pinquier[2], Stéphane Rabant[1]
1. Authôt, Ivry-sur-Seine, France | 2. IRIT, University of Toulouse, France

Automatic Speech Recognition (ASR) systems automatically convert the speech contained in an audio signal into text. However, the quality of the resulting text is highly dependent on many factors, such as the context of the recording, the speakers and the subject. In our work, we tried to predict the word error rate (WER) of automatic transcriptions before the speech is converted into text by an ASR system.

Among the various factors degrading the performance of ASR systems, which can be analysed a priori, we studied the sound environment to design this prediction before decoding.

Our approach consists of two steps. The first step is to extract signal parameters that are sufficiently correlated with the WER. The second step is to learn the relationship between these parameters and the WER by a supervised regression using a multilayer perceptron.

To analyse speech in noise, we differentiated three types of signal perturbation due to the sound environment: ambient noise (additive and stationary noise), signal superposition (with speech and music) and reverberation. For ambient noise, we designed a new parameter extraction method that consists in extracting statistics on noise and speech, after a separation by a binary mask.

For signal superposition, we designed new parameters exploiting the tracking of partials extraction in the spectrogram.

Concerning the reverberation, we designed a new parameter, named Excitation Behaviour, which exploits the residuals of the linear prediction.

The efficiency of our parameters has been compared with state-of-the-art methods: we obtain a better prediction of the WER, i.e. of the quality of the automatic transcription before decoding.

The different works presented are then tested in an industrial context: on a corpus of real data from the company Authôt. Our a priori prediction obtains excellent results: an average error of 5.26, omitting the case of regional accents (which are not currently processed by our method). Our method is reliable enough to inform a user as soon as possible about the quality of the automatic transcription of his speech recording. These results can also be used to estimate the time needed to correct automatic transcriptions.

Homophone processing during speech-in-speech situation and cognitive control implication

Samuel El Bouzaïdi Tiali[1], Fanny Meunier[2], Elsa Spinelli[1], Corentin Gonthier[3], Richard Palluel-Germain[1], Juan R. Vidal[4], Marcela Perrone-Bertolotti[1]
1. Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, Grenoble, FR | 2. Univ. Côte d'Azur, CNRS, BCL, Nice, FR | 3. Université de Rennes, LP3C EA 1285, Rennes, FR | 4. Catholic University of Lyon, Sciences and Humanities Confluence Research Center, 2 Place des Archives, Lyon, FR

Lexical ambiguity is ubiquitous in oral language and sometimes induces comprehension difficulties. Moreover, in ecological situations language is rarely processed without any surrounding noise and one of the more frequent situations of language comprehension is speech-in-speech situations. In the present study we evaluated the cognitive process at work during online sentences comprehension in which comprehension difficulties were manipulated by linguistic ambiguity (using French homophones) and the presence or absence of babble noise (multi-talker situations). Eye movements on a visual scene including four different images (correct target, an incorrect target, a semantic distractor and a phonological distractor) were measured to assess online effects of lexical ambiguity and noise processing during language comprehension. We hypothesized that looks to the incorrect target would increase with ambiguity and that this increase would be amplified with babble noise. Also, looks to the phonological and semantic distractors should be higher in the speech-in-speech situation than in silence. Furthermore, we hypothesize that in more difficult linguistic situations comprehension will be related with the better cognitive control of participants. Consequently, we also evaluated cognitive control performances during three cognitive control tasks (evaluating: attentional control, inhibitory control and updating in working memory abilities). Results showed significant effects of lexical ambiguity and multi-talker situations on behavioral performances and eye movement measures, suggesting that additional difficulty in language comprehension increased cognitive effort and processing. Combination of both difficult conditions induced the lowest performances but the interaction effect suggests a smaller impact of ambiguity in a speech-in-speech situation than in silence. Results also showed that the bigger the interference or difficulty cost in cognitive control tasks the bigger the ambiguity effect in the language comprehension task. Results are discussed in light of the literature linking cognitive control and language processing.

Speech recognition in quiet and in noise in a large group of bimodal (CI + HA) listeners

Ulrich Hoppe[1], Thomas Hocke[2], Frank Digeser[1]
1. Department of Audiology, ENT-clinic, University of Erlangen-Nürnberg, DE | 2. Cochlear Deutschland GmbH and Co. KG, Hannover, DE

Cochlear implantation is often performed on one side while the other side is provided with a conventional hearing aid (HA). For these subjects bimodal hearing (HA on one side and cochlear implant (CI) on the other side) is often the provision of choice. The aim of our study was to describe the benefit in terms of speech perception in quiet and in noise for a large group of bimodal listeners. Secondly, we aimed to investigate the influence of hearing thresholds on the HA side on bimodal hearing.

Sentence recognition with hearing aid alone, cochlear implant alone and bimodally were assessed in 148 experienced adult CI listeners. Data were analyzed for bimodal summation using measures of speech perception in quiet and in noise.

Most of the subjects showed improved sentence recognition in quiet and in noise in the bimodal condition compared to the hearing aid-only or cochlear implant-only mode. The large variability of bimodal benefit in quiet can be partially explained by the degree of pure tone loss on the HA side. Also, subjects with better hearing on the acoustic side experience significant benefit from the additional electrical input.

Bimodal summation shows different characteristics in quiet and noise. Bimodal benefit in quiet depends on hearing thresholds at higher frequencies as well as in the lower- and mid-frequency ranges. No correlation with hearing threshold in any frequency range was observed for bimodal benefit in noise.

Friday 21 January 2022, 15:10—15:35

Perception of dynamic pitch in speech

Hae-Sung Jeon
UCLan, Preston, UK

The perception of dynamic – constantly changing – pitch in speech has been extensively studied in psychoacoustics and linguistics. In psychoacoustic studies, listeners are usually presented with short stimuli such as vowels or syllables, and their ability to discriminate a pair of stimuli is assessed. On the other hand, linguistic studies concern intonation over an utterance. Intonation entails not only acoustic prominence realised by pitch, duration, and loudness, but also listeners' abstract knowledge about the relative prominence between syllables within a word or between words within an utterance.

Assuming that listeners' important task in speech communication is to allocate their attention to high-information sites in an utterance or over utterances, the linguistic approach has higher ecological validity than the psychoacoustic approach. But previous linguistic studies have limitations. First, many of them used psychoacoustic tasks, asking listeners to listen to pitch, rather than the intonational prominence. Second, the vast majority of them tested the perception of only pitch peaks (rise-falls), neglecting valleys (fall-rises). Third, participants were limited to young listeners speaking a standard dialect of the target language.

This talk will discuss a research project addressing these issues. In our experiments, we used both psychoacoustic and linguistic tasks; participants judged either relative pitch height or prominence between two pitch peaks or valleys in an utterance. Native English speakers in different age and dialect groups were tested. Outcomes indicate that first, listeners' pitch height discrimination in the utterance context seems to be more accurate than what previous studies report. Second, there is a robust perceptual asymmetry between pitch peaks and valleys, the valleys posing significant challenges as shown for both psychoacoustic and linguistic tasks. Third, listeners' perception of pitch height and prominence is disassociated. The findings taken together suggest an intricate interaction between the physical properties of the stimuli and listeners' top-down knowledge in the perception of speech intonation.

Friday 21 January 2022, 15:35—16:00

Impact of cochlear synaptopathy on speech-in-noise perception: Psychophysical and electrophysiological markers based on temporal fine structure coding fidelity

Emmanuel Ponsot, Sarah Verhulst
Ghent University, BE

Cochlear synaptopathy — i.e. damage to the synapses connecting inner hair cells to auditory nerve fibers, driven by aging or noise exposure — is thought to be an important factor contributing to speech-in-noise (SPiN) understanding difficulties in humans. Yet, it remains unclear which non-invasive biomarkers could be used to assess this pathology. Because of its crucial role for SPiN understanding, we reasoned that addressing temporal fine-structure (TFS) coding fidelity should provide a direct estimate of the impact of synaptopathy on SPiN perception deficits. In this talk, I will present recent psychophysical and electrophysiological measurements based on complex harmonic stimuli with different spectral shapes specifically designed to reflect TFS coding fidelity. We will discuss how several markers extracted from these measurements can account for SPiN intelligibility differences measured within and across various groups of listeners, young/older, with/without outer-hair cell damage.

Posters

Pxx: Session 1, Thursday 20 January, 13:15—15:15
Pxx: Session 2, Friday 21 January, 09:00—11:00
Pxx: Session 3, Friday 21 January, 13:00—15:00

P01 Speech-driven facial animations improve speech-in-noise comprehension in humans

Enrico Varano
Imperial College London, UK

Tobias Reichenbach
Friedrich–Alexander University Erlangen–Nürnberg, Germany

Comprehension of speech in noise can be improved by looking at the speaker's face. This effect is even more pronounced in people with hearing impairments and is thought to be linked to both the temporal and categorical cues carried by the visual component of speech. For instance, rhythms such as the amplitude modulations of a speech signal, which are known to play an important role in speech processing, correlate with the opening and closing of the mouth. Categorical cues are involved in audiovisual speech perception through the non-injective surjective mapping between phonemes and visemes and information about a speech signal can be obtained from other aspects of lip movement as well. However, the precise contributions of the different aspects of lip motion to speech comprehension, as well as the neural mechanisms behind the audio-visual integration, still remain unclear.

We considered both the natural video of a speaker as well as a variety of synthesized visual signals. The synthesized videos were designed to capture features of lip movements of increasing complexity, from the amplitude modulations of speech to realistic facial animations generated by deep neural networks. We then assessed the speech comprehension of participants for the different types of videos. We also recorded their brain activity through EEG while they listened to the audiovisual speech stimuli.

We found that simple visual features such as the size of the mouth opening, related to the speech envelope, modulated the neural response to the speech envelope. However, they failed to enhance speech comprehension. More complex videos including the realistic synthesised facial animations did improve the comprehension of speech in noise significantly, albeit not as much as the natural videos.

Taken together, our results suggest that categorical cues in the texture of realistic facial animations drive the audiovisual gain in speech-in-noise comprehension. Although the amplitude modulation of speech matters for speech processing, and although simplified visual signals that track these amplitude modulations influence the neural response, these signals do not aid in understanding speech in background noise.

P02 Audiotactile speech perception and neural mechanisms

Pierre H. Guilleminot
Imperial College London, UK

Tobias Reichenbach
FAU Erlangen-Nuremberg, Germany

Background: Speech involves a hierarchical structure of information, ranging from phonemes to syllables, words and sentences. These different units of information need to be segmented in order to be processed by the brain. The segmentation presumably relies on oscillations in the delta and in the theta frequency ranges (1-4 Hz and 4-8 Hz) in the auditory cortex, which track incoming speech at the rhythm of syllable and words. The tracking in the theta range plays a functional role in speech processing, as its modulation using transcranial current stimulation has been found to affect speech comprehension. Because these cortical oscillations can also be influenced by somatosensory stimulation, we wondered if such stimulation could impact speech comprehension as well.

Methods: We delivered sparse vibrotactile pulses to the hand of subjects while they listened to speech in background noise. The pulses were aligned to the centre of syllables and shifted in time to study the effect of different lags between the two sensory streams. We assessed the participants' comprehension of the speech signal. Furthermore, we studied the neural encoding of speech and the vibrotactile pulses through electroencephalographic recordings (EEG).

Results: We found that tactile pulses presented at the rate of syllables can modulate and even improve speech-in-noise comprehension. The enhancement occurred when the pulses were aligned with the perceptual centers of the syllables, without temporal delay. The neural responses to both speech and vibrotactile pulses were modulated by different delays between the two sensory streams, displayed audiotactile integration, and reflected the behavioural modulation of speech comprehension. Finally, we demonstrated that the comfort of subjects in understanding speech could be predicted from the electrophysiological markers of multisensory integration.

Conclusions: Our results provide evidence for the role of cortical oscillations and vibrotactile information for speech processing. The observed enhancement of speech comprehension suggests that such vibrotactile stimulation might be employed in auditory prosthesis to aid people with hearing impairment to better understand others in noisy environments.

P03 Modelling the effects of transcranial alternating current stimulation on the neural encoding of speech in noise

Mikolaj Kegler
Department of Bioengineering & Centre for Neurotechnology, Imperial College London, United Kingdom

Tobias Reichenbach
Department of Artificial Intelligence in Biomedical Engineering (AIBE) Friedrich-Alexander-University (FAU) Erlangen-Nürnberg, Germany

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

Based on previous work, we established a spiking neural network model for speech encoding. The model consisted of two coupled neuronal populations generating self-sustained oscillations in theta (4-8 Hz) and gamma (above 25 Hz) frequency ranges. The theta-generating module was designed to track the onsets of syllables and parse the faster gamma activity into smaller segments. To quantify the speech-in-noise encoding performance of the model, we simulated the encoding of spoken sentences in different levels of background noise and used the obtained spiking outputs to decode syllable identities. Furthermore, we subjected the model to a range of speech-inspired tACS waveforms and investigated their effects on the speech-in-noise encoding.

Both syllable tracking and decoding deteriorated in a sigmoidal fashion with increasing levels of background noise. The model performance in the presence of noise resembled typical human comprehension in the analogous speech-in-noise listening task. The simulated tACS interventions yielded phase- and time-dependent modulation of the speech-in-noise encoding in the model, similar to the effects of tACS on the speech-in-noise comprehension. The greatest facilitation of syllable tracking in the model was observed for tACS preceding the speech signal by 50-100 ms. Notably, this latency range corresponds to the typical neural delay associated with cortical auditory processing in humans and thus may explain why tACS applied without additional latency can influence comprehension. Stimulation waveforms filtered in the theta-band frequency range impacted the speech-in-noise encoding in the model the most, which reflected the effects of tACS in humans.

The proposed model provides biophysically-grounded estimates of neural encoding of speech in noise, suitable for studying the relationship between cortical processing of speech and comprehension. The effects of tACS interventions predicted by the model agree with recent experimental findings and shed light on neural mechanisms through which the behavioural effects may arise.

P04 Effects of noise and a speaker's impaired voice quality on spoken language processing in school-aged children: A systematic review and meta-analysis

Isabel S. Schiller
Institute for Psychology, RWTH Aachen University, Germany

Angélique Remacle, Nancy Durieux, Dominique Morsomme
Faculty of Psychology, Speech and Language Therapy, and Educational Sciences, University of Liege, Belgium

Classroom listening conditions are often characterized by high levels of background noise. Beyond that, every second teacher develops voice disorders during their career, meaning that many pupils are taught in impaired (e.g., hoarse or breathy) voices. This systematic review and meta-analysis aims at understanding how background noise and a speaker's impaired voice quality affect spoken language processing in 6-18-year-old children. We review 31 studies on the impact of noise and/or impaired voice on children's answer accuracy and response time (RT) in listening tasks. A classification of the findings is presented in the SPADE framework, featuring three processing dimensions: speech perception, listening comprehension, and auditory working memory. Statistical analyses reveal that noise compromises children's accuracy in listening tasks within each of these dimensions (Cohen's d between –0.67 and –2.65), and that listening to an impaired voice impedes children's accuracy in listening comprehension tasks (d = –0.35). RT data is too scarce to allow firm conclusions. Several factors related to the listener, task, environment, and type of exposure are identified as moderators of the impact of noise and impaired voice. Interaction effects between noise and impaired voice remain unclear and need further investigation. Our results highlight that acoustically adverse listening conditions disrupt children's spoken language processing and may potentially hinder their academic achievement at school.

P05 Effect of simple visual inputs on syllable parsing

Anirudh Kulkarni[1], Mikolaj Kegler[1], Tobias Reichenbach[1,2]
1. Imperial College London, United Kingdom | 2. Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Speech comprehension, especially in difficult listening situations, is affected by visual signals, such as those arising from a talker's face. From a neuroscientific perspective, this multisensory processing takes place at a stage as early as the primary auditory cortex. However, the neural mechanisms behind this audio-visual integration are poorly understood. Here we utilize a computational model of a cortical microcircuit for speech processing to understand how visual input can be incorporated into it. The model consists of a cross coupled excitatory and inhibitory neural population that generates a theta rhythm. This theta rhythm parses the syllable onsets of a speech input. To investigate the effect of visual input on parsing syllables, we add simple visual currents to the model that are proportional to one of the following: (1) the rate of syllables, (2) the mouth-opening area of the speaker or (3) the velocity of the mouth area of the speaker. We find that adding visual currents to the excitatory neuronal population

affects speech comprehension, either improving it or deteriorating it, depending on whether the currents are excitatory or inhibitory and depending on audio-visual timing. In contrast, adding visual input currents to the inhibitory population does not affect speech comprehension. Our results, therefore, suggest neural mechanisms for audio-visual integration and make testable experimental predictions.

P06 Isolating the locus of informational interference during speech-on-speech listening: when does masker intelligibility matter?

Sarah Knight, Sven Mattys
University of York, UK

Speech-in-noise research typically distinguishes between energetic masking (EM: interference between target and masker at the periphery) and informational masking (IM: interference higher in the auditory pathway). IM can itself be broken down into “lower-level” (e.g. acoustic, spatial) and “higher-level” (e.g. linguistic) IM. We use the term “informational interference” (inf-int) to refer to higher-level IM which is influenced by long-term linguistic knowledge.

Unlike EM, inf-int is poorly understood, in part because it is extremely difficult to manipulate inf-int without altering EM or lower-level IM. However, one apparently key feature of inf-int is that it often involves linguistic factors, such as whether or not a masker includes intelligible speech. It is generally believed that intelligible maskers, which involve both EM and IM, are more detrimental to target perception than acoustically-similar but unintelligible maskers, which involve only EM. However, this masker intelligibility effect has typically been demonstrated using connected speech maskers. As a result, it is difficult to determine which specific characteristics of the masker speech underlie any observed effects.

In the current series of studies (total N = 360), we tested the masker intelligibility effect using word list maskers. In their unmanipulated state, these maskers contain lexical and semantic information but lack syntax and sentence-level prosody. Our results suggest that intelligible word list maskers can actually be less detrimental to target perception (or at least no worse) than acoustically-matched, unintelligible equivalents, including time-reversed and noise-vocoded word lists. These findings suggest that the locus of inf-int is unlikely to be at the lexical-semantic level, and may instead reside in masker characteristics associated with connected speech, such as sentence-level syntax and/or prosody. By systematically varying our maskers along a range of parameters, we have taken steps towards isolating the specific characteristics contributing to inf-int. These results highlight the difficulty of consistently characterising and empirically quantifying IM during speech-in-noise listening.

P07 PCA based gesture and function selection for the control of assistive hearing devices

J. William Bailey, Manohar Bance
University of Cambridge, UK

Gestures for controlling audio algorithms may be selected arbitrarily or based on post-hoc analysis of preferences of users. In this study, a wearable device which uses head gestures to control a blind-source-separation-based assistive hearing device was developed. Gestures were drawn from a corpus of simple primitive movements and the most appropriate gestures to use to generate control signals were identified by analysing preference rankings using a principal-component-analysis-based approach. Subjects were asked to rank the appropriateness of head gestures for specific functions of the source separation algorithm. Subjects were also asked to rank social acceptability of gestures and the usefulness of individual functions of the device. Although naive analysis of ranks of mean responses indicate some preferences within the population, principal component analysis suggests that within the set of responses independent dimensions of preference and consensus could be identified. By weighting preference scores by consensus, functions of the assistive device can be assigned to control gestures with more confidence that they will be acceptable to users and allow for intuitive control.

P08 Development of attentive tracking of sound sources

Axelle Calcus
Université Libre de Bruxelles, Belgium

From lively playgrounds to busy classrooms, children communication usually happens in noisy settings. Perceiving speech in noisy is a complex task that requires an adequate combination of sensory perception and cognitive processing. In spite of their functionally mature auditory system, school-age children's perception of speech in noise remains poorer than adults'. The main aim of this study was to better understand the mechanisms underlying this protracted auditory development. In particular, we focused on auditory selective attention and its relationship with speech perception in noise throughout development. Participants were included in one of three groups based on their age: 8-11 years (children, n = 7); 12-18 years (adolescents, n = 37); 18+ years (adults n = 43). Participants were presented a selective attention task (Woods and McDermott, 2015, doi: 10.1016/j.cub.2015.07.043), as well as several speech perception tasks (in quiet, in speech-shaped noise and in the presence of one-talker interferer). Data are still being collected in the children group. So far, results indicate that adolescents remain poorer than adults at selectively focusing on a target auditory stream, with or without additional interfering stream. Additionally, our results confirm the protracted development of speech perception in noise until late adolescence. Interestingly, across all participants, there is a significant relationship between stream segregation and speech perception in noise. This is in line with previous studies showing that auditory scene analysis relies on selective auditory attention, an ability that develops until late childhood/adolescence.

P09 Does neural tracking of continuous speech indicate active distractor suppression?

Martin Orf[1], Ronny Hannemann[2], Malte Wöstmann[1], Jonas Obleser[1]
1. Department of Psychology, University of Lübeck, Lübeck, Germany | 2. Audiological Research Unit, WS Audiology-Sivantos GmbH, Erlangen, Germany

A listener's ability to deal with challenging multi-talker situations hinges on their attention resources. While the neural implementation of target enhancement is comparably well understood, processes that enable distractor suppression are less clear. Typically, distractor suppression is quantified by the difference of the behavioural or neural response to distractors versus targets. However, such a difference can be driven by either target enhancement, distractor suppression, or a combination of the two. Here, we designed a continuous speech paradigm to differentiate target enhancement (enhanced tracking of target versus neutral speech) from active distractor suppression (suppressed tracking of distractor versus neutral speech). In an electroencephalography (EEG) study, participants (N = 19) had to detect short repeats in the to-be-attended speech stream and to ignore them in the two other speech streams, while listening also to the content of the to-be-attended audio stream. The ignored speech stream was task-relevant (to-be-attended) in the previous trial and was task-irrelevant in the present trial. The neutral speech stream was always task-irrelevant. We used phase-locking of the EEG signal to speech envelopes to investigate neural tracking via the temporal response function of the brain. Behavioural detection of repeats indicated the suitability of the paradigm to separate processes of attending and ignoring. Sensitivity of behavioral responses according to the Signal Detection Theory revealed that the internal separation for attended versus neutral speech was larger than for attended versus ignored speech. Neurally, the attended stream showed a significantly enhanced tracking response compared to neutral and ignored speech. Unexpectedly, neural tracking did not reveal sizeable differences for neutral versus ignored speech. In sum, the present results show that the cognitive system processes to-be-ignored speech distractors differently from neutral speech. However, this is not accompanied by active distractor suppression in the neural speech tracking response.

P10 Does speech intelligibility depend on the spatial appraisal of the sound source in diffuse noise?

Nicola Prodi, Chiara Visentin, Matteo Pellegatti
Department of Engineering, University of Ferrara, Italy

In order to build up a spatial impression of the speech source the listener needs to extract several monaural and binaural cues from the signal. When the room impulse response is analyzed to this scope, it is known that some relevant cues are obtained by the specific elaborations of the early reflections, for instance exploiting the correlational mechanisms supporting binaural hearing. On the other hand, it is also known that the energetic balance between the early reflections and the reverberant tail has a strong impact on speech intelligibility, and that the ratio of the direct sound over reflections together with the direction of arrival of the reflections greatly affect spatial perception too. Thus, even in reverberant quiet conditions the interplay of spatial percepts and speech intelligibility is a complex issue, which is not much addressed

in the literature. When diffuse noise is added on speech, it is unclear whether and how the spatial appraisal of the source (e.g. percepts of distance, width, focus and envelopment) is altered, nor it is clear if this would affect speech intelligibility. In this work impulse responses with specular or scattered early reflections are used to manipulate the spatial percepts while keeping the energetic balance between early sound and the reverberant tail fixed according to the concept of sound clarity. Two different reverberant tails are employed, and several virtual sound fields are created. To this aim anechoic speech is convolved with suitable impulse responses; a quiet and a noisy background (SNR=-6dB) are used. Three sets of listening tests are pursued: (i) spatial percepts of distance, width, focus and envelopment are evaluated in quiet; (ii) the same percepts are evaluated in noisy conditions and (iii) speech intelligibility is measured in noisy conditions only. It is shown that the percept of distance does not differ significantly between quiet and diffuse noise, while width, focus and envelopment are significantly altered. The effect of noise on them largely depends on the nature of the early reflections. In some case diffuse noise restores spatial percepts that are unavailable in the reverberation-only (quiet) conditions. This finding is traced back to the unmasking of the early reflections which were masked by the reverberant tail. When speech intelligibility is measured, a modest improvement is obtained with scattered early reflections, but the direct association between the spatial percepts and speech intelligibility in diffuse noise is weak and limited only to focus.

P11 Speech in ‘Anti-Noise’: Intelligibility and listening-effort of shallow-envelope chimeras under active noise control headphones

Sohhom Bandyopadhyay, Nithin V. George
Indian Institute of Technology Gandhinagar, India

Active noise control (ANC) headphones are known to alter human speech perception abilities in certain noises, such as from aircrafts or traffic. Despite being studied in diverse functional contexts (e.g. pain perception), ANC headphones’ impact into the human auditory system have not been thoroughly studied. Speech intelligibility under various listening conditions is known to depend on the temporal envelope fidelity of the sound producing device.

in this study, we investigated the effect of ANC on speech envelope perception. We propose a type of ‘chimera’, which is a parametrically degraded speech-like signal constructed by combining the envelope at shallow modulation depths. Stimuli are created from a popular speech dataset called Coordinate Response Measure. By measuring the variation in intelligibility and listening effort ratings (NASA task load index) of these chimera speech under ‘ANC on’ and ‘ANC off’ conditions of a commercially available headphone, we provide the first evidence of the interaction between noise cancellation and envelope of the speech signal. Using a repeated measures statistical test on logit transformed aggregate intelligibility scores, small but significant change in intelligibility is seen across the ANC conditions at most of the modulation depths. Similarly the listening effort ratings also show some interesting changes in the various dimensions (effort, frustration, etc.) across ANC conditions.

Overall, the effect of the ANC control system on the playback signal is unknown. Past studies involving ANC headphones and human participants have largely ignored the change in shape (i.e envelope) or other attributes of the signals reaching the listeners' eardrum, and focused on end results alone. By systematically measuring the effect of ANC in a parametric degradation of the envelope, the present study opens up the investigation to a wider range of questions, starting with more specific changes to the envelope processing and also carrier (called "temporal fine structure") processing by the human brain. Given the rising popularity of ANC headphones among users living in noisy metro cities around the world, this work sheds some light on the potential advantages of using an ANC headphone over and above the comfort aspect. It also serves as a possible improvement to the assessment protocols for device manufacturers, to design more human oriented noise cancelling algorithms.

P12 Acceptable noise level comparison in pediatrics with cochlear implant and Bimodal fitting users

Samin Ashjaei, Hadi Hadipour
Shahid Beheshti Medical University, Tehran, Iran | Soroush Otolaryngology and Hearing Clinic, Tehran, Iran

Hamid Jalilvand
Shahid Beheshti Medical University, Tehran, Iran

Background and aim: Difficulty in understanding speech in the presence of noise, despite providing amplification, is one of the main complaints of hearing-impaired patients. Due to profound hearing loss in cochlear implant users, they are presumed to need higher signal-to-noise ratio for better performance. Wearing a hearing aid in opposite ear, which is called bimodal fitting, is a preferred way for cochlear implant users for the following reasons: (1) Prevents auditory nerve degeneration for possible next cochlear implant surgery; (2) As a complement in aided frequency range, more likely in lower frequencies; (3) Less invasive way to provide binaural hearing that can help speech-in-noise recognition. Acceptable noise level as a reliable test to assess noise tolerance while following speech, has not been used for bimodal vs. cochlear implant in pediatrics. The goal of this study was to determine whether bimodal using can help cochlear implanted pediatrics for more noise acceptance.

Subjects and methods: This descriptive-analytical study was conducted on 13 pediatrics with profound sensory neural hearing loss, with mean age of 8.5 years. Participants were cochlear implant users who wear a hearing aid for at least 4 hours a day in their opposite ear. Audiometry test was done in aided and unaided condition first. Then, Farsi Acceptable noise level test was conducted. Independent and paired t-tests and also repeated measures ANOVA were used to compare the results in two conditions.

Results: Most comfortable level is reduced by 1 dB in bimodal situations. Accepted background noise level while following continuous speech is increased for almost 0.5 dB. Acceptable noise level is reduced significantly.

Conclusion: The present study showed that cochlear implanted children with well aided hearing thresholds have acceptable noise level results almost like normal hearing ones. Bimodal fitting leads to more noise tolerance which can result in better speech in noise performance.

P13 The effect of voice familiarity via training on voice cue sensitivity and listening effort during voice discrimination of vocoder degraded speech

Ada Biçer, Thomas Koelewijn, Deniz Başkent
University Medical Center Groningen, University of Groningen, Netherlands

Understanding speech in real life can be challenging, such as in multiple talker listening conditions. Fundamental frequency (F0) and vocal-tract length (VTL) voice cues can help listeners segregate between talkers, enhancing speech perception in adverse listening conditions. Previous research showed that degradations of cochlear implant (CI) hearing reduce sensitivity to F0+VTL voice cues compared to normal hearing (NH), and in some listening situations, familiarity with a talker could provide an advantage. In this study, we investigated how voice familiarity could affect perceptual discrimination of voice cues, as well as listening effort, with or without vocoder degradations.

To establish voice familiarity, we implemented an implicit short-term voice training. Participants listened to a recording of a book segment that was presented for approximately 30 minutes, and to ensure engagement, they had to answer text-related questions. Following voice training, just-noticeable-differences (JNDs) for F0+VTL were measured with an odd-one-out task implemented as a 3 alternative forced choice adaptive paradigm. During the procedure, the reference voice either belonged to the trained voice or an unfamiliar voice, presented in both unprocessed and vocoder-degraded (12-band with low spread of excitation) versions. Effects of voice familiarity (trained and untrained voice), vocoding (non-vocoded and vocoded) and item variability (fixed or variable consonant-vowel triplets presented across three items) on voice cue sensitivity (F0+VTL JNDs) and listening effort (pupillometry measurements) were analyzed.

Results showed that F0+VTL JNDs were significantly larger for vocoded conditions than for non-vocoded conditions. With variable item presentations JNDs were significantly larger than fixed item presentations. Contrary to our expectations, voice training did not have a significant effect on voice cue discrimination. Peak Pupil Dilation response was significantly larger for vocoded conditions compared to non-vocoded conditions. Over the time course of pupil dilation response, analyzed with GAMM, there was a significant difference between untrained and trained voices while listening to vocoded speech. Specifically, pupil dilation was significantly larger during voice discrimination while listening to unfamiliar, vocoded voices than listening to trained, vocoded voices. However, there was no significant difference between conditions of voice training on the pupil dilation while listening to non-vocoded voices. These findings imply that, even in the absence of a clear benefit in behavioral measures of JNDs, voice discrimination among vocoded voices was less effortful with short-term voice training.

Funding: VICI grant 918-17-603 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw), the Heinsius Houbolt Foundation, and a Rosalind Franklin Fellowship.

P14 Background speech perception in the cocktail party

Orsolya Szalárdy[1], Brigitta Tóth[1], Dávid Farkas[1], Gábor Orosz[2], István Winkler[1]

1. Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Budapest, Hungary | 2. Universite Artois, Universite Lille, Universite Littoral Côte d'Opale, Unité de Recherche Pluridisciplinaire Sport Santé Société, Liévin, France

In the cocktail party situation, people with normal hearing can usually follow a single speaker among multiple concurrent ones. To this end, the auditory system must decompose the mixture of sounds into meaningful streams (auditory scene analysis) and select the one with the behaviorally relevant information. At the same time, processing the irrelevant streams should be suppressed to some degree in order to conserve capacities and prevent distraction. However, there is no agreement in the literature as to whether the background is segregated into multiple streams/speakers. The current study varied the number of concurrent speech streams and investigated target detection and memory for the contents of a target stream as well as the processing of distractors. A male-spoken target stream was either presented alone (single-speech), in parallel with one male-spoken (one-distractor), or with a male- and a female-spoken distractor (two-distractor). Behavioral measures of target detection and content tracking performance as well as target- and distractor detection-related ERPs were assessed. We found that the detection sensitivity and the target N2b amplitude decreased whereas the P3b amplitude increased from the single-speech to the concurrent speech streams conditions. Importantly, the behavioral distractor effect differed between the conditions with one- vs. two-distractor (distraction by the female speaker was lower than that of the male speaker in either condition) and the target N2b elicited in the presence of two distractors was significantly smaller than that elicited in the presence of one distractor. Further, the voltage in the N2b time window significantly differed between the one- and two-distractor conditions for the same (M2) speaker. These results show that speech processing was different in the presence of one vs. two distractors, and thus, the current data suggest that the two background speech streams were segregated from each other.

P15 Children's use of target/masker F0 contour differences during speech-in-speech recognition

Mary Flaherty

University of Illinois at Urbana-Champaign, US

This study investigated the extent to which children (n = 98, ages 5-17 yrs) can take advantage of differences in fundamental frequency (F0) contour to improve speech-in-speech recognition. F0 contour refers to the natural variation, or rise and fall, of F0 within an utterance. While talker differences in mean F0 can improve speech-in-speech recognition for adults, children's ability to use mean F0 differences remains immature into adolescence. One explanation for this age effect is that children may rely more than adults on dynamic, time-varying acoustic cues that contain redundant information. Examining F0 benefit in children, Flaherty et al. [2019, Ear Hear. 40(4):927; 2021, JSHLR 64(1):206] carefully controlled voice characteristics to isolate effects of mean F0, leaving F0 contours of the utterances unaltered. In natural speech,

F0 co-varies with duration and intensity. Children's ability to segregate target from masker speech was expected to improve as a function of the magnitude of time-varying differences in F0. In the present study, sentence recognition was measured adaptively in a two-female-talker speech masker. Both target and masker sentences were recorded with either neutral, flat, or exaggerated F0 contours. Adults (n=30) were also tested as a measure of mature performance. The results revealed that children's sentence recognition was impacted by differences in F0 contour depth between competing talkers, but the pattern of results differed between children and adults. While both children and adults benefitted to a similar degree when the target speaking style was Flat, age effects were observed in conditions with neutral and exaggerated speaking styles. Contrary to our hypothesis, children did not show a consistent benefit when the speaking style was exaggerated, but instead often showed a decrement in performance. This may reflect that the sentences with an exaggerated F0 contour have F0 trajectories that are less predictable, thus increasing stimulus uncertainty and making speech-in-speech recognition more difficult for children. Overall, the observed age effects in the current study do not appear to be due to limitations in children's ability to use F0 contour differences in general, but are likely related to the perceptual salience of the target contour relative to the masker. This suggests that the use of F0 contour depth differences as a segregation cue during speech recognition develops relatively early compared to the use of mean F0 difference between target/masker speech (Flaherty et al., 2019; 2021). However, the results indicate that the magnitude of benefit for children depends on the predictability of the F0 contours in question.

P16 Decoding EEG responses to the speech envelope using deep neural networks

Michael D. Thornton[1], Tobias J. Reichenbach[2], Danilo P. Mandic[1]
1. Imperial College London, UK | 2. Friedrich-Alexander-University (FAU) Erlangen-Nuremberg, DE

Background: Unimpaired human listeners are remarkably good at attending to a target speaker whilst filtering out background sounds. Distinct representations of attended and unattended speakers can be decoded from electrophysiological recordings, and it is anticipated that advances in so-called auditory attention decoding (AAD) methodologies will one day lead to a neuro-steered hearing aid for hearing-impaired listeners. For this application, auditory attention decoding from EEG recordings is required to run with a high accuracy, a low latency, and it needs to work in various listening conditions. We present an investigation into how the use of deep neural networks (DNNs) might address these challenges.

Methods: We compared the standard linear technique of regularised least-squares (ridge regression) against two distinct neural networks for reconstructing the envelope of the attended speech stream from a listener's EEG recordings. Our dataset included several listening conditions: clean speech in native English and foreign Dutch; native speech in background babble noise; and a competing-speaker scenario. An additional clean-speech dataset was used to train listener-independent decoders.

Results: For reconstructing the envelope of clean speech, listener-specific DNNs were shown to offer a considerable improvement over listener-specific linear methods. Even when listener-independent methods were used, the DNNs performed significantly better than ridge regression. Furthermore, whilst the listener-independent methods were trained using EEG recorded under native clean speech conditions, they generalised well to new listeners and different listening conditions. The pre-trained DNNs achieved a significantly greater reconstruction score than the pre-trained ridge regressor across all listening conditions.

Conclusions: We showed that linear methods and deep neural networks which were trained to reconstruct the envelope of native clean speech from EEG recordings can be applied effectively across a variety of listening conditions. DNNs are known to suffer from overfitting issues and can generalise poorly, so it is significant that the deep neural networks were capable of generalising between listeners and listening conditions. These results therefore suggest that the use of DNNs offer good prospects for real-world auditory attention decoding.

Funding: This work was supported by the UKRI CDT in AI for Healthcare http://ai4health.io (Grant No. EP/S023283/1).

P17 Influence of three auditory profiles on aided speech perception in different noise scenarios

Mengfan Wu[1,2], Oscar M. Cañete[3], Jesper Hvass Schmidt[2,4,5], Michal Fereczkowski[1,2], Tobias Neher[1,2]

1. Institute of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark | 2. Research Unit for ORL – Head & Neck Surgery and Audiology, Odense University Hospital, Odense, Denmark | 3. Hearing Systems, Department of Health Technology, Technical University of Denmark, Lyngby, Denmark | 4. Department of Otolaryngology, Head & Neck Surgery and Audiology, Odense | 5. OPEN, Odense Patient data Explorative Network, Odense University Hospital, Odense, DenmarkUniversity Hospital, Odense, Denmark

Hearing aid (HA) users differ greatly in their speech-in-noise (SIN) outcomes. This could be because the degree to which current HA fittings can address individual listening needs differs across users and listening situations. In two earlier studies, an auditory test battery and a data-driven method were developed for classifying HA candidates into four distinct auditory profiles differing in audiometric hearing loss and suprathreshold hearing abilities. This study explored aided SIN outcome for three of these profiles in different noise scenarios. Thirty-one older habitual HA users and six young normal-hearing listeners participated. Two SIN tasks were administered: a speech recognition task and a “just follow conversation” task requiring the participants to self-adjust the target-speech level. Three noise conditions were tested: stationary speech-shaped noise, speech-shaped babble noise, and speech-shaped babble noise with competing dialogues. Each HA user was fitted with three HAs from different manufacturers using their recommended procedures. Real-ear measurements were performed to document the final gain settings. The results showed that HA users with mild hearing deficits performed better than HA users with pronounced hearing deficits on the speech recognition task but not the just follow conversation task. Moreover, participants with pronounced hearing deficits obtained different SIN outcomes with the tested HAs, which appeared to be related to differences in HA gain. Overall, these findings imply that current proprietary fitting strategies are limited in their ability to ensure good SIN outcomes, especially for users with pronounced hearing deficits, for whom the choice of device seems most consequential.

P18 The ‘Wooftail’ party effect: Effect of Canis Familiaris barks on speech perception

Jim Saroj Winston
Nitte Institute of Speech and Hearing, Nitte University, Mangaluru, India

Akhila Rahul
Mangaluru, India

Dogs (Canis Familiaris) are an integral part of many households and hence aptly referred to as ‘man’s best friend. Such close interactions often lead to competing situations where human communication has to happen alongside dog communication. Though the effect of environmental noises on human communication has been studied before, there is a lacuna of reports on the effect of background cross-species vocalizations on human speech perception. Hence, the present study explored the effect of dog barks as a competing signal on human speech perception. Twenty-two normal-hearing native Malayalam speakers (18-25 years) participated in the speech identification experiment. Nine pseudo-words of CVCV syllable shape served as the speech identification stimuli. The syllable /Ba/ served as the first syllable for each word, and the nine most occurring consonants in Malayalam with vowel /a/ constituted the second syllable. Dog barks elicited and recorded in ‘stranger’ emotion served as the competing stimuli. The dog barks from different recording sessions mixed at equal levels made the three dogs and five dogs bark. The stimuli were presented in silence and three dog bark conditions at -10dB SNR (one dog, three dogs, and five dogs). Each of the nine CVCV’s was presented three times across four conditions, making 108 stimulus presentations. The participants responded using a closed set identification task wherein the CVCV’s were listed orthographically, and the participants selected the appropriate option. The results reveal a significant effect of dog barks on the perception of bi-syllabic pairs across all test conditions. The findings will be discussed based on dog bark-induced confusions in the auditory perception, implications of findings, and future directions.

P19 Perception of low pass filtered speech in Hindi: A cross-sectional study across age

Kiran Kumari, Divya Sethi, Ansila Raslan, Prashanth Prabhu
All India Institute of Speech and Hearing, Mysuru, India

Spectrally modified speech stimuli like filtered speech have been used as a monaural low redundancy test to assess auditory processing disorder (APD). Filtered speech helps to understand the contribution of different frequencies in the perception of speech. Approximately 490 million people speak the Hindi language across the world. There are limited studies that have attempted to determine the effect of age on the perception of low pass filtered speech. Hence, the current study aimed to observe the effect of age on low pass filtered speech scores. Ninety individuals having normal hearing (hearing sensitivity less than or equal to 15 dB HL) participated in the study. Thirty were children in the age range of 7-12 years, 30 were adolescents in the age range of 12-18 years, and 30 were adults in the age range of 18-30 years. Twenty-five phonetically balanced words in Hindi were low pass filtered with a cut-off frequency of 800

Hz and 1000 Hz. The scores obtained were compared across the three age groups for both cut-off frequencies. The study results showed that the scores increased across age, and children obtained the lowest scores, followed by adolescents and adults. However, there was no significant difference in scores between adolescents and adults. Similar results were obtained for the two cut-off frequencies. Thus, the results of the study show that the low-pass filtered speech scores increase with age and reach adult-like scores by the age of 12 years. This is consistent with the neurophysiological studies which report that temporal cortex is matured and adult-like by the age of 12 years. The results of the study are essential to understand the developmental trend for the perception of filtered speech in Hindi. This information can be used effectively while interpreting the test results in a disordered population.

P20 The impact of visual, acoustic and semantic cues on processing of face mask speech by children and adults

Julia Schwarz[1], **Katrina K. Li**[1], Jasper Hong Sim[1], Yixin Zhang[1], Elizabeth Buchanan-Worster[2], Brechtje Post[1], Jenny Gibson[1], Kirsty McDougall[1]

1. University of Cambridge, Cambridge, UK | 2. MRC Cognition and Brain Sciences Unit, Cambridge, UK

Emerging research indicates that face masks can cause language processing difficulties (Brown, van Engen & Peelle, 2021, Cogn. Research 6(1):49). However, it is still unclear to what extent these difficulties are caused by the visual obstruction of the speaker's mouth or by changes to the acoustic signal. Moreover, research in this area has so far concentrated on adults' masked speech perception, but not children's. The present study investigated the extent to which children and adults process masked speech more slowly than normal speech, whether this effect is due to missing visual cues or acoustic degradation, and whether the effect is reduced in sentences with high semantic predictability. Since children are somewhat less experienced in using semantic cues for predictive speech processing (Hahne, Eckstein & Friederici , 2004, JoCN 16(7):1302), they could be affected differently by these combined factors than adults. The study was conducted on the online experiment platform Gorilla. Videos of a female British English (BE) speaker were presented to BE children (age 8-12) and adults (age 20-60). Participants performed a cued shadowing task in which they had to repeat the last word of the English sentences presented in the videos. Target words were embedded in sentence-final position and manipulated visually, acoustically, and by cloze probability (high/low predictability of the target word; Kalikow, Stevens & Elliott, 1977, JASA 61(5):1137). In order to capture millisecond-accurate voice response times online, a sound signal was embedded at the beginning of each sentence and recorded together with participants' vocal responses. Reaction times were then extracted automatically (combined with manual corrections) and analysed with mixed effects models. First results from 16 adults and 16 children (half the sample size) showed that listening to speech through face masks slowed down processing in both groups. However, visual and acoustic mask effects individually were very small (10ms), and only in combination displayed a moderate effect size (40ms). Visual, acoustic, and semantic cues all significantly reduced adverse mask effects, suggesting that listeners can compensate for acoustic changes by utilising both visual and semantic cues. Although children were less proficient in predictive processing overall, they were still able to use semantic cues to compensate for face mask effects in a similar fashion to adults. These findings provide novel information on the integration of multiple linguistic cues in adverse listening conditions across the lifespan and have practical implications for improving communication with face masks in educational settings such as classrooms.

P21 Pupil measures predictive of hearing status-related listening effort as an output of the machine learning classification framework

Patrycja Książek[1,2], Adriana A. Zekveld[1], Dorothea Wendt[2,3], Thomas Koelewijn[4], Sophia E. Kramer[1]

1. Amsterdam UMC, Vrije Universiteit Amsterdam, Otolaryngology – Head and Neck surgery, Ear & Hearing, Amsterdam Public Health research institute, Netherlands | 2. Eriksholm Research Centre, Snekkersten, Denmark | 3. Department of Health Technology, Technical University of Denmark, Denmark | 4. Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

An increasing number of pupillometry studies have shown that hearing impaired (HI) listeners may differ from normal hearing (NH) peers in the effort spent during listening. While multiple pupil measures have been shown to be sensitive to the hearing status, it is currently not clear whether these measures relate to hearing-related changes in speech processing. Since changes in auditory processing are present every time HI listeners attempt to process the speech, it is important to identify which, if any, pupil measures (e.g., mean pupil dilation, principal components) are most distinct for HI versus NH listeners at the trial-level. The next step would be to test whether these measures are generalizable across listening situations. In this study, we tested feasibility to reliably distinguish HI versus NH listeners based on a collection of pupil measures recorded in adverse listening conditions. Besides, we investigated the relative predictive value of these measures in classifying hearing status. We used a machine learning classification framework to classify hearing status (NH; HI) based on trial-level pupil responses recorded during a speech-in-noise test. Data were collected by Koelewijn et al. (2012, doi: 10.1155/2012/865731; 2014, doi: 10.1121/1.4863198) in 32 NH (31-76years) and 32 HI (40-70years) listeners. We used commonly used measures of listening effort (Peak Pupil Dilation, Mean Pupil Dilation, Pupil Baseline) as well as temporal measures (Principal Component Analysis, Independent Component Analysis). To identify pupil measures specific to hearing status and furthermore determine measures' sensitivity to performance and signal-to-noise ratio (SNR), we performed three classification tasks on subsets of pupil responses. We either included all trials at two certain average intelligibility levels, or included only correct trials from the same average intelligibility levels, or included correct trials at a single SNR level. Lastly, we ranked pupil measures based on their importance in the classification process. We expected pupil measures to differentiate the hearing groups, especially when fixating performance and SNR. As hypothesized, classification performance was always above the baseline prediction. This indicates that pupil measures tap into differences in speech processing between the tested groups (HI, NH). Fixating performance or SNR did not increase the clas-

sification performance. Some measures (e.g., second principal component) was found to be important across classification conditions. Our results indicate that a machine learning classification framework might be able to aid in the automatic detection of hearing status based on the pupil responses recorded in a speech-in-noise test.

P22 Does reverberation have a worse effect on speech intelligibility in noise for hearing impaired listeners than for normal hearing listeners?

Raphael Cueille, Mathieu Lavandier
Univ. Lyon, ENTPE, LTDS, France

Nicolas Grimault
CRNL, UMR CNRS 5292, Univ. Lyon 1, France

Hearing-impaired people often complain that they have a hard time understanding speech in noisy environments and in rooms with high reverberation levels. There are several effects of reverberation which can reduce speech intelligibility in noise. Most of them have been studied with normal-hearing listeners. In the case of hearing-impaired listener these effects are often mixed, and it is hard to determine in which noise environments and rooms hearing impaired listeners have more difficulties to understand speech. The goal of this study was to determine the importance of two monaural effects of reverberation for normal hearing and hearing-impaired listeners. The first one is the temporal smearing of the target speech, which decreases its intelligibility. The second one is the temporal smearing of a modulated noise, which makes it more difficult to listen in the masker dips. To separately evaluate these two effects, reverberation was applied either on the target speech, on the noise masker or on both sources. Bayesian analyses were done to compare the intelligibility scores of the normal hearing and hearing-impaired listeners in these situations.

P23 Investigating the effects of temporal and semantic predictability on the cortical tracking of speech

Lars Hausfeld[1], Rosie Coppieters[2], Ahmad Jibai[2], Lars Riecke[1]
1. Dept Cognitive Neuroscience, Maastricht University, The Netherlands | 2. Maastricht Science Program, Maastricht University, The Netherlands

A common element of current theories of speech perception is speech-envelope tracking, the alignment of neural activity to the slow amplitude fluctuations of a speech signal. Cortical speech-envelope tracking has been shown to depend on various acoustic and linguistic speech-signal features as well as cognitive factors. In the present study, we investigate whether cortical speech-envelope tracking depends on the temporal and semantic predictability of the speech input. 23 normally-hearing participants listened to various 30-second passages of

continuous natural speech, while their cortical tracking of the speech envelope was measured using scalp electroencephalography (EEG). Predictability was experimentally manipulated by gradually increasing listeners' familiarity with the speech input over time; i.e., each speech passage was presented five times in direct succession. To disentangle effects of semantic and temporal predictability, a control condition allowing temporal, but not semantic, predictability was included (unintelligible noise-vocoded speech carrying the same envelope as the original speech). Analysis of EEG responses across speech-envelope repetitions reveals higher positive correlations among repetitions of natural speech than repetitions of vocoded speech. Relatively weak correlations are observed between repetitions of the same envelope in natural and vocoded form. These preliminary results suggest that speech intelligibility strongly contributes to the cortical representation of the speech envelope. Ongoing analyses test whether the observed correlation depends on predictability, as operationalised by repetition number. In addition, the repetitions allow analysing the performance (noise) ceiling observable in the current EEG data.

P24 Deep neural networks for speech enhancement in noise

Lars Bramsløw
Eriksholm Research Centre, Oticon A/S, Denmark

Gaurav Naithani, Tuomas Virtanen
Computing Sciences, Tampere University, Finland

Deep neural networks (DNN) have demonstrated substantial user benefits for speech-in-noise enhancement for hearing-impaired listeners and voice-on-voice enhancement. In a previous study [Bramsløw et al., 2018, JASA 144(1):172-185, doi:10.1121/1.5045322], two known talkers were separated, using different types of low-latency DNN algorithms. A 37%-point improvement in word recognition in 15 hearing-impaired listeners was shown when selecting one voice, and a 13%-point improvement in the same listeners when presenting the two separated talkers dichotically to the two ears. The present work uses similar DNN architectures for enhancing speech from more common noise types: a party noise and a shopping centre ambient noise. Speech from a 12-talker Danish HINT corpus was used for training; two male and two female talkers from the same corpus were used in the listening tests, recording the word recognition scores from 21 hearing-impaired listeners. Five different low-latency DNN architectures with time-frequency masking were tested, employing both talker-dependent and talker-independent training. In party noise, a statistically significant word -recognition improvement of 16%-points was found for a talker-dependent and a talker-independent DNN type, while two other DNN types showed a statistically significant improvement of 12%-points. In the more stationary shopping centre noise, no improvements were found. It was assumed that the more modulated party noise provided more glimpsing opportunities for the DNN algorithm compared to the shopping centre noise.

P25 Investigating auditory selective attention using musical scenes: effect of distraction by concurrent melodies

Mathilde de Kerangal
Imperial College London, London, UK

Tobias Reichenbach
Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany

Auditory scene analysis is a fundamental aspect of music perception. Polyphonic music consists of the superposition of independent melodic streams and can be used as a tool to study the perception of multi-source scenes. The current study aimed at imitating the challenges of a complex environment (like a cocktail party) and, for that purpose, used musical stimuli in a selective attention task. More specifically, the effect of distraction by concurrent melodies was investigated by (1) increasing the number of distractor melodies and (2) varying the distance in frequency between the distractor and the target melodies. In the first experiment, the participants (N=15) were presented with a selection of chorales of JS Bach containing 1, 2 or 4 parts playing concurrently. The task was to track a target melody - the “leading” part (highest-pitched melody) or the “bass” part (lowest-pitched melody) - and indicate each time two consecutive notes had the same pitch. The performance decreased as the number of distracting melodies increased. Additionally, this decrease was sharper when the subjects were instructed to focus on the bass part; in other words, participants were more distracted when they had to ignore higher pitched melodies compared to lower pitched ones. These effects could be explained by the variation of signal to noise ratio resulting from the addition of auditory sources. They could also be explained by an effect of density related to the addition of sources close to the target stream – a second experiment was designed to isolate this effect of density. In the second experiment, the participants (N=15) were presented with chorales containing 2 parts - the leading and the bass part. The distance in frequency between the two parts was systematically varied resulting in a “dense” and a “sparse” condition. In the dense condition, the distracting part was moved one octave up or down so that it was closer to the target melody; the sparse condition included the leading and bass parts initially present in the chorales. The performance decreased in the dense condition compared to the sparse one, demonstrating an effect of informational masking related to the distance to the target melody. The effect was similar when the participants had to focus on the leading and bass melodies. In summary, the current study shows that the number of sources as well as the frequency distance between sources impacts our ability to selectively attend to a target sound.

P26 Processing fast speech: Investigating challenges for children with hearing loss

Rosanne Abrahamse, Titia Benders, Katherine Demuth, Nan Xu Rattanasone
Department of Linguistics, Macquarie University, Sydney, Australia

Children with hearing loss (HL), wearing hearing devices, encounter spoken language at varying speaking rates in everyday life. Faster rates may contain altered or reduced acoustic cues, leaving them with less time to process language. Degraded hearing may also increase this challenge. While results of children with normal-hearing (NH) show that they can process fast and slow speaking rates with similar speed, it is unclear how children with HL might perform. The current study examines effects of speaking rate on sentence processing in children with HL vs. NH. We predicted that children with HL would process sentences more slowly than children with NH, especially at faster speaking rates. We also explored the effect of hearing characteristics (unilateral HL, bilateral cochlear implants (CIs), bilateral hearing aids (HAs) or bimodal fitting) on performance. Thirty-one children with NH (mean age = 10.0, s.d. 1.65) and 36 children with HL (mean age = 10.4, s.d. 1.6) participated in an Auditory Word Detection Task. They were instructed to press the spacebar when they heard a target word in sentences presented at naturally produced Normal (4.4 syllables/second) vs. Fast speaking rates (6.1 syll./s). Response time (RT) was taken as a measure of processing speed. Data were analysed using mixed-effects models. Model 1 included Age, Group and Rate as fixed factors. In Model 2, participants were divided according to hearing characteristics, comparing each sub-group against the NH group. Results of the first model showed that children with HL were slower to detect words than their NH peers. There was no effect of Rate, nor an interaction between Rate and Group. In the second model, although children with unilateral HL did not differ from NH children, children with bilateral CIs, bilateral HAs and bimodal fitting were significantly slower. No significant interactions with Rate were found. These results suggest that children with HL process sentences more slowly than their NH peers, an effect driven primarily by children with bilateral HL. However, contrary to our predictions, children with HL were not disproportionately affected by fast speech. This contrasts with reported delays in fast speech processing for elderly listeners with HL. Greater availability of cognitive resources might play a role in explaining these differences. Thus, while language processing ability of children with bilateral HL is not on par with NH peers, our results provide a positive outlook on their ability to deal with faster speaking rates in communication.

P27 The Neurogram — a quantification of real-life hearing impairments based on electrophysiology

Fabian Schmidt[1], Lisa Reisinger[1], Ronny Hannemann[2], Nathan Weisz[1]
1. Centre for Cognitive Neuroscience, University of Salzburg, Austria | 2. Audiological Research Unit, Sivantos GmbH, Germany

The current gold-standard for the diagnosis of hearing loss is pure-tone audiometry. Yet, the artificial pure tones used to assess hearing thresholds in pure-tone audiometry do not resemble real-life listening situations. Therefore pure-tone audiometry only provides an incomplete picture of individual hearing impairment, as disorders such as supra-threshold hearing loss (i.e. hidden hearing loss) can not be captured. Additionally, pure-tone audiometry is vastly dependent on subjective feedback. This can be problematic as giving informed feedback is challenging for some patient groups (e.g. babies that are born deaf or elderly people with dementia). Here we propose the "Neurogram", a possible way to overcome the shortcomings of pure-tone audiometry by using a combination of system identification approaches, magnetoencephalography and a naturalistic listening situation (a radio play). By subsequently fitting linear encoding and decoding models we regress features of an acoustic signal (e.g. spectrograms) from related measured brain activity. We find that the decodability of acoustic information decreases with individual hearing capacity/impairment measured using pure-tone audiometry. Furthermore, we found a stronger relationship between subjective reports of speech perception (assessed using the Speech, Spatial and Qualities of Hearing Scale) and the here proposed "Neurogram" compared to pure-tone audiometry. In the future we aim to further develop this approach and work towards a diagnostic procedure that allows clinicians to fit hearing aids optimally based on a characterization of individual hearing impairment without solely relying on subjective feedback.

P28 Sentential contextual facilitation of auditory word processing builds up during speech parsing

Min Wu[1], Hans Rutger Bosker[2,3], Lars Riecke[1]
1. Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Netherlands | 2. Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands | 3. Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, Netherlands

During the parsing of auditory speech, auditory input is processed more effectively near the end (vs. beginning) of sentences. It is still unclear from which processing level these temporal dynamics in auditory processing originate. We investigated whether auditory word-processing dynamics during sentence parsing can be driven exclusively by predictions derived from sentential context. We presented listeners with auditory stimuli consisting of word sequences (arranged into either coherent sentences or unstructured, random word sequences) and a continuous distractor tone. We recorded reaction times (RTs) and frequency-tagged neuroelectric responses (auditory steady-state responses, ASSRs) to individual words at multiple temporal positions within the sentences, and quantified sentential context effects at each position while controlling for individual word characteristics (i.e., phonetics, frequency, and familiarity). We found that sentential context increasingly facilitates auditory word processing on linguistic and acoustic levels as evidenced by accelerated RTs and increased ASSRs across words within the sentences. These purely top-down contextually driven auditory word-processing dynamics were modulated by the syntax of the speech and occurred only when the listeners focused their attention on the speech. Moreover, they did not transfer to the auditory processing of the concurrent distractor tone. These findings indicate that dynamic linguistic and acoustic processing of auditory input during speech parsing can be driven exclusively by sentential predictions. The predictions may be shaped by the syntax of the speech, require the listener to actively parse the speech, and affect only the processing of the parsed speech, not that of concurrent yet perceptually separate auditory streams.

P29 Mapping of adaptive SNR enhancement in modern hearing aids: Effects of compression and choice of noise reduction strategy

Marie Frederikke Garnæs[1], **Sébastien Santurette**[1,2]
1. Centre for Applied Audiology Research, Oticon A/S, Smørum, Denmark | 2. Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark

For people with sensorineural hearing loss, the ability to understand speech in the presence of noise can be significantly degraded. In noisy sound environments, the low signal-to-noise ratio (SNR) causes reduced speech intelligibility and increased listening effort. While modern hearing aid (HA) technology accommodates this by enhancing SNR in aided listening, individual listeners may have different needs for SNR improvement in a given listening situation. Modern premium HAs can be programmed to obtain a wide range of SNR improvement patterns to accommodate a multitude of individual hearing abilities. This study investigated the amount of SNR enhancement provided by noise reduction strategies consisting of a combination of beamforming (BF) and postfiltering (PF) algorithms. For this purpose, the SNR at the output of recent premium HAs was measured across a wide range of input SNRs and BF+PF activation settings, using the Hagerman & Olofsson phase inversion technique in an ecologically valid speech-test setup. The aims were to document the achievable SNR enhancement of current algorithms, to compare the SNR benefit of traditional vs deep-neural-network (DNN) based PF algorithms, and to estimate the effects of wide-dynamic range compression on output SNR. The results confirmed that a wide range of SNR enhancement patterns could be obtained depending on the desired level of help in noise provided to the user. SNR enhancements were more prominent with decreasing input SNRs and the tested BF+PF algorithms could provide up to 10 dB SNR enhancement at low input SNRs. The DNN-based PF was shown to provide a relatively stable SNR enhancement of up to 2 dB across all input SNRs on top of the enhancement provided by the BF, largely improving the enhancement provided by the PF relative to the traditional PF approach. In accordance with previous literature, applying compression decreased output SNR at positive input SNRs, which is an unintended consequence of the feature. These SNR-reducing effects of compression typically increased with higher BF+PF activation. Overall, these technical measurements demonstrate the wide range of possibilities for individualization of advanced noise reduction settings available in the latest HA technology.

P30 The impact of the COVID-19 pandemic on objective and subjective hearing ability in younger and older adults in the UK

Kate Slade[1], Christopher Plack[1,2], Charlotte Pennington[3], Robert Davies[1], Helen Nuttall[1]
1. Lancaster University, UK | 2. University of Manchester, UK | 3. Aston University, UK

During COVID-19, social interactions have reduced and become increasingly virtual. Older adults may be disproportionately affected by these changes. Outside a pandemic, socialisation may mediate the relationship between auditory and cognitive ageing. As such, it is important to understand how enforced isolation affects auditory functions in younger and older adults. The COVID-19 Social Hearing Study is an ongoing online study investigating whether auditory function in younger and older adults has changed during 12-months of the pandemic. The study additionally measures auditory engagement, loneliness, depression, and cognitive function, to determine which factors contribute to any changes in auditory function, and how this is affected by age. We present preliminary findings from a subset of the collected data (58 younger and 62 older adults, over 8-months). Preliminary analyses indicate a significant effect of Time ($p < .001$), but not Age group, on engagement in recreational auditory activities. No significant effects of Age or Time are observed for subjective hearing ability as measured online by the SSQ-12 ($p > .05$). However, there is a significant Time × Age effect on objective hearing ability, as measured by a novel online Speech-in-Noise task ($p < .001$). This appears to be driven by performance fluctuations in the younger ($p < .001$), but not older group, which may be due to differences in motivation for listening effort. These preliminary data present encouraging indications of auditory resilience during periods of auditory deprivation, as well as representing opportunities for successful online auditory assessments in multiple age groups.

P31 Rollover effects at above-conversational levels in speech materials with low but not high context

Michal Fereczkowski, Tobias Neher
University of Southern Denmark, Odense, DK

The relation between speech intelligibility and stimulus level typically takes the form of a non-decreasing function, for both normal-hearing listeners and listeners with sensorineural hearing loss (SNHL). A performance decrease at high presentation levels – so-called rollover (RO) in the performance-intensity function – has traditionally been interpreted as a sign of retro-cochlear hearing loss. In some recent studies based on speech stimuli presented at high levels, RO was also observed in young listeners with normal audiograms, possibly reflecting cochlear synaptopathy. Overall, RO measurements could therefore be a useful tool for characterizing suprathreshold hearing abilities in different listener groups. While RO has been observed in studies that employed monosyllabic words or low-context sentences, reports of RO in measurements performed with high-context sentences are lacking. Here, we hypothesized that RO in the performance-intensity function is related to the amount of context information available in the employed speech material. To test this, 22 young adults with normal audiograms and without any self-reported hearing problems were tested at two presentation lev-

els: 80 and 95 dB SPL (broadband). Three speech materials were used: monosyllabic words, low-context sentences and high-context sentences. To avoid ceiling effects and upward spread of masking, all measurements were performed in stationary speech-shaped noise with the stimuli bandpass-filtered from 1.4 to 7.4 kHz . RO effects were found in the speech scores collected with the two low-context materials but not with the high-context material. Overall, this suggests that low-context sentence materials allow for sensitive RO measurements in individual listeners.

P32 Revealing attentional mechanisms in static and dynamic cocktail party listening by means of error analyses

Moritz Wächtler[1], Josef Kessler[2], Martin Walger[3], Hartmut Meister[1]
1. Jean Uhrmacher Institute, University of Cologne, Germany | 2. Department of Neurology, University Hospital Cologne, Germany | 3. Clinic of Otorhinolaryngology, Head and Neck Surgery, University of Cologne, Germany

Cocktail party situations involve multiple concurrent talkers, making listening perceptually and cognitively demanding. Whereas in a static cocktail party situation the talker of interest remains the same, dynamic situations include unpredictable switches of the target talker (e.g., Brungart and Simpson, 2007). In static situations, listeners have to focus attention on the known target and ignore competing talkers. In addition to this, dynamic situations also require monitoring multiple potential targets and switching the attentional focus from one talker to another, hence they are assumed to be more cognitively challenging than static situations (Lin and Carlile, 2015). In order to shed light on the attentional mechanisms in static and dynamic cocktail party situations, we conducted an analysis of error types that occur during multi-talker speech recognition. The cocktail party situations involved three spatially separated competing talkers with different voice characteristics who uttered matrix sentences. Differentiating between target-masker confusions and random errors, we aimed at obtaining insights about how attention is focused on the talkers. Using a more fine-grained analysis of confusion errors, we examined the occurrence of talker biasing effects, that is, if the listeners have the tendency to focus their attention on the previous target after a switch of the target (cf. Lin and Carlile, 2019). Furthermore, to investigate potential effects of aging and hearing status, data of three listener groups were compared, including young as well as older listeners with and without hearing loss. Finally, based on the models of auditory scene analysis (Bregman, 1990) and auditory attention (Shinn-Cunningham, 2008), we discuss in how far the observed error types indicate the existence of the proposed mechanisms.

Acknowledgement: Supported by a grant of the Deutsche Forschungsgemeinschaft to HM (ME-2751.3-1).

Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound, Auditory scene analysis: The perceptual organization of sound, The MIT Press, Cambridge, MA, US, xiii, 773 pages.

Brungart, D. S., and Simpson, B. D. (2007). “Cocktail party listening in a dynamic multitalker environment,” Perception & Psychophysics, 69, 79–91. doi:10.3758/BF03194455

Lin, G., and Carlile, S. (2015). “Costs of switching auditory spatial attention in following conversational turn-taking,” Frontiers in Neuroscience, 9, 1–11. doi:10.3389/fnins.2015.00124

Lin, G., and Carlile, S. (2019). “The Effects of Switching Non-Spatial Attention During Conversational Turn Taking,” Scientific Reports, 9, 8057. doi:10.1038/s41598-019-44560-1

Shinn-Cunningham, B. G. (2008). “Object-based auditory and visual attention,” Trends Cogn Sci, 12, 182–186. doi:10.1016/j.tics.2008.02.003

P33 Sentence comprehension and word recall in noisy primary classroom: What is the effect of the individual cognitive competences?

Chiara Visentin[1], Matteo Pellegatti[1], Maria Garraffa[2], Alberto Di Domenico[3], Nicola Prodi[1]
1. University of Ferrara, Italy | 2. University of East Anglia, UK | 3. University of Chieti, Italy

Students spend a large part of their time at school while listening to a teacher speaking with a background of noise and in poor acoustic conditions. The effort that is required for the basic tasks of hearing and decoding what the teacher says may leave fewer cognitive resources for other, more complex processes such as comprehension and memorization of the oral message. Theoretically, the Ease of Language Understanding framework suggests that the effect of listening conditions on auditory tasks might be modulated by the individual characteristics of the listeners, such as their linguistic abilities and cognitive skills. This mediation effect should be apparent of both accuracy in performing the task and listening effort. This study aims to explore whether and how the effect of the listening condition on a complex academic task is mediated by individual (i) cognitive skills (selective attention, working memory capacity), (ii) linguistic competences (reading comprehension), and (iii) self-rated noise sensitivity. To the scope, primary school children (N=120, grades 3 to 5) were presented with a sentence comprehension and word recall task in the presence of a two-talkers background noise. The level of the masker was modulated so as to obtain three different signal-to-noise ratios (SNR: +1, +5, and +9 dB) which are highly representative of a typical classroom environment. Data on accuracy and response time (RT) were acquired; for each listening condition the children were also asked to provide a self-rating of the perceived effort. Finally, measures of the individual cognitive profile were obtained in quiet conditions a week after the experimental activity. The preliminary results indicate a main effect of the listening condition on both accuracy and response time in the sentence comprehension task, with the lowest SNR yielding a poorer performance and slower RTs. A significant effect of the grade was also found on both dependent variables in the comprehension task and in the number of correctly recalled items in the memory task, indicating a general disadvantage of the youngest students compared with their grade 5 peers. Concerning the effect of their cognitive profile, only a significant effect of the linguistic competences was found on the accuracy in the comprehension task. A more detailed statistical analysis will allow to better disclose the interplay of external and internal factors on the performance in the task, yielding additional information on the mutual relationship between sound environment and student's behavior.

P34 Pre-target α EEG power predicts intra-individual variability in Digit-in-Noise recognition

Thomas Houweling[1,2], Robert Becker[1], Alexis Hervais-Adelman[1,2]
1. Neurolinguistics, Department of Psychology, University of Zürich, CH | 2. Neuroscience Center Zurich (ZNZ), University of Zürich and ETH Zürich, CH

Features of spontaneous EEG activity predict individual variability in Speech-in-Noise perception tasks. For instance, resting-state MEG power predicts individual differences in signal-to-noise ratio (SNR) thresholds for correct word recognition of words in noise. Further, mean α power at baseline predicts individual differences in Digit-in-Noise (DiN) reception thresholds. Moreover, properties of the ongoing EEG are predictive of moment-to-moment, within-individual variability in performance on several speech recognition tasks, including syllable discrimination and lexical decision in noise. We set out to investigate the hypothesis that trial-to-trial fluctuations in brain activity during a pre-target, noise-filled period, affect the probability of accurately reporting a target item. To this end, we assessed the effects of pre-stimulus EEG (128-channel EGI HydroCel) power on both recognition and subjective clarity of monosyllabic German digits, at three levels of noise masking (108 trials each), calibrated to individual digit-wise performance levels in a sample of 25 young normally-hearing listeners. First, we computed differences in grand average pre-target spectral power between correct and incorrect trials, and between subjectively clearer and less clear ones. This suggested behavioural relevance of α band power within the. We performed mixed-effects regression of comprehension and clarity on source-reconstructed, parcel-wise power, averaged between 8.5 and 12Hz in 50ms time bins between –300ms and stimulus onset. This revealed a significant ($p<.05$ after non-parametric cluster-based permutation) interaction between α power and SNR on comprehension in a broad left-lateralised temporo-parietal region: higher α power was associated with increased probability of correct responses at higher SNRs, and decreased probability of correct response at lower SNRs. These results indicate that pre-target α power, in areas known to be involved in various stages of speech processing, predicts moments of relatively improved or decreased DiN comprehension. α band power is known to be associated with decreased cortical excitability and sensitivity to input. Increased α power during the noise-filled pre-target period has been previously interpreted as a noise suppression mechanism, such that α power-associated reduction in the processing of the noise leads to enhanced representation of the signal. In contrast, the trade-off between noise suppression and signal detection observed in the current study supports the more parsimonious view that α power indexes a relatively unspecific gain control mechanism that is engaged to suppress noise, with potentially detrimental consequences for signal detection at very low SNRs.

P35 Foreground speech robustness for audio scene classification

Siyuan Song, Brecht Desplanques, Nilesh Madhu
Ghent University, Belgium

Identification of the background acoustic scene can significantly benefit speech enhancement in hearables. Existing acoustic scene classification (ASC) techniques show competitive performance when the characterising sound, namely the background signal, is dominant. However, the classification accuracy degrades significantly when foreground speech is present. We present investigations of two classes of techniques to alleviate this degradation. The investigations were carried out on a classical iVector-based system for two reasons (a) this baseline shows competitive performance and (b) since the working of this approach is explainable, the benefits of the techniques can be better understood. Firstly, a noise-floor based iVector system was proposed, where the Mel frequency cepstral coefficient (MFCC) features were derived from an estimate of the power spectral density of the background signal. In combination with Multi-Condition Training (MCT), this system improved the ASC performance when the foreground speech was predominant, but at the cost of poorer performance in the absence of foreground speech and in low speech to background ratio (SBR) conditions. To improve this trade-off, we consider the integration of a soft Voice Activity Detector (softVAD) in the classical iVector system. Based on MFCC features extracted from the microphone signals, a frame-level speech-absence probability is calculated using a universal background model (UBM), respectively, for speech and background. Based on this probability, weighted Baum-Welch statistics are computed and used in the iVector extraction stage. Thereby the background-dominant frames are emphasised while speech-dominant frames are disregarded. Experiments show that this system outperforms the noise-floor based system in a wide range of SBRs. Yet, we believe, the information in the speech-dominant frames can be exploited by using the noise-floor-based features in these conditions. This can further improve performance. Therefore, we present a third system where the score of the noise-floor based ASC system is combined with the score of the second system in a weighted manner. To allow the noise-floor based ASC system to focus on the information in the speech-dominant frames, the system is modified to incorporate the speech presence probability when computing the Baum-Welch statistics. Further, the weights for the score fusion are obtained from the average background probability of each segment. This weighted score fusion system achieves overall the best accuracy in tested SBRs. These findings indicate that temporal frame attention is important for robust ASC solutions and this is applicable to DNN-based frameworks as well. Extension to DNN-based systems will be the focus of future work.

P36 Individual variability of theta-band cortical entrainment to speech in quiet predicts word in noise comprehension and is mediated by top-down processes

Robert Becker, Alexis Hervais-Adelman
University of Zurich, Neurolinguistics, Dept. of Psychology, Zurich, CH

Speech elicits brain activity that is time-locked to its amplitude envelope, a signal dominated by modulations at the syllable rate. The resulting Cerebro-acoustic coherence (CACoh) is thought to be an essential feature of a speech parsing mechanism crucial to comprehension and has been examined especially in the context of degraded or noisy speech, which showed evidence for a functional meaning of CACoh. The functional role of CACoh – and its variation - in the context of noise-free speech is underexplored and not well understood. Here, we set out to examine magnetoencephalography (MEG) data of participants (n=53) while listening to clear speech stories. We examined CACoh using Gaussian copula mutual information to establish the extent of tracking at time constants consistent with the phrasal, lexical and syllable rates of the stimuli. At the syllable rate (3-7Hz), bilateral superior temporal gyrus significantly tracked speech information. Notably, individual differences in tracking were positively linked to recognition accuracy in a completely independent words-in-noise (WiN) task. Intriguingly, left STG CACoh was also positively associated with higher fluid intelligence. Further examining potential sources of task-related top-down communication to left and right STG , using Granger causality, we found that a specific pattern of top-down connectivity from frontal areas was indeed associated to WiN performance. Subsequent mediation analysis confirmed that the relationship of CACoh and WiN was at least partially mediated by this top-down connectivity pattern, while in contrast, a mediation of CACoh by WiN through bottom-up connectivity was absent. Thus, results indicate that individual variability of speech tracking in the auditory system - even during clear speech - reflects the potential robustness of the speech system to noise, and that top-down influences from brain areas associated with higher-order linguistic processing and articulatory processes contribute to speech-brain entrainment.

P37 Neural speech tracking in age related hearing loss as a function of hearing aid use

Vanessa Frei, Nathalie Giroud
University of Zurich, Switzerland

Age-related hearing loss (ARHL) is associated with difficulty understanding speech, particularly in the presence of background noise. Hearing aids do not (sufficiently) filter background noise, while the use of audio-visual (AV) speech cues can contribute greatly to speech comprehension in noise. Nevertheless, the underlying neural processes of the facilitative role of visual speech cues is not well understood, especially in hearing aid users. The aim of this study was to investigate the effects of AV vs. audio only (A) listening conditions on neural speech processing in two groups of hearing-impaired older adults: 1) hearing aid users and 2) without hearing

aid experience. To test neural speech processing, we used neural speech tracking, a measure that reflects the synchronization of low frequency auditory cortex activity with the temporal regularity of a continuous speech signal, which has been shown to be enhanced in ARHL and was correlated with better speech comprehension. Seventy-eight older adults (aged 64-80 years) with ARHL (31.05 - 59.6 dB HL averaged from 0.5-8 kHz) participated. Hearing aid user and those without hearing aid experience groups (both n = 39) were carefully matched for age and hearing loss. They were presented with natural sentences with babble noise (8 overlapping sentences; SNR = 0) in A and AV (showing speaker's mouth and chin) settings while EEG was recorded. An intelligibility task and a comprehension task were performed. Speech tracking will be estimated by cross-correlation, which assesses the similarity between two time series across time lags. Using generalized mixed models, there was a significant improvement in both tasks for both groups as a function of visual cues. Hearing aid users performed significantly better in the comprehension task, but significantly worse in the intelligibility task compared to those without hearing aid experience. We are currently investigating to what extent these effects are reflected in the neural tracking. We hypothesize that speech tracking will decrease as a function of hearing aid use, enhanced by AV information, which would suggest that the hearing aid partially substitutes neural compensation of ARHL. Answering these questions can provide valuable insight for future optimization of hearing loss treatment, particularly the role of AV information in hearing aid implementation.

P38 Does reverberation affect the bimodal or audiovisual speech intelligibility benefit in cochlear implant users?

Peter Ihly, Tim Jürgens
Institute of Acoustics, University of Applied Sciences Lübeck, Germany

Daniela Hollfelder, Karl-Ludwig Bruchhage
Department of Otolaryngology, University of Lübeck, Lübeck, Germany

Hartmut Meister
Jean-Uhrmacher Institut for Clinical ENT-Research, University of Cologne, Germany

Cochlear implant (CI) users with access to contralateral acoustic hearing show better speech intelligibility in noise when using both ears simultaneously (bimodal benefit). Moreover, intelligibility is typically better when listeners have visual access to simultaneous articulatory movements of the speaker (audiovisual benefit). This study investigated the effect of reverberation on the amount of either bimodal or audiovisual speech intelligibility benefit. Therefore, seven CI listeners (five bimodal, two single-side deaf) and six normal-hearing listeners (as controls) were tested with the German HSM sentence test in noise at individually fixed signal-to-noise ratios using stationary noise with long-term speech spectrum. Speech stimuli were presented via loudspeaker either dry or reverberant, i.e., convolved with a room-impulse response with reverberation time of 1 ms. In addition to the acoustic stimuli, simultaneous visual stimuli were presented over a flatscreen monitor in half of the conditions. CI users listened to the stimuli in three different modalities: acoustic-only, electric-only and bimodal. Preliminary results showed significant average audiovisual benefits of 13.4% in CI users and 13.6% in normal-hearing listeners. Bimodal benefit in CI users was 31.6% with respect to electric-only listening, and 31.5% with respect to acoustic-only listening. In comparison to reverberant speech, non-reverberant speech was on average 26.9% better intelligible for CI listeners. Significant interactions between the factors reverberation and visual cues ($p = 0.017$), but not between reverberation and modality ($p = 0.892$) were found, indicating that audiovisual benefit was larger with dry speech, in particular with electric-only listening. In addition, this study found that bimodal benefit was governed by better-ear-listening in dry conditions. In contrast, speech intelligibility was significantly higher in bimodal conditions than in the better ear in the reverberant condition, indicating a combined benefit.

P39 Investigations on the optimal estimation of speech envelopes in speech enhancement

Yanjue Song, Nilesh Madhu
IDLab, Department of Electronics and Information Systems, Ghent University - imec, Ghent, Belgium

Speech envelope plays an important role in speech intelligibility. Model-based speech envelope enhancement approaches, when applied to the noisy input, require an estimation of both the speech and the noise envelopes. This results in a computationally complex search, since the noise envelopes can have a wide variety of shapes. Speech envelopes, on the other hand, exhibit a much smaller variance, imposed by the physiological nature of speech production. Recent research has attempted to reduce the search space of envelopes by exploiting this characteristic. Typically, it is done in a two-stage manner. First, a preliminary noise suppression is applied, yielding a rough estimate of the underlying speech. In the second stage, instead of estimating the clean envelopes from noisy spectra, a coarse estimate of the clean envelope of the underlying speech is recovered from this preliminary denoised signal. The search for the true underlying envelope can then be cast as a regression approach that predicts the speech envelope directly, or as a classification problem which finds the best candidate from several templates, based on this coarse estimate. We present our recent investigations into the achievable benefits of envelope reconstruction, using oracle methods for both classes of approaches. These define the upper bound of the achievable performance. Further, we benchmark two practical, real-time systems using the classification approach against the oracle classification. The first system is based on statistical modelling of the features using the well-known Gaussian mixture models, with temporal context being described by hidden Markov models (GMM-HMM). The second system is a DNN-based system with roughly the same number of parameters as the GMM-HMM. We also study two different feature sets for representing the envelope by both the oracle tests and the practical systems – those based on LPC-features (which implicitly assumes an auto-regressive model for speech) and those based on cepstral coefficients. Oracle results demonstrate that direct prediction of envelopes outperforms the classification strategy. Envelope representation using the cepstral coefficients seems most robust for a wide range of noise conditions, especially at low SNRs. Results of the practical systems confirm this advantage. The results also demonstrate that deep-learning-based systems match the oracle performance of the classification approach. In future investigations, we would like to explore how envelope enhancement will further benefit from a regression model.

P40 Listening in spatialized noise sentences (LiSN-S) test in Brazilian Portuguese

Bruno Masiero
University of Campinas, Fac. Ele. Eng., Campinas, Brazil

Maria Francisca Colella-Santos, Letícia Borges
University of Campinas, Fac. Med. Sci., Campinas, Brazil

Rationale: Spatial processing involves the ability to understand speech in a noisy environment which is directly related to the listener's ability to use binaural clues to differentiate sound source location from noise location. The spatial processing can be assessed by the Listening in Spatialized Noise Sentences (LISN-S) test, a binaural interaction test, applied by computer software using headphones, which produces a three-dimensional virtual auditory environment to evaluate spatial processing in individuals with complaints related to Central Auditory Processing Disorder. The main purpose of this study was to develop the LISN-S test (software and database in Portuguese), including normality criteria for the age group of 6 to 11 years. In addition, the results obtained in normal children will be analyzed and compared with those who had a history of otitis media in childhood.

Design: The research will be developed in four steps: 1) Develop a software to assess spatial processing and speech material for the Portuguese database to be inserted into the software; 2) Determine the relative intelligibility of sentences to make adjustments to the recordings and obtain intelligibility equivalence between them; 3) Determine normality criteria for a age group of 6 to 11 years, and 4) Compare the results between children with and without a history of otitis media.

Results: The research was approved by the Institution's CEP, under No. 3,462,572. A graphical interface was designed with MATLAB App Designer and it allowed the examiner to start the equalization procedure and, for each sentence, to input the number of correct words repeated by the subject. 188 phrases prepared by the researchers were recorded by a vocal actress in an anechoic chamber. Two children's stories were selected and recorded by three vocal actresses using the same recording procedures. Children are being selected from a public school to develop steps 2 and 3.

Conclusion: Research is ongoing and software and speech material are ready for application in the selected sample. It is hoped that the results obtained in this research may provide support for understanding the functioning of the central auditory nervous system structures involved in binaural interaction tasks, from the cochlear nucleus to the auditory cortex in Brazilian children. In addition, we hope to disclose the importance of studying spatial processing, especially in children with complaints related to hearing difficulties in noisy environments, to contribute for the diagnosis and help audiologists to plan a more fully and efficiently rehabilitation.

P41 Background noise effects on the frequency-following response (FFR) of newborns: A comparison between neonates and adults

Alejandro Mondejar-Segovia[1], Roser Llobet-Gil[1], Amineh Koravand[2], Sonia Arenillas-Alcón[1], Jordi Costa-Faidella[1], Adél Könczöl[1], Teresa Ribas-Prats[1], M. Dolores Gómez-Roig[3], Carles Escera[1]

1. Brainlab – Cognitive Neuroscience Research Group. Institute of Neurosciences. Department of Clinical Psychology and Psychobiology, University of Barcelona, Catalonia, Spain | 2. University of Ottawa, Ontario, Canada | 3. BCNatal – Barcelona Center for Maternal Fetal and Neonatal Medicine (Hospital Sant Joan de Déu and Hospital Clínic), University of Barcelona, Catalonia, Spain

Background: People across the age span show increased difficulty hearing speech in the presence of background noise. Developmental deficits in early life could exacerbate it and lead to language difficulties later. Impoverished speech-in-noise (SIN) perception has also been linked to other disorders like autism, where it is hypothesized that this difficulty occurs along with other perceptual impairments. The Frequency-Following response (FFR) is an auditory event-related potential elicited by periodic auditory stimuli that can faithfully phase-lock the spectrotemporal characteristics of the periodic sound wave presented. Considering that disruptions in the FFR has been seen in a wide range of language disorders, this response could emerge as an effective tool to assess this phenomenon. Nevertheless, despite its importance, noise background influence in speech perception has never been investigated in newborns. The present study aims to compare how background noise could modify speech perception in newborns and adults through the analysis of their FFRs.

Method: FFR was recorded in both healthy term newborns (aged <48 hours after birth) and normal-hearing adults (aged 20-35 years). The auditory stimulus was a /da/ syllable with a duration of 170 ms and a fundamental frequency (F0) of 113 Hz, reproduced at 65 dB. The /da/ syllable is divided into two segments: a consonant transition section (10-57 ms) and a steady vowel section (57-170 ms). The noise condition was assessed by playing a Spanish six-talker babble noise at 55 dB. Different FFR parameters were retrieved from the recordings in time and frequency domains for both sections of the stimulus separately.

Results: Results showed that both newborns and adults exhibited larger signal amplitude to the speech stimulus in silence than when presented in noise, albeit newborns obtained consistently lower FFR parameter values than adults. However, the adult group showed greater impairments of speech encoding mechanisms in the noisy condition compared to the newborn group.

Conclusion: Speech encoding in newborns seemed to be less affected by the noise condition than adults, perhaps due to their auditory system only being exposed to attenuated sounds during gestation. This study constitutes one of the first steps towards understanding the development of SIN perception from the first days of life. Detecting FFR impairments in the processing of SIN in an early stage might help to prevent the emergence of future language difficulties, like specific language impairments or dyslexia, or even improve academic experience of healthy individuals.

P42 The role of speech prosody in stream segregation and selective attention in a multi-talker situation

Petra Kovács[1], Brigitta Toth[2], Orsolya Szalárdy[2], Ferenc Honbolygó[3], Anna Kohári[4], Katalin Mády[4], Lilla Magyari[5], István Winkler[2]
1. Budapest University of Technology and Economics, Budapest, Hungary | 2. Institute of Cognitive Neuroscience and Psychology Research Centre for Natural Sciences, Budapest, Hungary | 3. Brain Imaging Centre, Research Centre for Natural Sciences, Budapest, Hungary | 4. Department of Phonetics, Research Institute for Linguistics, Budapest, Hungary | 5. University of Stavanger, Stavanger, Norway

To process speech in a multi-talker environment, listeners need to segregate the mixture of incoming speech streams and focus their attention on one of them. Potentially, speech prosody could aid either of these processes, but the contribution of prosody to the processing of one speech stream out of several is still largely unknown. To address these issues, we extracted functional networks connecting brain regions from brain electric signals while participants listened to two concurrent speech streams. Prosody manipulation was applied on the attended speech stream for one group of participants and on the ignored speech stream for the other group. Speech stimuli were either synthetically F0-flattened, naturally prosody-suppressed, or intact. Our results show that speech prosody - especially the parsing cues mediated by speech rate - facilitate attentional selection, while playing a smaller role in stream segregation. These conclusions will be discussed in detail in the poster.

P43 How hearing aid users shape their acoustic environment - implications for measuring hearing ability and hearing aid benefit in everyday life

Iris Borschke[1,2,3], **Nadja Schinkel-Bielefeld**[1,4], Tim Jürgens[5,3]
1. WS Audiology | 2. Universität zu Lübeck | 3. Lübeck, Germany | 4. Erlangen, Germany | 5. Technische Hochschule Lübeck

Measuring hearing aid benefit in everyday life is at least as important as assessing performance under controlled, but artificial conditions in the laboratory. Often ecological momentary assessment (EMA) is employed for this, a method where participants repeatedly fill out self-reports about how they subjectively experience specific situations. As self-reports describe situations that are happening now or happened recently, memory bias is reduced. However, individuals - with or without hearing impairment - often have many possibilities to modify their acoustic environment (e.g., by closing a window when it is too noisy outside or by adjusting the volume of the TV) to reach a pleasant listening situation. Compared to laboratory experiments, these modifications may lead to an underestimation of hearing problems, as well as lesser discriminative power with EMA, when comparing hearing technologies. The goal of this study was to assess the number of modified situations by hearing aid users in everyday life, how strong the improvement through modification was perceived, and if both of these depended on hearing aid processing. Therefore, we performed a three-week EMA study with 29 experienced hearing aid users to analyze how hearing aid users modify their acoustic environment to optimize the listening situation. Participants were on average 66.4 years old (SD

13.2 years) and had a pure-tone average (PTA) hearing loss of 46.5 dB HL (SD 12.2 dB HL). They were fitted with hearing aids consisting of two programs with strong and weak directionality and noise reduction, called 'focus' and 'awareness', respectively. Programs changed automatically each night and could not be controlled by the participants. In addition to self-reports, we continuously collected the broadband levels and the classification of the situation by the hearing aids. On average participants reported 2.4 (SD 1.9) situations per day they modified or would have liked to modify. The number of reported situations did not significantly differ between hearing programs. In 62.6% of situations the acoustic environment was modified, which improved their reported pleasantness by 2.2 points (SD 1.9) on a 7-point Likert scale. Data logging showed that the focus program was used on average 1.4 hours longer per day than the awareness program.

P44 The effects of lifetime occupational noise exposure and age on speech perception and self-reported hearing symptoms: An online study in Palestine

Adnan Shehabi, Garreth Prendergast, Hannah Guest
Manchester Centre for Audiology and Deafness, University of Manchester, UK

Chris Plack
Manchester Centre for Audiology and Deafness, University of Manchester, UK. The Department of Psychology, University of Lancaster, UK

Background: Workers in developing countries are typically exposed to unsafe levels of occupational noise throughout their lifespan due to the lack of enforcement of occupational health and safety measures. The effects of occupational noise exposure may be more apparent in the older adult worker population and may manifest as decreased ability to understand speech, especially in noisy environments, tinnitus, and hyperacusis. In this study, we tested the hypotheses that there are effects of occupational noise exposure and aging on (i) speech-perception-in-noise (SPiN) thresholds, (ii) the presence of chronic tinnitus, (iii) self-reported hearing ability, and (iv) severity of self-reported hyperacusis.

Methods: We recruited 251 adults (females: 152, age range: 18-70, mean age: 35.1) who are workers in either noisy or non-noisy industries in the Palestinian Territories, with no past diagnosis of hearing or memory impairments. Subjects completed a set of online instruments in Arabic including an otologic health and demographic questionnaire, the online forward and backward digit span test, a noise exposure questionnaire (which evaluates lifetime occupational, recreational, and firearm noise exposure), the Khalfa hyperacusis questionnaire, the short-form Speech, Spatial and Qualities of Hearing Scale (SSQ12), the Tinnitus Handicap Inventory (THI), and an Arabic online digits-in-noise (DIN) test. Multiple linear regressions, including both age and occupational noise exposure as predictors, were employed to test hypotheses (i) (iii) and (iv), while logistic regression was used to test hypothesis (ii). The covariates of sex, recreational noise exposure, and cognitive function, as measured by the forward and backward digit span test score and the highest qualification of formal academic attainment, were accounted for in all the statistical models.

Results: Preliminary analyses showed that both occupational noise exposure and age were significant predictors of SPiN (DIN thresholds). Only age significantly predicted self-reported hearing ability as reflected by the SSQ12 scores, while only occupational noise predicted the severity of hyperacusis. Neither occupational noise exposure nor age were significant predictors of the presence of chronic tinnitus.

Conclusions: There are significant differences in hearing ability due to age and occupational noise exposure among adult workers in Palestine. Many workers in Palestine seem to suffer from the auditory effects of aging and occupational noise damage despite no past formal diagnosis. These findings highlight the importance of occupational noise monitoring and the implementation of hearing-related health and safety regulations in developing countries like Palestine.

P45 The dynamic range of speech

John Culling, Molly Bishop
Cardiff University

The speech intelligibility index assumes that the dynamic range of speech is 30 dB, but it is unclear what the empirical basis of this assumption is. Leclère et al. (2016) tested the range by using speech-in-noise stimuli for which the speech-to-noise ratio was strongly frequency dependent. The speech or the noise was filtered using a step-function filter that attenuated the level by a specified number of decibels above or below 1400 Hz. The results showed that as the size of this step change was increased, the speech reception threshold (SRT) changed progressively up to a certain level of attenuation and then stabilised. The range of attenuations over which changes were significant was around 40 dB, indicating that up to this point, the signal-to-noise ratios both above and below 1400 Hz were affecting intelligibility. This result suggests that 30 dB may be an underestimate of the dynamic range relevant to intelligibility. The present experiment took an analogous approach using temporal modulation of the masker. Interrupted noise reduces SRTs by around 20 dB compared to continuous noise of equal long-term intensity. Rather than fully interrupting the noise, it was subject to square wave modulation of controlled depth. Consistent with Leclère et al. (2016), increases in modulation depth up to 40 dB produced progressive changes in SRT. At 40 dB attenuation or greater, the full 20-dB effect of interrupting the noise was observed.

Leclère, T., Théry, D., Lavandier, M. and Culling, J. F. (2016) "Speech intelligibility for target and masker with different spectra" in van Dijk, P., Başkent, D., de Kleine, A., Gaudrain, E., Lanting, C. & Woldhuis, R. (eds) Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing (Springer), 257-266.

P46 Phonological but not lexical processing alters the perceptual weighting of F0 and VTL cues for voice gender categorisation

Almut Naja Jebens, Laura Rachman, Deniz Başkent
University of Groningen and University Medical Centre Groningen, Netherlands

Two acoustic voice cues primarily define the perception of voice gender: the mean fundamental frequency (F0), related to the glottal pulse rate and perceived as a speaker's pitch, and the formant frequency distribution defined by the vocal tract length (VTL), related to speaker's size and adding to the characteristic vocal timbre. F0 and VTL are used beyond voice gender categorization and can enhance intelligibility in conditions where multiple talkers speak at the same time. Previous research has shown that voice perception, tested in speaker discrimination or identification tasks, could be altered by linguistic processing. Related to this, the advantage of the familiar language in voice perception tasks has been described as a robust effect. However, it is not clear at which linguistic level this familiarity effect arises, and how this effect influences the perceptual use of certain vocal parameters such as F0 and VTL for recognizing characteristics of a speaker. Here, we studied the effects of lexical and phonological processing on the weighting of F0 and VTL cues for perceived voice gender categorization in normal-hearing native speakers of Dutch. In a two-alternative forced choice paradigm, participants listened to monosyllabic words and matching nonwords and performed a voice gender categorization task. Responses were limited to "woman" or "man". F0 and VTL in three female reference voices were manipulated to produce different voice samples on a female-male continuum. The perceptual weighting of F0 and VTL cues for these categorizations was estimated by the coefficients for F0 and VTL factors in a logistic regression model. Linguistic effects were tested by comparing F0 and VTL weighting coefficients in three linguistic conditions: Effects of lexical processing were tested by comparing words with phonotactically plausible nonwords and for effects of phonological processing, nonwords were presented forward- and time-reversed, so that nonwords did not adhere to the phonotactic rules anymore. Listeners gave significantly more perceptual weight to F0 and VTL when listening to words and phonotactically plausible nonwords compared to time-reversed implausible nonwords, while there was no difference in weighting these cues between words and plausible nonwords. This advantage of linguistic stimuli that adhere to the phonotactic rules, regardless of whether they are meaningful words or not, indicates that the weighting of F0 and VTL cues is altered by phonological, but not by lexical processing when using these voice cues for categorizing the speaker's perceived voice gender.

P47 Comparing results from self-report and behavioural speech perception tests as means to measure outcomes for hearing aid users

Naomi Joannou
Sandwell and West Birmingham NHS Trust, UK

Antje Heinrich
Manchester Centre for Audiology and Deafness, University of Manchester, UK

Behavioural and self-report measures of speech perception often do not correlate well even though they claim to measure the same concept. To help reconcile previous conflicting results, this study investigated three potential reasons for the lack in correlation: 1) mismatch of International Classification of Functioning, Disability and Health (ICF) domains between measures 2) mismatch of listening situations, and 3) low test-retest reliability of each outcome measure. To do this, 26 new hearing-aid users, aged 53-83 years with bilateral mild-to-severe sensorineural hearing loss completed the following assessment tools on two occasions at least one week apart: (1) part two of the Client Orientated Scale of Improvement (COSI), (2) the Bamford-Kowal-Bench (BKB) sentence test presented in quiet, and (3) the Quick Speech-in-Noise (QuickSIN) test. Based on correlations, the COSI questions are more likely to reflect impaired hearing than activity limitations. Some high correlations between behavioural and self-report measures emerged when situations were matched more closely. Intraclass correlation coefficients (ICCs) demonstrated good to excellent test-retest reliability for overall COSI and speech perception scores. Test-retest reliability for individual COSI categories was mixed and may have contributed to some of the low correlations between measures. The behavioural speech perception tests assess only some very specific self-reported experiences, largely unrelated to the activity limitations reported as most important by many impaired listeners. Data for some individual COSI categories lacked reliability.

P48 Impact of masks on automatic speech recognition algorithm performance in noise

Arman Hassanpour[1], Thea Knowles[2], Vijay Parsa[1]
1. University of Western Ontario, Canada | 2. University at Buffalo, USA

Wearing face coverings to protect against COVID-19 has become mandatory or recommended in many locations. In certain cases, recommendations have included layering face masks with a second mask or with a face shield as an extra protective measure. Previous work has shown that certain types of masks impact the acoustic signal and speech intelligibility. These effects may be even more detrimental when face coverings are worn in a noisy environment. This project investigated the effect of various types of masks with or without a face shield in quiet and background noise. To this aim, pre-recorded Harvard sentences spoken by male and female speakers from the TSP corpus were played back through a Bruel & Kjaer head and torso simulator (HATS) and recorded in a sound treated room. Six mask conditions were considered in quiet and background noise conditions with different signal-to-noise ratios (SNRs). The distinct mask conditions were: no-mask, four types of face masks (cotton, surgical, N95,

and KN95 masks), and one double-layered mask (cotton and surgical masks), and these were assessed with and without a face shield over the top. The Amazon Web Service (AWS) was selected as the automatic speech recognition algorithm to transcribe speech into text. The transcribed text was scored corresponding keywords of the original sentence of the list, and the score of each sentence was normalized between 0 to 1. Normalized scores from male and female speech lists were averaged. To evaluate the impact of mask type in different levels and types of noise, two linear mixed effects models were conducted independently for the quiet and the background noise conditions. In the quiet condition, there were no significant differences across any of the face coverings compared to no mask. In the noise condition, with noise type and level held constant, all mask types except for the bare face shield were associated with significantly lower intelligibility. The layered presence of the shield was associated with significantly higher intelligibility, and this was observed to be true for most mask types. Significant two-way interactions with certain masks and noise variables suggested a complex relationship between the type of noise and type of mask. Thicker masks, such as the N95, cotton mask, and double-layered mask, may be more negatively impacted by noise levels and noise type, whereas the presence of a shield may actually create a resonance effect that facilitates intelligibility in some cases.

P49 Effect of hearing aid use on auditory temporal processing and speech-in-noise in conductive hearing loss

Merve Torun Topcu
İstanbul Medeniyet University, TR

Ayça Çiprut
Marmara University, Istanbul, TR

Objective: The chronicity of the problem of conductive pathologies, the lack of improvement with treatment, and the progression of hearing loss are seen as important criteria for using hearing aids. In the event of delay in the intervention of conductive hearing losses, the lack of sound transmission may cause auditory deprivation. In conduction pathologies, word discrimination scores are maintained in quiet while performance decreases in the presence of noise. The aim of this study is to evaluate the effect of hearing aid use on speech discrimination in quiet and noise in conduction pathologies by using the Turkish matrix test as well as its effect on temporal processing.

Methods: Sixty-eight patients with mild to moderate conductive hearing loss participated: 23 patients were bilateral hearing aid users, 22 patients were unilateral users and 23 patients who did not use hearing aids. Thirty-one normal-hearing subjects were also included as a control group. Acoustic immitancemetry (tympanogram and acoustic reflexes) and pure-tone audiometry (air and bone conductions) were measured. The Turkish matrix test was performed at +5, 0 and -5 dB signal-to-noise ratio (SNR) with supra-aural headphone (TDH-39). In addition, the gaps in noise, duration pattern and frequency pattern tests were also performed to assess temporal processing.

Results: In the Matrix test performed at all SNRs, a significant difference was found between the scores of the conductive-loss group that did not use hearing aids and the groups that used bilateral or unilateral hearing aids. When the ears of the group with unilateral hearing aids were compared with and without the device, no significant difference was found between the test scores. In addition, temporal processing performance for those with long-term conductive hearing loss who had never used hearing aids was significantly worse than those that used one or two hearing aids.

Conclusion: These results indicate that although the cochlea and auditory nerve can be initially intact in conduction pathologies, the existing hearing loss may lead to auditory deprivation and the speech discrimination performance in noise may be affected. Early intervention with appropriate amplification in conductive pathologies may contribute to speech in noise performance as well as basic temporal processing.

P50 Encoding of naturalistic speech with simulated hearing loss in fMRI Responses

Arkan Al-Zubaidi, Leo Michalke, Jochem Rieger
Applied Neurocognitive Psychology Lab and Cluster of Excellence Hearing4all, Oldenburg University, Oldenburg, Germany

In neuroimaging, voxel-wise encoding models are popular tools to characterize and predict brain activity from a given stimulus representation. In the auditory domain, hearing a naturalistic auditory stimulus leads to changes in brain activity that subserves sensory processing. Yet, little is known about cortical auditory processing of speech under clear and degraded naturalistic stimuli. We used a German audio description of the movie “Forrest Gump” soundtrack as a clear naturalistic stimulus (CS). However, the movie’s soundtracks were adjusted by hearing loss simulators to build two types of degraded naturalistic stimuli. One with low degradation (S2, steep sloping) and one with high (N4, moderately sloping) at higher frequencies. This study uses a data-driven approach to investigate how the acoustic information related to the CS, S2 and N4 stimuli is processed in the auditory cortex. We recorded the fMRI in 10 normal-hearing participants distributed over three sessions. The participants listened to the full movie in each session but in eight segments, i.e. eight runs. In each session, the eight movie segments were presented in chronological order. Each session began (run1) and ended (run8) with CS stimuli, while runs from 2 to 7 were presented as CS, S2, or N4 stimuli. The order of degradations was randomized over scan sessions, but the number of presentations was balanced across degradation levels. After every movie segment, participants were asked to rate their speech perception and answer two questions on the content of the preceding segment.

After preprocessing the fMRI data, we performed voxel-wise encoding models to predict BOLD activity elicited by an auditory movie envelope. We estimated encoding models for the sound envelope of CS, S2, and N4 stimuli separately and predicted left out runs in cross-validation schemes to test the generalization of the encoding models. The speech perception is best for CS, intermediate for S2, and worst for N4 stimuli. In contrast, the encoding models best predicted the BOLD responses in N4 compared to CS and S2 conditions. Primarily, in the N4 condition, we found the highest correlations between predicted and observed BOLD-responses concentrated in the core auditory areas on the superior temporal cortex (Heschl’s gyrus). However, high correlations were less focal in the CS and S2 conditions. Thus, our results concur with the assumption that degraded stimuli require an increase of attention which enhances activation in early auditory areas. Alternatively, the increased activation may indicate an improved prediction error in processing degraded stimuli.

P51 The influence of hearing impairment on the pupil dilation response to degraded speech

Adriana A. Zekveld, Hidde Pielage, Niek J. Versfeld, Sophia E. Kramer
Amsterdam UMC - location VUmc; APH Research Institute, The Netherlands

The measurement of pupil size to assess the influence of speech perception conditions on listening effort has been widely applied in the past decade. Several studies directly compared the pupil responses of listeners with hearing impairment with those of normal-hearing listeners. Although the results are not conclusive, the main evidence emerging from these studies is that hearing loss seems to be associated with a less dynamic pupil dilation response across intelligibility conditions. Also, the pre-trial baseline pupil response tends to be the same or slightly smaller than that of age-matched normal-hearing participants. It is currently not clear what the underlying mechanism causing these differences is. In the current study, we compared the baseline pupil size and the pupil dilation response between a group of 17 normal-hearing participants (mean age 46 years, age range 20 – 62 years) and a group of 17 age-matched hearing-impaired participants (mean age 42 years, age-range 20-63 years). The average pure-tone hearing thresholds at 1, 2, and 4 kHz were 6 dB HL for the normal-hearing group, and 50 dB HL for the hearing-impaired group. Participants performed speech perception tests in three degradation conditions: noise-vocoded speech, speech masked with stationary noise, and speech masked with interfering speech. They also listened to speech presented in quiet. Speech intelligibility of the degraded speech conditions was matched using an adaptive procedure targeting 50% sentence intelligibility. The results confirm that hearing impairment is associated with smaller increase in the pupil dilation response in difficult (degraded) conditions compared to easy conditions (speech in quiet). This finding was even observed in the noise-vocoded speech condition in which both the intelligibility level and the degradation level was similar between groups. We will present these findings, including the baseline pupil data. The results will be discussed in the context of the available evidence of the effect of hearing impairment in age-matched groups of participants. We will discuss several factors that may be associated with the mixed results found previously. These include the age-matching of participants, matching of performance level, and the baseline correction applied to the pupil dilation response. We conclude that more research is warranted to assess the effect of hearing impairment in larger, age-matched groups in a range of conditions.

P52 Effect of audiovisual lag on speech-in-noise comprehension in cochlear implant users and typical hearing controls

Cailey A. Salagovic, Muskaan Mann, Ryan A. Stevenson, Blake E. Butler
University of Western Ontario, London, Canada

When a person is seen speaking, a listener's ability to understand that speech is supported by both the auditory signal of the voice and complementary visual cues from mouth movements. The relative amount that these auditory and visual cues contribute to understanding multisensory speech signals varies depending on the individual's specific sensory abilities and characteristics of the listening environment. A great deal of foundational and contemporary literature has shown that visual speech cues facilitate speech-in-noise intelligibility in listeners with typical hearing. For individuals with hearing loss as well as users of assistive hearing devices such as hearing aids and cochlear implants (CIs), understanding target speech in the presence of background noise is especially challenging. Because the auditory signal produced by a CI is fundamentally degraded compared to that conveyed by the typically developed inner ear, CI users likely rely on visual speech cues more than individuals with typical hearing. While the role of visual cues in overcoming the challenge of background noise has been previously established, less is known about how CI users use visual cues when audiovisual speech material is compromised by temporal lag, as can occur with online video calling platforms. Here we show that that CI users benefit more from visual speech cues in an online speech-in-noise listening task but do so over a smaller range of asynchronies than typical hearing controls. For sentences presented in multi-talker babble, both groups benefitted from visual speech cues though the CI group showed greater multisensory gain. Additionally, both groups' ability to understand target speech decreased as temporal offset between the auditory and visual streams increased, with CI users showing a sharper decline. Generally, these findings are in keeping with established principles of multisensory integration including inverse effectiveness (that integration confers greater benefit when unisensory signals are ambiguous) and the existence of a temporal binding window (a range of asynchronies over which audiovisual signals are commonly bound perceptually). Taken together, these results are a first step towards characterizing the interaction between these principles and expanding this framework to populations with unique sensory experience. Given the rapid expansion of remote and hybrid schooling, careers, and even socializing in recent years, thorough understanding of the unique listening challenges of such virtual environments and their efficacy for various users is necessary and timely.

P53 Effects of age and musical expertise on speech-on-speech perception in adults

Laura Rachman[1], Eleanor Harding[1], Ryan Gray[1], Stefan Smeenk[1], Anastasios Sarampalis[2], Etienne Gaudrain[3,1], Deniz Başkent[1]

1. University of Groningen, University Medical Center Groningen, Netherlands | 2. University of Groningen, Netherlands | 3. Lyon Neuroscience Research Center, CNRS UMR5291, Inserm U1028, UCBL, UJM, Lyon, France

Understanding speech when multiple people are talking simultaneously relies on various perceptual and cognitive mechanisms, such as stream segregation, selective attention, and inhibition. Differences in voice characteristics, such as mean fundamental frequency (F0) and vocal-tract length (VTL), help discriminating between different voices and facilitate speech perception in the presence of competing speech. In older adults, speech perception difficulties become highly prevalent, particularly in noisy listening situations. Some studies on perceptual processing in older adults showed that older adults may be less sensitive to F0 differences, which could affect their ability to distinguish different speakers. Furthermore, age-related cognitive changes may lead to difficulties directing attention to relevant speech and inhibiting competing information. As a result, speech perception in the presence of interfering speech becomes particularly difficult for older adults. The first aim of this study was to investigate to what extent older adults, compared to younger adults, benefit from voice differences in a speech-on-speech (SoS) perception task. In another line of research, studies have suggested that musicians, compared to non-musicians, perform better in SoS tasks. However, this advantage has mostly been observed in younger adults, and it is not yet known if this advantage extends to older adults. It could be that musical training leads to improved processing of acoustic features such as F0, or improved stream segregation and selective attention abilities. This would protect older adults from declines in cognitive functioning and/or the ability to process acoustic features such as F0, such that age-related decline in SoS perception may be reduced in musically-active older adults. The second aim of this study was therefore to investigate whether a musician advantage in SoS perception exists for older adults. Together, this study explored the effects of aging and musical expertise on SoS perception by considering four experimental groups, namely younger (18-30 years) and older (60-80 years) adults with and without musical expertise, all with self-reported normal hearing. All participants performed an online Coordinate Response Measure test with a single target speaker and a single talker masker. Participants were asked to respond to keywords in the target speech, while target-to-masker ratios were varied, and differences in F0 and VTL voice cues between the target and competing voices were manipulated. We will report preliminary findings of this study and show to what extent aging effects and musical training may have an influence on speech perception in the presence of competing speech.

P54 The first clarity enhancement challenge for hearing aid processing

Michael Akeroyd[1], Jon Barker[2], **Trevor J Cox**[3], John Culling[4], Simone Graetzer[3], Rhoddy Viveros Muñoz[5], Graham Naylor[1], Eszter Porter[1]
1. University of Nottingham, UK | 2. Sheffield University, UK | 3. University of Salford, UK | 4. Cardiff University | 5. Universidad Austral de Chile, Chile

In 2021, the Clarity project ran the first ever open machine learning challenge for hearing aids (https://claritychallenge.github.io/clarity_CEC1_doc/). It was aimed at improving the processing of speech-in-noise. This paper will briefly outline this Enhancement Challenge. Competitors were tasked with improving speech-in-noise for cases where there was one target speaker, one noise interferer, a room with low reverberation, and the audiogram of the listener was known. The particular difficulties in running challenges for hearing aid processing with remote listening panels will be discussed. Innovations such as listening test subjects saying what they heard out loud, before scoring with ASR (automatic speech recognition) will be discussed, and compared to the more traditional approach of human transcription. Entrants were scored in two ways by: (i) Passing the improved speech-in-noise through a hearing loss model and then evaluating using the objective intelligibility metric MBSTOI (Modified Binaural Short-Time Objective Intelligibility). (ii) Listening tests of speech intelligibility using a panel of people with a hearing loss. For (i) the objective evaluation, 13 systems were evaluated, for (ii) the listening tests, 10 systems were used. The results from both evaluations will be presented. These demonstrated weaknesses in the objective evaluation. Consequently, a Perception Challenge (https://claritychallenge.github.io/clarity_CPC1_doc/) is currently running to improve the prediction of speech intelligibility, especially for people with a hearing loss listening through a hearing aid. For the Enhancement Challenge, a mixture of approaches were seen in the entrants. The static scenes and the limited range of positions for the target talker, made beamforming a good approach; this was used by six teams. The second enhancement challenge starting in Spring 2022 will have head movements to make the scenarios more realistic and more challenging for beam formers. A variety of approaches for (i) noise removal using Deep Neural Networks (DNNs), and (ii) Hearing Loss Compensation were used. The highest scoring systems were the ones that were robust across a wider range of signal-to-noise ratios.

P55 A data-driven distance metric for evaluating the effects of dynamic range compression in adverse conditions

Niels Overby, Torsten Dau, Tobias May
Technical University of Denmark

Dynamic range compression is one of the most essential building blocks in modern hearing aids and aims at restoring audibility for hearing-impaired listeners. However, the choice of suitable compression parameters, such as the time constants associated with the level estimation stage, depends on the acoustic conditions and the perceptual benefit of different parameter configurations is still controversial. Listening tests can provide an accurate assessment of the perceptual effects of compression in a limited set of acoustic conditions, but they are time-consuming and can therefore not be used to optimize the various compression parameters across experimental conditions. While several studies have attempted to link the perceptual outcomes of dynamic range compression to a set of objective metrics, there is no agreement on how to objectively quantify the effects of compression. In the current study, a data-driven distance metric based on objective metrics was developed to analyze different compression systems. This analysis included slow-acting, fast-acting, and 'scene-aware' compression that adaptively switched between fast- and slow-acting compression depending on the target source activity. In addition, a hypothesized 'ideal' system, termed 'source-independent compression', was used as a reference that had access to the individual signals and applied fast-acting compression to the target speech signal and slow-acting compression to the noise and reverberation. A comprehensive list of objective metrics was considered to evaluate the effect of the compression systems in a wide variety of acoustic conditions, including both interfering noise and room reverberation. Sparse principal component analysis was then applied to derive a compact set of interpretable features that explained the effects of compression as linear combinations of sparsely selected objective metrics. The reduced set of features corresponded to the amount of distortion of the noise and reverberation, the amount of compression of the target speech signal, and the relative amount of amplification of the target speech compared to the noise and reverberation. The Euclidean distance, within the reduced dimensionality representation, was used to compare the similarity between the compression systems. In this comparison, the adaptive 'scene-aware' compression system was consistently more similar to the 'source-independent' system compared to fast- and slow-acting compression for speech signals in both noise and reverberation. This newly developed distance metric allows a systematic analysis and optimization of the parameters of dynamic range compression systems by minimizing the Euclidean distance with respect to the source-independent compression system.

P56 Principal Components Analysis of amplitude envelopes from spectral channels: A preliminary comparison between music and speech

Olivier Y. Crouzet, Agnieszka Duniec, Elisabeth Delais-Roussarie
LLING (UMR6310, Nantes Université / CNRS), FR

Introduction: The efficient coding hypothesis predicts that perceptual systems are optimally adapted to natural signal statistics. Previous work provided statistics of speech signals for 8 languages based on Principal Components Analysis (PCA), arguing that 4

frequency channels would be sufficient to optimally represent clean speech signals for each of these 8 languages. Extending these data to cochlear implant simulations in english, it has been shown that 6 to 7 frequency bands would be sufficient to optimally represent vocoded speech.

However, research on music perception in cochlear implanted listeners sheds light on potential limits associated with these results. Performance observed on vocoded signal material in normal-hearing listeners as well as in CI users is systematically better for speech signals than for music. Our aim is to compare statistical properties of natural music signals with previous work on speech in order to evaluate their respective contributions to this theoretical proposal.

Method: Analyses were carried out using Matlab on music samples from the FMA open source database (Free Music Archive, https://github.com/mdeff/fma). Signal processing and statistical procedures were mirrored from previous studies on speech. The total sample duration was comparable. Sample signals were passed through a gammatone filterbank (1/4th ERB bandwidth, approx. 100-120 channels) and their energy envelope was extracted. This amplitude modulation matrix was then run through PCA and PCs were independently rotated. Channels that covary in amplitude envelope should be grouped as a single Principal Component. As our aim was to compare speech and music, for which typical signal bandwidths differ, two higher-frequency limits were compared (8000 Hz vs. 22000 Hz).

Results and discussion: Current graphical exploration of statistical results only provides partial descriptions. As should be expected, more PCs seem to be required in order to characterize music samples than was estimated for speech. The optimal number of PCs for music seems to stabilize between 24 and 32 channels (vs. 4 to 7 channels according to previous speech studies), at least for frequencies up to 12000 Hz. For more systematic comparisons, statistical analyses need to be enhanced in order to develop methods for automatically determining the optimal number of Principal Components as well as to estimate frequency boundaries between these PCs. Results for music and speech will be compared in order to identify possible discrepancies between optimal frequency channels.

P57 Assessing the generalization gap of a deep neural network-based binaural speech enhancement system in noisy and reverberant conditions

Philippe Gonzalez, Tommy Sonne Alstrøm, Tobias May
Technical University of Denmark

Noisy and reverberant speech signals are influenced by a plethora of factors, such as the spectro-temporal characteristics of the target speaker and the interfering noise, the room acoustics, the signal-to-noise ratio (SNR) and the position of the different sources in the acoustic scene. This large variability of acoustic conditions poses a major challenge for deep neural network (DNN)-based speech enhancement systems, since any mismatch between training and testing can substantially reduce their performance. In addition, the generalization capability of DNN-based systems is typically assessed by testing the system with an arbitrarily chosen speech, noise or binaural room impulse response (BRIR) database that was not seen during training. This poses a problem, as the difficulty of the speech enhancement task can substantially vary across databases, which strongly influences the results and complicates a comparison across studies. The present study systematically investigates the influence of six acoustic scene dimensions on the generalization capability of a binaural DNN-based speech enhancement system, namely the target speaker, the noise type, the room, the SNR, the target direction and the mixture level. We propose a new measure of generalization, which is referred to as the generalization gap. The generalization gap is expressed in percentage and is defined as the performance distance to a reference model trained on each test condition. To reduce the influence of the test condition on the generalization assessment, the generalization gap is measured using a cross-validation framework over multiple speech, noise and BRIR databases. We find that while a speech mismatch between training and testing affects generalization the most (generalization gap of 49% in terms of the mean squared error (MSE)), other dimensions such as the noise type (36%) and the room (30%) can also induce a substantial generalization gap. The SNR, direction and level dimensions can potentially induce significant generalization gaps, but these can be substantially reduced by training on diverse datasets that present a wide range of SNRs, directions and levels. The generalization gap can be measured for any learning-based system and facilitates a comparison across studies.

P58 The influence of linguistic properties on recognition of naturalistic speech in noise

Jing Shen[1], Lauren Calandruccio[2], Alyssa Liusie[2], Caitlyn Dececco[1]

1. Department of Communication Sciences and Disorders, Temple University, Philadelphia, PA, United States | 2. Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH, United States

Speech in noise outcome measures are typically based on the recognition of linguistically controlled sentences presented with background noise. These measures often lack ecological validity because they fall short in accounting for perceptual and psycholinguistic variables inherent in real-life conversations. To understand how lexical properties of conversational speech material influence recognition of speech in noise, new stimuli were created based on the Diapix task [Hazan and Baker, 2011, JASA 130(4):2139; Van Engen et al., 2010, Lang. Speech 53(4):510], a dialogue elicitation paradigm that was developed to study talker- and listener-related adaptation strategies. Sixty-eight short sentences from the Diapix corpus were selected as test items, and recordings of the sentences were made with 8 talkers. The talkers were instructed to read the sentences clearly and naturally following the transcription, which has repairs and interruptions preserved. The talkers were also instructed to produce non-speech sounds (e.g., laughter, lip smacks, sighs). The continuous recordings were spliced manually to extract each sentence. The linguistic properties of interest include the number of content words in each sentence, overall sentence length, word position, phonological word length, lexical frequency, and phonological neighborhood density. Speech recognition accuracy in spectrally shaped steady-state noise with 2 dB SNR was measured using an online protocol with 31 young listeners. The listeners were native speakers of English and had normal hearing by self-report. Each listener heard speech from only one talker and was asked to provided transcriptions for each sentence. Linear mixed effect models were used to assess the effects of lexical properties on word- and sentence-level recognition accuracy in noise. Preliminary analysis suggests the effects of linguistic properties of the speech material on recognition accuracy in noise. The results demonstrate the impact of psycholinguistic properties on speech perception in addition to the perceptual difficulty introduced by noise. These findings can inform future studies aimed at developing clinical outcome measures that better represent everyday speech communication and more accurately capture listeners' abilities in real-life scenarios.

Funding: Work supported by National Institutes of Health.

(Retracted)

P59 Predictive sentence processing of L2 speech in noise: Differential effects across linguistic cues

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

Incremental processing of speech leads listeners to build expectations and to make predictions about upcoming sentence information. Prediction is, however, subject to individual variation (Huettig & Janse, 2016, doi: 10.1080/23273798.2015.1047459) and may be compromised in L2 listeners (Kaan, 2014, doi: 10.1075/lab.4.2.05kaa). This eye-tracking study explores the time course of L2 sentence processing and the (anticipatory) integration of different types of linguistic information in different types of noise. Using a visual world paradigm, we tested predictive processing among 72 German L2 listeners of English across three acoustic conditions (quiet, stationary noise, multi-talker babble) on three linguistic conditions (1-3).

(1) The girl is laugh-ING/laugh-ED at (by) the boy. (morpho-syntactic)

(2) The man is looking at the BIG/SMALL green ball. (lexical)

(3) The baby is looking at A/THE red purse. (discourse)

Noise can impact information integration in differential ways. Processing of morpho-syntactic cues, such as in (1), tends to be more affected by noise than lexical cues (2) and low-salience discourse cues, like articles (3). At the same time, L2 listeners generally tend to rely more on discourse than morphosyntactic cues (Cunnings, 2017, doi: 10.1017/S1366728916001243). Hence, L2 listeners may differ in the types of information they continue to recover in noise as well as in their extent. In addition, different noise types mask speech differentially. Whereas stationary speech-shaped noise serves as an energetic masker, multi-talker babble functions as an informational masker. The informational masking effect is, however, modulated by language proficiency (Mattys et al., 2012, doi: 10.1080/01690965.2012.705006). We investigate (a) whether L2 listeners continue to predict during sentence comprehension even when noisy speech causes phonetic unreliability, and we examine (b) the degree to which different linguistic cues are affected by the two types of noise in L2 listeners. Preliminary results suggest that both noise types equally affect the processing of lexical information, whereas morpho-syntactic processing is differentially affected by stationary vs. babble noise. The processing of discourse information, by contrast, is not differentially affected by the acoustic conditions. These results suggest that stimulus-specific and listener-specific aspects interact in predictive processing. We discuss potential implications for models of L2 language processing and models of speech in noise.

P60 Normal-hearing listeners and cochlear-implant users benefit from voice-feature continuity at the Cocktail Party

Jens Kreitewolf[1,2], Daniela Hollfelder[3], Martin Orf[2], Julia Erb[2], Samuel R. Mathias[4], Karl-Ludwig Bruchhage[3], Barbara Wollenberg[5], Jonas Obleser[2]

1. Department of Psychology, McGill University, Montreal, Canada | 2. Department of Psychology, University of Lübeck, Lübeck, Germany | 3. University Hospital Schleswig Holstein, Campus Lübeck, Clinic for Otorhinolaryngology - Head and Neck Surgery, Lübeck, Germany | 4. Department of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, USA | 5. Clinic for Otorhinolaryngology, Head and Neck Surgery, MRI TUM, Technical University Munich, Munich, Germany

Most of our everyday listening happens under adverse conditions largely because natural auditory scenes often comprise a multitude of sounds heard at once. These "cocktail-party"-like situations pose a difficult problem for normal-hearing (NH) listeners and are particularly challenging for cochlear-implant (CI) users. We have previously shown (Kreitewolf et al., 2018, doi: 10.1121/1.5058684) that NH listeners are better at comprehending target speech when they can group sounds based on continuity in two prominent voice features: glottal-pulse rate (GPR) and vocal-tract length (VTL). Unlike NH listeners, CI users do not seem to benefit from voice cues. For example, recent work (El Boghdady et al., 2019, doi: 10.1121/1.5087693) suggests that CI users do not exploit differences in GPR and VTL to segregate target from masker speech. Here, we took a different approach to investigate the use of GPR and VTL for cocktail-party listening in N = 20 NH listeners and N = 20 CI users. In this experiment, participants heard a stream of spoken digits embedded in multi-talker babble noise and were asked to report each digit immediately after its presentation. To explore the contributions of GPR and VTL to cocktail-party listening, we manipulated continuity in GPR and VTL across consecutive digits. Overall performance was fixed at approximately 67%-correct. Our results showed that both NH listeners and CI users benefited from voice-feature continuity. Both NH and CI showed the greatest benefits when both voice features were continuous across consecutive target digits. When only one voice feature was continuous, both listener groups benefited from VTL but not GPR continuity. Once listeners successfully tracked the target stream, they were more likely to correctly report the subsequent digit. The magnitude of this so-called previous-digit-correct benefit (PDCB), however, differed across conditions and listener groups. Interestingly, CI showed greater PDCBs when they could rely on VTL versus GPR continuity. NH listeners seemed to benefit equally from GPR and VTL continuity. Sensitivity to differences in GPR and VTL was not correlated with benefits from voice-feature continuity. These findings provide evidence that CI users exploit continuity in voice features to improve listening at the cocktail party. Unlike NH listeners, CI users seem to rely exclusively on VTL continuity for perceptual grouping of target sounds, which may be due to their impaired processing of pitch cues. VTL is effectively fixed within natural talkers. Stronger reliance on VTL continuity may thus explain CI users' residual abilities to solve the cocktail-party problem in natural settings.

P61 Does the effect of sound sources diffuseness on speech perception change with age and linguistic experience?

Meital Avivi-Reich
City University of New-York (CUNY), NY, USA

Bruce. A Schneider
University of Toronto, Missassauga, ON, Canada

Studying the effect of sound source diffuseness level on speech recognition has increasing importance due to the growing use of amplification systems. When an auditory stimulus is amplified and presented over multiple, spatial-separated loudspeakers, the signal's timbre is altered due to comb filtering. In a previous study we examined how increasing the diffuseness of the sound sources might affect listeners' ability to recognize speech presented in different types of background noise. We found that listeners performed similarly when both the target and the masker were presented via a similar number of loudspeakers. However, performance improved when the target was presented using a single speaker (compact sound source) and the masker from three (diffuse sound source) but worsened when the target was diffuse, and the masker was compact. In the current study, we extended our research to examine whether the effect of timbre changes with age and linguistic experience. Twenty-four older adults whose first language was English (Old-EFL) and 24 younger adults whose second language was English (Young-ESL) were asked to repeat sentences presented in either noise, babble, or speech. Participants were divided into two experimental groups: (1) A Compact-Target Timbre group where the target sentences were presented over a single loudspeaker, while the masker was either presented over three loudspeakers or over a single loudspeaker; (2) A Diffuse-Target Timbre group, where the target sentences were diffuse while the masker was either compact or diffuse. The results indicate that the Target Timbre has a negligible effect on thresholds when the timbre of the target matches the timbre of the masker in all three groups. When there is a timbre contrast between the target and the masker a significant separation between the compact target and diffuse target conditions' thresholds is evident in all three groups when the masker is noise. However, for Speech and Babble Maskers, the advantage held by compact targets is severely diminished in the Young-ESL group compared to the two EFL groups.

P62 Simulated effects of deafferentation, deefferentation and their interaction

Jacques A. Grange, John F. Culling
Cardiff University, UK

MAPsim is a new simulator of sensorineural hearing loss that can be used to help differentially diagnose sensorineural pathologies because it is based on the MATLAB auditory periphery (MAP) model. The MAP-predicted sound encoding at the auditory-nerve is decoded back into an acoustic signal that is presented to participants in psychophysical tasks. MAP could predict dynamic range adaptation at the auditory nerve and reveal that efferent reflexes are the primary contributors to such adaptation, the acoustic reflex causing a shift of

rate-level functions towards a context level and the medial olivo-cochlear reflex sharpening auditory nerve sensitivity around that level. The simulated effect of general deafferentation on speech intelligibility in noise (SpIN) showed that simple stochastic under-sampling due to deafferentation underestimates the effect of deafferentation. Indeed, under-sampling alone suggested that over 90% of fibers could be knocked out without any appreciable effect on speech reception threshold (SRT). However, deafferentation also causes a reduction of efferent reflexes. The combined under-sampling and efferent signal reduction suggests that SpIN is sensitive to lower levels of deafferentation than previously thought. When testing for the effect of deafferentation on ITD discrimination, threshold elevation appeared more sensitive to under-sampling alone, the added effect of efferent signal reduction being more limited as a result. The simulated effect of deefferentation on SpIN was operationalized with both separate and combined, gradual acoustic reflex reduction and knockout of outer hair cells, such that interaction between efferent reflexes could be investigated. This manipulation suggested that at normal speech levels, one efferent reflex may somewhat compensate for the loss of another. Removing both efferent reflexes led to an SRT elevation (~ 4dB) comparable to that seen in moderately impaired patients. Overall, MAPsim opens the path to the differential diagnosis of sensorineural pathologies because it enables the testing of psychophysical tools in a search for the psychophysical signatures of specific pathologies.

P63 Effects of valence and priming on just-noticeable differences in signal-to-noise ratio

Janin Benecke, William M. Whitmer
Hearing Sciences - Scottish Section, University of Nottingham, Glasgow, UK

Difficulties understanding speech in background noise is often a pivotal reason to seek treatment for hearing loss. In the clinic, 'difficulty' and 'clarity' are commonly regarded as perceptually opposite valences; hearing aids should alleviate the perceived 'difficulty' by making sounds 'clearer'. Acoustically, speech intelligibility depends primarily on the signal-to-noise ratio (SNR). In the context of hearing aid evaluations, it is important to understand how and when differences in SNR are noticed. While just-noticeable differences (JNDs) in SNR have been studied for differences in clarity, it remained unclear if different valence or prior knowledge of the speech can affect JNDs. This study measured SNR JNDs when framed as 'clarity' or 'difficulty' and also with and without signal familiarity provided by auditory primes. In a two-by-two repeated-measures design, the SNR JNDs of twenty-five participants with varying hearing abilities were measured in a 2I/2AFC task. In each trial, two intervals with a randomly chosen sentence in speech-spectrum noise were presented in random order: one at each participant's speech reception threshold (SNR50), and the other at the SNR50 plus an SNR increment (0, 1, 2, 4 or 8 dB). In the prime conditions, participants listened to the same sentence without noise just prior to the two intervals. Participants were asked to select the interval they perceived as 'more clear' in the clarity frame or 'more difficult' in the difficulty frame. Mean SNR JNDs in the clarity frame were 2.8 and 3.1 dB with and without prime, respectively, and in the difficulty frame 3.3 and 4.0 dB with and without prime, respectively. Across prime conditions, framing had a significant effect on SNR JNDs; average JNDs were greater for the difficulty than the clarity frame. Individual JNDs for difficulty and clarity were positively associated but not significantly correlated in any condition. Although priming reduced individual

differences, this method had no significant effect on mean SNR JNDs. Whether the target was in the first or second interval affected participants' decisions in all conditions; without primes, there was a preference for the second interval in the clarity frame, and for the first interval in the difficulty frame. This bias was weakened and inverted by priming. The results suggest that the framing of speech intelligibility benefits as well as prior SNRs have an influence on how SNR differences are perceived, and hence should be considered in the clinic when discussing and addressing patients' hearing-in-noise needs.

P64 Functional hearing difficulties in blast-exposed service members with normal to near-normal hearing thresholds

Ken W. Grant[1], Sandeep A. Phatak[1], Lina K. Kubli[2], Kimberly A. Jenkins[1], Jennifer R. Myers[1], Douglas S. Brungart[1]

1. Walter Reed National Military Medical Center, Bethesda, US | 2. Veterans Affairs (VA), Washington DC, US

Background: Over the past decade, DOD and VA audiologists have reported large numbers of relatively young adult active-duty patients who have normal to near-normal audiometric thresholds (i.e., < 25 dB HL) but who report difficulty understanding speech in noisy environments. Many of these service members (SMs) also reported having experienced exposure to one or more explosive blasts as part of their military service. A two-phase, multi-site study was conducted to understand the functional hearing and communication deficits (FHCD) in blast-exposed SMs.

Methods: In the first phase roughly 3400 active-duty SMs were screened for FHCD on four auditory tests (two speech-in-noise and two tone-in-noise tests) and a brief 6-question survey about perceived speech understanding and sound quality. To investigate the potential sources of FHCD (Phase II), two groups of blast-exposed SMs — those with or without FHCD — were tested large of number of tasks with behavioral, cognitive, and electrophysiological measures. Performances of these two groups were compared with that of a control group consisting of SMs with normal hearing thresholds (≤ 20 dB HL) and no blast-exposure, who did not exhibit FHCD.

Results: The first phase of the study revealed blast exposure and mildly elevated hearing thresholds each increased the probability of FHCD by 2 to 3 times. SMs having both blast-exposure and mildly elevated hearing thresholds exhibited up to 4 times higher risk for FHCD. In the second phase, the 3-way comparison (i.e., among non-blast SMs, blast-exposed SMs with FHCD, and blast-exposed SMs without FHCD) indicated that blast-exposed SMs with FHCD performed significantly worse than the control group on metrics that measured peripheral auditory system performance such as pure-tone thresholds, high-frequency distortion product otoacoustic emissions, frequency-following response, and signal-to-noise ratio and stability of envelope-following responses. Performance of blast-exposed SMs without FHCD was not different that the control group SMs on most of these metrics. However, both blast- exposed

SM groups (with or without FHCD) had significantly worse scores than the control group on measures of peripheral auditory processing (click ABR wave V amplitude) and measures of central auditory processing (Warrington Recognition Memory Test and Speed and Capacity of Language Processing Test).

Conclusion: Roughly 33.6% of active-duty SMs with normal to near-normal hearing thresholds (or approximately 423,000 by the current military size) are at some risk for FHCD, and about 5.7% (i.e. blast exposed and with elevated hearing thresholds, approximately 72,000) are at high risk for FHCD. Phase II results indicated that blast-exposed SMs with FHCD tended to have increased neural noise resulting in a poorer internal signal-to-noise ratio when compared to the control group as well as to blast-exposed SMs without FHCD. These data are consistent with studies showing benefit of low-gain hearing aids for this population in that small amounts of amplification would at least partially compensate, for degradation in peripheral auditory processing.

P65 Temporal processing and speech perception in noise among children using cochlear implant

Anuj Kumar Neupane, Vanaja C.S.
Bharati Vidyapeeth (Deemed to be University) School of Audiology and Speech Language Pathology, Pune, India

Introduction: Temporal processing, one of the important constituents of auditory processing is affected in persons with hearing impairment. Attempts have been made to compensate for these deficits through hearing devices such as cochlear implants (CI) [Shannon, 1989, JASA 85(6):2587; Won et al., 2011, JASA 130(1):376].

Rationale: Most often the temporal processing and speech perception based studies are reported in adults using CI [Hochmair-Desoyer et al., 1985, Cochlear Implants (Raven Press NY), p271; Moore and Glasberg, 1988, JASA 83(3):1093; Winn et al., 2016, Ear Hear. 37(6):e377] and there is dearth of such studies in children using CI [Landsberger et al., 2019, Otol. Neurotol. 40(3):e311]. There is a need to study these in children to investigate the effect of maturation on the same.

Objective: The present study aimed at investigating the performance on speech perception in noise and temporal cues based speech discrimination tests in children using CI.

Methods: Participants included 10 children with normal hearing and 10 children using unilateral cochlear implant, in the age range of 4 to 8 years. Children using cochlear implant had pre-lingual hearing loss and the implant age of ≥1 year with CI assisted threshold ≤30 dB HL. Revised speech perception test in noise for Marathi speaking children (SPTN-M) was performed at 60 dB SPL. The test consists of bisyllabic words in Marathi presented with four talker babble at 5 dB SNR. Syllable categorization of a temporal cue-based task (/ba/ vs /pa/) was performed by manipulating voice onset time between -79ms to 26ms in ten steps continuum as recommended by Winn et al. (2016). Stimuli for both the tests were presented through speakers of a laptop. The participants were seated at 1 ft distance from the loudspeaker placed at 0-degree azimuth.

Results and Discussion: Non-parametric statistics was performed as Shapiro-Wilk test revealed non-normal distribution of data. The performance of children with CI was poorer than those of children with normal hearing for both the tests. Mann Whitney U test revealed significant difference observed was statistically significant for both SPTN-M and syllable categorization. This may be attributed to various limiting factors among children using CI and further needs to be studied to understand the causative factor/s resulting it [Caldwell & Nittrouer, 2013, JSHLR 36(1):13]. Among children using cochlear implant, a strong correlation was observed between SPTN-M scores and syllable categorization [$rs(8) = -0.822$, $p = 0.000$]. This may be because both the tasks are dependent on the usage of temporal cues (Landsberger et al., 2019; Caldwell and Nittrouer, 2013).

Conclusions: Perception of temporal cues and speech perception in noise is poorer in children using CI when compared to children with normal hearing. There is a strong correlation between ability to perceive temporal cues and speech perception in noise.

P66 The impact of communication possibilities on psychosocial difficulties in children with hearing loss

Tjeerd de Jong[1], Marc P. van der Schroeff[1], Laura Rachman[2], Luuk Stapersma[3], Bert Maat[2], Francien Coster[2], Rolien H. Free[2], Deniz Başkent[2], Jantien L. Vroegop[1]

1. Department of Otorhinolaryngology, Head and Neck Surgery, Erasmus University Medical Center, Rotterdam, Netherlands | 2. University of Groningen, University Medical Center Groningen, Netherlands | 3. Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC-Sophia Children's Hospital, Rotterdam, Netherlands

Rationale: Psychosocial difficulties (e.g., peer problems, anxiety, hyperactivity) occur more often in children with hearing loss than in the general population. Previous findings suggest that psychosocial difficulties may be related to poorer communicative abilities. Compared to normal hearing peers, children with hearing loss exhibit less adequate communication skills, both hearing related (e.g., speech perception, especially in complex listening environments) and language related (e.g., pronunciation, vocabulary, and syntax). Also, it is suggested that children with hearing loss are less likely to make use of the psycho-emotional elements of vocal communication, such as interpreting intonation. Children with hearing loss therefore may be more likely to become isolated from their social environment.

Objective: The primary objective of this study is to determine the effect of specific communicative abilities in children with hearing loss on their psychological and social outcomes.

Study design: This is a prospective observational study. As from June 2021, we aim to include a total of 120 children (6-18 years) with hearing loss, who have at least one year experience with a hearing aid or cochlear implant. During the clinical appointment with their treating audiologist, the following tests will be performed: pure tone audiometry, speech perception in quiet for soft speech measured with the NVA, speech intelligibility in noise, measured with the Digits In Noise test (DIN), and tests on vocal emotion recognition and expression: EmoHI emotion recognition test, and the vocal emotion expression test. Language proficiency will be investigated with the Children's Communication Checklist (CCC-2). Psychological and social functioning will be estimated by the hearing related quality of life questionnaire (HEAR-QL), and the Achenbach System of Empirically Based Assessment parent/teacher/child reported Behavior Check Lists (CBCL/TRF/YSR). Using the test results, and demographic factors, a prediction model will be composed to predict psychological and social outcomes. With this model, we aim to gain more insight in the effect of communication abilities on observed psychosocial difficulties.

Results: Preliminary results on the first ~45 included participants will be presented at the Speech in Noise Workshop.

P67 An open generic specification for the Digits-in-Noise test

Soner Türüdü[1], Thomas Koelewijn[1], Etienne Gaudrain[2,1], Gloria Araiza-Illan[1], Laura Rachman[1], Deniz Başkent[1]

1. University Medical Center Groningen, University of Groningen, Groningen, Netherlands | 2. Lyon Neuroscience Research Center, CNRS UMR5292, Inserm U1028, UCBL, UJM, Lyon, France

Smits and colleagues [2013, JASA 133(3):1693] developed the Digits-in-Noise (DIN) test, which requires individuals to repeat three spoken numbers (a digit triplet) presented in noise, for clinical usage. They determined the criterion validity of the DIN test by comparing results on the test with results on the sentence speech-recognition threshold (SRT) from Plomp and Mimpen [1979, Audiol 18:43]. Both tests strongly correlated with pure tone audiometry (PTA) averages. While PTAs require strictly calibrated apparatus, speech tests remain informative over a range of presentation levels. Moreover, in comparison to the sentence SRTs, the DIN test only requires understanding of short, closed set tokens. Therefore, its application is less restricted by severity of hearing impairment or linguistic skills, making it applicable to, cochlear implant users and children. Over the past years, the DIN test has grown to be the number one tool for easy assessment of hearing acuity both in the clinic and online (hoortest.nl). It has even been translated to English and promoted by the World Health Organization for mobile audiometric assessment (hearWHO).

While first validated as a screening tool for detecting incipient hearing loss, nowadays the DIN is also used to assess hearing aid and cochlear implant fitting, which resulted in a vast amount of normative data for normal hearing and different hearing-impaired populations across multiple languages. However, to accommodate the specific needs of these diverse populations, purposes and situations, slightly different variations of the DIN test have been used that differ, for example, in the way speech is presented (diotic or antiphasic), or in the starting signal-to-noise ratio of the staircase procedure.

Similarly, for our use across various projects, we have created different versions of the DIN test: in MATLAB, online, and in Robot Operating System (ROS, https://www.ros.org), using the original Dutch DIN material (Smits et al., 2013) that has been made available to us (Cas Smits from Amsterdam UMC). To allow for compatible implementations and comparable results, we established a specification defining the test options and results. Each project implementing a DIN test following this specification will be able to run the known variations of the DIN test observed in the literature and in the clinic; therefore, defining the results data structure, and simplifying data exchange across projects, across centers, or even across languages.

The specification is published with an open CC-BY license on the dB SPL GitHub repository at https://github.com/dbSPLab/din-specification. We invite comments from expert users of the DIN to make the specification evolve.

P68 The effects of lexical content, sentence context, and vocoding on voice cue perception

Thomas Koelewijn[1], Etienne Gaudrain[2,1], Thawab Shehab[1], Tobias Treczoks[1,3], Deniz Başkent[1]
1. University Medical Center Groningen, University of Groningen, Netherlands | 2. Lyon Neuroscience Research Center, CNRS UMR5292, Inserm U1028, UCBL, UJM, Lyon, France | 3. Carl von Ossietzky Universität Oldenburg, Germany

Speech perception in a cocktail party like situation can be challenging, especially for cochlear-implant (CI) users. Perceiving differences in voice cues, such as fundamental frequency (F0) and/or vocal-tract length (VTL), in difficult listening conditions, can facilitate speech communication tremendously. In a recent study, we showed the effect of lexical content on just-noticeable-differences (JNDs) in F0 and VTL voice cue perception. Specifically, within the context of high acoustic and linguistic variability, when presented with words, participants showed smaller VTL JNDs compared to time-reversed words, and this observation did not change when vocoding was applied. For F0 JNDs only in the non-vocoded condition with low variability, a lexical content benefit (words vs time-reversed words) was shown. These outcomes inspired two follow-up studies.

The first study expanded on the lexical content benefit effect on VTL perception, by comparing words, time-reversed words, and non-words. The purpose was to investigate if the lexical content benefit is related to lexical (words) and/or phonemic (non-words) content. In the second study, we investigated the effect of additional acoustic speech information and/or semantic context on F0 and VTL voice cue perception, by comparing words and sentences. In both experiments non-vocoded and vocoded auditory stimuli were presented, while participants performed an adaptive 3AFC task to determine the voice JNDs.

The outcomes of the first study showed a replication of the detrimental effect reversed words have on VTL perception. Additionally, VTL JNDs with non-words did not significantly differ from VTL JNDs with words, suggesting linguistic content benefits VTL perception at the phonemic level. Although there was a main effect of vocoding this did not interact with item type. Study 2 showed a benefit in processing a full sentence compared to a single word in both F0 and VTL perception, suggesting that the amount of acoustic speech information and/or semantic context led to increased voice cue sensitivity. There was a main vocoder effect and an interaction between voice cue and vocoding indicating a stronger negative effect of vocoding on F0 compared to VTL perception.

In addition to previous findings suggesting a lexical advantage effect, the current results show more specifically, that phonemic content available in both words and non-words improves VTL perception. Both F0 and VTL perception benefit from more content in sentences (possibly lexical or semantic) compared to words. These results may improve our understanding of speech and voice perception processes, and result in rehabilitation tools for populations with limited access to voice information, such as CI listeners.

P69 Comparison of a humanoid robot interface and a laptop interface used for auditory psychophysical tasks

Luke Meyer[1], Gloria Araiza-Illan[1], Laura Rachman[1], Etienne Gaudrain[2,1], Deniz Başkent[1]
1. University Medical Centre Groningen, University of Groningen, Netherlands | 2. Lyon Neuroscience Research Centre, CNRS UMR 5292, INSERM U1028, UCBL, Lyon, France

Maintaining engagement during long and arduous speech perception tasks can be difficult, potentially resulting in a loss of attention and reduced performance. However, conducting these tasks is beneficial to better understanding speech perception mechanisms. Currently, a laptop or computer interface is used to perform these tasks through which auditory stimuli are presented, and responses are given either verbally, or more interactively using a mouse click. To provide a higher level of interaction, we propose the use of a humanoid NAO robot as a potential interface for performing auditory psychophysical tasks.

Using four speech perception tasks evaluating voice cue sensitivity, voice gender categorisation, vocal emotion identification, and speech-on-speech perception; presented in the form of games, 28 normal hearing adults performed the tasks using both the laptop and the NAO robot. Stimuli were processed concurrently during the voice cue sensitivity task on both interfaces by altering the voice cues of fundamental frequency (F0) or vocal tract length (VTL), and pre-processed for all other tasks using predefined F0 and VTL values. Stimuli were presented through the loudspeakers on the laptop and the speakers on NAO. Responses were logged using the mouse and clicking corresponding buttons on the laptop screen, and using the tactile sensors on the hands and head of the robot.

As a first step towards determining if the robot could be used as a reliable psychophysical performance measure interface for audio stimuli, the results of the perception tasks obtained from both interfaces were compared. Results indicate that of the four tasks, three showed no significant difference between participants' performances on the two interfaces, and only the voice cue sensitivity task was significantly different. Many limitations could affect the ability of the NAO to reliably collect data such as the quality of the robot's speakers, delays in stimulus processing and response logging, or the location of the tactile sensors in a three-dimensional space compared to buttons on a laptop screen.

Despite the potential limitations, the robot was still able to provide similar data to the current laptop interface. Future studies will investigate the perception of the robot by the participants and what factors, if any, could contribute to observable performance measure differences between the two interfaces.

P70 Humanoid robot as an interface for autonomous auditory testing through speech audiometry

Gloria Araiza-Illan[1], Luke Meyer[1], Khiet P. Truong[2], Deniz Başkent[1]
1. University Medical Center Groningen, University of Groningen, Netherlands | 2. Human Media Interaction, University of Twente, Enschede, Netherlands

Human-robot interaction (HRI) is a multidisciplinary research field where some studied parameters are the acceptance, trust and engagement that humans show towards a robot while performing a specific task with it. Our previous studies show promising results with different participants (normal hearing adults and hard-of-hearing children) in relation to maintaining a constant engagement throughout auditory testing when interacting with a robot, which is a current challenge due to the repetitiveness of the tests.

We propose a humanoid NAO robot as a social interface, coupled with the open-source Dutch instance of the Kaldi speech recognition toolkit, to autonomously run two speech audiometry tests: the digits-in-noise (DIN) test and the Nederlandse Vereniging voor Audiologie (NVA) phoneme-in-word lists. In both tests, normally a clinician is in charge of controlling the presentation of each stimulus and scoring each answer. The DIN test assesses speech perception in noise, quantified by the number of correctly identified digits presented in background noise; whilst the NVA phoneme lists assess speech perception, quantified by the number of correctly identified phonemes embedded in 3-phoneme CVC words and presented at different intensity levels.

This study focuses on the evaluation of the robot as an autonomous testing interface, presenting all stimuli and scoring each of the participant's responses accordingly. The main aim is to offer a portable system and a positive experience through a social interaction, ensuring replicability and aiming to avoid work overload amongst clinicians.

Our HRI consists of an introductory participant-robot conversation, the explanation of the test, a training phase for participants to learn how to record their responses during the test and the speech audiometry test. The interaction will be video-recorded from two angles: one focusing on the participant's facial expressions and another on their body language. The participants' perception of the robot will be evaluated through the analysis of the verbal and non-verbal interactions, and through a questionnaire that shall be completed by the participant at the end of the interaction. The overall system (the speech recognition software and testing algorithm) will be assessed by comparing its performance (recognised words and computed scoring) to an offline evaluation by a clinician.

Funding: VICI grant 918-17-603 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw), and the Heinsius Houbolt Foundation and the Rosalind Franklin Fellowship from University Medical Center Groningen, University of Groningen.

P71 Cochlear synaptopathy in the ageing population

Heleen Van Der Biest[1], Sarah Verhulst[1], Ingeborg Dhooge[2], Hannah Keppler[1]
1. Ghent University, Belgium | 2. Ghent University Hospital, Belgium

Background: Pure-tone audiometry and otoacoustic emissions (OAEs) are the standard clinical hearing tests to evaluate hearing thresholds and outer-hair-cell (OHC) damage, respectively. However, these tests are insensitive to supra-threshold hearing deficits or auditory-nerve fiber loss [i.e., cochlear synaptopathy (CS)]. Since this pathology can hide behind a normal audiogram and OAEs, CS is known as "hidden hearing loss". Hitherto, animal studies and simulations using computational model of the auditory periphery have shown that supra-threshold auditory evoked potentials (AEPs) and speech intelligibility in background noise (SPIN) are promising measurements to assess CS in humans. Despite the recent progress in diagnostic CS markers, the relation between SPIN and AEPs has not yet been thoroughly investigated. This study incorporates CS-sensitive EEG markers into an auditory profile that was applied to a large group of aging participants to quantify age-related CS and its relation to speech recognition.

Method and results: A total of 69 Flemish subjects were tested with the test battery including questionnaires, pure tone audiometry (PTA) at conventional and extended high frequencies (EHFs), distortion product OAE (DPOAE), SPIN and AEP measurements [e.g., auditory brainstem response (ABRs) and envelope following responses (EFRs)]. Participants were divided into three groups: (i) young normal-hearing adults (18-25 years), (ii) normal hearing adults who complain of tinnitus or reduced speech intelligibility in noise (18-60 years), but have a normal audiogram and (iii) hearing impaired adults (18-60 years). We explored the effect of age, hearing sensitivity and self-reported hearing/tinnitus complaints on potential biomarkers of CS (e.g., EFR magnitude, ABR wave-I and V amplitudes) and speech reception thresholds, and studied their relation to metrics which are sensitive to OHC-loss (e.g., hearing thresholds and DPOAEs). We found that the EFR reductions observed in the older groups is consistent with age-related CS, but that it remains challenging to parse out the role of extended high-frequency hearing loss and CS to predict speech-recognition declines. We conclude that early markers of SNHL (EFRs or extended high frequency thresholds) are crucial for a timely diagnosis of speech intelligibility problems with age.

Funding: Work supported by ERC STG 678120 and ERC POC 899858.

P72 Who's behind the mask? Facemasks interact with talker and listener characteristics to affect speech intelligibility in noise

Paulina Swiatek, Charles Masterson, **Rachel Smith**
University of Glasgow, UK

Facemasks are a type of visual noise that can make speech harder to understand [Chládková et al., 2021, Psych. Bull. Rev 28(3):992; Toscano & Toscano, 2021, PLoS ONE 16(2):1]. We report two experiments to investigate how this visual noise affects speech intelligibility in combination with acoustic noise and with talker and listener characteristics: the listener's autism-spectrum traits (Experiment 1), and talker familiarity (Experiment 2).

Experiment 1 investigated whether listeners with greater autism-spectrum traits are more affected by the presence of facemasks alongside noise (multitalker babble), compared to those with lesser autism-spectrum traits. 40 participants aged 19-25 completed an online survey consisting of an Autism Spectrum Quotient (AQ) questionnaire [Baron-Cohen, 2001, J. Autism Dev. Disord., 31:5] followed by video recordings of 20 sentences from the SPIN test [Kalikow, Stevens & Elliott 1977, JASA, 61:1337], which they attempted to transcribe. Presence/absence of a facemask and of noise were manipulated in a factorial within-participants design and counterbalanced across sentences. Sentence intelligibility was scored in terms of words correctly reported. At the group level, both noise and presence of a facemask reduced intelligibility. For each participant we separately calculated the drops in intelligibility they experienced due to presence of a facemask, and to noise. Participants' AQ scores were significantly correlated with their intelligibility drop due to noise, and marginally correlated with the intelligibility drop due to a facemask.

Experiment 2 investigated whether the listener's personal familiarity with the talker affects intelligibility in cases where the talker is wearing a facemask. 16 participants each transcribed video recordings of 56 sentences from the SPIN test spoken by 4 talkers. Familiarity and presence of a facemask were manipulated in a within-participants design, and were counterbalanced across sentences. Each listener was personally familiar with two of the talkers and not with the other two; each talker appeared in both masked and unmasked conditions. Sentence intelligibility was again scored in terms of words correctly reported. While the unmasked conditions displayed high accuracy regardless of talker familiarity, in the masked conditions familiar talkers were more intelligible than unfamiliar talkers.

Taken together, the experiments show that the presence of a facemask interacts with and accentuates known adverse influences on speech intelligibility and allow us to identify situations where the wearing of facemasks is particularly likely to lead to comprehension difficulties.

P73 Continuous real-time rating of speech perception using connected mobile devices with emoTouch Web

Christoph Louven, Carolin Scholle, Fabian Gehrs
Universität Osnabrück, Institut für Musikwissenschaft und Musikpädagogik, DE

The production and perception of speech are dynamic phenomena that evolve and change over time. Researching these dynamic phenomena therefore also requires dynamic research instruments that make the development processes continuously observable in real time. This requirement for a valid research approach for time-bound phenomena is basically similar to the situation in other fields, e.g. music, theater, dance, or sports.

To meet the need of different disciplines for an up-to-date, easy-to-use research tools for real-time research we developed 'emoTouch Web', a web-based research system for continuous real-time evaluation of videos, music or live events of any kind. The development at Osnabrück University (Germany) was funded by the Volkswagen Foundation. The system is available free of charge for scientific purposes.

The system turns any networked smartphone and tablet as well as any desktop computer into a flexibly configurable and easy-to-use tool for real time research. For example, the audience of a lecture can participate in a previously designed continuous evaluation study just by accessing a website with the smartphones they carry anyway ('Bring-Your-Own-Device'). This makes it possible in a simple way to comprehensively conduct studies at live events with possibly hundreds of participants at the same time. However, it is of course also possible to conduct studies with desktop computers in a laboratory setting or as an online real-time survey.

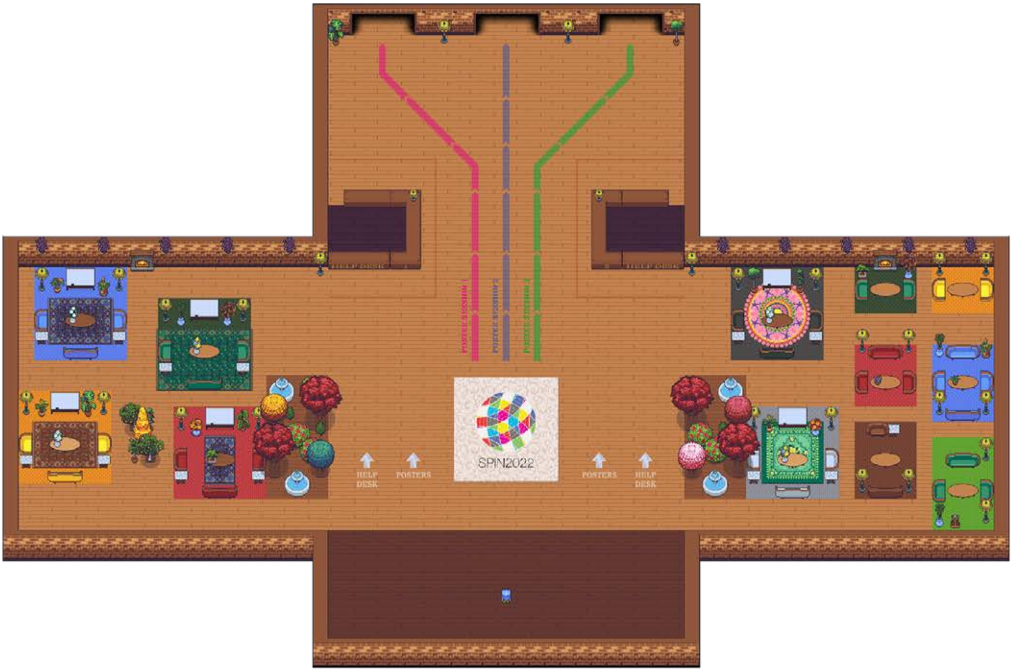
emoTouch Web was originally designed and developed for empirical audience research in the field of music psychology. However, the system is completely flexible and freely configurable and thus not limited to a specific research question or discipline. In the graphical editor of emoTouch Web, the study layout can be freely designed with numerous interactive elements (e.g. horizontal and vertical sliders, 2D rating areas, categorical scales, images, videos). The layouts dynamically adapt to the various mobile devices. The execution of a study can be controlled, monitored and observed by the researcher in real time. For the evaluation of the collected real-time data, coordinated tools for graphical and numerical analysis as well as interfaces to the scripting languages Python and JavaScript and flexible export options are integrated. At https://www.emotouch.de, a demo study shows the various possibilities of the study layout. Via a test access, the system can be tried out and own test studies can be designed, conducted and evaluated.

The poster presentation shows the possibilities of emoTouch Web by means of selected pilot studies and explains possible application scenarios in speech-related research, e.g. the continuous evaluation of the intelligibility during an evolving sound environment.

P74 Effect of semantically related and unrelated competing message on sentence production in neuro-typical adults

Abhishek Budiguppe Panchakshari
Nitte Institute of Speech and Hearing, Mangalore Karnataka, India

Lexical access is described as the retrieval of the most appropriate word from the lexicon. During this process, a set of words would compete for the process of selection/activation. The target word would be retrieved from the lexicon amidst this competition and this activation is generally based on the context. Few proponents in this direction propose that the competitor word may facilitate word activation (facilitation), while few other researchers postulate that these competitor words may impede the process of lexical activation (inhibition). These two mechanisms of lexical retrieval are tested even at the level of sentence production (using priming and the picture-word interference — PWI — paradigm). In the PWI paradigm, the competitor word is presented before the target word. This competitor word can be either semantically related or unrelated to the target. The current study used a paradigm similar to PWI but the basic difference was that the precursor was also presented in auditory mode. The reaction time and accuracy would be analyzed separately for the target words preceded by semantically related and unrelated words, based on which inference on facilitation and inhibition is deduced. The current study aimed at examining the mechanism of lexical retrieval at the level of sentence production. Semantically related and unrelated sentences acted as competitors to these sentences. The study was conducted in Kannada, a south Indian language. A total of 30 neurologically healthy young adults in the age range of 30-45 years served as participants. The stimulus was derived from the Kannada action naming test. The task required the participants to name the stimulus by using a noun and verb (example: “Two men are shaking hands”). Sixty such sentences were used in the current study. Each of these sentences were preceded by a precursor. Fifteen sentences had semantically related nouns as precursors, 15 sentences had semantically related verbs as precursors while 30 sentences had an unrelated noun and verb as precursors. The duration between the precursor and target was 500 milliseconds. The inter-stimulus duration between two stimuli was 500 milliseconds. Vocal reaction time and accuracy scores were the dependent variables. The vocal reaction time and accuracy scores were better for target sentences preceded by semantically related verbs followed by semantically related nouns. The participants experienced more difficulty in naming sentences with unrelated nouns and verbs. The results signify the role of semantically related competing messages on sentence production. The results will be discussed in details.



HELP DESK
POSTERS
SPIN2022
POSTERS
HELP DESK