# Organization

The Speech in Noise 2015 Workshop is hosted by the Technical University of Denmark and will take place in Copenhagen on January 8th and 9th.

## Organising Committee

Gaston Hilkhuysen (LMA)
Dorothea Wendt (DTU)
Sébastien Santurette (DTU)
Caroline van Oosterhout (DTU)
Rolph Houben (AMC)
Alexandre Chabot-Leclerc (DTU)

## The Workshop on the Web

| | |
|---|---|
| London '09 | Not available on the web |
| Amsterdam '10 | http://www.phon.ucl.ac.uk/events/Quality2010.html |
| Lyon '11 | http://www.phon.ucl.ac.uk/events/quality2011/ |
| Cardiff '12 | http://www.phon.ucl.ac.uk/events/quality2012/ |
| Vitoria '13 | http://www.phon.ucl.ac.uk/events/quality2013/ |
| Marseille '14 | http://www.phon.ucl.ac.uk/events/spin2014/ |
| Copenhagen '15 | http://www.spin2015.dk/ |

## Sponsoring

We warmly thank our sponsors: Carlsberg Foundation, Oticon, Phonak, Resound, Siemens, & Widex

# Table of Contents

## Oral Presentations

## Poster Presentations

# Oral presentations

# O1: Computational speech segregation based on an auditory-inspired modulation analysis

**Tobias May** and Torsten Dau

Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Kongens Lyngby, DK
tobmay@elektro.dtu.dk

One of the most striking abilities of the human auditory system is the capability to focus on a desired target source and to segregate it from interfering background noise. In this contribution, we consider the challenging problem of segregating speech from a noisy mixture by only exploiting monaural cues. More specifically, a speech segregation system is presented that estimates the ideal binary mask from noisy speech based on the supervised learning of amplitude modulation spectrogram (AMS) features. Instead of using linearly scaled modulation filters with constant absolute bandwidth, an auditory-inspired modulation filterbank with logarithmically scaled filters is employed. To reduce the dependency of the AMS features on the overall background noise level, a feature normalization stage is applied. In addition, a spectro-temporal integration stage is incorporated to exploit the context information about speech activity present in neighboring time-frequency units. It will be shown that auditory-inspired modulation processing can substantially improve the mask estimation accuracy of computational speech segregation systems in the presence of stationary and fluctuating interferers. Moreover, the ability to generalize to unseen acoustic conditions is evaluated by training the segregation system with a limited set of low signal-to-noise ratio (SNR) conditions, but testing it over a wide range of SNRs up to 20 dB.

# O2: Assessment of human speech intelligibility based on machine listening

**Bernd T. Meyer**

Medizinische Physik and Cluster of Excellence Hearing4all, Universität Oldenburg, DE
bernd.meyer@uni-oldenburg.de

One common approach pursued by engineers working on speech research is to improve speech processing algorithms by learning from the principles of the healthy human auditory system. In this talk, I will focus on the opposite direction: How can the technology developed for automatic speech recognition (ASR) be applied to create more capable models of speech intelligibility and to better understand the process of speech perception. The latter is investigated in experiments based on electrocorticography (intracranial EEG) data obtained from epilepsy patients listening to a sentence matrix test. The cortical measurements are compared to standard features borrowed from ASR, as well as to modulation features developed in our lab; both of these represent physical representations of the stimulus. Further, a typical ASR processing chain covers categorical representations such as phonemes and words. Hence, speech algorithms enable a comparison of cortical signals to a wide range of speech-specific representations. I will present early results on this analysis, as well as an approach to reproduce listener's errors with ASR for improving models of speech perception.

# O3: Utilization of the Lombard effect for the intelligibility enhancement of telephone speech

**Emma Jokinen** and Paavo Alku

Aalto University, FI
emma.jokinen@aalto.fi

In mobile communications, pre- and post-processing methods are used to enhance the quality and intelligibility of degraded speech. The degradation can be caused by quantization noise or acoustical background noise from the sending or receiving side of the communication channel, referred to as far end or near end, respectively. Noise reduction methods are typically used in mobile phones as a pre-processing step to remove far-end noise before speech is transmitted over the communication channel. In contrast, techniques that process speech in order to enhance its quality and intelligibility when listened to in noisy near-end conditions are implemented as a post-processing step in the receiving side of the channel. In high background noise levels, the goal of the post-processing is to make the speech stand out better from the background noise by enhancing its acoustic cues. While several post-processing techniques have been shown to improve intelligibility in adverse noise conditions, they are not capable of modelling the spectral changes that occur in natural speech, for instance, when vocal effort is increased in order to enhance loudness. An example of this is the Lombard effect which is observed when talkers modify their speaking style in an effort to increase the intelligibility of their speech in the presence of background noise. The increased intelligibility in Lombard speech is a combination of several factors, such as flattening of the spectral tilt, slower speaking rate, and increased vocal intensity. In this talk, the usage of adaptive post-filtering methods motivated by the natural Lombard effect is discussed. As a first step, the statistical dependencies of normal and Lombard speech were used to adjust the spectral tilt of the received speech frames. Subjective intelligibility tests indicate that with this kind of a statistical mapping, similar intelligibility improvement can be achieved as with fixed high-pass filtering.

# O4: Integrating beamforming with binaural sound reproduction using a spherical microphone array

**Michael Jeffet**[1], Noam R. Shabtai[1,2], and Boaz Rafaely[1]

[1] Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Be'er Sheva, IL
[2] Institute of Technical Acoustics, RWTH Aachen University, Aachen, DE
mickeyjeffet@gmail.com

Microphone arrays are widely used in speech enhancement systems for noisy and reverberant environments. Recently, a novel beamforming approach was developed, incorporating binaural sound reproduction in the beamforming process. This approach, called the generalized spherical array beamformer (GSB), spatially suppresses unwanted interferences, while maintaining the spatial information through the binaural cues; thus it improves both the speech intelligibility and the spatial realism of the output signal when reproduced over headphones. Practically, microphone arrays have a finite number of microphones; therefore, the captured acoustic scene suffers from a limited representation by the array signals. In previous work, this practical constraint was bypassed by using an unlimited representation of the acoustic scene. In this work, we constrain the representation of the acoustic scene to meet practical limitations and investigate the theoretical limits on performance. Theoretical results show a clear bound on the maximum order of the beamforming function and of the head-related transfer functions. In order to validate the theoretical results, a listening test was conducted and confirmed the theory, testing the GSB for the first time under the bounded conditions. The GSB in the listening test was simulated in a room environment that is typical of a standard video conferencing scene, thus adding a real-life dimension to the theoretical validation. Results show a clear trade-off between binaural reproduction and spatial selectivity performance, such that improving one is only possible at the expense of degrading the other. In order to benefit from the incorporation of both methods, despite the inevitable trade-off, a tunable GSB is proposed. By tuning between binaural reproduction and beamforming, the user is able to adjust the GSB to spatially suppress interference while maintaining some of the spatial information of the acoustic scene.

# *O5: Keynote Lecture*
# Predicting the intelligibility of connected speech and singing in adverse listening conditions

**Sarah Hawkins**

Centre for Music and Science, Faculty of Music, University of Cambridge, UK
sh110@cam.ac.uk

Much is known about speech perception in noise, yet predicting exactly what will be intelligible, and what mistakes will be made, tends to be inexact, presumably because of multiple potential influences. Major general influences can be summarised as the predictability of the speech, and the physical properties of the signal and ambient interference. However, this seemingly-simple dichotomy is in fact complicated because knowledge and signal/system factors affect both predictability and the perceptual systems interpretation of physical properties. Another complication is that background noise is not always undesirable and to be ignored. Ensemble singing conveys a linguistic message in highly desirable noise. I will explore results from experiments on intelligibility of spoken and sung text. The speech experiments mainly addressed word intelligibility due to fine phonetic detail signalling /r/, to elucidate (1) interacting influences of knowledge and physical signal, and (2) local vs long-domain effects of semantic or acoustic predictability. Singing experiments explored intelligibility in (3) different styles (medieval motets, lively jingles, modern Western classical), with (4) differing linguistic predictability and background sound. Some used meaningless sentences, one assessed word predictability using SPIN-type sentences in background spoken babble, sung vowels, or [ʃ], 'sh'. Conclusions are that linguistic predictability, ambient noise type and its relationship to the signals spectral properties, and signal-to-noise ratio, apply as much to singing as to speech. Fine phonetic detail seems to exert a stronger local than long-domain cumulative influence (though types of detail, and of meaning, should be distinguished), and while knowledge-based factors critically affect singing intelligibility, some acoustic factors are critically important. These consistently include discontinuities in f0, due either to phonetic or musical parameters, and rhythmic factors reflected in phonological vowel length. I speculate that understanding a linguistic message involves a form of pattern perception in which a fundamental process is a search for familiar auditory patterns (syllables, words, longer phrases) that fit most-likely meanings. In this process, the short-time physical signal is more important than expectation, but the overall pattern must fit spectro-temporal expectations over longer domains. When rhythm is as expected, local spectral detail can be interpreted appropriately by combining regions of high acoustic-phonetic certainty and pattern completion. Unexpected rhythm entails searches for less-likely interpretations, especially in adverse listening conditions. These suggestions seem broadly compatible with research on influences of amplitude envelope and temporal fine structure, which, for predicting intelligibility of connected speech, need to be related to linguistic structure and pragmatic function.

# O6: The access of mental representations of speech in face of signal degradation

**Anita Wagner**

University Medical Centre Groningen
a.wagner@umcg.nl

During speech comprehension normal hearing (NH) listeners can quickly map the heard signal to speech representations in their mental lexicon. This lexical access is an automatic process that is facilitated by the multitude of fine acoustic details present in natural speech (e.g., Salverda, Dahan and McQueen, 2003). Fine differences in, for instance, duration of syllables, facilitate listeners disambiguation of words embedded in other words (e.g., pain in painting), or across word boundaries (e.g. cancer in can sir). How does signal degradation affect lexical access? Can listeners make use of intact durational cues when processing spectrally degraded signal? Durational cues can be particularly important for users of cochlear implants (CI) since they are reliably transmitted through the device. We compare lexical processing by experienced CI users with NH listeners processing of sentences presented with and without acoustic CI simulations. We study how lexical competition between similar words is affected by spectral degradation. Our eye-tracking experiments record how manipulation of duration affects listeners estimate of a lexical target as the signal unfolds over time. Insight into lexical competition is gained through observing listeners gaze fixations to pictures displayed, which show a target (e.g., painting) and an embedded lexical competitor (e.g., pain) next to unrelated distractors. When listening to natural speech lexical access is instantly guided by the acoustic input: manipulation of duration shifts listeners gazes towards the object that is most likely given the duration. In cases where durational cues prove to be misleading, listeners can quickly recover from spurious competition. Spectral degradation of the signal reduces the effect of durational cues, and prolongs lexical competition. The gaze fixations of most, but not all, CI users reflect high sensitivity to durational cues, which then also govern lexical competition. In cases where these cues turn out to be misleading, however, CI users fail to quickly recover from spurious competition. Our results show that spectral degradation delays listeners access to durational cues, and holds back lexical competition, which comes at the cost of higher processing effort. Through experience CI listeners can regain access to durational cues, but among CI users there is individual variation in their sensitivity to duration as a cue that facilitates lexical access.

**Acknowledgments**
This research is funded by Marie Curie Actions (FP7).

**Reference**

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. Cognition, 90.

# O7: The eye as a window to the listening brain

**Adriana A. Zekveld**[1], Thomas Koelewijn[1], Dirk J. Heslenfeld[2], Barbara G. Shinn-Cunningham[3], Ingrid S. Johnsrude[4], Niek J. Versfeld[1], and Sophia E. Kramer[1]

[1] Section Audiology, Dept. of Otolaryngology-Head and Neck Surgery and EMGO Institute, VU University medical center, Amsterdam, NL

[2] Department of Psychology, VU University, Amsterdam, NL

[3] Department of Biomedical Engineering, Center for Computational Neuroscience and Neural Technology, Boston University, Boston, USA

[4] Department of Psychology, Queen's University, Kingston Ontario, CA

aa.zekveld@vumc.nl

Understanding speech in noise requires a considerable amount of effort. An objective measure of processing load is task-evoked pupil dilation. We assessed how cognitive processes like attention affect cognitive processing load. Furthermore, we examined the neural correlates of the pupil response in order to provide insight into the processes reflected by the pupil size. Study 1 investigated how divided attention influences cognitive load as indexed with pupillometry during speech recognition. The pupil response in 12 normal-hearing young adults was recorded while they focused on either one or both of two sentences that were presented dichotically and masked by fluctuating noise across a range of signal-to-noise ratios. Performance decreased when two target sentences were processed instead of one. Additionally, dividing attention to process two sentences resulted in larger pupil dilation than processing only one. In Study 2, we used functional magnetic resonance imaging (fMRI) in combination with pupillometry to identify the brain regions associated with pupil dilation evoked by comprehension of degraded spoken sentences in 17 normal-hearing listeners. Sentences were degraded in three different ways: the target female speech was masked by fluctuating noise, by speech from a single male speaker, or the target speech was noise-vocoded. Either 50% or 84% of the sentences were intelligible. Control conditions included clear speech in quiet, and silent trials. The peak pupil dilation was larger for the 50% compared to the 84% intelligibility condition, and largest for speech masked by the single-talker masker, followed by speech masked by fluctuating noise, and smallest for noise-vocoded speech. Activation in the bilateral superior temporal gyrus (STG) showed the same pattern. Larger peak pupil dilation was associated with more activation in the bilateral STG, bilateral ventral and dorsal anterior cingulate cortex and several frontal brain areas. A subset of the temporal region sensitive to pupil dilation was also sensitive to speech intelligibility and degradation type. By means of pupillometry and fMRI we show that the task-evoked pupil dilation reflects processes that are increasingly recruited when attention needs to be divided among separate speech streams. These processes are more relevant for degraded compared to clear speech, when the degradation type involves imposing a masker rather than distorting the target speech, and for speech-maskers as compared to noise-maskers. The pupil response is sensitive to speech segregation processes that are affected by attentional processes. The pupil response can be viewed as a "window" to the listening brain.

# O8: Cognition in hearing aid users

**Elaine Ng**[1], Mary Rudner[1], Thomas Lunner[1,2], and Jerker Rönnberg[1]

[1] Linnaeus Centre HEAD, Swedish Institute for Disability Research, Department of Behavioural Sciences and Learning, Linköping University, SE

[2] Eriksholm Research Centre, Oticon A/S, Snekersten, DK

hoi.ning.ng@liu.se

Cognitive abilities vary between individuals and have been shown to be related to hearing aid benefit. How individual differences in cognitive abilities interact with signal processing to reduce listening effort will be discussed in this presentation. Two studies were performed to investigate the effect of a hearing aid signal processing algorithm on free recall of speech heard in noise in hearing aid users, and the role of cognition. The specific aims were to develop a free recall test to measure this effect and to test whether the effect would interact with background noise and/or individual differences in cognitive capacity. Results demonstrated that noise impairs the ability to recall intelligible speech heard in noise. Noise reduction freed up cognitive resources and alleviated the negative impact of noise on memory when speech stimuli were presented in background noise consisting of speech babble. The possible underlying mechanisms are that noise reduction facilitates segregation of the auditory stream into target and irrelevant speech and reduces the capture of attention by the linguistic information in irrelevant speech. In both studies, the effect of noise reduction on free recall performance was modulated by individual differences in cognitive capacity, suggesting that the mechanism by which noise reduction facilitates free recall on speech heard in noise is dependent on working memory capacity.

# O9: Efficient SpiN testing for the routine evaluation of French cochlear implanted subjects

**Chris James**[1,2], Mathieu Marx[1], Marie-Laurence Laborde[1], and Olivier Deguine[1]

[1] CHU Toulouse Purpan, Toulouse, FR
[2] Cochlear France, Toulouse, FR
cjames@cochlear.com

We describe sentence test lists developed for the evaluation of speech recognition in noise by cochlear implanted (CI) subjects. The lists were originally developed to determine the performance gain for CI for candidates with marginal benefit from acoustic amplification; hence the moniker "MBAA" sentence lists. The 36 lists of 15 sentences were revised in 2003 and recorded by a female speaker with a standard accent. The lists are scored by word; 100 to 103 words in each list. Babble noise was constructed from recordings of 2 male and 2 female speakers reading aloud. The recorded "MMBA2" sentence lists were provided on compact discs each with a pre-mixed signal-to-noise ratio (Quiet, 10, 5, 2.5 and 0 dB SNR, with constant speech level). Each recorded list takes 2 minutes 6 seconds to present. The lists were distributed to a number of cochlear implant centres in France, and have been employed in a number of clinical and experimental studies. However, to date, little has been published on their characteristics. We present data on two studies comparing the scores on the MBAA2 at fixed SNRs with the adaptive French MATRIX SRT test; one with normal listeners and the other with CI subjects. In addition we present preliminary data for a sequence of 127 adult CI patients followed from the day of activation to 12 months post-activation. At each visit CI patients were tested with one list per fixed SNR, progressively reducing SNR until they scored below 50% correct. At 1 month post-activation 68% scored greater than 50% correct in quiet and 44% greater than 50% correct at 10 dB SNR. These proportions increased to 91% and 76% at 12 months. The data set provides information on the acquisition of speech recognition for CI users, as well as the range of performance-intensity (PI) functions encountered. We discuss the potential to formalise the testing protocol with the aim to optimise measurement of the entire PI function for individual subjects where intermediate SNRs may be chosen based upon initial results at say 10 dB SNR. A custom software tool was developed to mix the signals and score words in lists. The whole PI function approach may help us understand what determines the time course for acquisition of speech recognition in noise and which factors may limit speech recognition in noise for CI subjects.

# O10: Articulatory-motor regions in acoustically-degraded word processing – converging evidence

**Alexis Hervais-Adelman**

Alexis Hervais-Adelman, Brain and Language Lab, University of Geneva, CH
alexis.adelman@unige.ch

Speech we hear in our everyday lives is highly varied, be it as a result of the differences between speakers or between listening environments. The brain deploys powerful compensatory mechanisms to allow comprehension of speech that is, for example, heavily accented, masked by noise, compressed or otherwise non-canonical. I will present fMRI studies with different paradigms that provide converging evidence that part of this compensatory mechanism is at least partly based on an articulatory-motor network. In a first experiment, 15 participants listened attentively to clear and spectrally degraded words (6-band noise vocoded), while monitoring for occasional deviant stimuli. Degraded words were presented at two levels of degradation, one potentially comprehensible and the other never comprehensible. Results indicate that potentially comprehensible degraded words, but not incomprehensible ones, elicit greater activation in the left premotor cortex and the anterior insulae than clear speech. In a second experiment, nine listeners participated in an fMRI study on masked backward semantic priming, investigating the interaction between signal to noise ratio (SNR) and semantic relatedness on word recognition. Target words were presented at a range of SNR levels (-9dB, -7dB, -5dB and unmasked). Comprehension of masked words was tested with a two-alternative forced choice identification task. A significant correlation was found between masking level and activity in left premotor cortex, left supplementary motor area, and the anterior insulae. In a third study we employed EEG from 14 volunteers who listened to a series of noise vocoded (NV) and noise-vocoded spectrally-rotated (rNV) monosyllabic words, while they carried out a detection task (monitoring for animal names). We sought components of the EEG response that showed an interaction between spectral rotation and spectral degradation. Time-frequency analysis of the EEG signal in the alpha-band revealed a monotonic increase in event-related desynchronization (ERD) for the NV but not the rNV stimuli in the alpha band from 420560 ms, reflecting a direct relationship between the strength of alpha-band ERD and intelligibility. The effect was localized to the left superior posterior temporal cortex, an area known to be involved in audio-motor integration. These studies used different acoustic degradations and called for different cognitive efforts from the participants, but converge in implicating a network of speech-motor regions interacting with auditory association areas in the comprehension of degraded speech.

# O11: Consonant perception – sources of perceptual variability and modeling approaches

**Johannes Zaar** and Torsten Dau

Hearing Systems Group, Technical University of Denmark, DK
jzaar@elektro.dtu.dk

In consonant perception experiments, listeners are typically presented with nonsense speech stimuli consisting of combinations of consonants and vowels (e.g. /ka/, /ba/, etc.). Commonly, the speech stimuli are mixed with steady-state noise at various signal-to-noise ratios (SNRs) and the listeners responses are evaluated in terms of consonant recognition and consonant confusions. Compared to more "macroscopic" speech perception measurements using single words or entire sentences, this "microscopic" approach has the advantage that effects of higher-stage speech processing (e.g. lexical and context-related effects) are negligible, which facilitates an analysis of the mapping from the acoustic properties of the stimulus to the phone percept. Consonant perception data often show a large perceptual variability. Nevertheless, the data are usually averaged across stimuli of the same phonetic identity and across listeners. In the first part of the talk, the results from an experimental study investigating the influence of different sources of variability in consonant perception are discussed. It was distinguished between source-induced variability, referring to perceptual differences caused by acoustical differences in the speech and/or the masking-noise tokens, and receiver-related variability, referring to perceptual differences caused by within- and across-listener uncertainty. In the second part of the talk, different auditory-inspired modeling concepts are compared with regard to their suitability for consonant perception modeling. An audibility-based approach that corresponds to the Articulation Index (AI) and a modulation-frequency selective approach, as reflected in the speech-based Envelope Power Spectrum Model (sEPSM), were considered. Using the experimental data as a reference, the model outcomes are discussed regarding their respective suitability for the prediction of consonant perception.

# O12: The role of periodicity in perceiving speech in quiet and in background noise

**Kurt Steinmetzger**

UCL Speech, Hearing and Phonetic Sciences, UK
kurt.steinmetzger.12@ucl.ac.uk

Speech consists of a mix of periodic (voiced) and aperiodic (unvoiced) segments with very distinct acoustic properties. In a series of behavioral and electrophysiological experiments we have investigated how the presence and absence of periodicity in both speech and masker affects the ability of normal-hearing listeners to perceive speech and whether periodicity has an effect on cortical EEG signals in response to speech. In a first study we have found that in quiet, fully periodic speech was less intelligible than speech with a natural variation of voiced and unvoiced segments or completely aperiodic noise-vocoded speech. Secondly, Speech Reception Thresholds (SRTs) were measured adaptively and the same targets were combined with four different maskers: speech-shaped noise, harmonic complexes with a dynamically varying F0 contour, and amplitude-modulated versions of both. Performance was much better when the masker was periodic and also slightly improved with more periodicity in the target. Furthermore, high intelligibility rates of the speech materials in quiet were found to be necessary to enable a fluctuating-masker benefit and independently led to lower SRTs. Thirdly, we recorded EEG signals in response to target speech with varying amounts of periodicity. Sorting the single trials according to the spoken behavioural responses allowed us to examine acoustic effects while controlling for intelligibility, and vice versa. EEG waveforms were found to be consistently more negative with more periodicity and, to a slightly lesser degree, also for speech that is intelligible. An analysis of the power spectra of the EEG during the stimulus interval showed the same pattern of results in the delta band (1-4 Hz). These power differences were absent during the preceding baseline window but interestingly we instead found more alpha power (7-12 Hz) for unintelligible speech there. Finally, the neural oscillation pattern over time was found to strongly depend on the amount of periodicity in the speech signal. In summary, these results show that periodicity, particularly in the masker, might be an even more important factor then masker fluctuations when attempting to segregate speech from a background noise and that periodicity plays an important role in shaping the cortical representation of speech.

# O13: Neural oscillations reflect attentional challenges of understanding speech in noise

**Malte Wöstmann**

Research Group "Auditory Cognition", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, DE
woestmann@cbs.mpg.de

Understanding speech in noise is effortful. It critically depends on the attentional selection of the relevant speech signal from the mixture of sounds. I will present data from our lab, highlighting the role of neural alpha oscillations (~10 Hz) during effortful listening. Alpha oscillations reflect attentional challenges (e.g., acoustic degradations) as well as the use of contextual information (e.g., predictive cues) to overcome these challenges. Recently, we found that modulations of alpha oscillations also depend on the listening effort experienced with increasing age and with progressive hearing loss. Critically, alpha oscillations do not only indicate listening effort, but also listener intent (i.e., which auditory stream is in the listeners current focus of attention). In a dichotic listening study, we found that alpha oscillations lateralize to either hemisphere of the brain, depending on which speaker is attended. This alpha lateralization fluctuates at the speech rate and predicts participants recall of information from the attended stream. I will discuss the functional mechanisms of alpha oscillations for speech processing and possible implications for hearing aids.

# Poster presentations

# P1: Beyond Speech Intelligibility: using Response Times, Sound Quality, and Task Load to evaluate the benefit of Noise Reduction

C. Lesimple and J. Tantau

1 Bernafon AG, Bern, CH
cl@bernafon.ch

Noise reduction (NR) systems are found widely within digital hearing instruments. Objective benefits of this feature have been difficult to demonstrate using currently available speech-in-noise intelligibility tests. Alternative outcomes have been shown to be good indicators of NR benefit including response times (Gatehouse, 1990), rated sound quality (Alcantara, et al., 2003), and task load (Ahlstrom, et al., 2013). A word closed-set speech recognition test based on the Wallenberg & Kollmeier's Rhyme Test (WAKO) (Wallenberg and Kollmeier, 1989) combined with a modified version of the NASA Task Loading Index (Hart & Staveland, 1988) was developed in order to evaluate the effect of NR on speech intelligibility (SI), response time (RT), sound quality (SQ), and task load (TL). The TL combines different aspects of perceived workload regarding effort, frustration, performance, and mental demand. The following research questions were addressed in the investigation:

1) Does amplification improve the tested variables using a closed-set test?
2) Does the NR algorithm provide any measurable benefit with this test design?

The efficiency of a commercially available noise reduction was evaluated at 0 and +5 dB signal-to-noise ratio (SNR). A technical evaluation using the hearing instrument output SNR and a spectral comparison were initially measured. Subsequently, nineteen subjects with a mild to moderately severe bilateral sensorineural hearing loss participated in this investigation. The test conditions included unaided, aided, and aided with NR. It was found that amplification significantly improved SI, RT, and SQ (1). As expected there were no significant differences in SI when the aided conditions were compared. The NR significantly improved the SQ and effort scores which reflect the measured output SNR improvement (2). Speech scores from speech-in-noise tests only measure one aspect of speech understanding. Future research should examine the broader picture by further investigating the implications of other aspects of speech understanding.

### Declaration of interest

The authors are all employees of Bernafon AG. The authors alone are responsible for the content and writing of the paper.

### References

Ahlstrom, J. B., Horwitz, A. R., & Dubno, J. R. (2013). Spatial separation benefit for unaided and aided listening. Ear and Hearing, 35(1), 7285.

Alcantara, J. I., Moore, B. C. J., Kuhnel, V., & Launer, S. (2003). "Evaluation of the noise reduction system in a commercial digital hearing aid," Int. J. Audiol. 42, 3442.

Gatehouse, S. & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. British Journal of Audiology, 24(1), 638.

Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press

Hagerman, B. & Olofsson, A. (2004). A Method to Measure the Effect of Noise Reduction Algorithms using Simultaneous Speech and Noise. Acta Acustica, 90, 356361.

Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press

Sherbecoe RL & Studebaker GA. (2004) Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units. Int J Audiol, 43(8):442-8.

Svensson, E. (2003). Statistical methods for repeated qualitative assessments on scales. International Journal of Audiology, 42 Suppl 1, S1322.

Wallenberg, E.L. & Kollmeier, B., (1989). “Sprachverständlichkeitsmessungen für die Audiologie mit einem Reimtest in deutscher Sprache: Erstellung und Evaluation von Testlisten”, Audiologische Akustik 38, 50-65.

# P2: The effect of focussed attention on speech perception and listening effort

Thomas Koelewijn[1], Hilde de Kluiver[1], Barbara G. Shinn-Cunningham[2], Adriana A. Zekveld[1,3], and Sophia E. Kramer[1]

[1] Section Audiology, department of Otolaryngology-Head and Neck Surgery and EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, NL
[2] Department of Biomedical Engineering, Center for Computational Neuroscience and Neural Technology, Boston University, Boston, USA
[3] Linnaeus Centre HEAD, Department of Behavioral Sciences and Learning, Linköping, SE
t.koelewijn@vumc.nl

Talking with people in a group is often experienced as more effortful than having a one-to-one conversation. Research shows that knowing where, when, and who is going to talk improves speech intelligibility in more complex listening situations. In other words, being able to focus attention on the talker is beneficial for performance. However, how it affects listening effort is unknown. The pupil dilation response (PDR) is an index of cognitive load and is used as an objective measure for listening effort. In a recent study we showed an effect of attentional load on the PDR. The current study extends these results by investigating how the PDR is influenced by prior knowledge of target speech location, onset, and who is going to talk. In total 56 young adults with normal hearing performed a listening task in which sentences that were spoken by the target talker had to be recalled, while a simultaneously presented distracter sentence had to be ignored. Both sentences were independently masked by fluctuating noise. In three separate experiments, target location (left or right ear), speech onset, and talker variability were manipulated by keeping these features either fixed during an entire block or by randomizing these over trials. Pupil responses were recorded during listening and performance was scored after recall by the experimenter. The results showed an improvement in performance when participants were able to focus on the target location. Being able to focus on the target location and target talker decreased the PDR. These results suggest that prior knowledge of location and talker identity reduces cognitive load, without necessarily affecting behavioral performance in the case of target identity. The results suggest that target location and target identity influence cognitive load during speech processing. We conclude that communicating in a dynamic environment like a cocktail party, where multiple persons might talk at the same time and walk around, requires substantial listening effort because of the demands placed on attentional processes.

# P3: Is vocabulary size a reliable predictor for performance in speech intelligibility tasks?

Rebecca Carroll and Anna Warzybok

Cluster of Excellence "Hearing4all", University of Oldenburg, DE
rebecca.carroll@uni-oldenburg.de

Several researchers have suggested linguistic skills to correlate with speech intelligibility scores. In particular, vocabulary size has been considered a useful measure of linguistic skills or abilities. Benard et al. (2014) found significant correlations of vocabulary size, as measured by a Dutch version of the Peabody Picture Verification Test (PPVT), and intelligibility scores for interrupted speech using Versfeld sentences (Versfeld et al., 2000), with an interruption rate of 2.5 Hz. They reasoned that the restoration mechanism used to understand interrupted speech (somehow) involves vocabulary knowledge and verbal intelligence. Our aims were (i) to determine whether such correlations could also be observed for other acoustic conditions, and (ii) to replicate Benard et al. (2014) observations for German. We measured speech intelligibility of the Göttingen Sentence Test (Kollmeier & Wesselkamp, 1997) in test-specific, stationary noise (NOISE), in reverberation (REVERB), and in a combination of reverberation and noise (REVNOIS). For comparability with the Benard et al. (2014) study, we presented the Göttingen Sentences with interruptions at a 2.50 Hz rate (INTERRUPT). Signal-to-noise ratio and reverberation time were set to provide comparable intelligibility scores, according to STI predictions, in the conditions employed. Twelve young listeners (3 men; age 20 - 34; $\phi = 26.58$) with normal hearing participated in the study. In addition to the intelligibility tests, they also completed a standardized German version of the PPVT, another standardized test for receptive vocabulary knowledge (WST), and a lexical decision test measuring the processing times for word recognition (nonword vs. word) and lexical access (high vs. low frequency words). A significant correlation was only found for PPVT scores with the NOISE condition, and for word recognition times with REVERB scores. However, these observations appeared to depend on the choice of participants and listening conditions: Simply excluding one listener led to a similar correlation of PPVT with the INTERRUPT scores as Benard et al.'s (2014), or to a correlation with the REVERB condition. But 'outlier distribution' was not tied to single individuals, it rather varied across listening condition. It thus seems that of the different vocabulary/lexical tests, the PPVT could indeed be a more reliable correlate than the other tests. But such a conclusion would require more data points to identify, and eliminate the impact of, potential outliers. Our data suggest that different listening situations may involve lexical knowledge to different degrees, and that individual differences may also entail other auditory and/or cognitive measures yet to be tested.

# P4: Validation of the Spatial Fixed-SNR (SFS) test in anechoic and reverberant conditions

Niels Søgaard Jensen[1], Søren Laugesen[1], Filip Marchman Rønne[1], Christian Stender Simonsen[2], and Jens-Christian Britze Kijne[2]

[1] Eriksholm Research Centre, DK
[2] Oticon A/S, DK
nsj@eriksholm.com

Adaptive speech-intelligibility tests are widely used in the assessment of hearing aids. Typically, the administration of such tests is easy and the results are reliable. However, a problem associated with adaptive tests is the fact that different listeners are tested at different signal-to-noise ratios (SNRs). This introduces a possible confounding effect of the SNR on the outcome of the test. Furthermore, the test SNRs are often substantially lower than the typical daily-life SNRs, for which the hearing aids under test were designed. In such cases, the test result may not reflect the actual daily-life performance. In order to address these problems, the SFS test was developed. It is based on a paradigm where some of the parameters in a spatial speech-on-speech test are set individually in order to 'manipulate' all listeners in a given study towards the same given speech reception threshold (SRT). This will enable a comparison of hearing aids at the corresponding fixed SNR without encountering floor or ceiling effects when percent-correct words or sentences are scored. Based on findings from a previous experiment, the variable test parameters were scoring method, masker gender, and the target-masker spatial separation. The target speech material was HINT sentences, while the masker speech material was recordings of running speech. In two separate studies, the Danish SFS test was implemented and validated in an anechoic chamber and in a sound-treated listening room, respectively. N=26 and N=19 hearing-impaired test subjects took part in the two studies. Whereas the two test protocols differed on some aspects, e.g. by including two different experimental hearing-aid contrasts, both studies investigated the ability of the SFS test to bring listeners to perform within a desired percent-correct range at a given fixed SNR. Furthermore, both studies assessed the test-retest reliability. The results from the studies indicated that the SRT manipulations were reliable and similar in the anechoic and reverberant test conditions. In both test conditions, the test-retest standard deviation was estimated to be around 8.5%. Better overall test performance was observed in the anechoic chamber, probably due to the reverberation present in the listening room. In conclusion, the SFS test is a reliable method to compare different hearing aids at a given fixed SNR.

# P5: Binaural speech recognition for normal-hearing and hearing-impaired listeners in a competing voice test

Lars Bramsløw, Marianna Vatti, Renskje Hietkamp, and Niels Henrik Pontoppidan

Eriksholm Research Centre, DK
lab@eriksholm.com

Traditional speech recognition tests focus on one target talker in the background of a masker, which can be e.g. stationary noise or one or more competing talkers. In a complex real-world environment, the listener may need to attend to two or more simultaneous talkers. To simulate such a scenario, we have developed a new competing voices test where the listener must attend to two equally important targets at the same sound pressure level. The new test was used to evaluate binaural speech recognition scores in three binaural modes: 1) male and female talker in separate ears (dichotic); 2) male and female talker summed to both ears (diotic) and 3) female diotic and male phase-reversed across ears. The test person was required to repeat male, female or both sentences as indicated on a video monitor, and this visual cue was presented either before or after the playback of the sentences. The test was evaluated on four normal-hearing listeners and nine hearing-impaired listeners. Results for normal-hearing listeners showed near-perfect word scores (app. 100%) and statistically significant effects of both binaural mode and visual cue timing. The hearing-impaired listeners showed scores around 50%, also with statistically significant effects of binaural mode and visual cue timing.

# P6: Effect of residual hearing in bimodal users on top-down repair of interrupted speech

Jeanne Clarke[1], Etienne Gaudrain[1,2], and Deniz Başkent[1]

[1] University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Groningen, NL

[2] Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics Team, CNRS UMR 5292, INSERM U1028, University Lyon 1, Lyon, FR

j.n.clarke@umcg.nl

With speech cues being degraded due to limitations of the nerve-electrode interface, Cochlear Implant (CI) users do not perform as well as normal hearing persons in understanding speech in noisy environments. This may be due to the fact that pitch, a generally useful cue for perceptual organisation, is not well conveyed in CIs. A previous study from our lab indeed showed that CI users show different patterns than normal-hearing listeners for top-down restoration of degraded speech. A second study showed that the addition of pitch information to speech degraded spectrotemporally via an acoustic CI simulation enhances both the overall intelligibility and the top-down repair of interrupted speech. A similar pitch enhancement can occur in bimodal CI users (using a CI in one ear and a hearing aid  HA  in the other), where low frequency information can be transmitted in the contralateral ear (with the HA) if enough acoustic hearing is present. This motivates us to investigate how residual hearing in bimodal users can benefit top-down repair of speech. We hypothesise that the low frequency cues provided by the HA will contribute to a better pitch representation. Provided the complementary information from the two ears can be properly integrated by the brain, the additional use of the HA in combination with the CI should improve interrupted speech perception and restoration. Top-down repair of speech (or phonemic restoration) can be measured by the improvement of intelligibility of periodically interrupted speech once the silent intervals are filled with noise bursts. In this experiment, using such a paradigm, bimodal CI users will be tested with interrupted sentences (with or without a filler noise at 0 dB SNR), with three different duty cycles (proportion of on-and-off speech segments: 50 %, 62,5 %, 75 %) in two modalities (CI only and CI + HA). Speech intelligibility will be measured by the number of correctly repeated words in each sentence. We expect global intelligibility of interrupted speech to be better when both CI and HA are used compared to CI only. We further expect that phonemic restoration will improve as the duty cycle increases, providing more speech cues that can be used to better activate the top-down repair mechanisms. In general, we expect that access to pitch cues can help bimodal users to perform better for speech perception in adverse listening situations.

# P7: RM-ANOVA on RAUs vs mixed effects logistic regression: a ruling of the high court for statistics

Gaston Hilkhuysen

Laboratoire de Mécanique et d'Acustque, CNRS, Marseille, FR
Hilkhuysen@lma.cnrs-mrs.fr

Many speech-in-noise studies measure outcomes in percentage correct scores. Percentages are usually transformed into rationalized arc-sine units (RAU) and subsequently inspected with analyses of variance for repeated measurements (RM-ANOVAs). Such practice has various drawbacks. For example, each RM-ANOVA on RAUs violates the assumptions necessary for hypotheses testing from a theoretical perspective. Attempting to comply with these assumptions, most researchers reduce statistical power. And more, rejecting or accepting the hypotheses tested with a RM-ANOVA on RAUs frequently appears trivial. Additionally, ANOVAs are less suitable for analyzing unbalanced data. Mixed effects logistic regression (MELR) seems to overcome the drawbacks of RM-ANOVA on RAUs. Floor and ceiling effects are transformed into a lack of data problem, and the technique allows to simultaneously account for eventual differences between listeners. Including other random factors is easy, to inspect differences between utterances for example. MELR requires a more exploratory approach to data analysis than the one-shot RM-ANOVA and certain conditions need to be fulfilled before statistical inferences can be made. Both will be discussed, providing some basic guidance in the wonderful world of MELR.

# P8: Influence of hearing impairment on alpha power during retention of auditory stimuli

Eline Borch Petersen[1,2], Malte Wöstmann[3,4], Jonas Obleser[3], Stefan Stenfelt[2], and Thomas Lunner[1,5]

[1] Eriksholm Research Centre, Snekkersten, DK
[2] Technical Audiology, Department of Clinical and Experimental Medicine, Linköping University, Linköping, SE
[3] Max Planck Research Group "Auditory Cognition", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, DE
[4] International Max Planck Research School on Neuroscience of Communication, Leipzig, DE
[5] Linnaeus Centre HEAD, Swedish Institute for Disability Research, Linköping University, SE
ebp@eriksholm.com

Listening to sound in background noise is known to affect the working memory load in behavioral, as well as, electrophysiological experiments. The alpha activity (6-12 Hz) is a neural marker of working memory, known to increase with more difficult listening conditions. However, it has not previously been investigated how internal degradation of the auditory signal, in the form of hearing impairment, affect the neural mechanisms of working memory when audibility is ensured through hearing aid amplification. Applying an auditory version of the Sternberg task, participants were asked to remember 2, 4, or 6 spoken digits embedded in three different levels of background noise. After a stimulus-free delay period, the participants indicated whether a probe digit was among the preceding stream of digits. The participants were elderly (62-86 years) adults with hearing acuity ranging from normal to moderately impaired hearing. To ensure equal audibility, the background noise levels were individualized and all participants were wearing individually fitted hearing aids. Although the task performance showed no effect of hearing loss, analysis of the alpha power during the stimulus-free delay periods revealed a general increase in alpha activity with worse hearing. Interestingly, the alpha power was also dependent on the interaction between memory load, background noise level, and hearing loss: While alpha power increased with hearing loss during the low working memory conditions, it dropped for participants with the most severe hearing loss under the highest memory load and background noise level. These findings suggest that neural mechanisms for coping with adverse listening conditions break down for higher degrees of hearing loss, even when adequate hearing aid amplification is provided.

# P9: Good working memory capacity facilitates long-term memory encoding of speech in stationary noise


Mary Rudner[1], Sushmit Mishra[1], Stefan Stenfelt[1,2], Thomas Lunner[1,2,3], and Jerker Rönnberg[1]

[1] Linnaeus Centre HEAD, Swedish Institute for Disability Research, Department of Behavioural Sciences and Learning, Linköping University, SE

[2] Department of Clinical and Experimental Medicine, Linköping University, SE

[3] Eriksholm Research Centre, Oticon, Snekkersten, DK

mary.rudner@liu.se



Background noise makes listening more cognitively demanding, especially for persons with hearing impairment, and this seems to affect memory encoding. It is not clear whether this decrement can be restored by providing visual cues. In the present study, we investigated whether long term memory encoding of speech, in quiet and in background noise adjusted to retain intelligibility, improves when the talkers face is visible, and whether such an enhancement is associated with working memory capacity. Twenty adults with normal hearing in Experiment 1 and 24 adults with hearing loss in Experiment 2 listened to lists of 13 two-digit numbers, with or without seeing the talkers face, and then recalled as many numbers as possible in any order. The lists were presented in quiet as well in a steady-state speech-weighted noise and the International Speech Testing Signal at a signal-to-noise ratio individually adapted to give an intelligibility level of approximately 90%. Amplification compensated for loss of audibility. Working memory capacity was measured using the reading span test. Seeing the talkers face did enhance free recall performance. However, whereas the effect size for adults with normal hearing was large, for adults with hearing impairment it was small. Further, there was no evidence that visual cues specifically compensated for performance decrements due to noise or serial position and there was no evidence of an association between working memory capacity and performance with visual cues. However, good working memory capacity did improve performance for early list items, reflecting facilitation of long-term memory encoding, for both groups when stimuli were presented in steady-state noise. For participants with hearing impairment, good working memory capacity was associated with good performance on late list items in quiet, reflecting facilitation of working memory encoding. This pattern of results indicates that steady-state background noise reduces the cognitive capacity available for the long-term memory encoding of speech that is necessary for enduring retention of spoken information, irrespective of hearing status, but provides no evidence that this is specifically compensated for by visual cues. It also demonstrates that for individuals with hearing impairment, short term retention of speech heard even under the most favourable conditions is a function of individual working memory capacity. These findings support and extend the Ease of Language Understanding Model.

# P10: Measurement and prediction of speech intelligibility in noise and reverberation for different sentence materials

Anna Warzybok[1], Jan Rennies[2], Thomas Brand[1], and Birger Kollmeier[1]

[1] Cluster of Excellence Hearing4all Medical Physics, Department of Medical Physics and Acoustics, University Oldenburg, DE
[2] Project Group Hearing, Speech and Audio Technology, Fraunhofer Institute for Digital Media Technology IDMT, DE
a.warzybok@uni-oldenburg.de

The present study investigated the role of the speech material type for speech intelligibility in noise and reverberation. This work was motivated by the contradictory outcomes of previous studies examining the effect of reverberation and noise on speech intelligibility, which indicated that the detrimental effect of reverberation may depend on speech material (Rennies et al., 2014). In a current series of study, the effect of noise only, reverberation only and the combined effect of noise and reverberation were systematically investigated for two types of sentence tests. The hypothesis to be tested was that speech intelligibility is more affected by reverberation when using an open-set speech material consisting of everyday sentences than when using a closed-set test with syntactically fixed and semantically unpredictable sentences. Since the speech material of the open- and the closed-set sentence test was recorded with the same talker, it was possible to systematically examine if the different types of sentences lead to different results when measuring speech intelligibility in noise or reverberation. Furthermore, the reasons for the discrepancies between the tests were examined. The experimental conditions were established based on the predictions of the speech transmission index. The biggest differences between the speech tests employed were observed in the reverberant conditions. In the condition with a reverberation time of 4.1 s, intelligibility scores were 50 % higher for the closed-set sentence test than for the open-set sentence test. This difference increased to 75 % in the condition with reverberation time of 9.4 s. 20 % higher scores for the closed-set test than for the open-set test were found in the condition with noise only. This difference can be explained by the difference in the normative data, i.e., the closed-set speech test has lower speech reception threshold in noise than the open-set test. To examine the reasons for differences between the tests in reverberation, two lists of the closed-set sentence test were presented to naïve listeners. In contrast to the standard procedure, they were not trained prior the measurements and were not informed about the fixed syntax of the sentences and about the limited test vocabulary. Excluding this information, the differences between the closed- and open-set tests were not present indicating that the degree of familiarity with the speech material has a major impact on speech recognition in noise. This cannot be predicted by the speech transmission index.

# P11: Modeling speech intelligibility in hearing impaired listeners

Christoph Scheidiger and Torsten Dau

Center for Applied Hearing Research, Technical Univeristy of Denmark, Lyngby, DK
csche@elektro.dtu.dk

The work on speech intelligibility (SI) models started with the articulation index (AI). Models following the AI were extended to consider a larger variety of conditions. Recent studies predicted SI in normal-hearing (NH) listeners based on a signal-to-noise ratio measure in the envelope domain (SNRenv, Jørgensen and Dau, 2011 and Jørgensen et al, 2013). This framework showed good agreement with measured data under a broad range of conditions, including stationary and modulated interferers, reverberation, and spectral subtraction. Despite the advances in modeling SI in NH listeners, a broadly applicable model that can predict SI in hearing-impaired (HI) listeners is missing. This study investigates to what extent effects of hearing impairment on SI can be modeled in the sEPSM framework. Preliminary results indicate that, by only accounting for the sensitivity loss, the model seems to account for the reduced temporal masking release (MR) in HI listeners. The MR is defined as the speech reception threshold (SRT) benefit listeners obtain in fluctuating noise compared to stationary noise. However, the model does not correctly describe the poorer SRTs of HI listeners in stationary or modulated noise compared to NH listeners. The limitations of the model are analyzed and improvements for the band integration stage are proposed.

# P12: Room acoustic descriptors - room for more?

Mai-Britt Beldam

Saint-Gobain Ecophon, DK
mbe@ecophon.dk

Room acoustic descriptors  room for more? Reverberation (RT) time still remains the primary indicator of room acoustic response and RT is also known be the only demand/descriptor in building regulations for room acoustics in schools and day care centers in many countries. However, previous work and research since many years have shown that RT alone can be insufficient to describe the acoustic conditions in non-diffuse environments, especially in classrooms where the typical solution is a suspended absorbing ceiling; the majority of absorption is on one surface which normally leads to a non-diffuse sound decay. Both calculations on RT using the Sabine equation and measuring RT according to ISO 3382-1/2 will lead to results that dont exactly correspond to what takes place in a real room and what we actually hear. Alternative parameters have been presented to evaluate the acoustic response better of such rooms: Speech Clarity (C50)/ Deutlichkeit (D50) and Room gain. Despite the fact that previous work has shown the need for the above mentioned parameters we still know very little about how childrens performance practically is affected when C/D50 and sound strength (Room gain) are changed and RT stays the same. This poster will present a study on childrens performance in relation to both working memory and listening comprehension with/ without back ground noise in two rooms with same reverberation time but different values on C/D50 and sound strength. The research is made in situ in two rooms with different absorbing suspended ceilings. The subjects are 7 years old and attend public school. The research shows that there is a difference in performance in the two rooms despite the fact that they - according to the building regulation and standards  are exactly the same. The sound in the two rooms is very different and even small movements from the children increase background noise when fx Room gain is higher : We hear so much more than reverberation time and we need more acoustic parameters to evaluate rooms for teaching. Smaller children are highly affected by noise when it comes to listening and understanding speech and we should consider to reevaluate our standards and building regulations to secure optimum room acoustics in learning environments.

# P13: Speech intelligibility for target and masker with different spectrums

David Thery,[1] Thibaud Leclére[1], Mathieu Lavandier[1], and John F. Culling[2]

[1] Univ. de Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, France
[2] School of Psychology, Cardiff University, UK
mathieu.lavandier@entpe.fr

The speech intelligibility index (SII) calculation relies on the assumptions that intelligibility is not improved by increasing the signal-to-noise ratio (SNR) above 15 dB, nor impaired by decreasing the SNR below -15 dB. These assumptions were tested in four experiments. Speech reception thresholds (SRTs) were measured with a running speech target and a speech-spectrum noise masker, while attenuating the target or masker above or below 1400 Hz. Different levels of attenuation were tested, allowing the testing of different SNRs in the two bands. The results confirmed that, while plotting SRT as a function of attenuation in the filtered band, SRT varied linearly with attenuation at low-attenuation levels and an asymptote was reached for high-attenuation levels. However, the "knee-point" at which this asymptote was reached (and intelligibility was not anymore influenced by further attenuation) depended on the filtering condition: (1) While the -15-dB limit was confirmed for the high-pass filtered target, decreasing the SNR below -15 dB (down to about -30 dB SNR) significantly affected intelligibility for the low-pass filtered target. (2) The SRT obtained for the most-attenuated low-pass filtered target was about 10 dB higher than the one obtained for the most-attenuated high-pass filtered target (despite the cutoff frequency being chosen for each spectrum part to have approximately equal contribution to intelligibility regarding the SII frequency-weightings). (3) Increasing the SNR above 15 dB (up to about 40 dB SNR) significantly improved intelligibility for high- and low-pass filtered maskers. (4) Before reaching the knee-point/asymptote, a 10-dB increase of SNR obtained by filtering the masker had a larger impact on intelligibility than a corresponding 10-dB decrease of SNR obtained by filtering the target.

# P14: Speech intelligibility and speech detection in adverse monaural masking conditions: Comparison of empirical and model data

Wiebke Schubotz, Thomas Brand, and Stephan D. Ewert

Medizinische Physik and Cluster of Excellence Hearing4all, Universität Oldenburg, D-26111 Oldenburg, DE
wiebke.schubotz@uni-oldenburg.de

Speech is one of the most important factors for human communication. In everyday life, speech is often perceived in a background noise and speech intelligibility is hampered compared to situations without a noise. In such listening conditions several spectral and temporal properties of the noise masker, such as, frequency content, amplitude modulations, and duration of temporal gaps can affect speech intelligibility, but the amount of masking can hardly be attributed to one single parameter. In the literature, three "classical" types of masking are described, namely energetic (EM), amplitude modulation (AMM), and informational masking (IM). In arbitrary masker signals all three types of masking can occur to a different degree and it is difficult to disentangle their individual effects on speech intelligibility. In this study monaural speech intelligibility was assessed by measuring speech reception thresholds (SRT) and detection thresholds in masking conditions that ranged from stationary maskers to intact and noise-vocoded speech-like maskers. The aim was to gain further insight on the role of the masking types by systematically changing the spectro-temporal properties of the masker. Speech detection was measured in comparison to SRTs, since it is hypothesized that detection thresholds are described mostly by EM and AMM while higher level processes (as involved in speech perception) might play a smaller role. The target material were sentences from the Oldenburger Satztest, a matrix test providing non-sense sentences with a fixed sentence structure (subject, verb, numeral, adjective, object). The measurements were performed for three different combinations of male and female target and male and female maskers. The empirical data was compared to predictions of four different speech intelligibility models: speech intelligibility index (SII), extended speech intelligibility index (ESII), multi-resolution speech envelope power spectrum model (mr-sEPSM), and short term objective intelligibility measure (STOI). With this comparison further insight on the role of the three types of masking was gained, given that the models utilize different masker features. The results suggested that speech intelligibility is significantly improved by the presence of coherent and regular modulations across the masker spectrum (co-modulation). Intelligibility thresholds for different speech-like (vocoded and non-vocoded) maskers were similar, indicating no difference in informational masking. Comparison of detection and intelligibility data confirmed that for modulated maskers AMM, and for speech-like maskers IM was present in addition to EM. This was further supported by the predictions from the ESII.

# P15: Neurofeedback as a training tool for cochlear implant users

Annika Luckmann[1,3], Jacob Jolij[2,3], Etienne Gaudrain[1,3,4], and Deniz Başkent[1,3]

[1] University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology / Head and Neck Surgery, Groningen, NL
[2] University of Groningen, Experimental Psychology, Groningen, NL
[3] University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Groningen, NL
[4] Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics Team, CNRS UMR 5292, INSERM U1028, University Lyon 1, Lyon, FR
a.luckmann@umcg.nl

Cochlear implantees experience difficulties in voice discrimination, which can lead to problems during daily speech communication and pitch perception, especially in challenging noisy listening environments. Voice perception can be decomposed into two aspects: voice pitch (F0) and vocal tract length (VTL). Fuller et al (2014) have found that implantees rely almost entirely on F0 and do not use VTL when identifying a speaker's gender, and Gaudrain and Başkent (2014) have further confirmed that indeed cochlear implant (CI) users do not use VTL cues. The goal of our study is to investigate whether neurofeedback can improve perception of VTL in CI users. Neurofeedback is an online feedback method, in which brain wave patterns, measured with EEG, are used to give real-time feedback to individuals. Through the real-time feedback, patients can be trained to regulate brain activity to match specific cognitive states. In a pilot study with 28 first-year psychology students, we tested a pitch discrimination paradigm that later will be used in a neurofeedback setup with CI users. Two stimuli were presented, which differed in either musical pitch (pure tones), VTL, or F0, presented at levels just above the frequency difference limen The participants' had to rate how much the second stimulus differed from the first, whilst EEG was measured. We investigated both the acoustic change complex (ACC) and P300, as potential neural markers of near threshold pitch perception. The ACC is a cortical auditory evoked potential used to assess the neural detection of changes in auditory stimuli. The P300 is a positive posterior brain evoked potential around 250-500 ms after stimulus onset, known to be an electrophysiological correlate of decision making, evaluation and categorization. Using the EEG data as predictor in a regression analysis to predict perceived difference, and actual difference, we found that we could reliably and independently decode both actual and perceived difference from the ACC and P300, for musical pitch, as well as VTL and voice pitch stimuli. These results indicate that this paradigm is suited for a neurofeedback training paradigm. Testing the neurofeedback paradigm on CI users will give us insight into why they do not use VTL cues and answer the question whether this limitation is perception or cognitive related.

# P16: Musician advantage for speech-on-speech perception

Deniz Başkent,[1] Michael Chesnay[1], Nikki Tahapary[1], and Etienne Gaudrain[1,2]

[1] University Medical Center Groningen, Department of Otorhinolaryngology, Groningen, NL
[2] Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics, CNRS UMR 5292, Université Lyon 1, Lyon, FR
dbaskent@gmail.com

Musicians show advantages over non-musicians in a number of auditory perceptual tasks. Such advantages may be due to better processing of acoustic features, such as fundamental frequency (F0), which is also a primary dimension in music. These advantages may also be due to enhanced cognitive abilities, such as better attention abilities or extended working memory capacity, at least in the auditory modality. However evidence for a transfer of this advantage to better perception of speech in noise has been mixed. This could be due to the fact that previous studies have primarily used noise as background masker. Speech in noise is not a situation recognized as relying on fine F0 processing. Moreover, it is generally considered primarily driven by energetic masking, the release from which is thought to be less dependent on cognitive abilities than is informational masking. Speech on speech perception, on the other hand, has been shown to directly depend on F0 differences between the two competing voices and involves informational rather than energetic masking, which mobilizes more cognitive resources. Speech-on-speech could therefore be a more suitable test condition than speech-in-noise to investigate the potential musician advantage for speech perception. We measured intelligibility of concurrent sentences that were presented to musicians and non-musicians to test a potential musician advantage. To investigate what underlying factors may contribute to such advantage, if it exists, the competing voices differed in terms of vocal characteristics. F0, but also vocal tract length (VTL), were systematically varied to differentiate the target sentence from the masker sentence. The musicians showed overall better intelligibility performance than non-musicians, confirming a musician advantage. However, almost all the difference was attributable to the condition where the target and masker voices were identical. The two groups drew equivalent benefit from F0 differences between concurrent voices, while musicians seemed to benefit slightly less from VTL differences than non-musicians. Because most of the advantage was shown in a condition where there was no difference in average vocal characteristics of the talkers, the musician advantage is unlikely to be related to the processing of these vocal characteristics, and is perhaps more related to cognitive abilities. However, because the prosodic F0 contours of the concurrent voices differed, instantaneous F0 differences remained even when the average F0 were identical. The musician advantage could thus also be derived from an enhanced ability to process and disentangle fast changing F0 differences.

# P17: Improving blind reverberation time estimation on a two-microphone portable device by using speech source distance information

João F. Santos[1], Anderson R. Avila[1], Rachel E. Bou Serhal[2], and Tiago H. Falk[1]

[1] INRS-EMT, University of Quebec, CA
[2] École de Technologie Superieure, University of Quebec, CA
jfsantos@emt.inrs.ca

Room reverberation has been shown to significantly degrade speech intelligibility and automatic speech recognition (ASR) performance. For that reason, there is a great research effort in designing methods to enhance reverberant speech and ASR systems that are robust to the reverberation effects. Having access to an accurate reverberation time estimate allows these systems to adapt themselves to changes in the acoustic environment by, e.g., using matched acoustic models in the case of ASR. However, in most practical situations this estimate can only be obtained in a blind way, using only a single- or multi-channel reverberant speech signal as the input. These blind estimates suffer from variability issues due to the lack of a reference clean signal and lack of knowledge about other environmental conditions, such as the speaker localization. In this work, we evaluate the performance of a reverberation time estimation method based on a modulation spectral representation. We used a dual-channel portable voice-activated device (a prototype version of the Ubi) to record speech signals under a controlled reverberant environment and two different speaker distances (1m and 2m). We show that while reverberation time estimates are accurate for a single distance (Pearson correlations above 0.9), correlations decrease significantly if we consider both distances. We then propose a simple method of estimating the speaker distance using a magnitude-squared coherence (MSC) based approach, namely the Euclidean distance between the MSC of the two channels and a perfectly coherent pair of signals. We show that this measure can be used to improve the reverberation time estimates. Future work aims to incorporate the MSC distance in the reverberation time estimation model to improve estimations for any possible speaker distance.

# P18: Noise-induced neuro-degeneration - an Invisible noise-induced hearing loss

Pernille Holtegaard

Steen Østergaard Olsen (Rigshospitalet), DK
pernille_holtegaard@hotmail.com

Background: Studies on noise-exposure to mice and guinea pigs have shown evidence of primary permanent neural and synaptic damage to the afferent type I spiral ganglion neurons (SGNs). Furthermore this new research indicated that the degeneration is selective of the fibers with low spontaneous firing rate. These fibers are essential to our perception of sound/speech stimuli at supra-threshold level. Selective damage of these fibers can potentially affect speech recognition despite presence normal pure-tone sensitivity. Objective: The goal of my study (MSc thesis) was to investigate if evidence or signs of primary neuro-degeneration from noise-exposure could also be found in noise-exposed human individuals. Supra-threshold processing was studied to investigate: (1) If noise-exposed individuals, with thresholds within normal limits ( $\leq$ 20 dB HL), perform poorer in regards to speech recognition threshold in noise (SRTN) compared to non-exposed individuals; (2) If ABR reveal lower amplitudes of wave I in noise-exposed individuals compared to non-exposed individuals, when stimulating with supra-threshold stimuli; (3) discuss the clinical implications of the findings and whether these applied measurements should be included in the test battery, when screening for a noise-induced hearing loss (NIHL). Method: Two groups were recruited and tested: One test group of noise/music-exposed individuals, and one control group, considered a non-exposed group. All subjects were between 18 and 32 years of age and had pure-tone thresholds $\leq$ 20 dB HL from 125-8000 Hz. 19 subjects were recruited in the test group (15 men and 4 women) and 16 in the control group (12 men and 4 women). For both groups a basic hearing evaluation was completed (basic audiometry procedure). ABR using high-level click stimuli was tested, with a focus on wave I, in order to get insight into the functioning of the peripheral afferent auditory nerve fibers. Speech recognition threshold in noise (SRTN) was tested using DANTALE II. Results: The noise-exposed test subjects required a significantly better dB SNR to obtain SRTN, despite normal threshold sensitivity, compared to the control group. Signs of reduced wave I amplitudes, to supra-threshold stimuli, were also confirmed in the test group, when testing the ears that the test subjects experienced as most exposed. Conclusion: Overall, this study on supra-threshold processing ability of noise-exposed human subjects suggests that the results of noise-induced primary neural degeneration, found in experiments on noise-exposed mice and guinea pigs, also apply to humans. These results provide evidence that repeated excessive noise exposure to human individuals cause reduced speech recognition threshold in noise despite normal pure-tone thresholds. Such neuronal damage from noise exposure is invisible to the test of pure-tone thresholds, but it can still have significant consequences for the individual in their every-day communication tasks. It is therefore recommended to include the SRTN measurements as a part of the test battery when screening for a noise-induced hearing loss or noise-induced hearing impairment.

# P19: Preference judgments in the field and in the laboratory

Martin Dahlquist, Josefina Larsson, Sofia Hertzman, Florian Wolters, and Karolina Smeds

Widex A/S, ORCA Europe, SE
karolina.smeds@orca-eu.info

Currently, there is a certain focus in the hearing-device research community on how more realistic laboratory tests should be constructed. Another question is how we should collect reliable data from the field. In this study, two hearing-aid gain settings were compared in two hearing-aid programs. Twenty participants with impaired hearing evaluated the settings with regard to speech intelligibility and preference, both during a two-week field trial period and in a subsequent laboratory test. The study was designed as a double-blind trial, where neither participants nor test leaders knew which setting was used in each program. In the field, the following outcome measures were used: • A diary where the participants logged preference for hearing-aid setting 1 or 2 in various listening environments. • A structured interview, where overall preference was investigated. • A questionnaire, answered by the participants together with the test leader, focusing on preference in a number of pre-defined and self-selected listening situations • A hearing-aid log, where usage time and volume control settings could be studied for the two programs in a number of listening situations. In the laboratory, the following preference outcome measure was used: • A paired-comparison test with ratings of the difference between the two hearing-aid programs. Preference, speech clarity, and comfort were compared for the two programs in a number of listening situations. • Speech intelligibility in noise was tested using HINT sentences and a nonsense word test. In the presentation, the various types of preference data will be compared with each other and the speech test results. It was difficult to find one measure that could explain the "main preference" in the field, investigated in the structured interview.

# P20: Perception of vocal characteristics in cochlear implants

Etienne Gaudrain[1,2] and Deniz Başkent[1]

[1] University Medical Center Groningen, Department of Otorhinolaryngology, Groningen, NL
[2] Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics, CNRS UMR 5292, Université Lyon 1, Lyon, FR
etienne.gaudrain@cnrs.fr

Listeners use vocal characteristics of speakers to identify and discriminate voices in cocktail-party situations. Fundamental frequency (F0) and spectral shape parameters – and in particular vocal-tract length (VTL) – have been identified as the two main perceptual dimensions normal-hearing (NH) listeners rely on for this task. Previous studies on voice gender categorization have shown that cochlear implant (CI) users have more difficulties discriminating male and female voices than NH listeners, and do not benefit from speaker gender differences in concurrent speech situations. More recent data showed that while CI users are able to use gender-related F0 differences, unlike NH individuals, they are unable to exploit VTL differences for this discrimination. While F0 discrimination is widely studied with CI users, VTL discrimination is largely unknown. As a result, the reasons why CI users have difficulties perceiving VTL remain unclear. Do CI listeners not use VTL cues because they cannot detect VTL differences at all? Or is it because they cannot interpret the difference they hear as a change in VTL? The objective of the present study is to answer this question by characterizing VTL discrimination abilities in CI users, as compared to F0 discrimination abilities. VTL and F0 of recorded syllable sequences were manipulated using STRAIGHT, in order to measure just noticeable differences (JND) along these dimensions using an adaptive 3AFC method in postlingually deafened adult CI listeners. Although worse than in NH listeners, the F0 JNDs observed in CI users were consistent with the results of the gender categorization study. Namely, these were found to be sufficient to support discrimination of male F0 from female F0. However, CI listeners did not perceive VTL differences typically found between male and female speakers. Our results show that CI users cannot detect acoustic changes induced by gender-related VTL differences. More generally, these results indicate that present stimulation strategies do not allow CI users to benefit from VTL differences between male and female speakers. This may explain some of the difficulties CI listeners encounter in understanding speech in crowded environments. Furthermore, these underline that F0 is not the only problematic voice cue in CIs, and that VTL may be even more negatively affected by the poor spectral resolution of the device. New stimulation methods must therefore be developed to improve the representation of not only F0, but also VTL, in the implant.

# P21: The Perceptual Discrimination of Reduced and Clear Speech in Adverse Conditions

Terrin N. Tamati and Deniz Başkent

University Medical Center Groningen, NL
terrintamati@gmail.com

Real-life speech communication is complicated not only by background noise and competition from other talkers, but also natural variability encoded in the speech signal. Human listeners must adapt to multiple pronunciations for a single word in order to successfully understand the utterance, while also extracting information about the environment, context, and talker. This task may be extra difficult for hearing-impaired users of cochlear implants (CIs), since they also have to deal with degradations of the speech signal transmitted by a CI. While robust speech perception in CI users is mostly achieved for ideal speech, i.e., carefully controlled speech with clear pronunciations, our knowledge of CI perception of more realistic speech is still limited. We hypothesize that CI users perception of speech produced in well-controlled laboratory speech conditions may not reflect their actual real-life performance. To show this discrepancy, and to provide guidelines for best clinical practice, we seek to investigate the perception of reduced speech, a common form of real-life speech, and clearly articulated speech, characteristic of speech produced in formal situations or in the lab, in different adverse conditions. In order to begin to characterize CI perception of these speech forms, the current study examines NH listeners perception of reduced and clear speech in quiet conditions and under acoustic simulations of CIs. A series of discrimination tasks are used to assess the listeners ability to detect, discriminate, and classify reduced and clear speech in the different conditions. In these tasks, listeners are presented with unprocessed and CI simulated Dutch speech and are asked to make rapid judgements about the speaking style. We expect that the NH listeners will be able to perceive differences in the reduced and clear speech forms in the unprocessed condition to make accurate detection, discrimination, and classification judgements. Under CI simulation, NH listeners are expected to have more difficulty with the real-life speech variability since the simulation provides an approximation of CI hearing, with spectrotemporally degraded speech. However, performance should improve when more information is available to them with improvements in simulated CI device. Results will shed light on the perception, encoding, and representation of reduced speech in adverse conditions. Beyond the theoretical motivations, the findings will contribute to the development of clinical applications for the assessment and training of real-life speech perception performance, particularly for CI listeners.

# P22: Measuring the objective and subjective limens for speech intelligibility benefits

William M Whitmer, David McShefferty, and Michael A. Akeroyd

MRC/CSO Institute of Hearing Research - Scottish Section, UK
bill@ihr.gla.ac.uk

Numerous studies have shown statistically significant speech intelligibility benefits for speech enhancement strategies in hearing aids. But are these benefits significant outwith statistics? Here we attempt to answer this question using objective and subjective methods to see what change in speech-to-noise ratio (SNR) is necessary to be (1) detectable and (2) meaningful. First, we measured the smallest detectable change – the just-noticeable difference (JND) – in SNR. To measure an SNR JND, we modified the classic level discrimination paradigm using equalised sentences in same-spectrum noise with various controls to help focus listening to SNR, not level per se. Averaged across participants, and corroborated across different tasks, the SNR JND was 3 dB. JNDs were not correlated with hearing ability. Preliminary data comparing SNR JNDs to speech-recognition JNDs estimated from psychometric functions will also be discussed. Second, we measured the subjective import – the meaningfulness – of an increase in SNR by presenting paired examples of speech and noise: one at a reference SNR and the other at a variably higher SNR. In different experiments, different hearing-impaired adults were asked (a) to rate how much better or worse the change in SNR was in each paired example, (b) if they would swap the reference SNR for the better SNR example (e.g., their current device for another), or (c) if they would be willing to go to the clinic for the given increase in SNR. The results across these tasks showed that the mean ratings increased linearly with a change in SNR, but an SNR change of at least 6 dB was necessary to reliably motivate participants to seek intervention. Overall, the results indicate not only the difference limen for SNR, but also how large a change in SNR is needed for it to be meaningful to someone. While an SNR increase less than 3 dB may have relevance to speech-recognition performance, it may not be enough of an SNR improvement to be reliably recognized and, furthermore, may be too little benefit to motivate potential users. [Supported by intramural funding from the Medical Research Council (grant number U135097131) and the Chief Scientist Office of the Scottish Government]

# P23: The influence of talker- and language-specific aspects on speech intelligibility of bilingual talkers

Sabine Hochmuth and Thomas Brand

Cluster of excellence Hearing4all and Medical Physics, Carl von Ossietzky Universität Oldenburg, DE
sabine.hochmuth@uni-oldenburg.de

The purpose of this study was to investigate speaker- and language-specific aspects of speech intelligibility in noise and reverberation. The speech material of the German, Russian and Spanish matrix sentence tests was recorded with bilingual talkers of German and Russian, and German and Spanish. Talkers were selected to be highly proficient in both languages. The recorded speech material was used to measure speech reception thresholds of 50% intelligibility (SRT) with monolingual native listeners in the respective languages under different noise conditions (stationary and fluctuating noise, and speech in stationary noise with additional reverberation). Different acoustic-phonetic properties of the speech signal like aspects of fundamental frequency, long-term spectrum and vowel space size, found in the literature being related to speech intelligibility, were analyzed. Due to the test design inter-individual as well as intra-individual variation could be compared within and across languages. Differences in intelligibility between talkers of up to 6 dB were found. For the German/Russian bilinguals hardly any difference of one talker in the two languages was found. Considerably SRTs were measured for the German/Spanish bilinguals when speaking Spanish than when speaking German in each listening conditions. In the fluctuating noise condition the benefit from temporal noise gaps was similar across all languages. Spanish was significantly more disturbed due to reverberation than the other two languages, which showed similar a detrimental effect. The usage of a larger vowel space area and higher energy in the mid frequencies could be associated with lower SRTs for each of the employed languages. In general, the inter-individual variation of acoustic-phonetic properties was larger within languages than intra-individually across languages.

# P24: Effects of syntactic complexity on word recognition in noise

Line Burholt Kristensen[1] and Dorothea Wendt[2]

[1] Prominence and Grammar in Mind and Brain (ProGram), Department of Scandinavian Studies and Linguistics, University of Copenhagen, DK
[2] Hearing systems, Department of Electrical Engineering, Technical University of Denmark, DK
burholt@hum.ku.dk

Most Danish speech intelligibility tests use recordings of canonical subject-first word order and do not include object-initial sentences, thereby exposing the listeners to limited syntactic variability and complexity. However, a recent small-scale corpus study of conversational speech showed object-initial sentences are not infrequent: 1 out of 3 active declarative transitive clauses were object-initial. Out of noise, Danish object-initial sentences are known to show processing difficulties, but it has not yet been investigated whether this effect of word order complexity is strengthened in background noise and what parts of the sentence it affects. It may be hypothesized that the effect of noise is greater for object-initial clauses than for subject-initial ones e.g. due to an increased cognitive load. An investigation of the intelligibility and comprehension of different kinds of word order may contribute to increase the validity of speech in noise tests for Danish as well as for other languages. We here present a material based on a Danish translation of part of the German OLACS material. Two different sentence structures, a subject-verb-object (SVO) structure and an object-verb-subject structure (OVS) were created. Word order (the position of the full verb, e.g. ae = 'stroke') is the only cue to understanding who did what to whom. The Danish sentence corpus consists of 39 items recorded in 4 versions, exemplified by the item below:

| | |
|---|---|
| SVO1: | Det kloge pindsvin vil ae den søde hare ≈ 'The smart hedgehog will stroke the sweet hare' |
| OVS1: | Det kloge pindsvin vil den søde hare ae ≈ 'The smart hedgehog, the sweet hare will stroke' |
| SVO2: | Den søde hare vil ae det kloge pindsvin ≈ 'The sweet hare will stroke the smart hedgehog' |
| OVS2: | Den søde hare vil det kloge pindsvin ae ≈ 'The sweet hare, the smart hedgehog will stroke' |

Results of a preliminary test of word recognition in stationary noise are presented using this material. Twelve native speakers of Danish have been tested - each listened to full sentences from the material and to fragments of sentences. Significant differences were found between the number of correctly repeated words per sentence for OVS (5.0 correct words) vs. SVO sentences (5.3 correct words). In particular, OVS sentences had fewer correctly repeated main verbs (20 % correct repeats) than SVO sentences (70 % correct repeats).

# P25: The impact of open learning space in primary school classrooms on the cognitive load of children in Australia

Jeremy Marozeau[1], Shirley Lam[2], Scott Moore[2], and Michelle Moran[2]

[1] Hearing Systems Group, DTU Elektro, Lyngby, DK
[2] University of Melbourne, Melbourne, AU
jemaroz@elektro.dtu.dk

The primary school environment is of paramount importance in the development of a child. Teachers, parents and scientists are constantly trying to adapt it for the better good of the children. In Australia, one of the recent trends is to go away from the traditional "Chalk and Talk" classroom configuration towards "Agile Learning Spaces". Those large learning spaces are created so that teachers and students can collaborate, learn together and work together, and to provide better opportunities to address every childs special needs. However, it is important to consider the background noise that can be created by this large space. Children, seated in small groups, have to constantly separate the speech signals from their peers and teachers in the same group from the competing background noise generated by other groups. The aim of this project was to assess the effect of the sound environment of these new learning spaces on the students ability to understand speech in noise and to their learning abilities. First, the sound level was measured in two different primary schools around Melbourne during regular school days. Then this sound environment was reproduced in a sound proof booth with a circle of 12 loudspeakers and behavioral data were collected using a dual task paradigm. Seventeen normally hearing children aged 7-12 years participated in the study. Participants repeated AB words presented in babble noise (primary task) at different classroom signal to noise ratios (SNR: quiet, +4dB, 0dB, -4dB), while simultaneously memorizing a set of five digits to be recalled later (secondary task). Sound level measurements showed that children were in environments exceeding the recommended 60 dB (A) threshold for health. An average of 66 dB LAeq and peak sound levels of up to 124.9 dB (C) were recorded. Results of the single task conditions showed that, at 4 dB SNR, children can perceive on average 81% of the words accurately and can remember 85% of the digits. However, in the dual task condition the digital recall score drops to 67% for the same level. This study showed that despite that the positive aspects of the new learning spaces, it is important to consider the effect of noise on the children ability to learn and to understand the teacher.

# P26: Statistical uncertainty in speech intelligibility testing

Ellen Raben Pedersen and Peter Møller Juhl

University of Southern Denmark, DK
erpe@mmmi.sdu.dk

Objective: A person's ability to understand speech can be measured as the speech reception threshold (SRT). The SRT is the presentation level corresponding to a fixed percentage (often 50 %) of correctly answered words. It can be determined by presenting sentences at different presentation levels using an adaptive step procedure. Another measure of a person's speech intelligibility is the discrimination score (DS), which is the percentage of correctly answered words played at a fixed presentation level. In the scientific literature critical differences for the DS can be found. The critical differences are used for comparing two DSs, e.g. obtained with and without hearing aids for the same person. They give the intervals over which two scores are likely to range by chance alone and thereby how much the two scores have to differ in order to be statistically significantly different. Critical differences for SRTs are not reported in the scientific literature even though they would be a useful tool for comparing two SRTs. This study determines critical differences for the SRT. Methods: Critical differences for SRTs are determined by Monte Carlo simulations at a 5% significant level. The simulations were performed for five-word sentences such as the Dantale II sentences and two different adaptive step procedures, which in this study are denoted A and B. The two procedures differentiate by their step sizes and their determination of the SRT. For each test procedure 16 sets of simulations, each containing 50,000 runs, were performed for different combinations of the number of test sentences (10, 20, 30, and 40 sentences), and the slope of the discrimination function (0.05, 0.10, 0.15, and 0.20 1/dB). Results and conclusion: For both procedures it is found, that the more sentences, the narrower the critical difference, and the lower the slope, the wider the critical difference. For a low number of sentences (10 sentences) procedure B entails the narrowest critical differences, whereas for a low slope value (0.05 1/dB) procedure A entails the narrowest critical differences. The two procedures entail similar critical differences, when a test contains 20 sentences or more and the slope of the discrimination function is 0.10 1/dB or higher. For a test of 20 sentences and a slope of 0.10 1/dB the critical difference is 2.1 dB for both procedures, i.e. under these test conditions two SRTs have to differ by at least 2.1 dB in order to be statistically significantly different.

# P27: Hearing impaired speakers of tonal languages may be more affected by noise than speakers of non-tonal languages

Marianna Vatti, Lars Bramsløw, and Renskje Hietkamp

Eriksholm Research Centre, DK
mav@eriksholm.com

In tonal languages (TL), each monosyllable is pronounced with a distinctive tone which denotes a specific lexical meaning. Tones correspond to variations in pitch. Hearing Impaired (HI) listeners are often reported to be less sensitive in perceiving changes in pitch. In non-tonal languages (NTL), pitch is used for prosodic reasons and therefore speech recognition is not affected by degraded pitch perception caused by sensorineural hearing loss (HL). However, in TL difficulties in tone perception are likely to have more serious effects on communication as they can directly affect listeners speech understanding. Results from Mandarin Chinese, Cantonese Chinese and Thai speakers show that tone identification in HI listeners of moderate to severe HL is weakly impaired with respect to Normal Hearing (NH). However, performance in noise has not been explicitly examined across HL. Studies on TL NH speakers suggest that noise can greatly degrade sentence recognition when lexical tone perception is weak. Therefore, it is of interest to investigate how HL affects speech perception in noise as well as how its effect differs between TL and NTL HI speakers. Standardized speech intelligibility tests in TL, namely Mandarin, Cantonese and Taiwanese Mandarin Hearing In Noise Test (HINT), have been developed and allow norm-referenced results to be compared directly with results in other languages. The present work aims in highlighting the need for comparative studies in order to quantify the differences in noise performance between TL and NTL HI speakers and to reveal whether audibility compensation allows hearing in noise to an equal extent. The findings of such studies would strengthen or weaken the assertion that TL speakers may benefit from a specially-designed fitting rationale.

# P28: Single-trial EEG measures of attention to speech in a multi-talker scenario

Jens Hjortkjær[1] and Torsten Dau[2]

[1] Oticon Centre of Excellence for Hearing and Speech Sciences, DTU, DK
[2] Hearing Systems, DTU, DK
jhjort@dtu.elektro.dk

Normal hearing listeners can use selective attention to separate a single talker in a multi-talker context. Here, we used EEG to determine attentional selection during natural audio-visual speech perception. Normal hearing listeners were presented with two audio-visual talkers and were asked to shift attention between the talkers on subsequent trials. To identify the target talker from the EEG response, we used a spread spectrum technique that estimates neural response functions to speech envelope fluctuations via regularized least-squares regression. Response functions were estimated using a multiband representation of envelope processing in the auditory periphery. We assessed differences in the ability of these response functions to reconstruct the either the attended or the unattended speech signal based on single-trial EEG. We found a stronger correlation for the attended signal compared to the unattended one. In addition to previous results, we show that attentional selection can be estimated from EEG recordings below 30 sec in duration even when listeners switch attention between talkers on different trials. Our results also suggest that the classification may be aided by employing a realistic auditory model.

# P29: PAMBOX: A Python auditory modeling toolbox

Alexandre Chabot-Leclerc

Hearing Systems group, Department of Electrical Engineering, Technical University of Denmark, DK
alech@elektro.dtu.dk

Toolboxes for modeling auditory perception have a surprisingly long history, starting with the Auditory Toolbox, first written by Malcom Slaney for Mathematica, in 1993, and then ported to Matlab in 1998. Here we present the Python Auditory Modeling Toolbox (PAMBOX), an open-source Python package for auditory modeling. The goal of the toolbox is to provide a collection of components that can be easily combined and extended to solve auditory modeling problems. PAMBOX contains code for modeling cochlear filtering, envelope extraction, as well as modulation processing. The toolbox also includes speech intelligibility models. These models are commonly used to predict how well speech is understood in a given situation, such as in the presence of noise or reverberation. The intelligibility models use a simple and consistent "predict" API, inspired by the scikit-learn's "fit and predict" API (scikit-learn is a widely used machine learning toolbox). This simplifies comparisons across models. PAMBOX also includes a framework for performing intelligibility experiments compatible with IPython.parallel. Models that are not original to PAMBOX are validated against their original implementations, where available. PAMBOX is based on NumPy, SciPy, and Pandas. It is distributed under the Modified BSD License.

# P30: Lateralized speech perception, temporal processing and cognitive function in normal hearing and hearing impaired listeners

Gusztáv Lőcsei, Sébastien Santurette, Torsten Dau, and Ewen N. MacDonald

Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark
guloc@elektro.dtu.dk

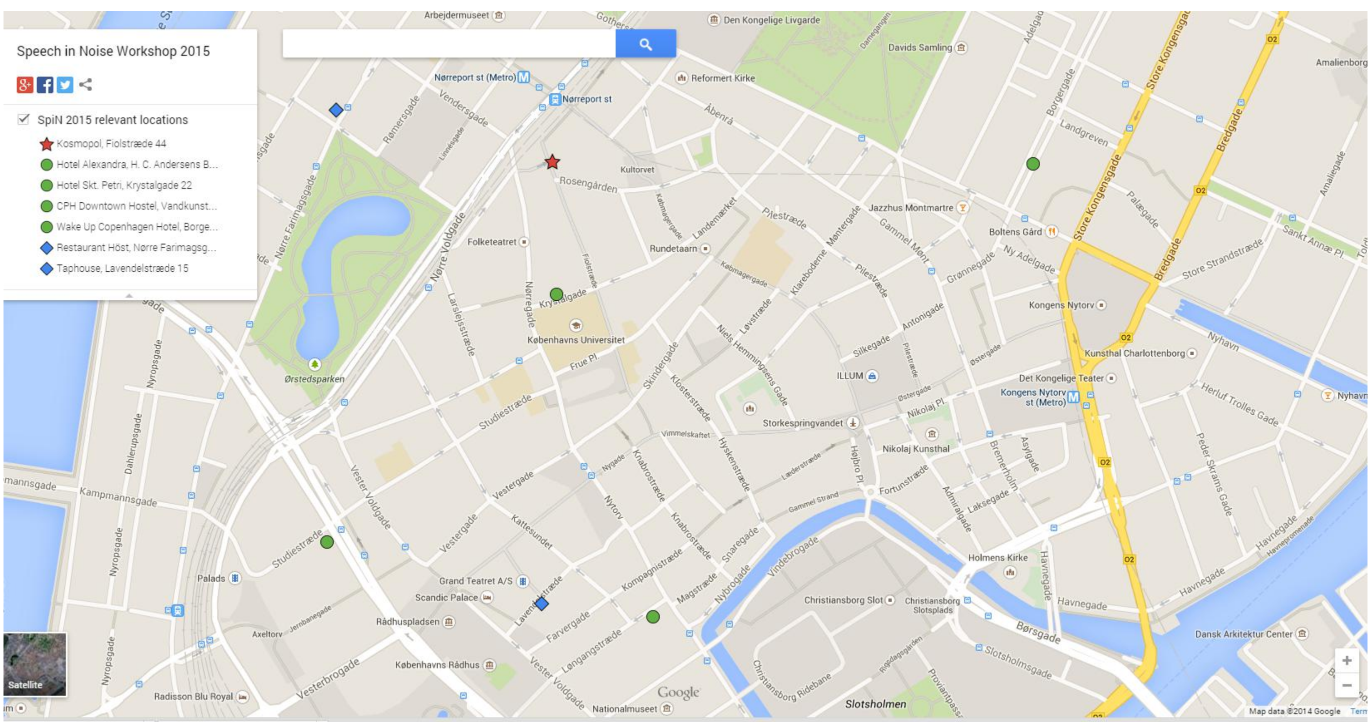
Solving the cocktail party scenario poses a serious problem for hearing impaired (HI) listeners, who often exhibit large inter-individual variance in everyday-life speech recognition tasks, even if they are fitted with hearing aids and share similar audiograms. This variability typically cannot be fully explained by the differences in audiograms, as these are poor descriptors of deficits that manifest themselves at above-threshold levels. The aim of the present study was to gain a better understanding of the relationship between speech intelligibility (SI) in lateralized stimuli, temporal coding abilities at low frequencies, and working memory capacity in HI subjects, once reduced audibility due to hearing loss has been compensated for. Speech reception thresholds (SRTs) were measured for normal hearing (NH) and HI subjects suffering from high-frequency hearing loss using five different types of interferers: speech shaped noise, reversed babble with two, four or eight talkers and normal babble with two talkers. The target and the interferers were presented as coming from the left or right side of the head by introducing interaural time delays between the ears for each stream. For a certain interferer type the number of maskers lateralized towards the side of the target was also varied (as all, some or none coming from the target side). Threshold estimations with the same type of interferers but with different lateralized distributions were run in an alternating manner, thus the listeners had to actively "tune in" into the noisy background in each trial in order to be able to follow the target talker. The speech stimuli were amplified to ensure partial audibility for the HI listeners up to 8 kHz. Frequency discrimination thresholds (FDT) and interaural phase difference (IPD) detection thresholds were measured at 250 Hz, and working memory capacity was assessed with a reading span test. The SI tests showed remarkable inter-individual differences in both NH and HI groups, presumably due to the high attentional complexity of the tasks. SRTs in the steady-state noise conditions were similar between subject groups. Contrary to NH listeners, however, HI listeners did not benefit from dip listening and showed only a slight SRT difference between the time reversed and normal two-talker babble. Although the HI listeners SI tests did not correlate significantly with the FDT, IPD, or reading span tests, but with high-frequency hearing loss (which had been compensated for), these results support the idea that SI performance was only partly limited by audibility.

# Author Index

# Dinner

Dinner on Thursday evening takes place at restaurant Höst, Nørre Farimagsgade 41, Copenhagen.

Cofoco, together with the design company Menu and Norm Architects have created a traditional New Nordic restaurant anno 2012, split over two floors there is space for up to 100 diners. Theres no interior decoration as such. It is the raw wooden furniture, the walls finished in a myriad of calming greys and the specially commissioned tableware that play the leading role. A role that the food, selected according to the produce available during the different Nordic seasons, of course shares and shines in.



# Time table $7^{th}$ SpiN Workshop, January 2015

| | Thursday $9^{th}$ | Friday $10^{th}$ |
|---|---|---|
| 08:00 - 09:00 | Welcome Coffee | Poster Session + Coffee |
| 09:00 - 10:00 | Introduction<br>Tobias May | |
| 10:00 - 11:00 | Bernd Meyer<br>Coffee Break | |
| 11:00 - 12:00 | Emma Jokinen<br>Michael Jeffet | Adriana Zekveld<br>Elaine Ng |
| 12:00 - 13:00 | Lunch (Kosmopol) | Chris James<br>Lunch (Kosmopol) |
| 13:00 - 14:00 | Keynote Lecture<br>Sarah Hawkins | Alexis Hervais-Adelman |
| 14:00 - 15:00 | Anita Wagner<br>Poster Session + Social | Johannes Zaar<br>Coffee Break |
| 15:00 - 16:00 | | Kurt Steinmetzger<br>Malte Wöstmann |
| 16:00 - 17:00 | | Discussion and feedback |
| 17:00 - 18:00 | | |
| 18:00 - 19:00 | Dinner (Restaurant Höst) | |