SPEECH IN NOISE WORKSHOP

7-8 January 2016

Groningen, NL

Programme

The 8th Speech in Noise Workshop is organised by selfless volunteers from the University Medical Center Groningen and University of Groningen, NL:

- *Annika Luckmann*
- *Enja Jung*
- *Nawal El Boghdady*
- *Elouise Koops*
- *Jefta Saija*
- *Jeanne Clarke*
- *Mirjan van Dijk*
- *Ria Woldhuis*

Coordination: Etienne Gaudrain, Lyon Neuroscience Research Center, CNRS — Université Lyon 1, FR.

Contact information: info@spin2016.nl

The Speech in Noise Workshop could not happen without the generous support of the following sponsors

Programme

Thursday, January 7

Friday, January 8

Talks

Thursday 7 January, 09:30—09:55

Meaningful improvements in speech intelligibility

W. Whitmer, D. McShefferty
MRC/CSO Institute of Hearing Research - Scottish Section, Glasgow, GB

M. Akeroyd
MRC Institute of Hearing Research, Nottingham, GB

Increases in speech intelligibility are conventionally reported as either (a) the change in relative levels of the target speech and noise(s) – the signal-to-noise ratio (SNR) – for a given percentage of utterances (e.g., 79%) or (b) the change in the percentage of utterances heard for a given SNR. What is considered an important change has been considered solely on account of the statistical distribution of scores, not what is a noticeable or convincing improvement in (a) SNR or (b) intelligibility. Through a series of experiments, we have determined what is the just noticeable difference (JND) in SNR and intelligibility, as well as inferred for both measures what is the just meaningful difference (JMD): the scale of improvement necessary to prompt an individual to seek intervention. For JNDs, participants of varying hearing ability and age listened to paired examples of sentences, words or digit triplets presented in various noises, and judged which of the two examples were clearer. The SNR JND ranged from 2.4 dB for digit triplets in noise to 4.4 dB for sentences in two-talker babble. The corresponding intelligibility JNDs, estimated from psychometric functions, ranged from 14% for words in same-spectrum noise to 33% for sentences in babble. JNDs were not correlated with hearing ability. For JMDs, participants also listened to paired examples of speech (sentences only) in same-spectrum noise: one at a reference SNR and the other at a variably higher SNR. In different experiments, different hearing-impaired adults performed various tasks: (a) better/worse rating, (b) better/worse rating of a corresponding level change, (c) ease-of-listening rating, (d) conversation tolerance, (e) device-swap, or (f) clinical importance. The SNR JMD was determined to be approximately 6 dB to reliably motivate participants to seek intervention. An intelligibility JMD was also estimated, albeit from mean psychometric functions. The SNR JMD did not correlate with hearing loss, age nor SNR JND, and only correlated weakly with hearing-aid usage and general quality-of-life. The SNR and Intelligibility JND provide perceptual benchmarks for performance beyond statistical relevance; the SNR JMD further adds clinical relevance to speech-intelligibility improvements.

Supported by intramural funding from the Medical Research Council (grant number U135097131) and the Chief Scientist Office of the Scottish Government

Clear speech strategies in adverse listening conditions

V. Hazan, O. Tuomainen
Speech, Hearing and Phonetic Sciences, UCL, London, UK

In order to communicate effectively in adverse listening conditions, talkers typically make acoustic-phonetic and linguistic adaptations to their speech and produce clear speaking styles. In a series of linked studies, we have analysed these adaptations in adolescents, young and older adults, in order to investigate how the ability to maintain effective communication in such conditions varies across the lifespan. In our studies, we naturally elicit clear speaking styles by recording pairs of speakers while they complete a 'spot the difference' picture task together (diapix task) either in good listening conditions or in a range of adverse listening conditions affecting one or both of the speakers. These adverse conditions include babble noise, hearing loss simulations affecting one of the speakers, or passing the speech of one of the speakers through a noise-excited vocoder. In this talk, I will review our findings on developmental changes across the lifespan. I will also discuss the degree to which speakers tailor their clear speech strategies to specific adverse conditions and individual differences in clear speech strategies.

Enhancement of ambient speech for robot audition – Noise reduction and dereverberation

P. A. Naylor
Department of Electrical and Electronic Engineering, Imperial College London, UK.

Whereas the impact of noise on speech quality and intelligibility has been studied and modeled extensively, only in the last 10 years or so have the effects of reverberation been taken into account with anything even approaching the same level of effort. However, in many real-world speech processing applications, noisy speech is only observed after being also subjected to significant reverberation. Commonly quoted examples of such applications include robot audition, hearing aids and hands-free terminals.

This talk will consider the effects due to combinations of noise with reverberation on speech quality, speech intelligibility and speech recognition performance. Using robot audition as an example, the problems addressed will first be outlined. Then, recent performance results will be given showing the state-of-the-art performance of various combinations of noise reduction and dereverberation processing. Conclusions will be drawn so as to identify research threads with strong potential with the aim of understanding what expectations are realistic of signal processing approaches to the enhancement of ambient speech.

Talker/listener interactions in speech on speech masking: Effects of age and sex

S. Rosen, L. Crook, L. Ball
UCL SHaPS, London, GB

Many factors determine performance when perceiving speech in the background of other talkers. We have recently found two examples of interactions between the characteristics of the listener and those of the target/masker talkers in such tasks. In both studies, we used a modified version of the Coordinate Response Measure to measure Speech Reception Thresholds (SRTs) from both adults and primary-school-aged children. Sentences were of the form 'Show the [animal] where the [colour] [digit] is' with 6 animals, 6 colours and 8 digits. The listener's task was always to report the coloured number from the target sentence (containing 'dog'). Speech maskers had the same form as the targets but with a different animal, colour and number.

In the first study, we investigated the impact of adult female and child interfering talkers (an 11-year old female) on the ability to perceive target speech from the same two talkers. SRTs were measured in four conditions, varying factorially target age (adult or child) and masker type (speech from the opposite age talker, or speech-spectrum-shaped noise matched to the target). Performance for children with the noise maskers was roughly similar to that of adults, with the adult talker more intelligible. Performance for children with speech maskers was, however, the same whether a child was supposed to attend to the child or adult, whereas for the adults, the adult target led again to better performance. It seems likely that this interaction arises from an attentional effect, with children more distracted by another child's voice than by an adult's (preferring fun to authority!).

In a second study, we used male and female adult target talkers in the presence of two simultaneous other talkers, one male and one female. We found a small, but robust interaction between the sex of the listener and the sex of the target, with listeners performing slightly better with talkers of their own sex. Note that this interaction cannot result from listeners preferring voices similar to their own, because target talkers were always adults, and the effect was as strong in children as in adults.

We are currently investigating the generality of this interaction by determining whether it occurs in the presence of 'uninteresting' maskers like speech-shaped noise, or whether it requires informational maskers like other speech.

Colin Cherry Award 2015 Thursday 7 January, 11:40—12:05

EEG decoding of continuous speech in realistic acoustic scenes

J. Hjortkjær
Technical University of Denmark

A number of studies have recently demonstrated that EEG measurements of cortical oscillations entrained to slow speech envelope fluctuations (<10 Hz) can be used to decode which of two competing talkers a listener is attending to. Although auditory cortex is thought to represent speech in a robust and noise-invariant fashion, it is still unknown whether EEG can be used to decode speech in more challenging real-life acoustic environments. To examine this, we have been recording EEG responses to natural speech in various realistic acoustic scenarios. Using a multi-speaker spherical array with ambisonics technology we simulated real rooms with varying degree of reverberation and competing noise sources. Natural speech targets to be detected were distributed at different positions in the virtual rooms. Analysis of single-trial continuous EEG recordings made in these virtual scenes suggest that low-frequency oscillatory activity can be used to decode both (a) what speech source the listener is attending to and (b) the spatial direction of the sound source, even in scenarios involving considerable reverberation and multiple interfering talkers.

Keynote Lecture 2016 Thursday 7 January, 13:30—14:30

Abstraction and adaptation: Key mechanisms listeners use to cope with speech variability

J. M. McQueen
Radboud University, Nijmegen, NL & Max Planck Institute for Psycholinguistics, Nijmegen, NL

Listeners have to cope with a highly variable input in order to be able to recognize speech sounds and spoken words. Two theoretically opposite accounts have been proposed: On the abstractionist view listeners make phonological abstractions about speech and discard variable information as noise; on the episodic view listeners store fully-detailed episodes of spoken language in long-term memory. In this talk I will present evidence from experiments on speech learning which lead toward a synthesis of these two positions. Episodic traces of speech are initially stored, but with time, phonological abstractions are formed, and it is these abstract representations that then mediate language understanding. The experiments (using a variety of behavioural and neuroscientific techniques) explore perceptual learning about speech sounds, how new words are learned, and how flexible lexical processing is. These experiments thus provide evidence that listeners use not only abstraction but also adaptation – the ability to adjust to different listening contexts – to cope with speech variability.

Speech after adoption: Relearning to perceive birth language contrasts by international adoptees

M. Broersma
Centre for Language Studies, Radboud University

Children who are adopted into a country where another language is spoken commonly stop using their birth language abruptly soon after adoption. Within only a few years or even months, they do not seem to remember anything about their birth language, even if they spoke and understood it well at the time of adoption. This is remarkable, given the special status of the first language, which is very robust and usually not forgotten even after many decades of disuse. In this talk I will present the results of two studies that used phonetic retraining of birth language contrasts to investigate whether international adoptees truly forget their birth language, or whether traces of the 'forgotten' birth language, in particular its phonology, remain and can be retrieved with re-exposure.

Study 1 investigates birth language attrition in progress, assessing Mandarin and Cantonese Chinese children in the Netherlands, who had been recently adopted, between six months and nine years prior to testing. Adopted children and a control group of Dutch children with no prior experience with Chinese were trained and tested on perception of Chinese affricate and tone contrasts which were difficult to distinguish for native Dutch listeners. Results show that even for these very recently adopted children, the birth language was not immediately accessible anymore. Re-exposure, however, led to a relatively fast improvement of perception of the birth language sounds, such that adoptees soon outperformed the Dutch control participants.

Study 2 addresses birth language memories several decades after adoption. It investigates whether Korean adoptees in the Netherlands still remember part of their birth language phonology by the time they reach adulthood. Korean adoptees and Dutch control participants were trained and tested on perception of Korean lenis/fortis/aspirated contrasts which are difficult to distinguish for Dutch listeners. Results show that initially (i.e., at the pretest), the Korean adoptees did not perceive the Korean phonemes better than the Dutch control participants. After several training sessions, however, the Korean adoptees did outperform the control group.

Both studies thus provide evidence that international adoptees do retain memories of their birth language phonology and that this aids them in relearning the sounds later in life.

A detailed look at speech recognition in realistic listening scenarios

T. Goverts
VUmc, Amsterdam, NL

H. S. Colburn
Boston University, Boston, MA, USA

In clinical audiology we are interested in optimal participation in social interactions of listeners with impaired hearing, hence we strive for optimizing auditory function. Speech recognition is a major component auditory function in the context of societal participation. A lot of research has been done in the field of speech recognition in realistic environments. However, in those studies the definition of "realistic" seems to be constructed based on theoretical considerations.

We are investigating and analysing bilateral recordings in realistic environments, e.g., at home, in public transport, in a shop, in a restaurant, outside . We will report on the nature of those recordings, in particular: 1) how "speech-like" are these recordings, and 2) how are binaural parameters distributed in these recordings.

The influence of noise or reverberation on the encoding of complex signals by neurons in the cochlear nucleus

I. Winter
Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, GB

Noise management for cochlear implants: Clinical state of the art and future perspectives

P. Hehrmann
Advanced Bionics GmbH, European Research Center, Hannover, DE

Cochlear implant (CI) users suffer particularly severe impairments when listening in noisy or otherwise adverse acoustic conditions. Modern CI speech processors are hence being equipped with a growing number of signal processing options for improved speech understanding and listening comfort in the face of environmental disturbances. Advanced Bionics' most recent generation of speech processor, for example, features algorithms for the suppression stationary and impulsive noise, wind noise and reverberation as well as program automation and a binaural four-microphone beamformer utilizing a bidirectional wireless audio link between two processors.

In a recent study with 15 experienced bilateral CI users, the benefit of the binaural beamformer was evaluated and compared to a monaural beamformer acting independently on either side. Speech reception thresholds (SRTs) were measured using the Oldenburg sentence tests with speech being presented from 0° and uncorrelated, speech-shaped noise from 7 loudspeakers surrounding the listener. SRTs with the binaural beamformer improved by 1.5dB over the monaural beamformers and by 6.0dB over the omnidirectional processor microphone (t-Test, $p < 0.001$ each), highlighting the feasibility and usefulness of binaural noise reduction schemes for CIs.

Linked audio signals shared between two bilaterally-worn opens up a wide range of new approaches to binaural signal enhancement, arguably even more so than in acoustic hearing. Promising concepts under current investigation will be further discussed, considering both bilateral electric and bimodal (electric-acoustic) applications.

The perception of speaking styles under cochlear implant simulation

T. N. Tamati
Department of Otorhinolaryngology / Head and Neck Surgery, University Medical Center Groningen, Groningen, The Netherlands

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. To deal with speech variability, listeners must identify the speech form and extract information about the environment, context, and talker. Further, they must make rapid perceptual adjustments and learn from systematic variation to facilitate speech recognition. These tasks may be very challenging for hearing-impaired users of cochlear implants (CIs), since limitations of the CI may prevent users from being able to reliably perceive and use subtle variations in speech. 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 detection and adaptation to real-life speech is still limited. Further, CI users' perception of speech produced in well-controlled laboratory speech conditions may not reflect their actual real-life performance.

In order to begin to characterize CI perception of real-life speech forms, and to provide guidelines for best clinical practice, the perception of different speaking styles common in real-life speaking environments was investigated in normal and CI-simulated conditions. In particular, normal-hearing listeners completed a perceptual discrimination task and a sentence recognition task using casual and careful speech in three CI noise-vocoder simulation conditions, including none, 12- or 4-channel simulation. The results indicate that the CI simulation had a significant impact on the perception of real-life speaking styles. In the discrimination task, NH listeners were unable to reliably make the categorization judgments under CI simulation. In the sentence recognition task, listeners' ability to recognize casual speech was disproportionately reduced as spectral resolution decreased, with listeners performing much worse on the casual speech than the careful speech under 4-channel simulation. Finally, performance on both tasks was compared to explore the relation between perceptual discrimination and word recognition across the speaking styles.

Taken together, the findings from the CI simulations suggest that perceptual adjustments to real-life speaking styles may be difficult for CI users, given that some important cues to speaking style, such as fine acoustic-phonetic detail, are not be available to them. Despite this, the results suggest that some CI listeners may still be able to use additional cues, such as durational cues related to speaking rate, and draw upon linguistic knowledge and experience to in their perception of real-life speaking styles. By characterizing how CI users perceive and encode speech information related to speaking style, we may be able to develop new clinical tools for the assessment and training of real-life speech perception performance for these listeners.

Using sung speech to evaluate the bimodal benefit to speech and music perception

J. Crew
University of Southern California, Los Angeles, CA, US

J. Galvin, Q. J. Fu
UCLA, David Geffen School of Medicine, Los Angeles, CA, US

Pitch cues provide important indexical and prosodic information for speech perception and are the basis for musical melody. Due to limited spectral resolution, cochlear implant (CI) users have great difficulty perceiving pitch. As a result, pitch-mediated speech perception and melody perception are poorer in CI users than in normal-hearing (NH) listeners. Combining a hearing aid (HA) with a CI ("bimodal" listening) has been shown to improve performance over the CI alone for many speech and music measures, presumably due to the better pitch perception afforded by the HA. However, bimodal benefit has been inconsistent across studies and patients, and seems to reflect interactions between the stimuli, test method, and hearing device, with the better ear for a given task and stimulus often contributing most strongly.

We recently developed the Sung Speech Corpus (SSC) to evaluate contributions of pitch and timbre to speech and music perception. The SSC consists of 50 sung monosyllable words, 10 for each of 5 categories (name, verb, number, color, and object). The words were sung at 13 fundamental frequencies (F0s) from A2 (110 Hz) to and A3 (220 Hz) in semitone steps and normalized to have the same duration and amplitude. The words can be used with a matrix test paradigm to test sentence recognition with fixed or mixed pitch cues across words. The words can also be used to measure melodic contour identification (MCI) with fixed or mixed timbre cues across pitch cues. As such, the SSC allows for the contribution of acoustic and electric hearing to be evaluated for speech and music perception, using the same stimuli that can be manipulated to provide different degrees of complexity for different listening tasks.

Sentence recognition and MCI with sung speech was measured in bimodal listeners; performance was measured with the CI-alone, the HA-alone, or with the CI+HA. Sentence recognition was measured with fixed or mixed pitch cues, as well as with spoken words. MCI was measured with fixed or mixed timbres (i.e., words), as well as with a piano sample. CI performance generally worsened as the stimuli became more complex. Thus, bimodal and CI sentence recognition declined from spoken word to fixed pitch to mixed pitch, and bimodal and HA MCI performance declined from piano to fixed word to mixed word. These preliminary data suggest that bimodal listeners still lack critical pitch processing abilities despite the low-frequency pitch cues provided by the HA.

Posters

Even numbered posters: Thursday 7 January, 15:15.
Odd numbered posters: Friday 8 January, 09:30.

01 Exploring the roles of auditory and cognitive factors and lexical difficulty in individual differences in word-in-noise perception by older adults

S. Knight, A. Heinrich
MRC Institute of Hearing Research, Nottingham, GB

Spoken word recognition relies on the ability to match representations derived from acoustic information to an existing item in the mental lexicon. Listeners' ability to do this depends upon a number of factors, including overall audibility and factors intrinsic to the target word. In particular, lexical characteristics of the target word, such as word frequency, neighbourhood density (number of phonemically similar words) and neighbourhood frequency (word frequency of phonemically similar words), have been shown to affect word recognition: lexically easy targets (words with high frequency and/or low density and/or low-frequency neighbourhoods) are recognised more easily than lexically hard targets (words with low frequency and/or high density and/or high-frequency neighbourhoods), in normal-hearing adults (Sommers & Danielson, 1999), children (Eisenberg et al, 2002), cochlear implant users (Kaiser et al, 2003) and native and non-native speakers (Bradlow & Pisoni, 1999). Moreover, one study investigating factors contributing to individual differences in young and old listeners' ability to recognise lexically hard versus easy words suggested a role for cognitive processes such as inhibition (Sommers & Danielson, 1999).

This study further investigated the relationship between lexical difficulty and individual differences in speech-in-noise intelligibility in older listeners. Fifty older adults (ages = 61-86; mean = 70, age-normal hearing) listened to 200 monosyllabic words whose lexical characteristics varied in terms of word frequency (WF; high vs. low) and neighbourhood density (ND; high vs. low), presented in a background of speech-modulated noise at two signal-to-noise ratios (SNR; +1dB and -2dB). Individual measures of hearing (PTA0.25-8kHz and temporal processing) and cognition (inhibition, working memory and linguistic skill) were also obtained.

Initial analyses show that hearing thresholds, but not age, influenced overall intelligibility scores. When accounting for hearing thresholds, main effects of SNR, WF and ND were as predicted, with participants performing better in the higher SNR, for high-frequency words and for low-density words. A significant interaction of SNR x ND showed that the effect of ND was larger in the more challenging SNR, while a significant interaction of WF x ND showed that participants performed better for high-frequency, low-density words than any other combination of WF and ND. Intelligibility scores were also significantly affected by inhibition as measured by a visual Stroop task, with better inhibitory abilities associated with higher intelligibility; inhibitory abilities may also account for the larger ND effect in the low SNR. Further analyses will examine the relationship between individual differences in intelligibility and cognitive abilities in more detail.

02 The contribution of cognition in a variety of speech-perception-in-noise tests in normal hearing listeners

A. Dryden
MRC Institute of Hearing Research, Nottingham, GB

H. Allen
The University of Nottingham, GB

H. Henshaw
NIHR Nottingham Hearing BRU, GB

A. Heinrich
MRC Institute of Hearing Research, Nottingham, GB

Speech perception is known to be effortful, particularly in adverse conditions. The importance of hearing sensitivity for speech-in-noise (SiN) perception has long been acknowledged, but cannot explain all individual differences. Cognition, and in particular working memory, has emerged as another key factor (Akeroyd, 2008). However, our understanding of cognition for speech perception has been limited by a lack of systematicity and theoretical rigor in their selection of speech stimuli and cognitive tests in many studies.

This study aimed to investigate the contributions of cognition to SiN perception in a systematic and theoretically guided way for a range of speech situations in a cohort of young adults (N = 44, aged 18-30 years) with normal hearing (<20dB HL PTA 0.25-8kHz). Participants were also tested for normal visual acuity (Landolt C test) and were self-reported as native English speakers with no known neurological, psychiatric or language disorders.

In the SiN tests, the speech situation varied concerning the target speech (from high to low predicable semantic content sentences to single words) and background noise (speech-modulated noise or 3-talker babble) with all combinations of fore- and background sounds being tested in the same experiment. Speech target stimuli were presented at 60dB SPL and two signal-to-noise ratios were set dependant on the background noise condition, -2dB SNR for 3-talker babble and -7dB SNR for speech-modulated noise. These different speech conditions were expected to invoke semantic processing and energetic versus informational masking to different extents, which in turn would be associated with the use of different cognitive abilities.

Cognitive abilities were assessed based on Baddeley's model of working memory (Baddeley, 2000) and included tests for verbal working memory, phonological loop, episodic buffer, visuo-spatial scratchpad, and central executive. Individual differences in cognitive abilities were related to performance on each SiN perception.

We expect different associations between intelligibility and cognition for sentences and words given differences in semantic processing required for intelligibility. We also expect a different association between intelligibility and cognition for babble background compared to speech-modulated noise because babble invokes informational as well as energetic masking while noise invokes only energetic masking. Lastly, by explicitly measuring and associating different cognitive components such as storage, verbal abilities and executive control (mental manipulation) we hope to understand why complex working memory measures such as the reading span are often found to correlate so highly with speech perception.

03 Development of a new test for determining binaural sensitivity to temporal fine structure

C. Füllgrabe
MRC Institute of Hearing Research, Nottingham, GB

A. J. Harland
Department of Psychology, University of Cambridge, GB

A. P. Sek
Institute of Acoustics, Adam Mickiewicz University, PL

B. C. J. Moore
Department of Psychology, University of Cambridge, GB

Speech identification in noise has been shown to be associated with a listener's sensitivity to the temporal-fine-structure (TFS) of sounds. Both hearing loss and age (Hopkins & Moore, 2011; Füllgrabe, 2013) seem to adversely affect the ability to process TFS information. Hence, there has been keen interest in the development of tests that could be used in the clinic or in large-scale research studies to assess sensitivity to TFS.

Moore and colleagues designed two such tests: the monaural TFS1 test that requires listeners to discriminate complex harmonic tones from the same tones with all components shifted upwards by the same amount in Hertz (Moore & Sek, 2009); and the binaural TFS-LF test that determines a listener's detection threshold for an interaural phase difference (IPD) for pure tones of a given low frequency (Hopkins & Moore, 2010). However, several studies have reported that some listeners were unable to reliably perform the tests, so a graded measure of sensitivity to TFS could not be obtained for all listeners.

To address this limitation, a new binaural TFS test was developed and validated, the so-called TFS-AF test (with AF standing for 'adaptive frequency'). In this test, listeners are required to detect an IPD in pure tones but, in contrast to the TFS-LF test, the IPD is fixed while the frequency of the pure tones is adaptively varied to determine the threshold.

Normative data for a range of IPDs were obtained for young (19-25 years) listeners with audiometrically normal hearing (≤20 dB HL). Results revealed thresholds between 950 and 1600 Hz for IPDs of at least 90º, and no evidence for improved sensitivity following protracted practice (Füllgrabe et al., 2015). In a follow-up study, IPD sensitivity was assessed in older (>60 years) adults with normal hearing to investigate if: (i) listeners can perform the TFS-AF test even when performance on the TFS-LF test is not possible, and (ii) familiarity with the test procedure and stimuli plays a greater role for listeners from an age range that is more representative of audiological patients. Results indicated that all older listeners could perform the TFS-AF test and there were no or only small practice effects.

The results suggest that IPD sensitivity is fairly stable across frequencies (on average, ≤1300 and 1000 Hz for the young and older normal-hearing listeners, respectively) and that the TFS-AF test is suitable for the rapid assessment of TFS sensitivity in untrained listeners.

04 Do individual differences in working memory predict speech-in-noise intelligibility in normal hearers?

C. Füllgrabe
MRC Institute of Hearing Research, Nottingham, GB

S. Rosen
UCL Speech, Hearing & Phonetic Sciences, London, GB

With the advent of cognitive hearing science, increased attention has been given to individual differences in cognitive functioning and their explanatory power in accounting for inter-listener variability in understanding speech in noise (SiN). The psychological construct that has received most interest is working memory (WM), representing the ability to simultaneously store and process information. Common lore and theoretical models assume that WM-based processes subtend speech processing in adverse perceptual conditions, such as those associated with hearing loss and background noise. Empirical evidence confirms the association between WM capacity (WMC) and SiN identification in older hearing-impaired listeners.

To assess whether WMC also plays a role when listeners without hearing loss process speech in acoustically adverse conditions, we surveyed published and unpublished studies in which the Reading-Span test (a widely used measure of WMC) was administered in conjunction with a measure of SiN identification. The survey revealed little or no evidence for an association between WMC and SiN performance.

We also analysed new data from 132 normal-hearing participants sampled from across the adult lifespan (18 to 91 years), for a relationship between Reading-Span scores and identification of matrix sentences in noise. Performance on both tasks declined with age, and correlated moderately even after controlling for the effects of age and audibility ($r = 0.39$, $p \leq 0.001$, one-tailed). However, separate analyses for different age groups revealed that the correlation was only significant for middle-aged and older groups but not for the young participants (< 40 years).

A possible explanation for this increasing cognitive involvement with age is the accumulation of age-related deficits in supra-liminary auditory processing (e.g. sensitivity to temporal-fine-structure and temporal-envelope cues; Füllgrabe, 2013; Füllgrabe et al., 2015), resulting in under-defined and degraded internal representations of the speech signal, calling for WM-based compensatory mechanisms to aid speech identification.

Füllgrabe, C. (2013). Age-dependent changes in temporal-fine-structure processing in the absence of peripheral hearing loss. American Journal of Audiology, 22(2), 313-315.

Füllgrabe, C., Moore, B. C., & Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. Frontiers in Aging Neuroscience, 6, 347.

Work supported by the Medical Research Council (grant number U135097130) and the Oticon Foundation (Denmark).

05 Investigating the role of auditory and cognitive factors for various speech-perception-in-noise situations in older listeners

A. Heinrich, S. Knight
MRC Institute of Hearing Research, Nottingham, GB

Understanding the causes for speech-in-noise (SiN) perception difficulties is complex, and is made even more difficult by the fact that listening situations can vary widely in target and background sounds. While there is general agreement that both auditory and cognitive factors are important, their exact relationship to SiN perception across various listening situations remains unclear. This study manipulated the characteristics of the listening situation in two ways: first, target stimuli were either isolated words, or words heard in the context of low- (LP) and high-predictability (HP) sentences; second, the background sound, speech-modulated noise, was presented at two signal-to-noise ratios. Speech intelligibility was measured for 50 older listeners (ages = 61-86; mean = 70) with age-normal hearing and related to individual differences in cognition (working memory, inhibition and linguistic skills) and hearing (PTA0.25-8kHz and temporal processing). The results showed that while the effect of hearing thresholds on intelligibility was rather uniform, the influence of cognitive abilities was more specific to a certain listening situation. By revealing a complex picture of relationships between intelligibility and cognition, these results may help us understand some of the inconsistencies in the literature as regards cognitive contributions to speech perception.

06 Listening effort: A consequence of continuous listening in noise?

V. Vianeva, H. Roebuck, A. Casaponsa, J. G. Barry
MRC Institute of Hearing Research, Nottingham, GB

Listening in noise is known to impact negatively on speech perception (Miller & Nicely, 1955). It also has consequences for cognitive abilities like short-term memory capacity (Rabbitt, 1968) possibly because of increased processing demands (i.e., listening effort) induced by the noise. This study aimed at understanding the role of listening in noise on ability to sustain attention. It was based on the hypothesis that listening in noise is resource-heavy and results increased lapsing of attention as processing resources are exhausted.

We tested this hypothesis in typically-developing children, aged 6 - 12 years, who completed a continuous listening task (Roebuck, Freigang, & Barry, in prep) presented in 4-speaker babble at +2dB SNR. The task involved listening to a story (16 mins) and pressing a button to intermittently occurring targets. The targets (n = 108) were either nonsense words (n = 36) or mispronunciations, which could be predicted (n = 36) or not predicted (n = 36) from the preceding context. Three noise conditions were used: (1) noise throughout (2) noise 1st half; (3) noise 2nd half.

Previous research (Roebuck et al., in prep) suggests a benefit when listening in quiet for predictable words compared with unpredictable or nonsense words. Reflecting increasing listening effort and consequent lapsing in attention, we predicted for Condition 1: A) an increase in errors due to the more difficult listening conditions, B) a decrease in benefit of context and, C) an increase in reaction time from the first to the second half of the task.

Predictions A and B were supported by the data for all age groups. Reaction times, however, decreased significantly from the first to the second half of the task. In addition to possible effects due to listening effort, the findings suggest influences from on-task learning, and possible adaptation to noise over time. Conditions 2 and 3 (in progress) explore these different possible contributions to continuous listening in noise.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 27(2), 323-352.

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. The Quarterly Journal of Experimental Psychology, 20(3), 241-248.

Roebuck, H., Freigang, C., & Barry, J. G. (in prep). Listening to continuous speech: the role of sustained attention in children referred for listening difficulties without hearing loss.

07 Development of normed speech-in-noise sentence lists for use with children

J. G. Barry, C. Freigang, S. Knight, A. Heinrich
MRC Institute of Hearing Research, Nottingham, GB

Listeners with equivalent pure tone audiograms can vary considerably in their ability to perceive and understand speech-in-noise. In addition to perceptual abilities, this variation may reflect differences in language or cognitive skills which support processing of the noise-degraded signal. The R-SPIN sentences (Bilger, et al., 1984) were developed to capture these differences and help researchers distinguish between perceptual and linguistic contributions to listening-in-noise. This is done by varying the extent to which the final word can be predicted from the preceding context. We recently developed a similar set of sentences (BESST-UK) which were additionally carefully matched for prosody, length and overall structure (BESST-UK, Heinrich, et al., 2014; Barry, et al., 2014). Here, we describe further work aimed at developing normed lists of sentences drawn from the BESST-UK database for use with children.

First, 250 children, subdivided into three age bands (4-6, 7-8, and 9-12 years), listened to and repeated sentences presented in 12-talker babble at up to three different signal-to-noise ratios (SNR). The aim was to identify the SNR resulting in scores in the range of 70 to 90 percent correct (unpredictable versus predictable sentences respectively) for each age band. As has been shown in previous studies, young children needed a more favorable SNR (+4 dB) to achieve the same level of performance as the oldest group of children (-1 dB).

In the next step of the project (in progress), the data at the age-specific preferred SNRs will be modelled using Rasch analysis. This is a form of latent trait modelling which is applied in measurement development. It estimates sensitivity of individual items to levels of 'difficulty' in a particular latent trait. Rasch analysis has not been applied to speech perception data. However, we predict that sentences drawing on cognitive / language abilities (i.e. predictable sentences) will be more amenable to modelling, than sentences where perceptual abilities are being tested (i.e., unpredictable sentences). We hope to use the results of the Rasch analysis to develop multiple sentence sets which offer equivalent measures of speech perception abilities.

Barry, et al. (2014). Sensitivity of the British English Sentence Set Test (BESST) to differences in children's listening abilities. BSA Conference.

Bilger, et al. (1984). Standardization of a Test of Speech-Perception in Noise. JSLHR, 27(1), 32-48.

Heinrich, et al. (2014). Assessing the effects of semantic context and background noise for speech perception with a new British English speech test. BSA Conference, Keele.

08 Binaural processing in hearing impairment

B. Bardsley, J. Culling
Cardiff University, United Kingdom

A typical complaint of those with hearing impairment is understanding speech in the presence of background noise. A laboratory measure used to assess performance in this situation is known as spatial release from masking (SRM). It has been established that normative SRM, using speech-shaped noise as the interferer, provides ~ 11 dB of benefit when the interferer moves from 0° to 112° azimuth (Plomp & Mimpem, 1981).

Many studies have reported a poorer SRM in those with hearing impairment (e.g. Noble et al. 1997). It would appear audibility is not a key predictor of SRM performance as there has been a varying level of correlation between SRM and PTA thresholds reported in the literature (e.g. Ter-Horst et al. 1993; Peissig & Kollmeier, 1997). Poor temporal fine structure (TFS) processing has been postulated as a reason behind the poorer performance of those with hearing impairment in SRM.

In this study we investigated the relationship between SRM, measures of binaural TFS processing and independent use of interaural time and level differences. Hearing loss has been reported to impact on binaural processing of TFS even when age is partialled out (King et al. 2014) so we would expect poorer SRM linked to poorer binaural TFS processing. In analysing results on an individual basis it is now emerging that our results are heterogeneous and do not support our predictions for this association. The independent use of different interaural cues and the link to SRM performance will help elucidate this matter.

King, A., Hopkins, K. & Plack, C. (2014). The effects of age and hearing loss on interaural phase discrimination. J. Acoust. Soc. Am. 135, 342-351.

Noble, W., Byrne, D. & Ter-Horst, K. (1997). Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners. J. Acoust. Soc. Am. 102, 2343-2352.

Peissig, J. & Kollmeier, B. (1997). Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. J. Acoust. Soc. Am. 101, 1660-1670.

Plomp, R. & Mimpen, A. M. (1981). Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences. Acustica. 48, 325-328.

Ter-Horst, K., Byrne, D & Noble, W. (1993). Ability of hearing-impaired listeners to benefit from separation of speech and noise. Australian Journal of Audiology. Aust J Audiol. 15, 71-84.

09 Phase modification for increasing the loudness of telephone speech in adverse noise conditions

E. Jokinen
Aalto University, FI

H. Pulakka
Microsoft Phone Technologies, Tampere, FI

P. Alku
Aalto University, FI

Near-end noise conditions, where the background noise is in the listener's environment, are common in mobile communications. In such conditions, post-processing techniques can be used in the receiving mobile device to improve the speech signal's intelligibility and reduce its listening effort by increasing speech prominence, clarity, and loudness. While many intelligibility enhancement methods have been proposed for this problem scenario, they are commonly based on either modifying the magnitude spectrum or the time domain signal directly.

In this study, two loudness increasing methods based on the modification of the phase spectrum are studied. One of the algorithms aims to reduce the dynamic range of the signal and take advantage of the energy gain resulting from amplitude normalization to increase the loudness, while the other algorithm is designed to sharpen the high-amplitude peaks in the time-domain signal generated by the periodic glottal excitation to make the speech sound more clear. Both methods are based on first modifying only the phase spectrum, after which the time-domain signal is computed using the inverse Fourier transform. Finally, the time-domain signal is amplitude normalized by scaling its sample values so that they occupy the original amplitude range of the processed frame. The performance of the proposed methods was compared to unprocessed speech using subjective loudness and quality evaluations as well as objective quality measures. The results suggest that the phase modification methods increase both the loudness and clarity of telephone speech, thus reducing the listening effort while generally maintaining the subjective speech quality.

11 Cross-talk cancellation and its potential benefits to speech understanding in background noise to patients with bilateral bone-anchored hearing aids

R. W. J. Mcleod, J. F. Culling
Cardiff University, School of Psychology, GB

Bilateral bone-anchored hearing aids (BAHAs) suffer from cross-talk within the skull, but two BAHAs could be utilised to create a cross-talk cancellation system. In order to achieve this, sound vibrations from each BAHA reaching the contralateral cochlea must be cancelled by an out-of-phase signal of the same level from the ipsilateral BAHA. We have developed a psychoacoustic method for calculating the cancellation level and phase needed for cross-talk cancellation, shown how these values can be employed to increase speech understand in background noise and demonstrated the potential benefits to patients. Participants with normal hearing wore two B71 bone transducers (BT) (one on each mastoid) and bilateral ER2 earphones. Both BTs were stimulated with the same pure tone and the level and phase adjusted in the right BT in order to cancel all perceived sound at the right ear. Participants could reliably pinpoint the level and phase necessary for cancellation between 1500 and 8000 Hz. Speech reception thresholds (SRTs) were obtained with and without cross-talk cancellation based on these measurements, which were used to modify the noise signal and add it to the speech signal presented to the right BT. The noise from the left BT was cancelled at the right cochlear by the modified noise signal on that side, resulting in a mean benefit of 15.6 dB in SRT. In order to get an indication of real-world benefit, head related impulse responses (HRIR) at 0° and 90° from a BAHA 4 were used to simulate a simple listening situation that was then presented either using earphones or using the BTs. The difference between the two gives an indication of the deleterious effect of cross-talk. SRTs were measured with speech at 0°and noise at 0° or 90°. Because the cross-talk cancellation only addresses the distortion of ILDs by cross-talk, the relative spatial unmasking benefit from ILDs and ITDs was measured; the level and phase components of the HRIR from the 90° and 0° were exchanged to produce ILD-only and ITD-only conditions. Spatial release from masking due to ILDs recorded from BAHA microphones was 4.4 dB better with the ER2s, than with the BTs, indicating the maximum potential benefit of cross-talk cancellation.

12 Simulation of a cochlear implant with current spread and varied number of activated electrodes

J. A. Grange, J. F. Culling
Cardiff University, School of Psychology, GB

Cochlear implant (CI) users' spectral resolution is inherently restricted by a limited number of excitatory spectral channels/electrodes (up to 20). Thresholds of speech intelligibility in noise (SpIN) are consequently higher in CI users than in normally hearing listeners. CI current spread along the spiral ganglia further reduces spectral resolution, which lowers the number of "effective channels" to around 8 (Friesen et al., 2001), beyond which no significant SpIN improvement is found. Simulations of CIs through tone or noise vocoding traditionally employ the same numbers of analysis bands (electrodes) and carriers (tones or noise bands) used for resynthesis. A recent extension to traditional vocoding incorporates current spread as an exponential function weighting the contribution of envelopes from each analysis band to each carrier (Oxenham & Kreft, 2014). When measuring percent-correct recognition of words in sentences as a function of the number of activated electrodes, a knee point should appear where current spread starts to counteract the benefit of further increase in activated electrode number. This occurs at the number of effective channels. Increasing the simulated current spread should shift the knee point to a lower number of effective channels. However, a vocoder incorporating current spread as per Oxenham & Kreft fails to produce the expected outcome. We propose a novel vocoder with a fixed, large number of carrier bands that more adequately models the fixed ganglion cells. The exponential weighting function is integrated over the carrier bands to ensure an improved representation of coupling between electrodes and ganglion cells. The weighting can equally be applied to carrier bands or tones. Where using a variable number of carriers fails to demonstrate the expected effect of current spread on the number of effective channels, the proposed novel approach succeeds.

Friesen, L., Shannon, R., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am., 110, 1150.

Oxenham, A., and Kreft, H. (2014). "Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing," Trends Hear., 18, 1–14.

13 Binaural room impulse responses for speech-in-noise testing in virtual restaurants

J. Culling, L. Ramage, A. Shields
Cardiff University, School of Psychology, GB

Binaural room impulse responses (BRIRs) were recorded in two real restaurants using the tone-sweep method (Farina, 2007). In a semi-automated procedure, ten-second logarithmic tone sweeps were played from Cambridge Minx satellite speakers and recorded binaurally using a B&K 4100 head and torso simulator. These were convolved with an inverse sweep to derive BRIRs between different seating positions. In each restaurant, a speaker position and an opposing listener position were selected for each table. BRIRs were recorded from each table to every other. In Mezza Luna, with 18 tables, three listener head positions (-30°, 0°, +30° with respect to the speaker position on the same table) were also recorded, making 18×18×3=972 BRIRs. In A Caverna, with 15 tables, 15×15=225 BRIRs were recorded. Predicted speech intelligibility was explored using the Jelfs et al. (2011) model of spatial release from masking for speech. Predicted SRTs (pSRTs) improved when the head was turned to ±30°. In Mezza Luna, a long narrow restaurant, pSRTs were higher in the middle of the room. In A Ceverna, a restaurant on two levels, pSRTs were much lower on the mezzanine floor. This set of predictions provides a powerful test for the model that can be evaluated using virtual acoustic listening tests based on the same BRIRs. The BRIRs could also be used to assess the acoustic capacity of the restaurants, the largest number of simultaneous voices in the room that listeners can tolerate.

Farina, A. (2007). Advancements in impulse response measurements by sine sweeps. Proceedings of 122nd AES Convention, Vienna.

Jelfs, S., Lavandier, M. and Culling, J. F. (2011). Revision and validation of a binaural model for speech intelligibility in noise. Hear Res, 275, 96-104.

14 Perception of American English consonants by Japanese listeners in background noise, reverberation, and background noise + reverberation

H. Masuda
Waseda University, Tokyo, JP

Numerous studies have shown that both native and L2 listeners' phoneme identification are negatively affected by background noise and reverberation, and that the challenge is larger for L2 listeners. However, past research have often discussed the various listening environments separately. Although direct comparison between background noise and reverberation is difficult, it is important to examine the similarities and differences of environmental impact on L2 speech perception. The present study aims to overlook L2 speech perception in various listening environments (quiet, background noise, reverberation, and background noise + reverberation), and to reexamine the relationship between L2 speech perception in adverse environments and L2 proficiency. In particular, we focus on Japanese listeners' identification and confusion patterns of American English consonants in background noise, reverberation, and background noise+ reverberation. Twenty-three Japanese and 12 American English listeners participated in Experiment 1 (published), and 22 Japanese and 24 American English listeners participated in Experiment 2 (new experiment). Listeners were presented with 23 English consonants in intervocalic context under quiet, background noise, reverberation, and background noise + reverberation. Experiment 1 consisted of four listening environments; 1) quiet, 2) SNR = 10 dB, 3) SNR = 5 dB, and 4) SNR = 0 dB. Experiment 2 consisted of five listening environments; 1) quiet, 2) RT = 0.78 s, 3) RT = 1.12 s, 4) RT = 1.43 s, and 5) SNR = 10 dB + RT = 0.78 s. Correct identification rates and confusion patterns were analyzed against Japanese listeners' TOEIC(R) scores and LOR. Mean identification rates pooled across 23 consonants showed a significant difference between native and non-native listeners in Experiment 2 ($p = 0.03$) but not in Experiment 1 ($p = 0.58$). Correlation coefficient r showed that TOEIC(R) scores and listening environments had marginal correlations in Experiment 2, suggesting higher TOEIC(R) scores result in better identification in reverberation. Experiment 1 showed the opposite case, where correlation became weaker as listening environment became more adverse, suggesting even high TOEIC(R) score holders suffer identifying consonants with increased background noise. Weak to moderate correlation ($r = 0.31 \sim 0.50$) was observed between LOR and identification rates only in Experiment 2, with the lowest correlation in quiet environment, suggesting that experience of residing overseas may be beneficial for L2 perception in reverberation but not necessarily so in background noise. Further analysis is performed on the confusion patterns of higher and lower Japanese groups and compared against native listeners.

15 Beta range responses in thalamic CM/Pf complex indicate a role in speech cue signaling

I. M. Schepers
Oldenburg University, Oldenburg, DE

A. K. Beck
Hannover Medical School, Hannover, DE

S. Bräuer
Oldenburg University, Oldenburg, DE

K. Schwabe, G. Lütjens, J. K. Krauss
Hannover Medical School, Hannover, DE

J. W. Rieger
Oldenburg University, Oldenburg, DE

Speech processing often requires cognitive control to focus on certain cues and initiate adequate responses to the speech input. The centromedian-parafascicular complex (CM/Pf) of the thalamus has been described as part of a cognitive control loop and may focus cortical and basal ganglia processing on task relevant information. It may also have an important function in the cognitive control of speech processing. Postoperative recordings of local field potentials (LFPs) in patients implanted with electrodes for deep brain stimulation (DBS) allow to obtain direct neurophysiological information with high spatial and temporal accuracy from the CM/Pf in humans. Here, we characterize the neural responses in the CM/Pf in speech processing and processing of action cues.

We used a multi-speaker paradigm with task relevant auditory and visual cues. In each trial, an auditory cue word was uttered in one of two concurrent speech streams that signaled which speaker would provide task relevant information later on in the speech stream. After speech offset a visual cue indicated that the participant should press buttons that corresponded to the task relevant information the speaker with the cue word had used. LFPs were obtained from two patients with chronic neuropathic pain (one woman, one man, aged 57 and 55 years, both right handed) implanted with quadripolar DBS electrodes in the right CM/Pf. Task performance was high (94% and 95% correct) indicating that participants paid attention to the cue and switched attention to the task-relevant speech stream. Average time-frequency responses showed transient increases in the beta range (20-30 Hz) after the cue word in the speech stream and in the theta band (4-8 Hz) after the visual cue indicating the beginning of the response interval. Single trial latencies of the beta-band responses correlated with the cue word onsets and latencies of the theta-band responses correlated with the visual cue onsets. Beta responses after the cue word were mostly greater than after the task relevant words.

Our results support the notion that the CM/Pf is part of a cognitive control loop involved in the initiation of goal relevant selection based on speech and visual information. Importantly, these signals are not only related to an externally triggered orienting response as they are elicited by both, the semantics of speech as well as the visual cue in the task context. Furthermore, they do not reflect the attentional orienting to a speech stream because the CM/Pf responses are transient after the cue.

16 A model-based evaluation of speech perception in noise for electro-acoustic listeners

L. Zamaninezhad, V. Hohmann, M. R. Schädler, T. Jürgens
Medizinische Physik, Forschungszentrum Neurosensorik, and Cluster of Excellence "Hearing4All", Carl-von-Ossietzky Universität Oldenburg, DE

Due to recent advances in surgical techniques, there is a possibility to preserve existing residual acoustic hearing in many cochlear implant (CI) candidates even after implantation. This group of listeners who receives both electric and acoustic stimulation in the same ear is called electro-acoustic (EA) listeners. For these listeners, many clinical studies reported a benefit in EA-listening condition in comparison with electric-only or acoustic-only listening conditions (termed EA-benefit) for speech intelligibility in noise.

The goal of this study is to introduce a physiologically inspired auditory model of speech intelligibility that can predict the speech-in-noise perception of EA-listeners. In addition to the assessment of the effect of different physiological factors (residual acoustic hearing, spatial spread of electric field) on speech intelligibility, the model could help to investigate the underlying mechanism for EA-benefit.

Two different auditory models are used to simulate combined electrically/acoustically stimulated auditory nerve (AN) spikes. The auditory model of Fredelake and Hohmann (2012) simulates the AN spikes in response to electric stimulation. As this model mimics the CI signal processing strategy with a constant electric pulse rate, only the information about the envelope of the speech signal is extracted. The Meddis (2006) model is used to produce AN spikes in response to acoustic stimulation, mostly restricted to low frequencies for EA-listeners. This acoustically stimulated AN spiking pattern contains vital information about the fundamental frequency and in some cases up to first formant, which may not be available to conventional CI users in this form. The AN spiking patterns are further processed by a central auditory processing stage (Fredelake and Hohmann, 2012). This results in an internal representation (IR) of the stimuli. Speech reception thresholds in stationary noise were predicted by simulating the German matrix sentence test with an adapted automatic speech recognition system (Schädler, IJA, in press), using the individual or concatenated IRs as front-ends.

The model predicts an EA-benefit of up to 3 dB, a result which is in line with clinical studies. Changing the upper frequency boundary of residual acoustic hearing in the model showed that even a very restricted frequency range (which contains only information about the fundamental frequency) can still result in EA-benefit. Increasing the electric field spatial spread resulted in higher SRTs both in electric-only and electro-acoustic listening conditions. Overall the model could reproduce most of the typical SRTs and EA-benefits reported by clinical studies.

Supported by DFG cluster of excellence EXC 1077/1 "Hearing4All"

17 Are deep neural network speech recognizers still hearing-impaired?

C. Spille, B. T. Meyer
Medizinische Physik and Cluster of Excellence Hearing4all, University of Oldenburg, DE

Previous comparisons of human speech recognition (HSR) and automatic speech recognition (ASR) focused on monaural signals in additive noise, and showed that HSR is far more robust against intrinsic and extrinsic sources of variation than conventional ASR. The difference in performance between normal-hearing people and ASR was in about the same order of magnitude than the difference between normal-hearing (NH) and hearing-impaired (HI) listeners (listening experiments were performed during the HearCom project, www.hearcom.eu), leading to the saying that "ASR systems are hearing-impaired". Recent developments in ASR, especially the use of deep neural networks (DNNs), showed large improvements in ASR performance compared to standard recognizers based on Gaussian Mixture Models (GMMs). The aim of this study is (A) to compare recognition performance of NH and HI listeners in monaural conditions with different noise types with state-of-the-art ASR systems using DNN/HMM and GMM/HMM architectures and (B) to analyze the man-machine gap (and its causes) in more complex acoustic scenarios, particularly in scenes with two moving speakers and diffuse noise. The overall man-machine gap is measured in terms for the speech recognition threshold (SRT), i.e., the signal-to-noise ratio at which a 50% recognition rate is obtained. For both scenarios we also investigate the effect of auditory features on the performance of ASR systems and measure the similarity between different ASR systems and NH listeners.

For (A) we compare responses of 10 normal-hearing listeners to different ASR systems with the identical speech material utilizing the Aurora2 speech recognition framework. Besides, compare data collected from normal-hearing and hearing-impared listeners during the HearCom project using the Oldenburg sentence test (OlSa) with additive stationary noise. Results show that state of the art ASR systems can reach performance of normal-hearing listeners in terms of SRT under certain conditions.

For (B) responses of nine normal-hearing listeners are compared to errors of an ASR system that employs a binaural model for direction-of-arrival estimation and beamforming for signal enhancement. The comparison shows that the gap amounts to 16.7 dB SRT difference which exceeds the difference of 10 dB found in monaural situations. Based on cross comparisons that use oracle knowledge (e.g., the speakers' true position), incorrect responses are attributed to localization errors (7 dB) or missing spectral information to distinguish between speakers with different gender (3 dB). The comparison hence identifies specific ASR components that can profit from learning from binaural auditory signal processing.

18 Can hearing aids speed up speech comprehension in noise? Insights from eye-tracking measurements and speech-evoked potentials

J. Habicht
Oldenburg University, DE

M. Finke
Hannover Medical School, DE

B. Kollmeier, T. Neher
Oldenburg University, DE

The effects of hearing aid provision on cognitive-linguistic processes are still rather underexplored. In an earlier study, we compared the performance of two matched groups of participants either with or without hearing aid experience on an eye-tracking task tapping into speech comprehension (cf. Wendt et al, PLoS ONE 2014) under different aided conditions. We found that, despite comparable speech recognition performance, the participants with hearing aid experience were considerably faster at grasping the meaning of the presented sentences than the participants without any hearing aid experience, irrespective of the aided condition. In the current study, we followed up on this by investigating the effects of auditory acclimatization to bilateral amplification on performance on the eye-tracking task. To that end, we tested groups of novice and long-term hearing aid users (N >= 15 each) before and after several months of acclimatization to, or continued use of, bilateral amplification. To explore any neuroplastic changes induced by the provision of amplification we also measured speech-evoked potentials. From the results, we expect that acclimatization to bilateral amplification will result in faster speech comprehension and concomitant changes in late auditory potentials.

19 The differential role of auditory and non-auditory measures for predicting speech-in-noise intelligibility: A comparison between hearing-impaired listeners with and without hearing aid supply

A. Gieseler, M. Tahden, H. Colonius, C. M. Thiel
Department of Psychology, University of Oldenburg, DE

There exists a large variability among hearing aid users (HAUs) in rehabilitation success as reflected in measures of speech-in-noise (SIN) intelligibility, even when accounting for age and hearing loss (HL). Research has shown that standard audiological measures alone cannot comprehensively explain this variance, suggesting factors beyond these measures to contribute to the differences. Particularly the role of cognitive abilities is currently being investigated. Widely neglected so far has been the subclinical population (SCP), i.e. individuals with an age-related HL who are not yet provided with hearing aids (HA). To address these issues, the current study aims at identifying relevant factors beyond audiological ones which contribute to an improved prediction of SIN performance, and furthermore at differentiating between hearing-impaired listeners with and without HA supply at comparable levels of HL.

A multivariate analysis was conducted on data from n=333 hearing-impaired subjects from the database of the Hörzentrum Oldenburg comprising auditory measures, cognitive measures and a questionnaire with subjective ratings. These measures were entered as explanatory variables in a stepwise linear regression model for predicting speech-in-noise reception thresholds (SRTs) on group level for HA users (HAUs) (n=216) vs. non-users (SCP) (n=117). For group comparison, HAUs were measured unaided and both groups were categorized into slight (26-40 dB HL) and moderate (41-60 dB HL) degree of HL.

Preliminary results indicate that relevant predictor variables differ substantially as a function of group and level of HL. At moderate levels of HL, the audiogram was the best predictor for SRTs in both groups, accounting for ~60% variance. At slight levels, the best model in the SCP group explained ~30% variance and comprised audiological measures alone. In contrast, in the HAU group, verbal intelligence and subjective ratings accounted for two thirds of the total variance (~60%) with the audiogram explaining another third.

These results suggest that not only the degree of HL plays a role in models of performance prediction but, moreover, the provisioning with a HA. At moderate levels, audiological measures are the main predictors for both groups. For hearing-impaired subjects with a slight HL who are not provisioned with a HA, audiological measures were the best predictors as well, although leaving much unexplained variance. In contrast, for HAUs at slight levels, verbal intelligence was a better predictor than hearing ability for unaided speech understanding. This suggests that including non-audiological measures can improve patient's performance prediction depending on HA provisioning and degree of HL.

20 Long-term effects of hearing impairment for word processing and speech recognition in noise

R. Carroll, A. Warzybok, E. Ruigendijk
University of Oldenburg, DE

Speech recognition in noise has been proposed to correlate with a listeners' vocabulary size (McAuliffe et al., 2013; Benard et al., 2014). One reason for such a relation is that a larger lexicon requires a more detailed representation in the mental lexicon to successfully distinguish phonological neighbors. Some researchers (Arlinger, 2003; Classon et al., 2013) have suggested untreated long-term hearing loss to lead to cue deprivation that may render detailed lexical information (e.g., phonetic details) less reliable. Carroll et al. (submitted) showed age effects of lexical access times and speech recognition scores in younger versus older listeners with normal hearing. Measures of vocabulary size and working memory did not show group effects. Effective lexical access time was associated with speech recognition scores. Lexical access time depends on the size of the mental lexicon and the relative efficiency with which a lexical candidate is found when boundaries between phonological neighbors become less distinct due to phonetic blurring caused by impaired hearing. We therefore asked whether hearing impaired listeners show long-term effects regarding the mental lexicon that relate to speech perception in noise. We tested 22 older listeners with normal hearing (60-78 yrs.) and 20 older listeners with mild-to-moderate hearing loss (60-80 yrs.). We measured speech recognition thresholds (SRT) in noise correct using a German everyday sentence test with adaptive procedure to establish the individual SNRs yielding 50% and 80% correct recall. Speech was presented with NAL-R and without amplification. Both groups also completed a small battery of selected individual differences measures with a focus on the mental lexicon. Data collection of hearing impaired listeners is ongoing. First data show strong differences of about 4 dB SNR in speech recognition scores without amplification. With NAL-R amplification, SRTs yielding 50% correct were comparable, with only 1 dB difference, but SRTs yielding 80% correct still showed a 3 dB benefit for the group with normal hearing. Whereas measures of vocabulary size were comparable between groups, working memory showed a small disadvantage for hearing impaired listeners. Hearing impairment was also associated with much longer lexical access times. The combination of slow lexical access, comparable vocabulary size and somewhat lower verbal working memory is expected to contribute to relatively poorer speech recognition scores.

21 Real time hearing impairement simulator for speech intelligibility measurements

N. Grimault
Auditory Cognition and Psychoacoustic team, Centre de Recherche en Neurosciences de Lyon, CNRS, Université Lyon 1, FR

E. Parizet
Laboratoire Vibrations Acoustique, INSA-Lyon, FR

A. Corneyllie, S. Garcia
Auditory Cognition and Psychoacoustic team, Centre de Recherche en Neurosciences de Lyon, CNRS, Université Lyon 1, FR

L. Brocolini
Laboratoire Vibrations Acoustique, INSA-Lyon, FR

The AIDA projects aims at improving speech intelligibility in cars for people suffering from mild hearing losses. Signal processing will be used to modify speech messages provided by a car (e.g. from the navigation system), in order to make them more intelligible to the driver.

This will necessitate many experiments using hearing-impaired people. In order to ease such experiments, it was decided to use a hearing-loss simulator. This simulator will make it possible to evaluate the effectiveness of some signal processing techniques using normal hearing people, while simulating a given threshold profile.

The simulator reproduces the enlargement of the auditory filters as well as the loss of recruitment at low level of presentation. The procedure is basically an inverse gammachirp auditory filter bank as first described by Irino. The algorithm is implemented in Python and parallelized using the graphic card resources. Such a design enables the algorithm to work in real time, so that the signal produced by a software conducting the intelligibility experiment is modified before being presented to the listener.

This paper will present the simulator as well as an experiment conducted to evaluate its accuracy. First of all, Speech Reception Threshold has been measured in three typical car interior noises, using a french version of the Four Alternative Auditory Feature test (Foster and Haggard, 2005). Participants were 76 people, recruited either as customers of an audioprothesist, either according to age criterion (over 55). Tonal audiometry allowed to separate participants in three groups, from normal hearing to a mild hearing loss. SRT was measured within these three groups.

Then, twenty young normal-hearing subjects were recruited. The simulator allowed to recreate the hearing ability of a typical member of the mild hearing loss group, as defined in the first step. The SRT was measured and compared to the one obtained with the previous subjects.

This device allows an easier recruitement of subjects, for which the effectiveness of speech emphasizing techniques will be evaluated.

22 Longitudinal assessment of spectral ripple discrimination and speech perception evolution in cochlear implant users

A. Lopez Valdes, M. Mc Laughlin
Trinity Centre for Bioengineering, Trinity College Dublin, IE

L. Viani, P. Walshe
National Cochlear Implant Programme, Beaumont Hospital, IE

J. Smith
National Cochlear Implant Programme, Beaumont Hospital, IE

L. Flood
National Cochlear Implant Programme, Beaumont Hospital, IE

R. B. Reilly
Trinity Centre for Bioengineering, Trinity College Dublin, IE

Introduction — Psychoacoustic studies have shown that cochlear implant (CI) users' ability to discriminate spectrally-rippled noise stimuli correlates reasonably well with their speech perception. Non-linguistic tests utilizing spectral ripple have been proposed to be useful for evaluation of CI performance. However, little is known about the evolution of spectral ripple discrimination (SRD) over time after the implantation and its relationship with speech perception rehabilitation. Here, we evaluate longitudinally the evolution of SRD and speech perception in quiet and noise for adult CI users.

Methods — Nine adult CI users attended research sessions, during their first year of rehabilitation, at dates indicated by the clinic (switch-on, one week, one month, two months, three months, six months, nine months and one year after switch-on). Behavioural SRD thresholds were measured using a two-alternative forced-choice paradigm. Speech perception in quiet and in talker-babble noise (10dB SNR and 5dB SNR) was measured using AzBio sentences. Stimuli were presented unilaterally, at most comfortable level as indicated by the participant and sent directly to the CI speech processor via the auxiliary input.

Results — Repeated measures ANOVA indicates that there is a significant time effect in the evolution of SRD ($F(7,56)=6.65$, p-value<0.0005) and speech perception in quiet and in noise at 10dB SNR ($F(7,16.21)=4.45$, p-value<0.01; $F(6,55.82)=3.16$, p-value$=0.01$). However, speech perception in noise at 5dB SNR showed a slight improvement over time that did not reach statistical significance ($F(7,41.66)=1.54$, p-value>0.1). Post-hoc tests using the Bonferroni correction revealed that there is an increase in SRD at all sessions compared to the switch-on session, however, statistically significant changes only occur two months after implantation and onwards. Statistically significant changes for speech perception occur one year after implantation for the quiet condition and in noise at 5dB SNR. Pearson's correlation analysis revealed that SRD at one week after switch on has a strong and significant correlation with speech perception in quiet ($r=0.878$, p-value$=0.002$) and in noise at 10dB SNR ($r=0.707$, p-value$=0.05$). SRD at one week after switch-on also correlated with speech perception in quiet up to nine months after implantation ($r=0.835$, p-value$=0.038$).

Conclusion — Longitudinal assessment of SRD and speech perception indicated that the SRD progression is faster than speech perception in quiet and in noise for adult CI users. SRD even at one week after switch-on showed promising potential estimating speech perception abilities longitudinally. This opens the prospect of using SRD as an objective metric to estimate future speech perception abilities in CI users.

23 Do speech-in-noise scores in normal-hearing humans correlate with amplitude modulation depth detection abilities?

S. M. Waechter, A. Lopez Valdes
Trinity College Dublin, IE

C. Simoes-Franklin, L. Viani
National Cochlear Implant Programme, Beaumont Hospital, Dublin, IE

R. B. Reilly
Trinity College Dublin, IE

Introduction — Extensive research has been performed to explore human temporal auditory processing abilities and their links to speech perception. This study aims to investigate the influence of temporal processing abilities, in form of amplitude modulation depth (AMD) detection abilities, on speech-in-noise recognition scores.

Methods — Seven young, normal-hearing adults participated in two psychoacoustic (PA1 and PA2) tasks and in a speech-in-noise test. Both PA tasks consisted of amplitude modulated and unmodulated noise presented monaurally to the left ear at 65dB SPL. The modulation frequency was set to 8Hz and the AMD was varied.

The PA1 paradigm determined behavioural AMD detection thresholds with a three-alternative forced-choice two-down/one-up task. The PA2 paradigm probed the AM detection ability for specific AMDs (10%-100%) and monitored the percentage of correct responses. Stimuli were presented individually and the participant determined whether it was modulated or unmodulated.

AzBio sentences were tested at three SNR levels (10dB, 5dB and 0dB) with respect to ten talker-babble noise. The percentage of correctly identified words was recorded as the speech-in-noise score.

Results — The PA1 AMD detection threshold indicating 70.7% correct responses ranges between 8.9% and 15.1% (mean of 12.49%) revealing overall similar discrimination abilities for this group.

Results for the PA2 task show high-pass characteristics with the drop-off located below 25% AMD. A one-way ANOVA revealed statistically significant differences between AMD detection scores ($F(5,54)=37.78$, $p<0.001$). Post-hoc Tukey comparisons determined that the 10% and 12.5% AMD conditions were significantly different from all other conditions (25%-100%) and from each other.

A one-way ANOVA revealed significant differences for the different SNR level speech test scores ($F(2,33)=31.20$, $p<0.001$). Following Tukey comparisons revealed that the 0dB SNR level is significantly different to the other conditions. Furthermore, a significant correlation was found between speech test scores at 0dB and PA2 scores for the 12.5% AMD level ($r2=0.9077$, $p=0.0047$), but not for the 10% level ($r2=0.1626$, $p=0.7275$) or the PA1 AMD detection thresholds ($r2=0.0323$, $p=0.9440$).

Conclusion — Speech-in-noise recognition scores correlate well with the ability to discriminate modulated and unmodulated sounds at 12.5% modulation depth, which also marks the average AMD detection threshold for this group at 70.7% correctness level. However, this finding has to be corroborated with an increased sample size. If validated, this method could be applied to evaluate or forecast speech recognition ability in clinical populations such as cochlear implant users.

24 Disentangling effects of aging and hearing loss on speech perception in different background noises

T. Goossens, C. Vercammen, J. Wouters, A. van Wieringen
University of Leuven, Department of Neurosciences, Research Group Experimental ORL, BE

Speech perception problems are highly prevalent in the adult population, especially in the presence of background noise which is common in daily conversations. This well-known communication problem originates from a combination of factors that are closely intertwined: aging, peripheral hearing loss, temporal processing deficiencies and declining cognition. The extent to which speech perception is degraded depends on the type of background noise as well. Informational maskers are considered confusing and therefore more difficult to deal with than energetic maskers. Amplitude modulated energetic maskers show temporary increases in the signal-to-noise ratio (noise dips) that facilitate speech perception. However, accurate temporal processing is required to achieve release from masking by the presence of noise dips.

This study includes speech identification tasks with three types of background noise: steady-state and amplitude modulated speech weighted noise (energetic maskers) and the International Speech Test Signal (informational masker). Three age groups are included in the study: young (20-30 yrs.), middle-aged (50-60 yrs.) and older adults (70-80 yrs.). In each age group, persons with clinically normal audiometric thresholds as well as persons with peripheral hearing loss are included. All participants passed a cognitive screening to prevent differences in cognition from confounding the results.

This careful participant selection enables us to disentangle the complex interplay of aging and peripheral hearing loss with regard to speech perception abilities. Effects of aging are investigated by comparing speech perception performance across the normal hearing age groups whereas effects of hearing loss are mapped by examining the normal hearing and hearing impaired age-matched adults. We examine to what extent effects of aging and hearing loss differ in the three types of background noise. In addition, we investigate release from masking. By this, we gain insight into the temporal processing efficiency of the different subgroups and whether this auditory processing ability is related the speech perception performance in the different types of background noise.

25 Large-scale evaluation of the Digit Triplet Test for school-age hearing screening of 5th & 9th grade school children

S. Denys, H. Luts
University of Leuven, Department of Neurosciences, Research Group Oto-Rhino-Laryngology, Leuven, BE

C. Guérin
University of Leuven, Department of Public Health and Primary Care, Environment and Health, Leuven, BE

K. Van Hoeck, A. Keymeulen
Flemish Scientific Society for Youth Health Care, Leuven, BE

M. Hofmann, A. van Wieringen
University of Leuven, Department of Neurosciences, Research Group Oto-Rhino-Laryngology, Leuven, BE

K. Hoppenbrouwers
University of Leuven, Department of Public Health and Primary Care, Environment and Health, Leuven, BE

J. Wouters
University of Leuven, Department of Neurosciences, Research Group Oto-Rhino-Laryngology, Leuven, BE

Given that some types of hearing impairment (HI) are only acquired in the early years of life, and given the well-known negative consequences of (even minimal) HI, accurate hearing screenings at a time when adequate hearing is of educational importance are needed as a follow-up to the newborn hearing assessment. Unfortunately, pure-tone screening (PTS) protocols — which are most commonly used — lack accuracy in a school screening environment. Better screening protocols need to be developed.

In Flanders, school-age hearing screenings (SHS) are mandatorily organized by regional school health services in 1st graders (5-6 yrs), 5th graders (10-11 yrs) and 9th graders (14-15 yrs). In these last two cohorts, PTS has recently been replaced by a speech-in-noise (SPIN) screening protocol using the Flemish Digit Triplet Test (DTT). This automated self-test has a very high sensitivity and specificity (> 90%) to detect mild HI in adults, as well as a high test-retest accuracy. Furthermore, it has been shown that bilateral speech reception thresholds (SRT) can be obtained within 4-6 minutes. In the current study, large-scale applicability for SHS was investigated and age-appropriate pass/fail-criteria were determined.

Eleven school health service centers participated. Children who failed the test, based on preset cut-off SRTs of -7.2 dB (5th graders) and -8.5 dB (9th graders), were referred for a diagnostic evaluation with a full audiogram assessment. We aimed for a referral rate of 5% to receive sufficient audiograms enabling pass/fail-criteria to be fine-tuned.

Out of the 6499 children tested (3304 5th graders, 3195 9th graders), 4.1 and 3.6% failed the test, respectively. We obtained audiograms of half of the referred children (48% in 5th graders, 52% in 9th graders). Elevated audiometric thresholds were confirmed in only 1/5 (5th graders) and 1/3 (9th graders) of the children. Because of this low positive predictive value and relatively high referral rate, pass/fail-criteria of the DTT must be tuned less strict. When a fail is only registered in case of an SRT being 2.5 interquartile ranges above the median group-SRT (-9.7 dB in 5th graders, -10.5 in 9th graders), referral rates drop to 2.6 and 3%, respectively, whilst positive predictive values rise by 10 and 14%.

Effective data registration will enable monitoring of the SRT over the school years. Furthermore, this will enable to track the performance of the screening program. By collecting epidemiological data, we will be able to estimate the prevalence of acquired HI in school children.

26 Cortical evoked potentials of speech in noise for normal hearing listeners

E. Brint, P. Iverson
University College London, GB

Normal hearing listeners have a wide range of speech recognition outcomes in noise, and it is not clear why this variation occurs. In order to investigate the cause of this variation between listeners we need to use objective tests that can predict speech in noise performance. This study evaluates cortical evoked potentials for this purpose, more specifically, the acoustic change complex (ACC), which is an event related potential (ERP) that reflects cortical auditory processing. The aim is to see if speech perception performance in noise is reflected in the ACC. Thirty normal hearing listeners (aged 19-38 years) were tested on their speech in noise abilities and their ACC responses to random continuous sequences of vowels and fricatives at three different noise levels. The ACC stimuli consisted of four vowels, four fricatives and silence, each lasting 300-400 ms, in quiet, at -3 dB SNR and +4 dB SNR. Participants listened to 20 minutes of these sequences at each noise level whilst their ACC was measured using a Biosemi EEG system. Their speech in noise abilities were assessed using a sentence recognition task at -3, -6 and -9 dB SNR. Preliminary results show that with increasing noise level the amplitude of the ACC decreases and the latency increases. Furthermore, the latencies of the P1 and N1 peaks decrease with increasing speech in noise ability for certain stimulus pairs at +4 dB SNR. It therefore appears that the ACC response has the potential to be used to predict speech in noise performance in normal hearing listeners.

27 Your eyes and brain reveal your hearing ability

A. A. Zekveld
Section Ear & Hearing, Dept. of Otolaryngology-Head and Neck Surgery and EMGO Institute, VU University Medical Center, Amsterdam, NL

D. J. Heslenfeld
Department of Psychology, VU University, Amsterdam, NL

N. J. Versfeld, S. E. Kramer
Section Ear & Hearing, Dept. of Otolaryngology-Head and Neck Surgery and EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, NL

A relevant aspect of listening is the degree of effort required to understand the message. We assessed the neural correlates of the pupil dilation response, a measure of listening effort, in two samples of individually age-matched and educational-level matched participants. One group (n=17) included listeners (M age 45.9 years) with normal hearing; the other group comprised listeners (n = 17; M age 45.4 years) with sensorineural hearing loss (mean PTA 46.9 dB HL). Participants repeated sentences that were degraded by noise-vocoding the speech, by imposing stationary noise, or by imposing interfering speech. Speech Reception Thresholds (SRTs) aiming at 50% intelligibility were adaptively estimated, pupil dilation responses were assessed and functional magnetic resonance imaging (fMRI) data were acquired.

As expected, hearing impairment was associated with relatively poor SRTs, especially for interfering speech. Brain activation and pupil dilation responses were largest for this condition. Effects of degradation type were observed in bilateral superior temporal and superior and middle frontal gyri, bilateral precentral gyrus, and left inferior frontal gyrus. Pupil dilation responses of the hearing impaired participants tended to be smaller and a frontal area tended to be less active than in the normal hearing controls. Activation in a variety of frontal, temporal and medial brain regions was associated with the pupil response. Part of these regions reflected condition-specific effects.

The results extend previous findings showing the effect of hearing acuity and speech degradation on speech perception and the pupil and brain responses during listening. We will discuss the relevant processes underlying the results.

28 The contribution of auditory attention to processing load in normal hearing and hearing impaired adult listeners

T. Koelewijn, H. de Kluiver, A. Zekveld
Section Ear & Hearing, department of Otolaryngology-Head and Neck Surgery and EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, NL

B. G. Shinn-Cunningham
Department of Biomedical Engineering, Center for Computational Neuroscience and Neural Technology, Boston University, Boston, US

N. Versfeld, S. E. Kramer
Section Ear & Hearing, department of Otolaryngology-Head and Neck Surgery and EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, NL

A major complaint of people with hearing impairment is the high level of processing load they experience (listening effort) when having a conversation in a noisy environment. From previous studies we know that the pupil dilation response, an objective measure of listening effort, is affected by working memory capacity and linguistic abilities. These studies also suggested the involvement of attention-related processes, but it is unclear how these processes work.

In this project we investigate the effect of auditory attention on processing load in normally hearing and hearing-impaired listeners. Research shows that knowledge about where, when, and whom is going to talk benefits speech perception. We suggest that this knowledge allows a listener to better focus attention, resulting in better filtering of irrelevant information. This should lead to less use of working memory capacity and linguistic resources, and therefore, less processing load.

In the first study we investigated the effect of focused versus divided attention on processing load. In this study 12 normal hearing young adults (21 to 26 years) had to focus on either one or both of two sentences that were presented dichotically and masked by fluctuating noise. They showed a larger pupil dilation response, indicating higher processing load or listening effort, when processing two sentences simultaneously instead of one.

In the second study 56 normal hearing young adults (18 to 28 years) participated in three experiments. Dichotic tasks similar to those in the first study were used. In these experiments target location (left or right ear), speech onset, and talker variability were manipulated by keeping these features either fixed during an entire block of sentences or by randomizing these over trials. The results showed a performance benefit when they were able to focus on the target location. Additionally, a decrease in pupil dilation response was observed when they were able to focus on location or talker.

In a third study we examined the effect of focused attention and location information on processing load in adults with a moderate sensorineural hearing impairment. Preliminary results will be presented.

Based on the results of all studies we conclude that communicating in a cocktail party like environment requires substantial processing load because of the demands placed on attentional processes.

29 Effects of masker type on spatial release from masking in acoustic simulations of bilateral cochlear implants

T. Green
University College London, GB

Bilateral cochlear implant (CI) users show reduced benefit from spatial separation of target speech and background noise compared to normal hearing listeners. This is consistent with the fact that the availability of important spatial cues is severely limited by characteristics of typical CI speech processing strategies, such as the elimination of temporal fine structure (TFS). Additionally, the ability to use those spatial cues that survive CI speech processing may be limited because neural processing of binaural information has been degraded due to long-term deprivation of auditory input. The use of acoustic simulations of CI processing may help to isolate limitations due to CI speech processing from those associated with auditory deprivation. However, previous simulation studies have provided conflicting evidence. Garadat et al. (2009) observed differences in speech reception thresholds (SRTs) of up to around 8 dB due to spatial separation of target and masker, suggesting that even after elimination of binaural TFS cues, sufficient interaural envelope cues remained to allow substantial spatial release from masking (SRM). In contrast, Schoof et al. (2013) found only very limited evidence of SRM. The present work assessed the role of differences in target material and type of masker in this discrepancy. SRTs were measured for male target speech (spondees or BKB sentences) masked with either 20-talker babble or a single competing male talker. Spatial separation of speech and masker was simulated by applying a spherical head model prior to 6-channel noise-excited vocoding. Target speech was always presented with 0° azimuth. Masker azimuth was either 0° or 90°. SRM was significantly affected by masker type but not by target material. Mean differences in SRT due to masker azimuth were substantial for single-talker maskers (3.4 dB for BKBs, 6.7 dB for spondees) but were only around 1 dB in babble for both types of material. Differences in masker type therefore appear to underlie the contrast in previous findings, perhaps reflecting greater potential for SRM in the presence of informational masking.

30 Impact of stimulus-related and listener-related factors on cognitive processing load as indicated by pupil dilation

B. Ohlenforst[1,2], A. Zekveld[2,3,4], T. Lunner[1,5,4], D. Wendt[1,6], G. Naylor[7], Y. Wang[1,2], S. Kramer[2]

1. Eriksholm Research Center, Oticon A/S, Snekkersten, DK

2. VU University Medical Center and EMGO Institute for Health Care Research, Amsterdam, NL

3. Department of Behavioral Sciences and Learning, Linköping University, Linköping, SE

4. Linnaeus Centre HEAD, The Swedish Institute for Disability Research, Linköping and Örebro Universities, Linköping, SE

5. Department of Clinical and Experimental Medicine, Linköping University, SE

6. Hearing Systems group, Department of Electrical Engineering, Technical University of Denmark, Lyngby, DK

7. Institude of Hearing Research Scottish Section, Glasgow, GB

This study investigated the influence of stimulus-related factors such as speech intelligibility level and masker types and listener-related factors, including cognitive abilities and hearing status, on the pupil dilation during speech perception. Pupil diameter was recorded during the speech recognition task across a wide range of intelligibility levels in hearing-impaired (N=29) and normal-hearing (N=34) listeners. Sentences were presented in quiet, in a stationary masker, and in the presence of a competing talker. Two aged-matched groups of listeners, one with symmetrical, sensorineural, mild-to-moderate hearing loss and one with normal-hearing thresholds, participated in this study. Hearing-impaired listeners received frequency-specific amplification to compensate for reduced audibility. We measured individual working memory capacity and linguistic abilities. Final results will be presented at the conference as this study is still ongoing. Based on findings from previous research, an interaction between speech-intelligibly, masker types, hearing status and peak pupil dilation relative to baseline (PPD) is expected (Kramer et al., 1997; Zekveld et al., 2011). A systematical increase in PPD with decreasing intelligibility is hypothesized. We expect hearing-impaired listeners to show larger pupil dilations for speech recognition in the interfering talker condition compared to the stationary noise background (such as Koelewijn et al., 2014). Moreover, a smaller task-evoked PPD is expected to be found for the hearing-impaired listeners compared to listeners with normal-hearing and for listeners with poorer cognitive abilities compared to those with better cognitive skills (as proposed by Grady, 2012; Zekveld et al., 2011). The results of this study will help to better understand how stimulus-related factors, such as speech intelligibility level and masker type, and listener-related factors, like cognitive skills and hearing abilities, influence cognitive processing load as indexed by the PPD.

31 Predicting the intelligibility of noisy and enhanced binaural speech

A. H. Andersen
Oticon / Aalborg University, Smørum, DK

J. M. de Haan
Oticon, Smørum, DK

Z. H. Tan
Aalborg University, DK

J. Jensen
Oticon / Aalborg University, Smørum, DK

Modern hearing aids make use of increasingly sophisticated signal enhancement techniques such as noise reduction, de-reverberation and binaural beamforming. While the availability of such techniques allows for increasing the performance of hearing aids, it also complicates the task of designing them by introducing many additional design decisions, each of which has to be investigated and validated. Such validations are often done through speech intelligibility tests with many human subjects. These are time consuming and expensive to carry out. Such considerations have recently led to an increasing interest in the application of objective speech intelligibility measures in the context of hearing aids. If such measures can be used to predict the advantage of different signal processing techniques, the required number of speech intelligibility tests can be greatly reduced. For a speech intelligibility measure to be applicable for evaluation of hearing aids, it must be able to take account of the factors which influence the intelligibility of a hearing aid user. In particular, this includes the influence of 1) non-linear processing as carried out by the hearing aid and 2) binaural advantage from spatial separation between target and masker (which must be expected to depend on hearing aid processing as well). We propose and evaluate a binaural intelligibility measure for noisy and enhanced speech. The measure is based on combining the short term objective intelligibility (STOI) measure with an equalization cancellation (EC) stage. The STOI measure is a monaural intelligibility measure, which has been shown to predict well the influence of several speech enhancement techniques. The EC stage is a simple model of how the brain obtains a binaural advantage. By combining the EC stage and the monaural STOI measure, we obtain a measure which is theoretically capable of predicting the full effect, on a normal hearing subject, of wearing a set of hearing aids. We evaluate the proposed measure against a range of listening tests and show that the measure can accurately predict 1) the binaural advantage of spatial separation between a frontal target and an additive noise source in the horizontal plane, 2) the impact on binaural advantage obtained by processing the left and right ear signals independently with ideal binary masks (IBMs) and 3) the impact of wearing a hearing aid with beamforming in conditions with multiple interferers.

32 The role of working memory capacity in processing demand during speech comprehension

D. Wendt, T. Dau, J. Hjortkjær
Hearing Systems group, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, DK

The relationship between working memory capacity (WMC) and processing demands was investigated in a group of listeners with normal hearing. Processing demands were measured with either subjective ratings of perceived effort, or with pupil dilations as a physiological correlate of processing effort. Both measures were tested in an audio-visual picture matching paradigm, where the participant's task was to match a spoken sentence with a picture that was presented prior to the sentence. The paradigm was tested using either syntactically simple or complex sentence structures. Sentences were presented at low and high noise levels, where speech intelligibility was still high. It was found that a higher WMC of the participants was correlated with an increase in pupil dilation. This finding was consistent with the resource hypothesis stating that people with higher WMC allocate more available resources that are utilized in the task (e.g. van der Meer et al., 2010; Zekveld et al., 2010). Furthermore, it was found that those participants with a larger WMC reported speech comprehension to be less effortful than participants with a lower WMC. This finding is in line with the ease of language understanding (ELU) model, which predicts that people with more cognitive resources require lower processing demands to achieve similar or better performance than people with reduced resources (Rönnberg et al. 2003; 2008). Our two findings indicate that WMC is in different ways related to perceived effort and processing effort, respectively. The two metrics represent different, potentially independent, components of cognitive processing demands.

33 Factors influencing word recognition for self- and other-produced speech in noise

W. L. Schuerman, A. Meyer
Max Planck Institute for Psycholinguistics, Nijmegen, NL

J. M. McQueen
Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, NL

Sensorimotor experience modulates perception; expert basketball players have been found to be more accurate at predicting shots than expert watchers (Aglioti, Cesari, Romani, Urgesi, 2008). The common coding hypothesis (Prinz, 1990) argues that this indicates parity between the representations accessed for production and perception, and predicts that actions that match our own sensorimotor experience will be more accurately recognized. This is supported by research demonstrating that people are more accurate when lip-reading their own speech compared to that of others (Tye-Murray et al., 2013).

In this ongoing study, we investigate how phonetic, lexical, and indexical factors may modulate how talkers perceive their own speech and that of others. Groups of seven gender-matched participants produce Dutch sentences containing two semantically unrelated words (e.g. “de kwal is boven de deur” -- “the jellyfish is above the door”). The sentences are prompted by serial presentation of the orthographic labels of two objects, followed by visual display of the referenced objects in a simple spatial configuration. Items consist of 112 words, sorted into four categories based on phonological neighborhood density (PND) and word frequency (high vs. low). Each item appears in first and second position in three sentence lists, totaling 336 sentences.

In sessions spaced approximately one to two weeks apart, the same participants attempt to recognize these sentences, evenly divided amongst the seven talkers, under three conditions: 1) 6-band noise-vocoded speech (NVS), 2) sentences embedded in speech-shaped noise at a ratio of -7 dB (SPIN), and 3) sentences filtered to approximate the signal generated by a combination of air and bone conduction and then embedded in -7dB noise (FilSPIN). Order of SPIN-FilSPIN sessions was counterbalanced across participants. For half of the participants, each sentence is preceded by a talker-label, consisting of either a common Dutch name or “jij” (“you”). These No-Label participants are not informed about the number of talkers or that some of the stimuli were based on their own recordings.

Preliminary results suggest that overall accuracy is greater for self-produced stimuli. This self-advantage is greater in the SPIN and FilSPIN sessions compared to the NVS session. However, the proportion of accurate responses appears to be modulated by the presence/absence of a label, whether or not the sentences had been filtered, and word frequency/PND. Taken together, these findings suggest that sensorimotor experience and lexical properties may interact to shape representations and that indexical cues guide access to different representations.

34 Visible speech enhanced: What do gestures and lip movements contribute to degraded speech comprehension?

L. Drijvers
Centre for Language Studies & Donders Institute, Radboud University, Nijmegen, NL

A. Ozyurek
Centre for Language Studies, Radboud University, Nijmegen, NL

Face-to-face communication involves an audiovisual binding that integrates information from multiple inputs such as speech, lip movements, and iconic gestures. Previous research showed that these kinds of visual inputs can enhance speech comprehension, especially in adverse listening conditions. However, the contribution of lip movements and iconic gestures to understanding speech in noise has been mostly studied separately. The current study aimed to investigate the contribution of both types of visual information to degraded speech comprehension in a joint context.

In Experiment 1, we investigated the contribution of iconic gestures and lip movements to degraded speech comprehension in four auditory conditions (clear speech, 16-band, 10-band and 6-band noise-vocoding) to determine the noise level where these visual inputs enhance degraded speech comprehension the most. Participants were presented with video clips (speech/lips or speech/lips/gesture), of an actress uttering a Dutch action verb, followed by a cued-recall task. This cued-recall task included the target verb, a semantic competitor, a phonological competitor and an unrelated distractor. Over all noise-vocoding levels, visual input significantly enhanced degraded speech comprehension. This enhancement was largest at 6-band noise-vocoding, as indicated by the largest difference in response accuracy in speech/lips/gesture vs. speech/lips trials. In addition, the error analyses revealed that information from lip movements was used for phonological disambiguation, whereas gestural information was used for semantic disambiguation.

In Experiment 2, we investigated the individual contributions of lip movements and iconic gestures to this audiovisual enhancement. Participants watched videos in 3 speech conditions (2-band noise-vocoding, 6-band noise-vocoding, clear speech), 3 visual conditions (speech/lips blurred, speech/lips visible, speech/lips/gesture) and 2 non-audio conditions (lips only and lips/gesture), to understand how much information participants could get from visual input alone. Participants showed significantly higher response accuracy for speech/lips/gesture conditions over speech/lips and speech/lips blurred conditions, over all noise-vocoding levels, and in lips/gesture videos compared to lips only. Additionally, the difference between speech/lips/gesture and speech/lips was significantly larger for 6-band noise-vocoding compared to 2-band noise-vocoding and compared to the difference between the two non-audio conditions. However, there was no difference between the two non-audio conditions vs. speech/lips/gesture and sleeps/lips at a 2-band noise-vocoding level.

Our results indicate that when degraded speech is processed in a visual context, listeners benefit significantly more from gestural information than from just lip movements alone, especially when auditory cues are moderately reliable (6-band noise-vocoding) compared to listening situations where auditory cues are no longer reliable (2-band noise-vocoding).

36 Investigating the effects of noise on arousal during a communication task

N. Cvijanovic, P. Kechichian, K. Janse
Philips Research Laboratories Eindhoven, NL

A. Kohlrausch
Philips Research Laboratories Eindhoven, NL
TU Eindhoven, NL

The performance of speech communication systems depends highly on the acoustic characteristics of the environment in which they are being used. In a natural environment, noise is inevitable and ubiquitous and has detrimental effects on communication. During a conversation, it can not only reduce speech intelligibility and lead to an increase in listening effort but also cause annoyance and stress. A lot of work has been invested in the past in methods reducing ambient noise without degrading the speech signal. Various solutions exist, and the benefit of such noise reduction algorithms in communication systems is usually quantified using metrics like the signal-to-noise ratio (SNR) or the perceptual evaluation of speech quality (PESQ). Through our study we want to examine the possibility of evaluations based on physiological changes in users of a communications system.

The aim of our study is to investigate the effects of noise on arousal. For this, we examine physiological changes in speakers in a communication setting for different levels of background noise. In our communication setting, two speakers are seated in separate rooms. They communicate with each other using headsets and microphones on the tables in front of them and try to solve collaborative tasks in different noise conditions. The goal of this setup is to emulate a hands-free communication scenario (e.g., a Skype conversation). Each pair of speakers solves the same types of tasks in low, moderate and high noise conditions with relaxation periods in between. Throughout the conversation, we measure and record the heart rate and skin conductance of each speaker. These physiological signals are later analyzed and extracted features are used to determine arousal states throughout the experiment.

Here, we explain our design, discuss the relevant physiological signals and their characteristics and show some preliminary results and insights. On the long run, quantifying the effects of noise on arousal could enable us to determine the benefits of a noise reduction algorithm on arousal and thus well-being of the users of speech communication systems.

37 Recording and evaluation of a multiple-talker version of the WAKO rhyme test

C. Lesimple, J. Tantau, B. Simon
Bernafon AG, Bern, CH

Human speech recognition shows great robustness despite a wide range of speech production variability including talkers, speaking rates, and dialects. Word recognition accuracy decreases and response latency increases when this variability is introduced in word recognition tests compared to a single talker condition (Mullennix et al., 1988; Kirk et al., 1997). The use of a single professional speaker within a commercially available speech test underestimates the real difficulties met in daily life by hearing impaired listeners and limits the ability to generalize any findings on speech perception in general (Clark, 1973).

A multi talker version of the WAKO rhyme test (von Wallenberg & Kollmeier, 1989) was recorded by four non-professional talkers (two male and two female native German speakers with either German or Swiss accent) for this experiment. The quality of the recordings was first evaluated by ten normal hearing subjects.

Nine hearing aid wearers were tested in aided and unaided conditions with the original and the newly recorded versions of the WAKO test. Word recognition, response time and subjective sound quality ratings were simultaneously measured for each test condition. Statistical analysis used a logistic mixed effect regression model on the word recognition scores and a linear mixed effect regression model on the response times and the sound quality ratings. Test condition and test version were treated as fixed effects and their p-values were obtained by likelihood ratio tests comparing the full model against the model without the effect. Introducing variability in the test material led to a significant decrease in performance for word recognition accuracy, response times and sound quality ratings. These results suggest that using speech materials closer to that experienced in daily life makes word recognition more difficult compared to material intended for clinical use. Applications of these findings are suggested for the use of test designs for researchers who have difficulties interpreting results due to celling effects.

38 Pure linguistic interference during comprehension of competing speech signals

B. Dai
Max Planck Institute for Psycholinguistics, Nijmegen, NL

A. Kösem
Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, NL

J. M. McQueen
Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, NL
Max Planck Institute for Psycholinguistics, Nijmegen, NL

P. Hagoort
Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, NL
Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6500 HB Nijmegen, NL

Human listeners have more difficulty in understanding speech in a multi-talker environment than in the presence of non-intelligible noise. The costs of speech-in-speech masking have been attributed to informational masking, i.e. to the competing processing of the target and the distractor speech's information. Yet, it remains unclear what kind of information is competing, as intelligible speech and unintelligible speech-like signals (e.g. reversed, noise-vocoded, and foreign speech) differ both in linguistic content and in acoustic information. Thus, intelligible speech could be a stronger distractor than unintelligible speech because it presents closer acoustic information to the target speech, or because it carries competing linguistic information. In this study, we intended to isolate the linguistic component of speech-in-speech masking and we tested its influence on the comprehension of target speech.

To do so, participants performed a dichotic listening task in which the interfering stimuli consisted of noise-vocoded sentences that could become intelligible through training. The experiment included three steps: first, the participants were instructed to report the clear target speech from a mixture of one clear speech channel and one unintelligible noise-vocoded speech channel; second, they were trained on the interfering noise-vocoded sentences so that they became intelligible; third, they performed the dichotic listening task again. Crucially, before and after training, the distractor speech had the same acoustic features but not the same linguistic information. We thus predicted that the distracting noise-vocoded signal would interfere more with target speech comprehension after training than before training. To control for practice/fatigue effects, we used additional noise-vocoded sentences that participants were not trained on, as interfering signals in the dichotic listening tasks. We expected that performance on these trials would not change after training, or would change less than that on trials with trained noise-vocoded sentences.

The first results are consistent with our predictions. The trained noise-vocoded speech interfered more with target speech intelligibility after training than before training, but only for low SNR (-6 dB). Crucially, the interference was significantly stronger for trained noise-vocoded sentences than for untrained noise-vocoded speech, ruling out a fatigue effect. In line with past reports, the present results show that intelligible distractors interfere more with the processing of target speech. These findings further suggest that speech-in-speech interference originates, to a certain extent, from the parallel processing of competing linguistic content.

39 Acoustic analysis of communication disorders within Moroccan students

B. Sabir, M. Moussetad, B. Touri
University Hassan II Mohammedia, Faculty of science Ben M'Sik, Physics department, Casablanca, MA

Objective — Communication disorders negatively affect the academic curriculum for students in higher education. Acoustic analysis is an objective leading tool to describe these disorders; however the amount of the acoustic parameters makes differentiating pathological voices among healthy ones not an easy task. The purpose of the present paper was to present the relevant acoustic parameters that differentiate objectively pathological voices among healthy ones.

Methods — Pathological and normal voices samples of /a/, /i/ and /u/ utterances, of 400 students were recorded and analyzed acoustically with PRAAT software, then a feature of acoustic parameters were extracted. A statistical analysis was performed in order to reduce the extracted parameters to main relevant ones in order to build a model that will be the basis for the objective diagnostic.

Results — Mean amplitude, jitter local absolute, second bandwidth of the second formant and Harmonic-to-Noise Ratio (HNR) are relevant acoustic parameters that characterize pathological voices among healthy ones, for the utterances of vowels /a/, /i/ and /u/. Thresholds of the acoustic parameters of pathological /a/, /i/, and /u/ were calculated. A training model was built and simulated on Matlab, and a comparison between HMM (Hidden Markov Model) and KNN (K-Nearest Neighbors) classification methods were done (HMM had a rate of recognition of 95% and KNN within the reduced acoustic parameters reached a recognition rate of 97%).

Conclusion — Through the identified parameters, we can objectively detect pathological voices among healthy ones for diagnostic purposes. As a future work, the present approach is an attempt toward identifying acoustic parameters that characterize each voice disorder.

40 Prosody and semantics are separate but not separable channels in the perception of emotional speech: test of rating of emotions in speech (T-RES)

B. M. Ben-David
Interdisciplinary Center (IDC), Herzliya, IL

Our aim is to explore the complex interplay of prosody (tone of speech) and semantics (verbal content) in the perception of discrete emotions in speech. We implement a novel tool, the Test for Rating of Emotions in Speech (T-RES). Eighty native English speakers were presented with spoken sentences made of different combinations of five discrete emotions (anger, fear, happiness, sadness, and neutral) presented in the prosody and semantics. Listeners were asked to rate the sentence as a whole, integrating both speech channels, or to focus on one channel only (prosody/semantics).

Our results show three main trends: Supremacy of congruency — a sentence that presents the same emotion in both speech channels was rated highest; Failure of selective attention — listeners were unable to selectively attend to one channel when instructed; Prosodic dominance — prosodic information plays a larger role than semantics in processing emotional speech.

Conclusions: Emotional prosody and semantics are separate but not separable channels and it is difficult to perceive one without the influence of the other. The findings indicate that T-RES can reveal specific aspects in the processing of emotional speech and in future use may prove useful for understanding emotion processing deficits in pathological populations.

41 The time course of steam segregation in nonnative listeners

B. M. Ben-David, **M. Avivi-Reich**
Interdisciplinary Center (IDC), Herzliya, IL

Younger nonnative speakers with good hearing find it more difficult to comprehend speech than do younger native speakers of the language, especially in noisy situations. Older native speakers of the language also appear to find it difficult to understand speech in their native language in noisy environments, although it is quite likely that the reasons for these difficulties differ from those responsible for the problems younger nonnative listeners experience. In a previous study (Ben-David, Tse & Schneider, 2012), even though Speech Recognition Thresholds (SRTs) were, on average, higher (worse) for older adults than for younger adults, both groups benefited from a delay between the onset of a noise masker and the presentation of the speech target, with SRTs improving as the onset delay increased. However, when the masker was a babble of voices only the younger listeners benefitted from an onset delay between the babble and the speech target. In the current study, we used the same task with two groups of 30 younger nonnative English speakers: recent arrivals to Canada, and long-term residents of Canada. The results showed that the speech recognition thresholds for all younger listeners improved at the same rate regardless of their linguistic status as a function of the delay between the onset of the target speech and the onset of both noise and babble maskers. When the masker was babble, speech recognition thresholds were, on average, lower (better) for both groups of younger nonnative speakers than for older native speakers. But, with a noise masker, the function relating thresholds to onset delay was higher (worse) for recent arrivals (younger adults who arrived in Canada after the age of 15) than for older native speakers. The relevance of these findings to age-related changes in auditory scene analysis will be discussed.

42 Effects of exposure to noise during second language consonant acquisition

M. Cooke, M. L. Garcia Lecumberri
University of the Basque Country, ES

The process of acquiring the sounds of a first language takes place in everyday settings that are potentially noisy and uncontrolled, yet most formal training in a second language is based on exposure to exemplars heard in quiet conditions in the language laboratory, leading to the question: could sound acquisition in noise be beneficial? Exposure to noisy tokens might help in one of two ways: by forcing a listener to focus on robust cues that on average survive masking, or through the formation of noisy exemplars. On the other hand, it could be harmful, by masking relevant information on some trials, or increasing attentional load and fatigue.

The current study addressed these issues in the context of English consonant acquisition by Spanish learners. A cohort of 86 learners undertook a pre-test involving 24-alternative forced choice consonant identification in quiet and in two masked conditions (SSN: speech shaped noise; BAB: 8-talker babble). The cohort was then divided into 4 groups. One (CON-NOISE) underwent consonant in noise training; another (CONS-QUIET) was exposed to the same consonants in quiet; a third (VOW-NOISE) heard vowels in noise; a fourth (VOW-QUIET) heard vowels in quiet. The two noise-trained groups heard exemplars in SSN at a range of SNRs. Each group underwent 10 training sessions over 5 weeks, after which they performed a post-test identical to the pre-test. The vowel-trained groups acted as a control to measure any benefits from simultaneous curriculum activities on consonant acquisition.

All groups showed gains from pre- to post-test. However, gains for vowel-trained groups were 2-4 p.p., while those for the consonant groups ranged from 10 to 14 p.p. suggesting only a small effect of other activities. The CONS-QUIET group produced larger gains when tested in quiet than the CONS-NOISE group. Conversely, the CONS-NOISE group displayed higher gains than the CONS-QUIET when tested in SSN. The noise advantage did not transfer to the BAB condition, where the CONS-NOISE and CONS-QUIET groups showed near-identical gains. No clear evidence was found of noise exposure during vowel training transferring to the consonants in noise test. These outcomes suggest that exposure to noise during training helps in identifying consonants both in noise and in quiet, with a benefit for matched noise exposure and test conditions and little evidence of a transfer to other maskers (SSN/babble) or sound type (vowels/consonants).

43 Stimulus-brain activity alignment between speech and EEG signals in cochlear implant users, more than an artifact?

A. Wagner, N. Maurits, D. Başkent
University Medical Center Groningen, NL

A number of studies (e.g. Luo & Poeppel, 2007) suggest that synchronization between neural oscillations, as they are observed between different bands of electroencephalographic (EEG) signals, and the temporal amplitude modulations of speech is foundational to speech processing (Giraud & Poeppel, 2012). Speech intelligibility appears to vary with the strength of such alignment between the brain and acoustic signals. These studies are based on vocoded speech materials, and hence stimuli that are similar to the signal transmitted via cochlear implants (CI). EEG studies with CI users suffer from the presence of electrical artifacts induced by the device. This study investigates the phase alignment between EEG signals recorded with CI users and envelopes of natural sentence stimuli, and queries how much such alignment reflects brain signals engaged in speech processing or a CI induced artifact.

EEG signals within the theta range (3-8 Hz) of eight CI users with their own device, and eight normal hearing (NH) participants, recorded while they were listening to naturally spoken sentences, were compared in terms of their alignment with the signal envelopes. The analysis involved a cross-correlation between the envelopes and the EEG channels to extract the lag between the signals. Coherence between aligned signals was then measured in terms of correlation and phase coherence. Preliminary results show for CI users correlations between 0.29 and 0.69, with higher values observed for channels close to the CI but also for contra-lateral temporal and central channels. For NH listeners, correlations were found on left temporal and central channels (0.16-0.33).

The EEG signal lagged behind the speech signal for 120 ms to 380 ms for CI users, and for 200 till 450 ms for NH listeners. The correlation between speech envelopes and signals recorded in their respective trials was generally greater than the correlation found between EEG signals correlated with randomly chosen speech envelopes.

Greater coherence between the speech signal and the channels in vicinity to the CI, together with the absence of coherence for these channels in NH listeners, suggest that signal transmitted via the device is at the source of this alignment. The question of whether greater coherence reported for NH listeners in studies with vocoded stimuli reflects the less natural amplitude modulations in these signals is currently being tested. The coherence found for the central channels for NH and CI listeners, however, suggests that this alignment may indeed reflect a step in speech processing.

44 The development of SpiN recognition by French cochlear implant users

C. J. James, C. Karoui
Cochlear France and CHU Toulouse, FR

M. Marx, C. E. Molinier, B. Lepage, O. Deguine
CHU Toulouse, FR

Understanding the time-course of development of speech recognition in noise in cochlear implant users may allow us to better discover factors limiting performance and ways to improve cochlear implant design and sound coding, and to develop and customise rehabilitation strategies. Patient factors such as the history and pathology of deafness were investigated as well as CI brand and type and position of electrode array.

We discuss longitudinal results for the entire sequence of 136 adults unilaterally implanted in our centre in Toulouse between 2010 and 2014. These cochlear implant (CI) users were followed from the day of activation to 12 months post-activation. French MBAA2 lists of 15 sentences containing 100 words were used for all tests with eight-talker babble noise. At each visit CI users were tested using implant alone with one list in quiet and at fixed SNRs (10, 5, 2.5 and 0 dB). SNRs were reduced until word-in-sentence scores were <50% correct.

A program of post-operative CT imaging was introduced to determine the position of implanted electrode arrays. Complete data-sets were available for sub-group of seventy-eight CI users.

Mean scores in quiet after one month's use of the CI were ~70% correct, and in 10 dB SNR ~55%, with a large range (0 to 100). Sentence scores progressed ~10% points per three months from one to nine months of use. After twelve months most CI users developed high scores in quiet and >50% of users obtained scores >80% correct in 10 dB SNR. Multivariate analysis of the subgroup with CT data showed that implant brand had some effect on scores in quiet and in noise; however one of the three brand sub-groups was small (N=5) and there was strong co-variation of brand with insertion depth. There was no significant difference in scores between CI users due to perimodiolar versus straight design or the presence of scala dislocation. Further step-wise multiple regression revealed cause of deafness to be a significant factor: Subjects with any particular pathology could obtain high scores in quiet whereas those with a history of chronic otitis, Meniere's, and otosclerosis generally had lower scores in noise ($p<0.05$). Those with late onset genetically-related deafness had wide-ranging scores in noise compared to those with deafness caused by meningitis, noise exposure, ototoxicity or trauma (all >75% correct in 5 dB SNR). Some pathologies were limited to certain age groups which may have contributed to some of these differences.

45 Combined effect of reverberation and noise on binaural speech recognition and vocal effort in real classroom acoustics

G. E. Puglisi
Department of Energy, Politecnico di Torino, Turin, IT

A. Warzybok
Cluster of Excellence Hearing4all and Medizinische Physik, Carl von Ossietzky Universität Oldenburg, Oldenburg, DE

A. Astolfi
Department of Energy, Politecnico di Torino, Turin, IT

A. Carullo
Department of Electronics and Telecommunications, Politecnico di Torino, Turin, IT

T. Brand, B. Kollmeier
Cluster of Excellence Hearing4all and Medizinische Physik, Carl von Ossietzky Universität Oldenburg, Oldenburg, DE

Good acoustics in classrooms is necessary to guarantee appropriate teaching and learning practices. The majority of classrooms in Italy does not comply with national or international standards since they take place in historical buildings with big volumes and vaulted ceilings, generating unfavorable environments. In fact, long reverberation times and excessive noise levels have been proved to be factors that negatively influence the speech communication process. However, most of the past studies refer to laboratory conditions with artificial added reverberation and speech-shaped noise.

This work focuses on investigating the influence of reverberation and noise in real classrooms on speech recognition. Speech recognition was measured adaptively converging to signal-to-noise ratio yielding 80% correct recognition scores (SRT80). Since the noise level was fixed at 60 dB SPL, the speech level can be estimated from the measured SRTs and linked to the vocal effort that a teacher would need to maintain to guarantee a good level of speech communication.

Five experiments were designed based on realistic receiver positions in two representative Italian classrooms, one with an acoustical treatment and one without, where room impulse responses were measured at a head and torso simulator ears. Receiver positions in axis with a speech-source were two in the room with good acoustics, at 1.5 m and at 4 m, and three in the room with bad acoustics, at 1.5 m, 4 m and 6.3 m. In each room noise-sources were placed at different distances and azimuths with respect to the receivers to account for binaural cues in the cocktail party phenomenon, namely at 0°, 120° and 180°. Babble noise was measured in real classrooms during a break between lessons, acquiring the noise produced by children moving and talking. The respective impulse responses were convolved with speech and noise signals of Italian matrix test.

Preliminary results show that lower SRT80s were measured for good room acoustics, indicating the detrimental effect of reverberation on speech recognition. A major effect of the reflections was also evident in the increase in SRT80s when the distance between speech-source and receiver increased. The angular separation of the noise-source was evaluated as speech release from masking, resulting up to 4 dB when noise came from 120° instead of 0° or 180° in both room acoustics. Estimations of the speech level will be performed and compared to the indications of ISO 9921, since a high vocal effort can be a risk factor for vocal health.

46 Oscillatory EEG activity – a valid measure of aided listening effort in noise

R. L. Fischer
Sivantos GmbH, Erlangen, DE

P. Schäfer
Saarland University, Faculty of Medicine, Systems Neuroscience & Neurotechnology Unit, DE

M. Serman
Sivantos GmbH, Erlangen, DE

D. J. Strauss
Saarland University, Faculty of Medicine, Systems Neuroscience & Neurotechnology Unit, DE

F. I. Corona-Strauss
Systems Neuroscience & Neurotechnology Unit, Saarland University of Applied Sciences, Saarbrücken, DE

The recognition of a single speaker in a multi-speaker environment is a difficult listening task. To solve that kind of task one has to spend effort, in the following referred to as listening effort (LE). The gain and compression of the hearing instruments (HI) are thought to affect the LE required by the hearing impaired listener in such listening situations. However, there is a lack of a suitable, standardized method to quantify LE.

Consequently, the first goal of this work was to verify the recently proposed objective measure of LE by using electroencephalography (EEG) in a unified framework utilizing the EEG phase-reset hypothesis (Strauss et al., 2013; Bernarding et al., 2014) in a demanding, as well as realistic listening environment. The second goal was to utilize the proposed measure to observe the effect of different HI settings on the exerted LE. To do so, the impact of four different HI settings and two different listening task difficulties (LTD) on the LE of thirty hearing impaired subjects was observed in a selective, real-speech listening task – the Numbers in Babble Paradigm (NIB). NIB task consists of an auditory number comparison task introduced by Wöstmann et al. (2015) that has been embedded into an eight-speaker free-field cafeteria-noise environment. As a result, NIB offers a well-controlled but nonetheless ecologically valid selective listening situation. The participants have to perform an auditory number comparison task masked by a distracting talker while they are sitting in a multi-speaker environment.

HI setting A, B and C all had an adaptive compression with static characteristic, but differed in gain and compression. Setting D had an adaptive and therefore situation-dependent compression. For the quantification of LE the on-going oscillatory EEG activity has been recorded. Based on those recordings, the proposed objective measure was calculated, i.e. the Objective Listening Effort based on oscillatory EEG data (OLEosc). For comparison, the subjects also performed a subjective LE rating on a seven-point scale. Additionally, response time, decision accuracy and decision confidence were recorded as well. The results show that the OLEosc correlates with LTD, as well as with the response time, decision accuracy and processing time, in all tested conditions. These findings lead us to the assumption that the applied objective measure is a good indicator for LE. Furthermore, the results also suggest that OLEosc might be more sensitive to small variances in LE than the subjective LE rating scale.

47 The influence of aided speech in noise performance on hearing aid setting preference in hearing impaired listeners

M. Serman
Sivantos GmbH, Erlangen, DE

P. Schäfer
Systems Neuroscience and Neurotechnology Unit, Neurocenter, Faculty of Medicine, Saarland University, Homburg/Saar, DE

R. L. Fischer
Sivantos GmbH, Erlangen, DE

F. Corona-Strauss, D. Strauss
Systems Neuroscience and Neurotechnology Unit, Neurocenter, Faculty of Medicine, Saarland University, Homburg/Saar, DE

The fitting of a new hearing instrument (HI) will almost invariably take place in a speech in quiet (SiQ) situation. An acoustician and the hearing impaired patient may engage in a short conversation as to the patient's preferred HI settings. This may result in an individual adjustment of gain/compression and adaptive HI features for this patient. Unsurprisingly, it is quite common for the hearing impaired patient to come back and report difficulties in real-life speech in noise (SiN) situations. At the first glance, performing a SiN test with different HI settings might help the acoustician achieve a more accurate HI fitting. But is this really so? Firstly, such a procedure would be extremely time and effort demanding. Secondly, it is unclear whether current standard SiN tests sufficiently reflect the difficulties of a real-life SiN situation. Lastly, there remains the question of whether the hearing impaired person would accept the HI settings that resulted in his/her best SiN performance. In the study reported here we investigated the influence of SiN performance with specific HI settings on the preference for these settings.

30 impaired subjects were tested with four HI settings in terms of their SiN and SiQ performance as well as regarding their overall spontaneous acceptance of the settings. The SiN test was a selective, real-speech listening task requiring an auditory number comparison rendered more difficult by a distracting talker and an eight- speaker free field cafeteria-noise environment, the so called Numbers in Babble (NiB) test. The SiQ test was the standard monosyllabic word test whereas the spontaneous acceptance was calculated as the sum of subjective evaluations of various sound attributes perti-

nent to SiN, SiQ, environmental noise, music and own voice sound examples. Subjective and objective listening effort were also measured during the NiB task. We found no significant correlation between the performance in NiB task for a given HI setting and subject's preference for that setting according to the spontaneous acceptance. In other words, subject's ability to perform well in SiN task, with a specific HI setting, was not reflected in his/hers preferences for this setting. Equally, there was no correlation between the listening effort involved in the SiN task and the subjective preferences, for the given setting. One simple interpretation of this finding is that subjectively preferred HI setting is chosen on the basis of sound dimensions other than that of SiN understanding. Implications of these results will be discussed.

48 Cochlear implanted children's perception and comprehension of grammatical cues in speech

A. Schouwenaars, E. Ruigendijk
University of Oldenburg, DE

P. Hendriks
University of Groningen, NL

This study investigates the processing of grammatical cues in speech (case and verb-agreement) of normal hearing children (NH) and children with a cochlear implant (CI). The aim is to find out whether both groups of children not only perceive the different cues, but also make use of them in processing.

Therefore, we examined children's comprehension of subject and object questions. In order to correctly interpret object questions (in which the object precedes the subject) in German, case (e.g. Welchen Esel fängt der Tiger 'Which donkey is the tiger catching?') and/or -in some events- verb-agreement cues need to be used (e.g. Welche Maus fangen die Frösche? 'Which mouse are the frogs catching?').

The acquisition of object questions is a long ride for NH children as they heavily rely on word order and interpret object questions incorrectly as subject questions. For CI children, the subtle cues may even be harder to detect and mastered, since their linguistic input is different in terms of length (less years) and reliability. Even though CIs provide relevant cues for the perception of speech (Svirsky, Robbins, Kirk, Pisoni & Miyamoto, 2000), previous research suggests that the comprehension of grammatical aspects of speech is more difficult for children with a CI (Nikolopoulos, Dyar, Achbold & O'Donoghue, 2004; Friedmann & Szterman, 2011).

Participants were 36 NH children (age 7;05-10;09, Mean: 9;01) and 33 CI children (7;01-12;04, Mean:9;07, bilateral < 3ys). As the main task, a picture selection task with eye-tracking was carried out to test children's comprehension of subject and object questions. Two additional tasks were carried out; a picture selection task to test children's comprehension of verb-agreement in sentences with a standard word order and an auditory discrimination task to test whether children perceive the differences with respect to case marking (e.g. 'der' vs 'den').

Overall the children performed well on the additional task on verb-agreement (CI: 86%; NH: 96% correct responses) and on case (CI: 90%; NH: 99%). These scores correlated with 'Hearing Age' and 'Age at Implantation'. Considering only those CI children how scored like NH children on the additional tasks, their comprehension of object questions was still worse (CI: 66%; NH: 86%) and more time was needed to detect the correct interpretation.

To conclude, even though CI children perceive case and verb-agreement, their development of syntactic capacities to use these cues for comprehension still lacks behind.

49 The cocktail party effect revisited in older and younger adults: When do iconic co-speech gestures help?

L. M. Schubotz, **L. Drijvers**, J. Holler, A. Özyürek
MPI for Psycholinguistics, Nijmegen, NL

Understanding speech in noisy surroundings is notoriously difficult, especially for older adults. Previous research suggests that visual cues such as the speaker's articulatory lip movements can significantly improve comprehension in both older and younger adults. Another potentially valuable visual cue for understanding speech in noise are iconic co-speech gestures, hand movements that depict semantic information related to the content of speech, e.g. tracing a circle while saying "ball".

The current study investigates whether such gestures improve older and younger adults' comprehension of speech presented in multi-talker babble noise beyond the benefit of visible lip movements, and whether this is modulated by the signal-to-noise ratio (SNR). Older and younger adults were presented with video clips of an actress uttering an action verb. After each video, participants had to select the uttered verb out of four verbs presented in a cued recall task. Videos were presented in three visual conditions (mouth blocked/audio only, visible lip movements, visible lip movements + co-speech gesture) and four audio conditions (clear speech, SNR -18, SNR -24, no audio).

Response accuracies showed no age-related differences in trials where either only auditory or only visual information was presented. Hence, older adults perform as well as younger adults for speech comprehension, lip reading, and gesture interpretation. However, older adults performed significantly worse than younger adults in those trials with combined visual and auditory input. Yet, both age groups benefitted equally from the presence of gestures in addition to visible lip movements as compared to visible lip movements without gestures; this benefit was significantly larger at the worst noise level.

These results suggest an overall age-related deficit in comprehending multi-modal language in noisy surroundings, potentially due to age-related declines in cognitive abilities (processing speed, working memory). Yet, iconic co-speech gestures provide additional semantic cues that can help both younger and older adults in disambiguating visible speech, particularly as the level of background noise increases.

50 Speech perception studies with matrix sentence tests: A comparison across languages

M. Zokoll, A. Warzybok, S. Hochmuth
Medizinische Physik and Cluster of Excellence „Hearing4all", Carl von Ossietzky Universität Oldenburg, DE

M. Boboshko
St. Petersburg Pavlov State Medical University, RU

R. Bentler
Hearing Aid Lab for Basic & Applied Research, University of Iowa, Iowa, US

B. Kollmeier
Medizinische Physik and Cluster of Excellence „Hearing4all", Carl von Ossietzky Universität Oldenburg and HörTech gGmbH, Oldenburg, DE

It is known that testing a listener in a background of noise helps to assess information in addition to the pure-tone audiogram, such as possible supra-threshold distortions that occur in the auditory system as a result of the hearing impairment. By now, speech recognition tests in noise have become increasingly important in audiological diagnostics. Most of those tests use everyday sentences presented in an open-set format. So-called 'matrix sentence tests', however, use a closed-set format and comprise syntactically fixed, semantically unpredictable sentences (e.g. "Peter kept two green toys") composed from a vocabulary of 50 words (10 alternatives for each word group). Matrix sentence tests are suitable for repeated speech perception testing in a multilingual society even when experimenters are not proficient in the test language (by using the closed-set response format). So far, matrix sentences tests are available for at least 15 different languages (e.g., in American English, French, German, Russian, Spanish, Turkish, and Italian) together with a varying degree of supportive data.

This contribution presents matrix sentence test data of multi-center studies in the USA, Canada and Russia investigating the reliability and comparability of the American English and Russian matrix sentence test, as well as the influence of the hearing ability on speech perception in quiet and noise. Data include adaptively estimated speech reception thresholds (SRTs), i.e. the sound pressure levels or signal-to-noise ratios (SNR) yielding 50% speech intelligibility, as well as correlations between hearing ability (pure tone average, PTA, for 0.5, 1, 2 and 4 kHz) and SRT. Results are that both tests provided reliable values with high comparability across languages. Individual SRTs in quiet correlated closely with audiograms. A comparatively poor relation was found between the SRT in quiet, or PTA and the SRT in noise. This supports the notion that the matrix sentence test in noise assesses an individual auditory factor that is separate from the audiogram and can help to disentangle the contribution of possible supra-threshold distortions to a certain hearing loss from that of the pure loss in sensitivity. Further, it can be concluded that matrix sentence tests are a sensitive diagnostic tool suitable for multilingual comparisons.

51 Identification of oral and nasal segments in band-vocoded speech : Specific difficulties associated with nasal vowels in French.

O. Crouzet

Laboratoire de Linguistique de Nantes, Université de Nantes, FR

Recognition of french oral and nasal segments (consonants and vowels) was investigated in 6 normal-hearing listeners using noise-band vocoded speech from naturally produced VCV and V tokens. Our aim was to determine whether nasal segments would behave specifically in comparison to oral sounds when processed through spectrally degraded algorithms in order to investigate the influence of band-limited envelope modulations on their perceptual classification. This issue was addressed in two parallel N-Alternate forced-choice experiments in which participants were required to classify either consonants (VCV sequences, 18 consonants in 3 different vowel contexts) or vowels (natural steady-state extracts, 13 vowels) within the full French segment inventory.

Naturally produced VCV sequences and steady-state vowels were processed through a noise-band vocoder in order to produce various degrees of spectral (number of frequency bands among {1, 2, 4, 6, 8} bands) and temporal (low-pass modulation frequency cutoff among {4, 16, 128} Hz) resolution. Performance were compared to chance levels using binomial tests for each combination of number of bands and frequency modulation cutoff in each experiment (consonant identification vs. vowel identification), specifically comparing participants perfomance between oral and nasal segments.

As expected, classification performance showed a gradual increase in performance in relation to the increase in both spectral and temporal resolution. In the consonant classification experiment, nasal segments displayed an evolution of performance that closely mimicked the results of oral consonants. In the vowel classification experiment however, a strong discrepancy was observed between oral and nasal segments: Though oral vowels showed performance patterns that were relatively similar to the consonant identification performance, nasal vowels showed very poor performance over all conditions of acoustic degradation. In the specific case of nasal vowels, performance never significantly out-performed chance levels.

We are now addressing issues related to the interpretation of this observation through two different directions. Confusion matrices were extracted from the observed data to provide information concerning the various strategies and feature confusions involved. Though there is a small under-representation of nasal responses, this difference is (1) not massive (9% obs. vs. 11% exp. for consonants, 17% obs. vs. 23% exp. for vowels) and (2) displayed over both consonants and vowels. Further investigations will help identifying tendencies in the classification errors. In parallel, acoustic analyses and control experiments are performed in order to target which cues (static or dynamic) may be lacking that prevent nasal vowels from being correctly identified.

52 Neural network based speech enhancement applied to cochlear implant coding strategies

F. Bolner
Cochlear Technology Centre Belgium, Mechelen, BE

T. Goehring
ISVR - University of Southampton, GB

B. van Dijk
Cochlear Technology Centre Belgium, Mechelen, BE

J. Wouters
ExpORL - KU Leuven, BE

S. Bleeck
ISVR - University of Southampton, GB

Traditionally, algorithms that attempt to significantly improve speech intelligibility in noise for cochlear implant (CI) users have met with limited success, in particular in the presence of a fluctuating masker. Motivated by previous intelligibility studies of speech synthesized using the ideal binary mask [1] and its estimation by means of machine learning [2], we propose a framework that integrates a multi-layer feed-forward artificial neural network (ANN) into CI coding strategies.

The algorithm decomposes the noisy input signal into time-frequency units, extracts a set of auditory-inspired features and feeds them to the ANN to produce an estimation of which CI channels contain more perceptually important information (higher signal-to-noise ratio, (SNR)). This estimate is then used accordingly to suppress the noise and retain the appropriate subset of channels for electrical stimulation, as in traditional N-of-M coding strategies.

Speech corrupted by speech-shaped and BABBLE noise at different SNRs is processed by the algorithm and re-synthesized with a vocoder. Evaluation has been performed in comparison with the Advanced Combination Encoder (ACE™) in terms of classification performance and objective intelligibility measures. Results indicated significant improvement in Hit – False Alarm rates and intelligibility prediction scores, especially in low SNR conditions.

These findings suggested that the use of ANNs could potentially improve speech intelligibility in noise for CI users and motivated the collection of pilot data from CI users and simulations with normal-hearing listeners. The results of this ongoing study will be presented together with the objective evaluation.

[1] Y. Hu and P. C. Loizou, "A new sound coding strategy for suppressing noise in cochlear implants.," J. Acoust. Soc. Am., vol. 124, no. 1, pp. 498–509, Jul. 2008.

[2] Y. Hu and P. C. Loizou, "Environment-specific noise suppression for improved speech intelligibility by cochlear implant users.," J. Acoust. Soc. Am., vol. 127, no. 6, pp. 3689–95, Jun. 2010.

The work leading to this deliverable and the results described therein has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n° PITN-GA-2012-317521.