SPEECH IN NOISE WORKSHOP

11-12 JANUARY 2018 GLASGOW UK

THURSDAY 11 Jan 2018

900	Registration & Coffee
930	Welcome
945	**Gaston Hilkhuysen** *The past, present and future (if any) of speech intelligibility metrics: a review and analysis*
1015	**Ning Ma** *Auditory scene enhancement for hearing impaired listeners*
1045	Coffee + Poster set up
1115	**Riikka Möttönen** *Auditory-motor speech perception in young and older adults*
1145	**Tim Griffiths** *Figure-ground analysis relevant to speech in noise perception in health and disease*
1215	**Tobias Reichenbach** *Towards a smart hearing aid: decoding the brain's response to speech*
1245	Lunch
1400	**Edmund Lalor** *The effects of attention and visual input on non-invasive electro-physiological indices of natural speech processing at different hierarchical levels*
1500	**Michael Richter** *The impact of motivation on listening effort: Effects of task demand and success importance on cardiovascular correlates of listening effort*
1530	Coffee
1600	**Posters**
1730	End of Day 1
1900	**Dinner** : *Fratelli Sarti* 121 Bath Street, G2 2SZ

FRIDAY 12 Jan 2018

900	Coffee
930	**Posters**
1030	Coffee
1100	**Posters**
1230	Lunch
1330	**Brechtje Post** *Intonation in speech perception*
1400	**Outi Tuomainen** *Effects of mild age-related hearing loss and background noise on speech communication*
1430	**Silke Paulmann** *Don't use that tone with me: How emotions and motivations are processed from speech*
1500	Coffee
1530	**Lea-Maria Schmitt** *Semantic predictability and brain state modulate neural representations of speech in noise*
1600	**Carine Signoret** *Semantic representations involvement during degraded speech perception*
1630	Closing remarks
1700	End of Day 2 & Workshop

PLENARY SESSIONS

THURSDAY 11.01.2018 09:45 – 10:45

The past, present and future (if any) of speech intelligibility metrics: A review and analysis

Gaston Hilkhuysen [g.hilkhuysen@ucl.ac.uk], Mark Huckvale

Speech, Hearing & Phonetic Sciences, University College London, UK

Seventy years ago, French and Steinberg (1947) presented a computational method that employed 'the intensities of speech and unwanted sounds received by the ear' to predict speech intelligibility in noise. Over the years, their articulation index has been fine-tuned; adapted to more listening condition; crystallized into a still frequently reaffirmed ANSI standard; and inspired scholars to develop new types of speech intelligibility metric. Recent literature in particular shows an avalanche of such metrics, each reporting higher correlations with intelligibility scores than their predecessors. But as a consequence, users now find it hard to choose the most appropriate metric for their application: lost in the maze created by the overwhelming number of possible variants.

This presentation will put forward a taxonomy to structure the current types of speech intelligibility metrics. A genealogical tree of types helps identify similarities among and differences between various metrics. A resulting analysis suggests a way to find the best 'fit for purpose' metrics and illuminate their current limitations. Because these shortcomings provide challenges and opportunities for new metrics, they may give direction to future developments.

THURSDAY 11.01.2018 10:15 – 10:45

Auditory scene enhancement for hearing impaired listeners

Ning Ma [n.ma sheffield.ac.uk], Guy Brown, John Barker

Department of Computer Science, University of Sheffield, UK

Michael Stone

School of Psychological Sciences, University of Manchester, UK

Age-related hearing loss affects 55% of the UK population over 60. It is a degenerative condition that not only reduces the quality of the sound that is heard but also changes a person's ability to communicate with others. Hearing aids are the most common form of treatment for mild to moderate hearing loss. However, adoption of hearing aids so far is low (15-20% of patients), largely due to the relatively poor performance in noise.

Despite of some advances that have been made in the field, current algorithms adopted in modern hearing aids have had limited success in improving speech intelligibility. The strategy most state-of-the-art hearing aids adopt is via microphone array beamforming to amplify speech while filtering out sounds coming from different directions. While this strategy is effective in removing noise, it is precisely why the experience of wearing hearing aids is often poor in a noisy environment. First, beamforming cuts out sounds outside the "beam" which makes it difficult for the wearer to stay aware of the acoustic environment. For example, a hearing aid user often finds it challenging to discern specific voices in a crowd. More importantly, there is no physical difference between a noise and a sound that a listener intends to hear. Imagine that a hearing aid user is watching TV, when someone on the side starts talking. Which sounds should the hearing aid consider as noise and remove? The brain is an amazing organ that can automatically focus in on the sounds that a listener wants to hear from a mixture of sounds. It is therefore important that a hearing aid delivers to the brain as much sound as possible. It cannot and should not think for the listener.

The goal of this study is to introduce an "auditory focus" to hearing aids. In vision, objects out of the focus range become blurred so that those in focus appear sharper. However, the background objects are not completely removed and one can still perceive some primitive characteristics, e.g. colour and shape, and choose to shift focus to them. Akin to vision, the proposed hearing aid processing strategy enhances the auditory objects within the auditory focus (defined as a directional beam), whereas the spectral details of those out of focus are simplified so that they have less interference but can still be heard by the listener.

This method is analogous to pulling the auditory objects in focus closer to the listener while moving those out of focus further away. This helps the brain interpret the signal delivered from damaged cochlear hair cells. The key difference to beamforming is that the "noise" is not removed but is blurred spectrally in the acoustical sense. The primitive characteristics of the sounds outside the beam (focus) can still be perceived by the listeners. For example, if a hearing-impaired listener wearing such hearing aids is watching TV, he/she would be able to hear somebody starts talking at the side and choose to shift the auditory focus in order to understand the voices. The aim of this study is to determine the benefit of the proposed methods on speech intelligibility for hearing-impaired listeners. The objectives include:

- To investigate key auditory features that hearing impaired listeners can use for hearing out speech in noise;
- To develop suitable machine learning techniques that can enhance sounds so that they are simplified spectrally for easier perception while maintaining primitive sound properties;
- To measure the benefit of different strategies for introducing the auditory focus – from head movement, eye movement, to a brain computer interface.

THURSDAY 11.01.2018 11:15 – 11:45

Auditory-motor speech perception in young and older adults

Riikka Möttönen [Riikka.Mottonen@nottingham.ac.uk]

School of Psychology, University of Nottingham, UK

Speech communication relies on both auditory and motor systems. It has been proposed that the involvement of the articulatory motor cortex in speech processing increases when intelligibility of speech signals decreases. Thus, according to this hypothesis the articulatory motor cortex has a compensatory role in speech processing. We tested this hypothesis by measuring excitability of the articulatory motor cortex during listening to sentences across a wide range of signal-to-noise ratios (SNR) using Transcranial Magnetic Stimulation (TMS) and electromyography. Excitability of the articulatory motor cortex was facilitated during listening to clear sentences relative to non-speech baselines. Decreasing SNR of the sentences decreased their intelligibility, but had no effect on motor excitability. These findings do not support the motor compensation hypothesis. I will also present evidence that the articulatory motor cortex is involved in speech processing in older adults with normal hearing, but its involvement decreases in order adults with mild-to-moderate hearing loss. Future studies are needed to investigate whether the decline in auditory-motor processing is associated with speech perception difficulties in adults with age-related hearing loss.

THURSDAY 11.01.2018 11:45 – 12:15

Figure-ground analysis relevant to speech in noise perception in health and disease

Tim Griffiths [tim.griffiths@ncl.ac.uk]

Institute of Neuroscience, University of Newcastle, UK / Wellcome Trust Centre for NeuroImaging, UCL / Departments of Neurosurgery and Otolarngology, University of Iowa

I will describe work on generic mechanisms for recognition of auditory targets in noisy backgrounds using stochastic stimuli [1]. The work specifies a general system relevant to speech-in-noise (SIN) perception, deficits in which are a ubiquitous symptom of cochlear damage and central auditory disorder.

Non-invasive human MEG [2] and fMRI [3] work specifies a system involving auditory cortex and intraparietal sulcus. Invasive human local field potential recordings from auditory cortex demonstrate high-frequency oscillatory activity corresponding to figure emergence.

Work in a macaque model demonstrates similar behavioural performance in macaques and man and macaque fMRI demonstrates a network that involves similar areas of high level auditory cortex in the macaque, in rostral parabelt [4]. This definition of network organisation in the macaque provides a basis for neurophysiological work in progress to systematically define mechanisms for generic figure-ground analysis at the level of neurons and neuronal ensembles that is not possible in man.

Data from >100 normal-hearing subjects establishes that the detection of generic figure ground tasks correlates significantly with speech in noise perception using single CVCs, based on a modification of the California Consonants Task to incorporate noise. Data from hearing impaired subjects using conventional amplification or cochlear implants demonstrate significant correlations that are much stronger, accounting for more than a quarter of the variance.

The relationship between the generic figure-ground measure has a likely component related to cochlear processing constituting a common source of variance in the generic figure-ground and SIN task performance, especially in the hearing impaired. But the imaging and recording work on normal subjects demonstrates a central mechanism for the figure-ground analysis that is another factor that contributes to SIN performance. Such central mechanisms represent further determinants of SIN listening success in normal and hearing-impaired listeners.

1. Teki S, Chait M, Kumar S, Shamma S, Griffiths TD. Segregation of complex acoustic scenes based on temporal coherence. Elife 2, e00699 (2013).
2. Teki S, Barascud N, Picard S, Payne C, Griffiths TD, Chait M. Neural Correlates of Auditory Figure-Ground Segregation Based on Temporal Coherence. Cereb Cortex 26, 3669-3680 (2016).
3. Teki S, Chait M, Kumar S, von Kriegstein K, Griffiths TD. Brain bases for auditory stimulus-driven figure-ground segregation. J Neurosci 31, 164-171 (2011).
4. Schneider F, Dheerendra P, Balezeau F, Petkov CI, Thiele A, Griffiths TD. SfN (Washington, DC, 2017).

THURSDAY 11.01.2018 12:15 – 12:45

Towards a smart hearing aid: Decoding the brain's response to speech

Tobias Reichenbach [reichenbach@imperial.ac.uk]
Department of Bioengineering, Imperial College London, UK

Understanding speech in noise is a main problem for people with hearing impairment, and it persists for hearing aid users. The problem is impeded by an often imperfect fitting of the settings in a hearing aid that is informed by pure-tone audiometry but not directly by speech-in-noise comprehension. Moreover, algorithms for enhancing the intelligibility of speech in noise exist, but their usage in a hearing aid requires knowledge of the user's target sound, such as a particular voice amongst competing speakers that the hearing aid wearer wants to listen to. Here we present recent progress on decoding speech comprehension as well as the attentional focus of a listener to one of two competing voices from non-invasive EEG recordings. The decoding is based on both cortical and subcortical neural activity in relation to different acoustic as well as linguistic features of speech. The developed methods may be applied in a smart hearing aid that measures brain activity from electrodes within the ear canal to better fit the hearing aid's settings as well as to inform its noise-reduction algorithm.

THURSDAY 11.01.2018 14:00 – 15:00

The effects of attention and visual input on noninvasive electrophysiological indices of natural speech processing at different hierarchical levels

Edmund Lalor [edmund_lalor@urmc.rochester.edu]

Lalor Lab for Computational Cognitive Neurophysiology, University of Rochester Medical Centre, NY, US / School of Engineering, Trinity Centre for Bioengineering and Trinity College Institute of Neuroscience, Dublin, IE

How the human brain extracts meaning from the dynamic patterns of sound that constitute speech remains poorly understood. This is especially true in natural environments where the speech signal has to be processed against a complex mixture of background sounds. In this talk I will outline efforts over the last few years to derive non-invasive indices of natural speech processing in the brain. I will discuss how these indices are affected by attention and visual input and how attentional selection and multisensory integration can be "decoded" from EEG data. I will outline work showing that EEG and MEG are sensitive not just to the low-level acoustic properties of speech, but also to higher-level linguistic aspects of this most important of signals. This will include demonstrating that these signals reflect processing at the level of phonetic features. And, based on our most recent work, it will also include evidence that EEG is exquisitely sensitivity to the semantic processing of natural, running speech in a way that is very strongly affected by attention and intelligibility. While showcasing these findings, I will outline a number of paradigms and methodological approaches for eliciting non-invasive indices of speech-specific processing that should be useful in advancing our understanding of receptive speech processing in particular populations.

THURSDAY 11.01.2018 15:00 – 15:30

The impact of motivation on listening effort: Effects of task demand and success importance on cardiovascular correlates of listening effort

Michael Richter [M.Richter@ljmu.ac.uk]

Natural Sciences & Psychology, Liverpool John Moores University, UK

In my talk, I will present an adaption of Brehm's motivational intensity theory – a psychological theory that predicts effort mobilisation in instrumental tasks – to listening effort and cardiovascular correlates of listening effort. According to this theory, listening effort should be a direct function of listening demand: the more difficult the listening task, the higher the invested effort. However, this relationship between listening demand and listening effort should be limited by success importance and only hold if listening demand is fixed. If success importance does not justify the required listening effort, individuals should disengage and not invest any effort. If listening demand is unclear or unfixed, success importance should be the direct determinant of listening effort: the higher the success importance, the higher the invested effort.

I will present three studies that tested these predictions using cardiovascular indicators of effort. Study 1 examined the impact of listening demand on effort-related cardiovascular activity manipulating the difficulty of a speech-in-noise task across four levels. Study 2 varied both task demand and success importance of an auditory discrimination task to examine the joint impact of both variables on listening effort. Study 3 tested the predicted effect of success importance on effort-related cardiovascular activity in a speech-in-noise task with unclear demand. The results of all three studies corroborated the predictions of motivational intensity theory demonstrating that it is important to consider motivational variables beyond task demand in the research on listening effort.

FRIDAY 12.01.2018 13:30 – 14:00

Intonation in speech perception

Brechtje Post [bmbp2@cam.ac.uk]

Phonetics Laboratory, University of Cambridge, UK

Intonation plays a central role in human communication, since it provides immediate cues that facilitate word recognition in speech comprehension, but it also provides crucial cues to the meaning of utterances and their role in the wider discourse context. Research in second language learning of intonational properties as well as aphasic speech, speech synthesis and automatic speech recognition has confirmed that when the intonation is 'wrong', communication tends to break down.

Unfortunately, intonation is quite elusive from a modelling perspective: It is continuous in nature, and it is signalled by multiple interacting phonetic cues which simultaneously map to different linguistic as well as non-linguistic functions.

In this paper, I will briefly review what we know about the 'substance' of intonation, how we can measure and model it, and what its main functional roles in language comprehension are. My main objective will be to argue that intonation has a crucial role to play in speech perception more generally, but also more specifically in speech recognition in adverse listening conditions.

FRIDAY 12.01.2018 14:00 – 14:30

Effects of mild age-related hearing loss and background noise on speech communication

Outi Tuomainen [o.tuomainen@ucl.ac.uk]

Speech, Hearing & Phonetic Sciences, University College London, UK

Our ability to communicate successfully with others can be affected by the presence of noise in the environment and becomes increasingly difficult with advancing age. Much research considers the effect of aging on the ability to understand speech but there is less attention on speech production and the adaptations talkers make to overcome adverse listening conditions, especially in more ecologically valid situations such as during interactive speech. Our aim was to investigate what speech modifications talkers make in background noise and which parameters predict communication efficiency, and if these are influenced by the hearing status of older talkers.

Our study included 83 talkers of Southern British English: 57 older adults aged 65-84, of which 30 had normal hearing (OANH) and 27 (OAHL) had mild age-related hearing loss but did not wear hearing aids (mean PTA .250-4kHz: 27.7 dB, better ear), and 26 younger adults (YA) aged 18-26. Talkers were recorded while they completed an interactive 'spot the difference' picture task (diapix) with a younger conversational partner when i) they could each other normally (NORM); ii) they were both in multi-talker babble noise (BAB2). We collected background sensory and cognitive measures (speech-in-noise thresholds, working memory, verbal fluency). From the diapix recordings, we measured acoustic-phonetic features (articulation rate, relative energy in the 1-3 kHz region [ME13], f0, vowel hyperarticulation) and for a subset of talkers (YA=24, OANH=21, OAHL=21) dysfluency rates (filled pauses, false starts, repetitions). Communication efficiency was measured as time to find eight differences between the pictures.

The results showed that OAHL talkers had poorer speech-in-noise thresholds, lower verbal fluency scores and it took them longer to find the eight differences in the diapix tasks (NORM, BAB2) than for YA or OANH talkers. OAHL talkers were also more dysfluent than YA and OANH talkers. Background noise induced slower speech and modifications that resemble Lombard speech (i.e., increase in f0 and ME13) that were greater for OA than for YA talkers. However, when speaking in noise, OAHL talkers made adaptations more consistent with an increase in vocal effort. Only speaking rate, ME13 and frequency of dysfluent repetitions predicted how effectively the two talkers found the differences between the pictures.

Our results suggest that even mild levels of age-related hearing loss can affect communication efficiency and fluency regardless of listening condition. However, background noise can influence speaking effort in OAHL talkers.

FRIDAY 12.01.2018 14:30 – 15:00

Don't use that tone with me: How emotions and motivations are processed from speech

Silke Paulmann [paulmann@essex.ac.uk]

Department of Psychology, University of Essex, UK

It's been said that 10% of conflicts are due to differences in opinion and 90% are to due to wrong tone of voice. This nicely outlines how important non-verbal cues are in our daily communication. In this talk, I will give an overview about my own work on how social intentions and emotions in particular are communicated through tone of voice alone. I will refer to work looking at acoustic correlates of emotional speech, how easy or difficult it is to recognise emotions from speech and will also show some electro-physiological data from listeners in response to emotional speech. Moreover, I will outline which factors we have identified that might influence how emotions and attitudes are processed from speech.

Semantic predictability and brain state modulate neural representations of speech in noise

Lea-Maria Schmitt [l.schmitt@uni-luebeck.de], Malte Wöstmann, Lorenz Fiedler, Jonas Obleser

Institut für Psychologie, Universität zu Lübeck, DE

When listening to speech against competing talkers, the human brain draws on compensatory cognitive resources. I will demonstrate that both cognitive resources provided by a listener's brain state as well as cognitive resources demanded by signal quality modulate electroencephalographic (EEG) signatures of speech processing. In the first study, we revisited an effect as old as the EEG itself: closing the eyes increases the power of parieto-occipital alpha oscillations (8–12 Hz), which are – coincidentally? – also a proxy of focusing attention to the auditory modality. To test whether closing the eyes benefits auditory attention to speech, participants attended to one of two streams of spoken numbers in darkness. Whereas eye closure per se did not enhance perceptual sensitivity, closing the eyes increased the tendency to judge probes as attended in participants exhibiting stronger attentional modulation of alpha power under closed eyes. Our findings demonstrate that closing the eyes has the potency to enhance the neural dynamics of auditory attention by supporting the adaptive inhibition of the dominant visual system. Next, I will give an outlook for my future research on the neural underpinnings of building up predictions for forthcoming speech segments informed by semantic context (e.g., "The ship sails the sea" for a highly predictable final keyword). To test how semantic context alters speech comprehension under adverse listening conditions, participants attended to one of two narrative speech streams both varying in sound intensity over time. We expect the signal-to-noise ratio to moderate the mapping of semantic predictability on the power of neural oscillations related to auditory attention and semantic predictability. In sum, I will argue that a listener's brain state as well as the use of predictions efficiently balance the recruitment of limited cognitive resources during speech comprehension.

FRIDAY, 12 January 2018 16:00 – 16:30

Semantic representations involvement during degraded speech perception

Carine Signoret [carine.signoret@liu.se]

Linnaeus Centre HEAD, Swedish Institute for Disability Research, Department of Behavioral Sciences and Learning, Linköping University, SE

The perceptual clarity of speech is not entirely dependent on the acoustic quality of the sound. Other resources, such as linguistic representations, are involved during degraded speech perception. For example, presentation of the written version of a degraded sentence before hearing it will enable prior knowledge on the exact speech content, which will make the degraded sentence seems clearer. This phenomenon has been explained by top-down influence of phonological and lexical representations on acoustic processing. Another example is the influence of semantic representations on the intelligibility of degraded speech: degraded sentences are better reported if they are meaningful than meaningless. The question now is whether the semantic representations could further influence the perceptual clarity of degraded speech for both normal-hearing (NH) and hearing-impaired (HI) listeners. In the reported set of three experiments, grammatically correct Swedish spoken sentences were presented at different sound quality levels, from clear to unintelligible. The sound quality levels were manipulated by using noise vocoding (NV) method in which the number of bands reflects intelligibility: more bands for more intelligibility. HI listeners were provided with amplification according to the Cambridge formula. The sentences had either high (e.g. “His new clothes were from France.”) or low (e.g. “His red school was from the newspaper”) semantic coherence and were matched at the word level. The written version of each spoken word (matching text) or a string of consonants (non-matching text) was presented 200 ms beforehand in a rapid serial visual paradigm. The task of the listeners was to rate the clarity of each spoken sentence on a 7-point Likert scale. Results revealed significant interactions between coherence and text for both groups, showing a benefit of coherence with matching and non-matching text for NH listeners but only with matching text for HI listeners. Significant three-way interactions including sound quality level modified this finding to some extent. Indeed, NH listeners benefitted from semantic coherence with non-matching text at 6 and 12 band NV (but not 3 band) while HI listeners benefitted at 12 band (but not 3 and 6 band). Preliminary fMRI results obtained for NH listeners indicated that processing of semantic coherence with non-matching text is supported by right middle temporal gyrus. The overall pattern of results suggest that NH listeners successfully utilize semantic representations in spoken sentences that are moderately degraded and when no prior knowledge is available. What prevents HI listeners to do the same?

Posters Galore!

1 Development and evaluation of the Coordinate Response Measure speech-in-noise test as a hearing assessment tool for the Armed Forces

2 Sociolinguistic effects of speaking in masking noise

3 Challenging listening conditions make understanding Glaswegian /r/ even harder

4 The influence of educational attainment on the relationship between auditory and visual inhibition measures and speech-in-noise perception

5 The contribution of cognition and hearing loss to individual differences in speech intelligibility in a variety of speech-perception-in-noise tests in younger and older adult listeners

6 What we talk about when we talk about speech intelligibility

7 Listening effort and difficult listening conditions: Lexical processing and neural entrainment with noise and vocoders

8 Effects of noise suppression on spatial release from masking in simulated realistic listening environments

9 Do you hear the noise? Influence of background noise level on speech reception

10 Speech in noise threshold measurements in cochlear implant users

11 An information-theoretic analysis of auditory features in noisy environments

12 Revision of a binaural model predicting speech intelligibility against envelope-modulated noise interferers

13 Laboratory paired comparisons based on the CoSS framework

14 Do noise-induced latency shifts of the auditory brainstem response to speech reflect degradation in neural synchrony?

15 Relating speech perception in noise to temporal-processing auditory capacities in childhood: effects of typical development and of sensori-neural hearing impairment

16 Speech-perception-in-noise deficits in dyslexia: a possible method to improve speech perception abilities

17 Caught by surprise: the effect of unpredictable varying talker location on speech intelligibility and listening effort

18 Can biophysically inspired features improve neural network-based speech enhancement ?

19 The pupil dilation response of adults with acquired brain injury during speech processing in noise

20 Decoding Attention at Higher Levels of Linguistic Processing using EEG

21 The role of offset sensitivity in consonant discrimination in noise

22 An assessment of objective intelligibility metrics for signals with low mixture signal-to-noise ratios after enhancement using Ideal Binary Masks

23 On the dynamics of the preference-performance relation for hearing aid noise reduction

24 Periodicity is not a factor in making harmonic complexes less effective maskers of speech than noise

25 A binaural model predicting speech intelligibility in noise for hearing-impaired listeners

26 The role of periodicity in the perception of masked speech with simulated and real cochlear implants

27 Analysis of the individual listening effort reflected by the pupillary responses during speech perception in noise

28 Improving localization in binaural beamforming for hearing aid wearers

29 Pupil dilation during the speech understanding task in dark and light - potential influence of hearing impairment to the parasympathetic nervous system

30 How do users and non-users of hearing aids differ?

31 Automatic scene classification improves speech perception of CI users in simulated real world listening scenarios

32 The effects of SNR driven amplitude compression in hearing aids on output SNR and signal envelope distortion

33 Slope of the performance-intensity function and reaction time for speech in different noise types

34 Segregation enhancement for hearing impaired listeners using a deep neural networks separation algorithm

35 Subjective listening effort: Influence of background noise direction and speaker's gender

36 A Model of Concurrent Vowel Identification Without Segregation Predicts Perceptual Errors

37 Acoustic analyses of vowel variation for the investigation of perceptual adaptation to speaker properties in channel-vocoded speech: preliminary data.

38 Biological inspired MEMS acoustic sensors

39 The contribution of salient localizable glimpses on speech intelligibility in a multitalker setting with spatially diffuse signals

40 Data-driven discovery of general mechanisms of cortical processing of natural sounds

41 3D printed acoustic metamaterials for small-scale noise control applications

42 Bio-inspired Frequency-adaptive Acoustic System

43 Preattentive Processing in the Spatial Unmasking of Speech

44 Speech-in-noise recognition abilities are associated with vocal pitch perception abilities in controls but not in high-functioning autism spectrum disorder

45 Who are you listening to? Towards a dynamic measure of auditory attention to speech-on-speech

46 Effects of global brightness on salience and auditory foreground perception

47 Encoding of Mid-Level Speech Features in MEG Responses

48 How speech statistics limits the number of effective channels in cochlear implants. Implications for sound-coding strategies.

49 The spatial speech test of real-world listening for assessing binaural hearing

50 How the tongue and lips produce clear speech: CVC words with randomised vowels, transcribed by listeners in normal and noisy conditions

51 Speech perception under eye-controlled and head-controlled directional microphones in a dynamic 'cocktail party'

52 The effect of the language proficiency of bilingual adults on the Canadian Digit Triplet Test

53 Simulating hearing loss in neural networks: Does pre-training on intact speech boost performance on degraded input?

Development and evaluation of the Coordinate Response Measure speech-in-noise test as a hearing assessment tool for the Armed Forces

Hannah Domenica Semeraro [h.d.semeraro@soton.ac.uk], Daniel Rowan
Institute of Sound and Vibration Research, University of Southampton

The ability to listen to and understand commands in noisy environments, whilst maintaining situational awareness, is an important skill for military personnel and can be critical for mission success. Accurately measuring auditory fitness for duty (AFFD) within the Armed Forces ensures that personnel have sufficient hearing ability to be effective in operational scenarios. Pure-tone audiometry (PTA) is currently used by the UK military but it is known to be a poor predictor of overall listening ability. In addition, conducting accurate audiometry presents some practical challenges in occupational environments. Here we suggest a speech-in-noise test as an alternative AFFD assessment tool and explain how simulations of speech-communication mission-critical auditory tasks (MCATs) were developed to explore the predictive validity of potential AFFD tools.

Firstly, we developed a British English version of the CRM using call signs from the NATO phonetic alphabet. After ensuring all the target words had similar intelligibility, the CRM was implemented in an adaptive procedure in stationary speech-spectrum noise to measure speech reception thresholds (SRTs). The SRTs of normal-hearing civilians and hearing-impaired military personnel were assessed. The CRM is sensitive to hearing impairment and the results display good test-retest reliability (95% confidence interval < 2.1 dB) and good concurrent validity when compared to the Triple-Digit-Test ($r \leq 0.65$).

Secondly, we developed simulations of the speech-communication MCATs in order to explore the relationship between performance on potential AFFD tools and performance when listening to ‘real world’ stimuli. Participants included normal-hearing civilians ($n = 28$) and military personnel ($n = 28$). We also explored the influence of military experience when listening to commands. Participants listened to commands recorded over a military radio, presented in armoured vehicle engine noise at a fixed SNR and processed through a hearing loss simulator. The percentage of correctly repeated commands was scored. Participants also completed the CRM adaptive procedure. Results showed that both the simulated commands and CRM are sensitive to simulated hearing impairment. Military personnel outperform civilians on the commands but not on the CRM, suggesting that personnel are, to some extent, able to use their knowledge and experience of command structure and vocabulary to overcome adverse listening conditions and compensate for hearing impairment.

Further work is required to determine whether PTA or the CRM, when combined with additional information about non-psychoacoustic factors that may influence performance, such as military experience, best predict AFFD.

Funded by the Royal Centre for Defence Medicine.

2

Sociolinguistic effects of speaking in masking noise

Sophie Meekings [sophie.meekings@ncl.ac.uk], Danielle Turton, Joel Wallenberg
Newcastle University

A classic study by Mahl (1972) found that when participants spoke in loud masking noise, they used 'more colloquial' language variants, such as glottal stops in place of intervocalic /t/. This was interpreted as demonstrating that participants' attention had been diverted from self-monitoring their speech by the masking noise, causing them to revert to less prestigious language variants rather than using 'more formal', Received Pronunciation forms to increase their intelligibility. However, this conclusion was based on speculation rather than analysis of participants' intelligibility, and the sample size was very small—results were drawn from close analysis of one participant, and anecdotal reports of others.

Here, we attempt a partial replication of Mahl (1972), looking at phonetic variation in masking noise. Participants were recorded as they spoke in silence and over white noise played at 60, 70 and 80dB SPL through circumaural closed-back headphones. Recording speech in a range of masker intensities allows us to correlate noise intensity with the degree of phonetic change. The transcribed interviews were then analysed for th-fronting and intervocalic /t/ glottalisation. Results are discussed with reference to linguistic and psycholological theories of auditory self-monitoring and speech production in noise.

3

Challenging listening conditions make understanding Glaswegian /r/ even harder

Robert Lennon [r.w.lennon@leeds.ac.uk]
University of Leeds

Non-ideal listening conditions come in many forms. One challenge to speech perception may exist when words are acoustically very similar, possibly resulting in the wrong message being conveyed to the listener. Another potential difficulty is when the listener is uncertain about the identity or the accent of the talker.

Both of these challenging listening conditions can be examined using Glaswegian speech as a case study. Previous research has found that Glaswegian can be hard to understand for unfamiliar listeners (Smith et al. 2014; Adank et al. 2009), but this paper focuses on the difficulty that may arise for native Glaswegian listeners when asked to identify which word from a pair, e.g. 'hut/hurt', the speaker produced. Due to extreme /r/ weakening over time (Lawson et al. 2017), many working class (WC) Glaswegians produce these words so that they are acoustically (Anon 2015) and perceptually (2016) almost identical. Middle class (MC) speakers, in contrast, produce e.g. 'hurt' with a very strong /r/ (Anon 2012, 2015; Lawson et al. 2011) so their 'hut-hurt' distinction is not difficult to perceive – however, if the listener is unaware that they are hearing a MC speaker, MC 'hut' could still be confusable with WC 'hurt'. To explore the perceptual consequences of uncertainty about speaker accent, this experiment manipulated whether words like 'hut' and 'hurt' were presented in single-talker or mixed-talker blocks.

Glaswegian listeners heard single words from 12 pairs presented in three blocks: Single MC, Single WC, and Mixed (MC and WC stimuli were randomised together). Their task was to use the mouse to click the word they heard, out of two onscreen options per trial – e.g. 'hut/hurt', 'bud/bird'. MouseTracker (Freeman&Ambady 2010) recorded cursor trajectories as participants moved towards their chosen word, in the top-left or top-right of the screen.

Trajectories were analysed using a suite of measures (RT; Area-Under-the-Curve (ibid); Discrete-Cosine-Transformation (Watson&Harrington 1999)), finding significant processing costs when listeners tried to distinguish WC 'hut-hurt'. Furthermore, listeners experienced more perceptual difficulty when word pairs were heard in the Mixed block, meaning that when words from two accents and speakers were randomised together the intended message was much harder to decode. Interestingly, even the 'easier' MC pairs were more difficult in the Mixed block. These results are discussed in terms of some general principles which underlie exemplar theories and Bayesian inference, concerning how listeners resolve simultaneous uncertainty about the linguistic unit being produced and the speaker.

4

The influence of educational attainment on the relationship between auditory and visual inhibition measures and speech-in-noise perception

Sarah Knight [sarah.knight.34@gmail.com]

Department of Psychology, Royal Holloway, University of London, UK

Antje Heinrich

Manchester Centre for Audiology and Deafness (ManCAD), University of Manchester, UK

Inhibition – the ability to suppress goal-irrelevant information – is thought to be an important cognitive skill in many situations, including speech-in-noise (SiN) listening [1;2]. It is also suggested to worsen with age [3]. Researchers have therefore begun to investigate the ability of age-related inhibitory declines to account for the difficulties older listeners have in noisy environments.

Inhibition is often assessed using Stroop-type tasks, in which one stimulus dimension must be named while a second, more prepotent dimension is ignored. The to-be-ignored dimension may be relevant or irrelevant, and inhibition scores are traditionally derived from the reaction time difference between relevant and irrelevant conditions. Both visual and auditory Stroop tasks are used, but individual studies typically employ only one type—although equivalence between Stroop tasks in different domains cannot be assumed [4].

Some studies show a clear relationship between Stroop scores and SiN performance [1]; others do not [4]. One possible influencing factor is educational attainment. Older adults with higher educational attainment perform better on Stroop tasks, perhaps because education partially compensates for cognitive decline; this suggests that the predicted Stroop/SiN relationship may not be observed for this group.

In this study, 50 older adults (ages=61-86, mean=70; age-normal hearing) performed two Stroop tasks (visual and auditory) and two SiN tasks (with targets either isolated words or words in low- and high-predictability sentences, presented in speech-modulated noise at two signal-to-noise ratios). Individual measures of hearing (PTA0.25-8kHz) and educational attainment were obtained.

Results showed a clear effect of education on the relationship between visual Stroop and SiN scores, with the predicted relationship observed only for listeners with lower educational attainment. The auditory Stroop results were less clear-cut, with complex interactions between multiple variables.

These findings suggest that educational attainment modulates the role of cognition in SiN perception.

Supported by BBSRC grant BB/K021508/1.

[1] Sommers, M.S., & Danielson, S.M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. Psychol Aging, 14(3), 458-472.
[2] Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. Aging Neuropsychol C, 19(6), 741-758.
[3] Hasher, L., & Zacks, R.T. (1988). Working memory, comprehension, and aging: A review and a new view. Psychol Learn Motiv, 22, 193-225.
[4] Knight, S., & Heinrich, A. (2017). Different measures of auditory and visual stroop interference and their relationship to speech intelligibility in noise. Front Psychol, 8, 230.

The contribution of cognition and hearing loss to individual differences in speech intelligibility in a variety of speech-perception-in-noise tests in younger and older adult listeners

Adam Dryden

MRC Institute of Hearing Research, School of Medicine, University of Nottingham

Harriet Allen

School of Psychology, University of Nottingham

Helen Henshaw

NIHR Nottingham Biomedical Research Centre, School of Medicine, University of Nottingham

Antje Heinrich [antje.heinrich@manchester.ac.uk]

Manchester Centre for Audiology and Deafness (ManCAD), University of Manchester, UK

Successful Speech-in-Noise (SiN) perception can be difficult, particularly for older listeners. Changes in hearing and in cognition probably both play a role in understanding these difficulties. To fully understand unique and shared contributions of the two, a systematic and theoretically rigorous approach in selecting speech stimuli and cognitive tests is essential. We measured SiN perception in a total of six listening situations that varied in the sematic predictability of the target sound and the extent of informational masking in the background. We selected cognitive tests based on the amodal and verbal components of Baddeley's model of working memory (central executive (CE), episodic buffer (EB), phonological loop (PL)) and assessed each component by multiple tests (three for CE, and two for EB and PL). We then combined test scores to compute latent variable scores for each component, reducing variation due to surface test characteristics.

To examine potential interaction effects between hearing loss and cognition we tested younger (N=50, age range 18-30 years) and older listeners (N=50, age range 60-85 years). Neither were hearing aid users. Young listeners, who all had pure-tone averages (PTAs) within the clinically normal range (

We analysed the behavioural results from the six speech perception tests (3 target types x 2 background types), three latent cognitive variables (CE, EB, PL), hearing sensitivity and age (young, old) in a linear mixed model. Here we concentrate on reporting those results that involve the predictive roles of hearing loss and cognition for speech intelligibility. Hearing loss: The predictive role of hearing level interacted with type of background masker such that it was comparable between younger and older listeners for three-talker babble but was substantially increased for older compared to younger listeners in the presence of signal-modulated noise. Cognition: EB: better scores predicted better speech intelligibility overall, regardless of age group or hearing status. CE: better scores predicted better overall speech intelligibility in older but not younger listeners. PL: better scores predicted better speech perception performance in three-talker babble for younger listeners but worse performance for older listeners.

These results suggest that younger and older adults may employ different listening strategies in order to understand speech in noise in the same situation.

6

What we talk about when we talk about speech intelligibility

William M Whitmer [bill.whitmer@nottingham.ac.uk], David McShefferty
MRC/CSO Institute of Hearing Research - Scottish Section

When discussing speech intelligibility benefits, it is common to refer to the signal-to-noise ratio (SNR) where a listener's ability to repeat the signal correctly 50% of the time (SNR50). If performance has been measured robustly, there should be objective equivalence in difficulty across any signal and noise pairs presented at a listener's SNR50. It is reasonable to assume that the listener's perception of difficulty will also be equivalent across stimuli presented at their respective SNR50s. We found this assumption of subjective equivalence to be false.

Twenty adult (median age of 67 years) listeners (nine female) of varying hearing ability (median better-ear average 29 dB HL) participated. In different blocks of trials, listeners first were tasked with repeating back IEEE sentences in same-spectrum or two-talker babble noise at various SNRs. Individual SNR50s were then estimated from the psychometric functions. Thereafter, in a modification to our previous method to measure the SNR JND, listeners heard on a given trial two intervals: a sentence presented in babble and the same sentence presented in same-spectrum noise. One interval would be presented at its SNR50 and the other at its SNR50 plus an increment varying from 0-8 dB in 2 dB increments. Listeners were asked to choose which sentence was clearer. All stimulus combinations and orders were counter-balanced and repeated 12 times.

The result of note was when there was a 0 dB increment (i.e., when both stimuli were presented at their SNR50). It was initially expected that listeners would choose each stimulus 50% of the time. Listeners on average, however, chose sentences in babble to be clearer 64% vs. 36% for sentences in same-spectrum noise [$t(19) = 7.81$; $p < 0.0001$]. In the rest of the conditions, this preference or "clarity gap" persisted. There was no correlation between the clarity gap and individual SNR50s nor individual differences in SNR50s. This particular result indicates a difference between objective, perceptual benefits and subjective, perceived benefits. If equivalent performance is not perceived as being equivalent in clarity across stimuli, perhaps an altogether different measure, such as effort, could yield subjective equivalence.

[Work supported by the Medical Research Council (grant number U135097131) and the Chief Scientist Office of the Scottish Government.]

7

Listening effort and difficult listening conditions: Lexical processing and neural entrainment with noise and vocoders

Anna Exenberger [a.exenberger@ucl.ac.uk], Paul Iverson
UCL, London

Listening effort, the increased attention associated with hearing under non-optimal conditions, has been measured in three ways: subjective measures, behavioural measures, and physiological measures. In this study, we examined all three with two aims in mind: first, to explore the relationship between attention, signal degradation types, and neural entrainment; and second, to relate different ways of measuring listening effort with one another. We created stimuli for a dichotic listening experiment with different types and levels of degradation. We prepared five different stimuli conditions: no signal degradation, three levels of SNR (+3 dB, -0.5dB, -4dB), and one vocoded condition (14 channels). Noise was speech-shaped noise matched to the spectrum of the target talker. SNR levels were determined in a pilot experiment and fixed to match different intelligibility levels. Participants were presented with a dichotic listening task and were instructed to focus on the female (target) speaker, while ignoring the male (distractor) speaker. Stimuli consisted of single sentences, and were either semantically predictable, unpredictable or semantically incoherent. In order to ensure participants' attention to the task, and to gain a behavioural measure of listening effort, they were instructed to press a button whenever they heard a semantically incoherent sentence. Subjective measures of listening effort were collected through a questionnaire administered at the end of the experiment. Results of this study indicate that cortical entrainment to the target speaker is higher than the distractor, which is in line with previous research. Entrainment and lexical processing are robust with increasing noise, then fall when the stimulus becomes (almost) unintelligible (-4 dB SNR). Unexpectedly, results from the vocoded condition indicate no entrainment advantage for the target speaker, even at an intelligible level. In future studies, we intend to explore the issue with different stimuli and listener groups.

8

Effects of noise suppression on spatial release from masking in simulated realistic listening environments

Tim Green [tim.green@ucl.ac.uk], Gaston Hilkhuysen, Mark Huckvale, Stuart Rosen
UCL London UK

Signal processing methods that eliminate noise from noisy speech signals have the potential to be beneficial in adverse listening environments, particularly for hearing-impaired listeners. However, existing noise suppression techniques have typically been developed to operate on monaural signals. Applying such methods independently to the two ears seems likely to diminish the advantages for speech recognition in noise that are derived from interaural timing and level differences. The present study examines the effect on spatial release from masking of the application of noise suppression based on minimal mean squared error spectral estimation independently to the two ears of normal hearing listeners in different simulated listening environments. These environments are implemented using head-related impulse responses and include an anechoic room, a university meeting room, a school classroom and a cafeteria. Speech reception thresholds (SRTs), defined as the signal-to-noise ratio at which 50% of words are correctly identified, are obtained for IEEE sentences. Target speech is always presented from 0° azimuth, while maskers, which include speech shaped noise and interfering speech, are either co-located with the target or spatially separated. Spatial release from masking, i.e., the improvement in SRT with spatial separation of target and masker, will be compared across conditions with and without noise suppression. The data presented will be informative regarding the extent to which non-linear noise suppression processing is detrimental to the effectiveness of binaural spatial cues and will be used in the development of binaural metrics that can predict the intelligibility of processed speech in realistic environments. This, in turn, will contribute to the development of binaural signal processing methods for hearing aids that reduce noise while preserving important spatial cues.

9

Do you hear the noise? Influence of background noise level on speech reception

Nina Wardenga [wardenga.nina@mh-hannover.de]

Department of Otolaryngology, Hannover Medical School, Hannover, Germany

Melanie A. Zokoll

Hörzentrum Oldenburg GmbH, Oldenburg, Germany

Birger Kollmeier

Medizinische Physik, Carl von Ossietzky Universität Oldenburg, Germany

Hannes Maier

Department of Otolaryngology, Hannover Medical School, Hannover, Germany

Objective

The aim of this study was to determine the relationship between hearing loss and speech reception threshold (SRT) at different fixed background noise level conditions using the German Oldenburg Sentence Test (OLSA). In a previous study (Wardenga et al., 2015), 177 subjects with various hearing abilities were tested at a fixed noise level of 65 dB SPL. With this standard setting, SRT in noise could be determined for listeners with pure-tone averages (PTA, 0.5, 1, 2, 4 kHz) below about 45 dB HL. Above this PTA, the SRT was affected to an increasing degree by problems with listening in quiet. The present two-center study investigated the effect of changing the level of the background noise to lower or higher levels.

Design

All subjects were trained with two lists of the OLSA in easy listening conditions. The SRTs were

determined monaurally with headphones using the standard noise of the OLSA (Olnoise) at different levels and a standard adaptive procedure converging to 50% speech intelligibility. The order of the noise levels was randomized.

At the first center (Hörzentrum Oldenburg), listeners with normal to moderate hearing loss were tested in quiet and with fixed noise levels of 55, 65, 75, and 85 dB SPL (in total N=41 ears). At the second center (Medical University Hannover), listeners with normal to severe hearing loss were tested in quiet and with fixed noise levels of 65, 85, and 95 dB SPL (N=52 ears). In total, data for 93 ears with hearing losses ranging from 0 to 90 dB HL PTA were obtained.

Results

For all noise levels, two domains could be identified with a linear dependence of SRT on PTA. For PTAs < noise level – 20 dB HL, the SRT increased with slopes of approximately 0.09 dB SNR/dB HL. For higher PTAs, the identified domain includes very heterogeneous data, thus the corresponding regression differ in slope for the different background noise level.

Conclusion

The OLSA can be applied to listeners with a wide range of hearing losses. Preliminary results indicate that for real speech in noise testing, the selected background noise level should be 20 dB higher than the PTA - otherwise, the SRT is influenced predominantly by a reduced hearing ability in quiet.

10

Speech in noise threshold measurements in cochlear implant users

Chris J James [cjames@cochlear.com]

Service ORL, CHU-Toulouse and Cochlear France, France

Chadlia Karoui

Centre de recherche Cerveau et Cognition, University Paul Sabatier, Toulouse, France

Mathieu Marx, Marie-Laurence Laborde, Marjorie Tartayre, Carol Algans, Olivier Deguine, Bernard Fraysse

Service ORL, CHU-Toulouse, France

We previously established that the sentence recognition scores of cochlear implant (CI) users one month after activation are similar to those for normal listeners listening to 8 12 channel acoustic (vocoder) simulations: Scores were near 100% in quiet and in 10 dB SNR as long as aetiology and duration of deafness were taken into account; that electrode insertion depths were between 300-400 degrees and that there was no scala dislocation of the electrode array (James et al., CIAP, Lake Tahoe, July 2017).

We reduced the results of fixed-SNR-level testing to SNR50 values to avoid ceiling effects in long-term sentence-in-noise data. We validated an analytic sigmoid fit to scores obtained from two SNRs per visit with an additional maximum, in-quiet score parameter. Fifteen CI users were tested twice in one session using French MBAA2 lists presented at progressively reducing SNRs

(10, 5, 2, 0) until scores were < 50/100. Test-retest reliability was 0.86 dB using a full Levenburg-Marquardt fit based on 3 or more fixed SNRs, and 0.84 dB using the simplified analytic fit. The mean difference in SNRs test to re-test was 0.14 dB ($p=0.65$, $t[14]=0.462$).

The simplified analytic method is useful for clinical testing since it allows the standard collection of scores for fixed SNRs for longitudinal follow-up but also allows expression of results in terms of dB SNR50 to accommodate the large range of subject performance. For example, we will present preliminary analysis of patient- and device-related factors which may influence SpiN scores for CI users over time.

We will also present some clinical uses of the method to determine critical differences; for example to determine the spatial unmasking benefit of a hearing aid used with a CI (i.e. bimodal listening) and to compare measures of speech recognition performance for CI-alone in implanted unilateral hearing loss patients where acoustic leak may be an issue.

11

An information-theoretic analysis of auditory features in noisy environments

Lotte Weerts [lw1115@ic.ac.uk]
Imperial College London

Current engineering efforts such as automatic speech recognition (ASR) systems and cochlear implants tend to perform much worse than the human auditory system in noisy environments. We aim to address this problem by developing auditory-inspired features that are robust to noise. Here, we review several general principles that appear to be important for the noise robustness of the auditory system, such as precisely timed inhibition. These principles are combined with a basic model of the auditory nerves (a half-rectified Gammatone filterbank) to create a new set of auditory features. To assess the quality of these features, we use information theoretic measures. Altogether this new class of features may allow us to both improve current understanding of general principles in the auditory system, as well as finding new features that could be directly applied to ASR systems.

12

Revision of a binaural model predicting speech intelligibility against envelope-modulated noise interferers

Thibault Vicente [thibault.vicente@entpe.fr], Mathieu Lavandier
Univ Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, Rue M. Audin, 69518 Vaulx-en-Velin Cedex, France

Collin and Lavandier (2013) developed a preliminary model predicting binaural speech intelligibility against non-stationary noise interferers in rooms for normal-hearing listeners. The model has four parameters (frequency resolution, SNR ceiling, temporal resolutions of the better-ear listening and binaural unmasking components), the influence of which was not thoroughly tested. The aim of the present work was to realize a parametric study - based on a sensitivity analysis - on the model parameters to optimize their values using several experiments with conditions critically testing the model.

This study used the data from five experiments, four from the literature and one realized during this work. The data from the literature were taken from the two experiments of Culling and Mansell (2013) and from two of experiments of Collin and Lavandier (expt 1 and 4, 2013). Culling and Mansell used noise interferers with an envelope artificially modulated by a square wave with different modulation rates. Their stimuli isolate the two components of spatial unmasking, which is interesting to test their temporal resolutions separately. Collin and Lavandier used noise interferers with speech-modulated envelope. The first experiment is relevant to test the model predictions when reverberation is filling in the interferer gaps for different modulations depth. The second experiment allows testing the better-ear glimpsing component of the model for speech modulations in the interferer. The additional experiment realized during this study further tested the influence of reverberation, varying the modulation depth of speech-modulated noise interferers, in binaural conditions involving better-ear glimpsing and binaural unmasking in combination and in isolation.

The sensitivity analysis allows quantifying the impact of each model parameter on the predictions for the five experiments. It also highlights the potential interactions between these parameters. The main criterion used to evaluate model performance was the mean absolute error between data and prediction. The maximum absolute error and the correlation between data and prediction were also considered. The parameter values were revised to describe the different perceptual effects involved across the five experiments. Finally, the revised model will be tested on data not used to define its parameter (Ewert et al, 2017). This data set is interesting because it involves speech-like masker and maskers based on a stationary speech-shaped noise, while isolating the two components of spatial unmasking.

13

Laboratory paired comparisons based on the CoSS framework

Karolina Smeds [karolina.smeds@orca-eu.info]
ORCA Europe, Widex A/S, Sweden

In a previous study, we compared two hearing-aid settings in the field and in the laboratory using paired comparisons. We found that the correlation between the data collected in the laboratory and in the field was low. Specifically, the laboratory data could not predict the individual overall preference in the field (Dahlquist et al 2015). Potential explanations for the low predictive value of the laboratory test included the selection of the laboratory sound stimuli, and the artificial task used in the laboratory.

In another study, we investigated the listening situations people encounter in real life. Based on data from a literature study, a framework called the Common Sound Scenarios (CoSS) was developed. Three intention categories were formed: “Speech communication”, “Focused listening" (without own speech), and "Non-specific" (including monitoring surroundings and passive listening) (Wolters et al 2016).

When studying the CoSS framework, it becomes obvious that most laboratory tests only tap into the "Focused listening" intention category. Neither real speech communication, nor more passive listening situations are usually included in the laboratory.

In the current study, a new laboratory paired-comparisons paradigm was tested. The method focused on a test participant's intention and the task the test participant had to solve in a scenario. Five mandatory scenarios (representing all three intention categories in CoSS) and up to six individual test scenarios were included. The individual test scenarios were selected from a list of situations experienced in an accompanying field trial. Real conversations between the test participant and one or two test leaders was central to the method, but we also included ecologically valid scenarios with focused listening and scenarios including sound monitoring and passive listening.

Test participants judged the new laboratory test to be ecologically valid. When the results from the field and the laboratory were compared, the correspondence was satisfactory, even though the background sounds experienced in the field were not strictly matched in the laboratory, and the laboratory loudspeaker setup was simple. Laboratory testing with a focus on scenarios with a variety of intentions and tasks and on commonly experienced sounds can of course be used with more acoustically accurate setups.

The pros and cons of the method will be discussed.

References

Dahlquist M, Larsson J, Hertzman S, Wolters F, Smeds K. (2015) Predicting individual hearing-aid preference in the field using laboratory paired comparisons. In: 5th International Symposium on Auditory and Audiological Research (ISAAR). Nyborg, Denmark.

Wolters F, Smeds K, Schmidt E, Christensen EK, Norup C. (2016) Common Sound Scenarios: A context-driven categorization of everyday sound environments for application in hearing-device research. Journal of the American Academy of Audiology 27: 527-40.

14

Do noise-induced latency shifts of the auditory brainstem response to speech reflect degradation in neural synchrony?

Jessica De Boer [Jessica.deBoer@nottingham.ac.uk]
MRC Institute of Hearing Research

Helen E. Nuttall
Department of Psychology, Lancaster University

Katrin Krumbholz
MRC Institute of Hearing Research

Understanding speech in noisy environments is crucial for human communication. Noise is known to increase the latency of the auditory brainstem response (ABR) to speech sounds, and this has been suggested to reflect noise-induced degradation in the neural temporal processing of speech. However, speech ABR latencies are also strongly influenced by the distribution of response contributions from different cochlear regions, tuned to different frequencies. This is, because cochlear regions tuned to lower frequencies have considerably slower responses than regions tuned to higher frequency regions. Thus, if noise masking changed the cochlear distribution of the speech ABR, then this might provide an alternative explanation for the noise-induced increase in speech ABR latency. The aim of the current experiment was to investigate this by examining the effect of noise masking on speech-evoked ABRs from frequency-restricted cochlear regions. We used the 'derived-band' technique to obtain speech ABRs from octave-wide regions centred at 0.7, 1.4, 2.8 and 5.6 kHz. Frequency-restricted 'derived-band', and unrestricted ('broadband') speech ABRs were recorded both in quiet and in noise. Consistent with previous findings, noise masking caused a significant increase in the latency of the broadband speech ABR. In contrast, however, the noise effects on the latencies of the derived-band speech ABRs were invariably small, and also inconsistent across bands. Instead, the predominant effect of noise on the derived-band speech ABRs was to change the distribution of their amplitudes: with increasing noise level, the predominant amplitude moved from the 2.8-kHz band down to the 0.7-kHz band. These results suggest that the latency increase of the speech ABR as a result of noise is primarily caused by a cochlear place mechanism.

15

Relating speech perception in noise to temporal-processing auditory capacities in childhood: effects of typical development and of sensori-neural hearing impairment

Laurianne Cabrera [laurianne.cabrera@gmail.com], Lorna Halliday

Department of Speech, Hearing & Phonetic Sciences, UCL, London, UK

Christian Lorenzi

ENS

Stuart Rosen

Department of Speech, Hearing & Phonetic Sciences, UCL, London, UK

Temporal cues (e.g., amplitude modulation, AM) play a crucial role in speech intelligibility for adults. This study aims to characterize the development of sensory (AM encoding) and non-sensory mechanisms (i.e., processing efficiency) constraining auditory temporal processing and their relationship with speech in noise perception (SIN) for normal-hearing (NH) and hearing-impaired (HI) children.

Eighty-one NH children from three groups of age ranging from 6 to 10 years participated in the study, and pilot data were collected from eight 5-to-7-year-old children with sensorineural mild-to-moderate HI.

Three 3I-3AFC adaptive tasks were designed. The first task assessed AM sensitivity using pure tone carriers and 3 modulation rates (4, 8, 32 Hz). The second task assessed AM masking by comparing AM detection thresholds at 8 Hz modulation using 3 carriers varying in their inherent AM fluctuations: tones, narrowband noises with small inherent AM fluctuations and noises with larger fluctuations. The third task assessed temporal integration, the effect of increasing the number of AM cycles (between 2 and 8 cycles) on AM detection using tones modulated at 4 Hz. Finally, identification thresholds were measured in speech-shaped noise using fricative consonants (/f/-/v/-/ʃ/-/ʒ/-/s/-/z/).

For the NH children, overall the youngest group (6-7-years) had poorer AM detection thresholds in all tasks. However, their thresholds were affected by AM rate, large carrier fluctuations and number of AM cycles in the same way as older children, suggesting that sensory mechanisms may be mature by 6 years. Nevertheless, only 6-7-year-olds did not show any difference between tone and noise with small fluctuations, reflecting perhaps a change in the magnitude of internal noise with age for AM processing, and lower processing efficiency. Regarding SIN, regression analyses indicated that better AM detection with 8 cycles predicted better SIN thresholds (R2adj=20.9%). Moreover, the slope of the temporal-modulation transfer function (an estimate of temporal acuity) predicted an additional 3.5% of the variance.

Preliminary results with HI children show good AM detection abilities and similar effects of rate, carrier fluctuations and AM cycles on AM detection, but not surprisingly, worse SIN thresholds than NH children. Regarding the relationship between AM tasks and SIN, the same trend as NH was observed. These results suggest that SIN perception in childhood is related to temporal processing, and better AM processing might contribute to better SIN perception for children with HI. Computational modelling will help to better disentangle the relationship between sensory and non-sensory processing of AM in SIN during childhood.

16

Speech-perception-in-noise deficits in dyslexia: a possible method to improve speech perception abilities

Tilde Van Hirtum [tilde.vanhirtum@kuleuven.be], Jan Wouters

ExpORL, Department of Neurosciences, KU Leuven, Belgium

Pol Ghesquière

Parenting and Special Education, KU Leuven, Belgium

The exact cause of developmental dyslexia, a specific learning disorder characterized by severe reading and spelling difficulties, remains widely debated. Growing evidence exist that dyslexia is related to a temporal processing deficit. Hence, a low-level auditory dysfunction to process temporal information, might result in a subtle speech perception deficit and in turn interfere with the development of phonological representations and literacy skills. Previous results from our ongoing work indeed reveal low-level temporal deficits, mainly with processing specific onset cues. Moreover, these transient parts of the speech signal, such as onsets, are most important for speech intelligibility in normal-hearing listeners. Therefore we hypothesize that enhancing these particular cues in the speech signal might benefit speech processing in adults with dyslexia. An Envelope Enhancement strategy (EE, Koning and Wouters, 2012) was implemented to amplify the onsets of the speech envelope without affecting other parts of the speech signal.

In the present study we investigated (1) speech-in-noise abilities in a group of dyslexic and normal reading adults and (2) the potential of the EE strategy in dyslexia research. We tested speech understanding in four different conditions: natural speech, vocoded speech and their enhanced versions. A constant procedure was followed using four fixed signal-to-noise ratios for each condition. Additionally, cognitive test of phonological awareness, language skills, verbal short term memory and working memory were administered to investigate possible confounding effects. The results of this study will be discussed at the conference.

17

Caught by surprise: the effect of unpredictable varying talker location on speech intelligibility and listening effort

Annelies Devesse [annelies.devesse@kuleuven.be], Alexander Dudek, Astrid van Wieringen, Jan Wouters

ExpORL, KU Leuven, Belgium

Background - Having a successful conversation in a crowded environment can be challenging, especially if one has to divide his/her attention over multiple conversation partners. In such complex listening situations, cognition and multitasking skills are of particular importance. Furthermore, head orientations play an important role in improving speech understanding. On the one hand, listeners try to face the speaker to lip read the message. On the other hand, moving the head away from the speaker allows the listener to use interaural differences to improve speech intelligibility (SI), by means of head shadow effects, redundancy and binaural squelch. Research showed that both lip reading and binaural advantages can be combined for head movements up to 30° (1). Nevertheless, not knowing upfront who will speak can hamper appropriate head orientations, possibly impairing both SI and listening effort (LE). In this study, auditory-visual multitalker situations were administered to investigate the effect of unpredictable varying talker locations on both SI and LE.

Methods - 30 normal hearing, Dutch speaking adults (18-30 years) participated. A challenging listening situation was mimicked by means of an audio-visual environment and five virtual human-like characters. In this scenario, participants performed a behavioural speech-in-noise task with sentences either coming from a predefined direction (0°, fixed condition) or from randomly assigned directions between -45° to +45° in the horizontal field (random condition). A uniform noise field was created via a loudspeaker right above the participant's head. Speech was presented auditory-only or auditory-visually (AV) and free head movements were allowed. In an extra condition, the AV speech-in-noise task was combined with three secondary tasks to investigate LE. It is known that a drop in performance on the secondary tasks, compared to their respective baseline conditions, reflects an increase in LE.

Results - SI was significantly better in the random listening conditions with regard to the fixed conditions. Performance on one of the secondary tasks significantly dropped when combining four tasks together, compared to its baseline condition. However, performance on the secondary tasks did not differ between the random and fixed conditions. First analysis of these results suggest good performance when speech was presented from varying talker locations compared to a fixed location, possibly due to binaural listening benefits. Furthermore, both listening conditions required an equal amount of LE.

1. Grange & Culling (2016) JASA, 140(6), pp. 4061–4072.

Funding bodies: the Oticon Foundation and a TBM-FWO grant from the Research Foundation-Flanders (nr. T002216N).

18

Can biophysically inspired features improve neural network-based speech enhancement?

Deepak Baby [deepak.baby@ugent.be], Sarah Verhulst
Gent University

Recent advances in neural network (NN)-based speech enhancement schemes outperform most of the conventional techniques and are less sensitive to the increase in input dimensionality and correlation between features. However, the performance of the state-of-the-art speech enhancement systems in adverse noisy conditions such as negative signal-to-noise ratios (SNRs) and unseen noise conditions has not been extensively studied. In addition, the performance of these systems is still far from that of humans especially in such difficult noisy conditions. Therefore, there is a growing interest in feature-related research that focuses on applying our knowledge about human auditory processing into the NN framework.

This work investigates the use of various biophysically inspired cochlear models as input to a NN for speech enhancement in adverse and unseen noise conditions. The various cochlear models investigated are Mel frequency spectral coefficients (FBANK), gammatone energies (GT), dynamic-compressive gammachirp (DCGC), dual resonance nonlinear filterbank (DRNL), cascade of asymmetric resonators with fast acting compression (CARFAC) and non-linear transmission-line (TL) model. The DNN setting is comprised of 3 hidden layers with sigmoid activations. The NNs are trained such that they minimize the mean square error between the NN output and the desired mask for optimal noise suppression.

The NNs are trained using noisy speech containing babble noise with positive signal-to-noise ratios (SNRs) and are evaluated using noisy speech containing babble and ICRA noise types with negative SNRs. The quality of noise suppression is evaluated using various objective measures such as short-time objective intelligibility measure (STOI), cepstral distance (CD), log-likelihood ratio (LLR) and segmental SNR (segSNR).

19

The pupil dilation response of adults with acquired brain injury during speech processing in noise

Thomas Koelewijn [t.koelewijn@vumc.nl], Sophia E. Kramer

Section Ear & Hearing, department of Otolaryngology-Head and Neck Surgery and Amsterdam Public Health research institute, VU University Medical Center, Amsterdam, The Netherlands

A significant group of adults with acquired brain injury (ABI) that have a normal pure tone audiogram, report difficulties and high levels of effort when listening to speech in noise. Listening difficulties in this group seem to be due to disturbed cognitive functioning when processing auditory information, independent of the peripheral hearing system. Notably, almost one third of people with ABI experience high levels of fatigue. According to the Framework for Understanding Effortful Listening (FUEL), fatigue may have a mediating effect on listening effort. To investigate this, we examined how ABI affects speech processing in noise and listening effort.

Twenty adults with ABI (aged 26-62 years) participated. All had a normal pure-tone audiogram (PTA) but reported difficulties with speech processing in noise. The participants listened to sentences masked by fluctuating noise or a single-talker at fixed intelligibility levels of 50% and 84% full sentence correct performance using a staircase procedure. The pupil diameter was recorded during each trial and used as an index of listening effort. Additionally, participants performed the Text Reception Threshold (TRT) task - a visual sentence completion task - measuring language related processing. Finally, data were compared to the results obtained for normally hearing and hearing impaired age-matched groups in earlier studies using the same design and setup.

A direct comparison between the current data and data recorded in a previous study including normal hearing participants with no neurological problems, revealed significant worse Speech Reception Thresholds for participants with ABI, which confirmed their reported hearing difficulties. This, while PTA as well as the performance on the TRT task was the same for both groups. While the same pattern of results for the pupil dilation response was shown over groups, self-rated effort scores were significantly higher for the ABI group compared to the normal hearing and hearing impaired groups.

Based on the outcomes of this study, we conclude that ABI and related fatigue affected speech perception most probably at a central auditory processing stage, which seemed to impact the participants listening effort.

Decoding Attention at Higher Levels of Linguistic Processing using EEG

Michael P Broderick [brodermi@tcd.ie]
Trinity College Dublin, Ireland

Andrew J Anderson
University of Rochester, New York

Giovanni M Di Liberto
PSL Research University, Paris France

Edmund C Lalor
University of Rochester, New York

Recent discoveries have shown that single trial (~60s) unaveraged EEG data can be decoded to determine attentional selection in a multi-speaker environment. This is achieved by using recorded neural data to reconstruct an estimate of the speech envelope and then comparing that reconstruction with the attended and unattended speech streams. This result is particularly impressive given that it is based on the envelope of speech. There are two reasons for this: 1) the envelope is a very simplified representation of speech meaning that the corresponding EEG measures may be relatively poor in terms of reflecting speech-related activity; and 2) attention effects tend to be stronger at higher levels of sensory-perceptual hierarchies meaning that attention measures based on the envelope may be quite small. Here, we sought to investigate whether the addition of EEG measures reflecting higher-level linguistic processing could lead to improved decoding of attention in a "cocktail party" experiment.

We recorded EEG as subjects listened to one of two concurrently presented stories. We represented each story in terms of its semantic content by using computational language models to quantify the meaning of each word in a sentence in terms of how semantically dissimilar it was to its preceding context. We then regressed the EEG data against this semantic representation to produce Temporal Response Functions (TRFs; i.e., beta regression weights) for both the attended and unattended story.

These TRFs display a prominent negativity at time-lags of ~200-600ms over centro-parietal electrodes, sharing similar characteristics to those of the classic N400 response. Importantly, this negativity is consistent across subjects and is exquisitely sensitive to whether or not subjects were understanding the speech they heard. As such, it is consistently present across subjects for attended speech and absent for unattended speech. By including measures of this semantic TRF along with envelope reconstruction, we show improved decoding of cocktail party attention using unaveraged EEG.

The role of offset sensitivity in consonant discrimination in noise

Fatima Ali [fatima.ali.15@ucl.ac.uk]
UCL Ear Institute, London, UK

Doris-Eva Bamiou
The Royal National Throat, Nose and Ear Hospital, London, UK

Stuart Rosen
Department of Speech, Hearing & Phonetic Sciences, UCL, London, UK

Jennifer F Linden
UCL Ear Institute, London, UK

Sound offsets are important cues for recognising, distinguishing and grouping sounds, but the neural mechanisms and perceptual roles of sound-offset sensitivity remain poorly understood. In particular, while it is known that troughs in amplitude modulation are essential to consonant perception, there is a gap in the literature relating physiological studies of sound-offset responses in the auditory brain to the psychophysics of speech perception.

Recent studies in a mouse model of developmental disorder have reported the discovery of an auditory deficit specific to the processing of sound offsets (Anderson & Linden, 2016). This finding raises the possibility that deficits in sound-offset sensitivity might contribute to listening difficulties associated with developmental disorders. Difficulty perceiving speech in noise is the characteristic feature of central auditory processing disorder, and is also associated with other developmental or language disorders.

Here, we used mathematical modelling to investigate how sound-offset sensitivity relates to discrimination of vowel-consonant-vowel (VCV) stimuli in multi-talker babble noise. We used a phenomenological model introduced by Anderson and Linden (2016), based on the assumption that auditory brain activity arises from a sum of inputs from independently weighted onset-sensitive and offset-sensitive channels. By reducing the weighting of the offset-sensitive channel, we simulated reduced offset sensitivity and assessed its influence on the discriminability of model outputs for 48 non-sense VCV speech stimuli in varying levels of multi-talker babble noise (-12, -6, 0, 6, 12 dB SNR). We show that offset salience in noise can be used to categorise phonetic consonants into three groups of high, moderate and low salience, and we identify particular consonants for which discrimination in noise is more strongly or more weakly affected by offset sensitivity. We also report the results of an on-going psychophysical study of offset sensitivity and VCV perception in normal healthy subjects aged 18-60, comparing ratios of sound-onset to sound-offset reaction times with thresholds for gap-in-noise detection and VCV discrimination in noise. Consistent with model predictions, our preliminary results show that differences in consonant discrimination performance for consonants with markedly different offset salience ('w'-'d') varied more strongly with our measure of offset sensitivity than

differences in performance for consonants with similar offset salience ('f'-'sh'). We anticipate further testing will reveal greater insight into the role of sound-offset sensitivity in auditory processing.

Reference

Anderson, L.A. & Linden, J.F. 2016. Mind the gap: two dissociable mechanisms of temporal processing in the auditory system. Journal of Neuroscience, 36 (6), 1977-95.

22

An assessment of objective intelligibility metrics for signals with low mixture signal-to-noise ratios after enhancement using Ideal Binary Masks

Simone Graetzer [s.graetzer@liverpool.ac.uk], Carl Hopkins
Acoustics Research Unit, University of Liverpool, England

This study concerns how well objective indicators of speech intelligibility correlate with the percentages of words that are correctly identified when speech is mixed with white Gaussian noise at low signal-to-noise ratios (SNRs) and subsequently enhanced with Ideal Binary Masks (IBMs). Such masks require a priori knowledge of both the target signal and the masker. The objective indicators under consideration include Short-Time Objective Intelligibility (STOI), which is suitable for both noisy and degraded speech, including non-linearly processed or time-frequency weighted speech. STOI involves the correlation of the envelopes of clean and degraded (or processed) speech signals that have been divided into overlapping short-time (384 ms) segments. In this study, we mixed speech produced by two male and two female speakers of British English with white Gaussian noise at SNRs as low as -25 dB. Signals were subsequently enhanced using IBMs with a Local Criterion equal to zero or to the mixture SNR. Listening tests involving normal-hearing human listeners were carried out, where each signal was presented to the listener three times. The results characterise the relationship between the objective indicators and the percentages of words correctly identified by the listeners in the context of low mixture SNRs.

23

On the dynamics of the preference-performance relation for hearing aid noise reduction

Rosa-Linde Fischer [rosa-linde.fischer@sivantos.com]
Sivantos GmbH

Kirsten C. Wagener, Matthias Vormann
Hörzentrum Oldenburg GmbH

Tobias Neher
University of Southern Denmark, Institute of Clinical Research

Previous research has shown that hearing aid users can differ substantially in their preference for noise reduction (NR) strength, and that preference for and speech recognition with NR processing typically are not correlated (e.g. Neher 2014; Serman et al. 2016). In other words: hearing aid users may prefer a certain NR setting, but perform better with a different one.

The aim of the present work was to investigate the influence of individual noise sensitivity, hearing aid experience and acclimatization on the preference-performance relation for different NR settings. For this purpose, a longitudinal study with three consecutive laboratory assessments distributed over a 12-week period was conducted. Two experimental groups of experienced and inexperienced hearing aid users (N = 20 each) participated. These subjects were bilaterally fitted with hearing aids (HA) with individual NAL-NL1-based hearing loss compensation (HL). All adaptive HA parameters were set to the default values in the fitting software. A control group of experienced hearing aid users (N = 10) completed the study with their own HAs.

All participants were selected based on their preferred NR strength as assessed during an initial screening visit (N = 100). Care was taken to ensure that the groups were comparable in terms of age and hearing loss, and that there were notable differences in preferred NR strength ("NR haters" vs. "NR lovers") within each experimental group.

The laboratory assessments of preference and performance were conducted with four different NR settings: (1) only HL compensation, no further signal processing, (2) HL compensation and single-channel NR, (3) HL compensation and directional microphones (DIR), and (4) HL compensation combined with DIR and NR. Preference was assessed with a spatial dynamic speech-in-noise task after Getzmann et al. (2015) that required the participants to attend to a target speaker while ignoring simultaneously occurring distractor talkers. In addition, speech understanding and recall was assessed using a listening span test (Fischer et al. 2017).

Here, we report on the data collected during the first laboratory assessment of the study. In particular, the influence of hearing aid experience and individual noise sensitivity on the preference-performance relation will be presented and discussed.

24

Periodicity is not a factor in making harmonic complexes less effective maskers of speech than noise

Stuart Rosen [stuart@phon.ucl.ac.uk], Kurt Steinmetzger

Department of Speech, Hearing & Phonetic Sciences, UCL, London, UK

David M Perry

Ear Institute, UCL, UK

Harmonic tone complexes, whether dynamic or static, are much less effective maskers than a noise with the same overall spectral envelope, but the reasons for this are not yet clear. For one thing, the periodicity of the complex may allow it to be more effectively segregated or 'cancelled out' than an aperiodic noise. On the other hand, the masking of modulations in the speech by modulations in the masker may be important, and the modulation spectra of noises and harmonic complexes are very different.

We compared the relative masking effectiveness of static and dynamic complexes in which the discrete spectral components form either a harmonic or inharmonic series. The harmonic dynamic complexes have continuous modulations in fundamental frequency (F0) modelled on genuine F0 contours found in connected speech. Static complexes varied in F0 from trial to trial to match the distribution of F0s in the dynamic ones. In addition, median F0s were varied in relation to the target sentences spoken by a relatively high-pitched male speaker (F0 ≈ 150 Hz), to be low (≈ 100 Hz), medium (≈ 150 Hz) or high (≈ 225 Hz).

Inharmonic complexes were created in two ways, either by shifting all the components in the harmonic series up or down by 25% of the median F0, or by spectrally rotating the harmonic complexes around a centre frequency near 2 kHz. For static contours, these two methods are equivalent for an appropriate choice of parameters. For dynamic contours, the actual shift in component frequencies changes throughout the stimulus, sometimes more and sometimes less than 25%, but the resulting sound is typically inharmonic. Crucially, the modulation spectra of all three variants of the static complexes are essentially identical, but are very different for the frequency-shifted and spectrally-rotated dynamic complexes.

Speech Reception Thresholds (SRTs), determined adaptively, revealed the following: SRTs tend to decrease with increasing masker F0. SRTs are generally worse for dynamic as opposed to static contours. Inharmonic and harmonic complexes lead to similar SRTs, except for the rotated dynamic complexes, which are more effective maskers than the other types, especially for the two higher masker F0s. It thus appears that an adequate theory for explaining the difference in masking effectiveness between harmonic complexes and noises must consider the role of modulations.

This work was supported in part by the Medical Research Council, UK (Grant Number G1001255).

25

A binaural model predicting speech intelligibility in noise for hearing-impaired listeners

Mathieu Lavandier [mathieu.lavandier@entpe.fr]
Univ. Lyon, ENTPE, France

Jörg M. Buchholz, Baljeet Rana
National Acoustic Laboratories/Macquarie University

A binaural model is presented which predicts the effect of audibility on the intelligibility of speech in the presence of speech-shaped noise and vocoded speech maskers. It is based on the short-term binaural speech intelligibility model described by Collin & Lavandier [J. Acoust. Soc. Am. 134, 1146-1159 (2013)] and takes the calibrated target and masker signals (independently) at

each ear as inputs along with the listener hearing thresholds in order to calculate a binaural "effective" signal-to-noise ratio. Differences in ratio across conditions can be directly compared to differences in speech reception threshold (SRT).

Model predictions are compared to SRTs measured in the presence of two speech-spectrum noises or two vocoded-speech maskers, which were either (artificially) spatially separated or co-located with the frontal speech target. The spatial separation was realized by presenting each masker on a different single ear using headphones, while the target was presented diotically as coming from the front. Comparing the co-located and separated configurations allows evaluating a spatial release from masking (SRM) which was based here primarily on better-ear glimpsing (no realistic ITDs simulated). Normal-hearing and hearing-impaired listeners were involved in the data collection, which mostly used stimuli spectrally shaped for equalized audibility across listeners. Audibility was varied during the experiment by testing four broadband sound levels for the combined maskers (while the target level was varied relative to these reference levels to measure the SRTs).

For both group of listeners, the model allows a good prediction of the decrease of SRT as well as the increase of SRM with increasing audibility/levels. The averaged absolute error between data and prediction was generally below 1 dB across tested conditions.

26

The role of periodicity in the perception of masked speech with simulated and real cochlear implants

Kurt Steinmetzger [kurt.steinmetzger.12@ucl.ac.uk]
University of Heidelberg

Stuart Rosen
University College London

In normal hearing, periodic sounds are much less effective maskers of speech than aperiodic ones [Steinmetzger and Rosen (2015). J. Acoust. Soc. Am. 138, 3586–3599]. Here, it is shown that this masker-periodicity benefit is strongly reduced in noise-vocoder simulations of cochlear implants (CIs) and almost absent with real CIs, which helps explain why CI users face such difficulties when attempting to understand speech in noisy environments. Furthermore, when there were no F0-related periodicity cues in the target speech, CI users performed markedly worse with maskers whose envelopes were adjusted to be the inverse of the target sentence, while no such effect was observed with simulated CIs. Moreover, neither group of listeners showed a masking release with these maskers, although they were designed to maximise glimpsing opportunities. In both cases, the difference between simulated and real CIs is thought to be caused by current spread across the CI's electrode array: Firstly, CI users seem unable to perceive the random envelope modulations characteristic for aperiodic sounds, which diminishes the contrast between the aperiodic and periodic maskers. Secondly, the segregation of speech and

masker is complicated further by spectral smearing, explaining the more prominent role of periodicity cues in the target speech.

27

Analysis of the individual listening effort reflected by the pupillary responses during speech perception in noise

Patrycja Książek [paks@eriksholm.com], Dorothea Wendt, Emina Alickovic, Thomas Lunner

Eriksholm Research Centre, Snekkersten, Denmark

Pupillometry has been applied to investigate listening effort involved in speech recognition in adverse listening situations. It has been demonstrated that the pupil dilation can be sensitive to the cognitive resources that are allocated to perform a task such as speech recognition in noise. Commonly, pupillary response is analyzed based on the measure of the maximum diameter or the mean diameter of the dilating pupil while performing a task. Recent studies performed a Growth Curve Analysis (GCA, multilevel regression technique designed for analysis of time course or longitudinal data) on the pupillary data in order to account for time-dependent changes of the pupillary response. So far, analysis of the pupillary responses aimed to examine changes in listening effort at a group level. The objective of the current study was to analyze and classify individual pupil traces in order to examine individual's listening effort. The present work provides some preliminary results on the classification of the high/low listening effort as reflected by the individual pupil traces recorded during a speech-in-noise test. The approach was based on defining a multilevel regression to model pupil responses. Furthermore machine learning (ML) algorithms were applied to derive a deeper knowledge on changes in pupil response due to changes in the listening effort. In a first step, GCA, with a polynomial functional basis, was applied to fit the pupil traces from a limited set of 22 individuals performing a task with either high or low demands (recorded by Wendt et al., 2017). By employing the GCA, time-dependent, differentiable features of the individual's pupil responses were extracted. In a second step, it was studied how these estimated GCA parameters can be used to build a 'high/low listening effort' classifier by assessing the performances of different classifiers found in ML research. Preliminarily finings indicated that the best obtained detection accuracy of listening effort was higher than 80% (36/44). Our preliminary results demonstrated that linear and cubic parameters were highly relevant for separating the pupil curves. Moreover, both sensitivity (correctly identified high effort) and specificity (correctly identified low effort) of the classifier were above 80%. Analysis of the individual pupil traces showed furthermore that some pupil responses were too weak or too irregular within the class to be modelled with the chosen 3rd order polynomial basis. In general, this study provides a first approach to quantify and classify individual's pupillary response recorded within a speech-in-noise test. First results are encouraging that this approach can give an insight into the individual's listening effort involved in speech recognition in adverse listening situations.

Improving localization in binaural beamforming for hearing aid wearers

Nadja Schinkel-Bielefeld [nadja.schinkel-bielefeld@sivantos.com], Christos Oreinos, Homayoun Parsi Kamkar

Sivantos GmbH, Erlangen, Germany

Hearing impairment often goes along with diminished frequency selectivity which can lead to speech understanding problems in noisy environments. In hearing aids this is typically addressed by using an advanced binaural beamforming algorithm to effectively attenuate background noise as well as non-target interfering speakers in "cocktail party scenarios" (Kamkar-Parsi et al., 2014). This improves speech understanding and decreases listening effort (e.g. Mejia et al., 2017; Bernarding et al., 2014). In specific laboratory situations, it can even result in better speech understanding of hearing impaired subjects compared to normal hearing unaided subjects (Fröhlich et al., 2015).

Another important aspect of binaural hearing is the ability to localize sound in order to be aware of one's surrounding and speakers' location. However, this is more difficult to maintain in binaural beamforming than in the omnidirectional mode. Binaural beamforming improves the Signal-To-Noise Ratio (SNR), but it also distorts spatial binaural cues to some extent. Here, we report on a study investigating a new algorithm (implemented in the Signia Nx platform) that improves the preservation of spatial cues by restoring phase information in the low frequencies while maintaining target speech understanding.

In a laboratory test with 19 hearing impaired subjects we compared Speech Reception Threshold (SRT) and localization ability for hearing aids with three algorithms: TrueEar program (i.e. mimicking the natural directivity of head and pinna), the beamforming algorithm used in a predecessor implementation and the improved beamforming algorithm (Signia Nx) with better preservation of spatial cues. Subjects were on average 69.6 years old (SD: 15.6 years) with a mean pure tone average of 54.3 dB. Hearing aids were individually fitted for each subject.

The localization performance was best in TrueEar mode. When using the binaural beamforming with phase cue restoration (Signia Nx implementation), localization performance was significantly improved compared to the predecessor implementation. Most importantly, a speech understanding in noise test showed no significant difference between the two beamforming modes.

It is also notable that subjects seem to adapt their localization strategy to the beamforming mode, i.e., their performance improved with time in the predecessor beamformer and decreased in True Ear. Such an adaptation effect was not observed for the enhanced (Signia Nx) beamformer.

This could indicate that in real life where hearing aids change between TruEar and beamforming modes, it is beneficial to have similar localization cues in both modes so that less adaptation is necessary and overall localization performance remains more stable.

Pupil dilation during the speech understanding task in dark and light - potential influence of hearing impairment to the parasympathetic nervous system

Yang Wang [y.wang@vumc.nl], Sophia E. Kramer

Section Ear & Hearing, Dept. of Otolaryngology-Head and Neck Surgery, VU University Medical Center and Amsterdam Public Health research institute, Amsterdam, The Netherlands

Dorothea Wendt

Technical University of Denmark, Department of Electrical Engineering, Lyngby, Denmark

Thomas Lunner

Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark

Graham Naylor

Medical Research Council/Chief Scientist Office Institute of Hearing Research - Scottish Section, Glasgow, UK. Part of The University of Nottingham

Barbara Ohlenforst, Adriana A. Zekveld

Section Ear & Hearing, Dept. of Otolaryngology-Head and Neck Surgery, VU University Medical Center and Amsterdam Public Health research institute, Amsterdam, The Netherlands;

Pupil dilation response during the speech understanding in noise task has been widely used as a measurement to evaluate listening effort. Larger pupil dilation is associated with higher level of effort during the speech in noise task. One may intuitively assume that listeners with hearing impairment would experience more effort than normally hearing listeners when intelligibility levels are similar for both groups. Consequently, we should then expect that hearing-impaired listeners to show a larger pupil dilation than their normally hearing peers. However, previous studies have repeatedly reported the opposite findings that in challenging listening conditions, the pupil dilation was smaller in hearing-impaired participants than normal hearing controls. One of the potential explanations for this contradiction between hypothesized and observed effects of hearing impairment involves interactions with the parasympathetic nervous system. Reduced parasympathetic inhibition during cognitive processing could lead to smaller peak pupil dilation when testing in light conditions, evidence indicates that people with hearing impairment might have reduced parasympathetic activity. Therefore, the aim of the current study was to assess the influence of hearing impairment to the parasympathetic nervous system by comparing the pupil dilation response during the speech understanding in noise tasks in dark (the activity of the parasympathetic nervous system is minimum) and light (the inhibition of the parasympathetic nervous system helps to dilate the pupil) conditions.

Two age-matched groups of listeners participated in this study: 19 with hearing impairment and 27 normal-hearing listeners. They performed SRT tests with a single competing talker, to estimate the speech-in-noise-ratio required for 50% correct responses. The tests were conducted in both dark and light conditions with randomized order, and pupil responses were recorded during the listening task.

Results showed that the peak pupil dilation of the two groups showed no difference in dark condition, whereas the peak pupil dilation was significantly larger for normally hearing participants in light condition, indicating a stronger parasympathetic inhibition for the normally hearing group. Latency to peak pupil dilation in light condition was also lower in hearing impaired participants. The results on peak pupil dilation in different light conditions were in line with our hypothesis, this finding suggests that people with hearing impairment have a reduced parasympathetic inhibition comparing to their normally hearing peers. The present study may help to provide more insight into the complex underlying the pupil response.

30

How do users and non-users of hearing aids differ?

Maike Tahden [maike.tahden@uni-oldenburg.de], Anja Gieseler

Cluster of Excellence 'Hearing4all' & Cognitive Psychology Lab, University of Oldenburg, Oldenburg, Germany

Markus Meis, Kirsten C. Wagener

Hörzentrum Oldenburg GmbH & Cluster of Excellence 'Hearing4all', Oldenburg, Germany

Hans Colonius

Cluster of Excellence 'Hearing4all' & Cognitive Psychology Lab, University of Oldenburg, Oldenburg, Germany

Among hearing-impaired listeners there exist large differences in speech recognition in noise and in the individual benefits obtained from a hearing aid. For rehabilitation success, explanation of these variabilities is fundamental. Since auditory measures alone cannot comprehensively account for these differences, the role of other factors is currently being investigated.

Moreover, despite available devices and their benefits, only a minority of hearing-impaired individuals are using hearing aids. Given the possible negative side effects of not wearing hearing aids, it is of major interest to target this group of untreated hearing-impaired individuals with hearing support. Therefore, it is necessary to first characterize these individuals and contrast them to their supplied counterparts to obtain a descriptive consumer profile.

Differences between hearing aid candidates (the so-called hearing aid non-users, HA-NU) and hearing aid users (HA-U) might also provide additional information in terms of possibly relevant factors for predicting speech recognition in noise and individual hearing aid benefits.

Here, we compare elderly hearing-impaired hearing aid non-users with elderly hearing aid users matched for age, sex, and the degree of hearing impairment. Results indicate that HA-NU perceive their hearing problem subjectively as less severe than their supplied counterparts. Furthermore, HA-NU showed lower values than HA-U in the socioeconomic status and technology commitment.

Automatic scene classification improves speech perception of CI users in simulated real world listening scenarios

Anja Eichenauer [anja.eichenauer@kgu.de], Uwe Baumann, Tobias Weissgerber
University Hospital Frankfurt, Audiological Acoustics, Germany

Introduction

Speech perception in everyday listening conditions is demanding due to the presence of multiple noise sources and reverberation in rooms. Cochlear implant (CI) users in general experience deteriorated speech perception compared to normal hearing subjects. Current CI audio processors apply signal processing algorithms to improve speech perception in such complex listening scenes. However, measurements that determine the benefit of those algorithms are regularly performed under free field conditions in setups consisting of only a few sound sources. Consequently, the impact of signal processing on hearing performance in real world scenarios is not yet determined. Aim of the present study was to assess the benefit of an automatic scene classification algorithm (Cochlear SCAN) on speech perception in CI users.

Methods

10 unilateral CI users with Cochlear Nucleus 6 speech processors and a control group of 5 normal hearing (NH) subjects participated in the study. The Oldenburg sentence matrix test (OLSA) was used to assess the speech reception thresholds (SRTs) in an “everyday life simulation” task consisting of acoustic scenes of different acoustic complexity (e.g. multiple sources, reverberation). A 128-channel loudspeaker setup in an anechoic chamber was used for sound reproduction. Reverberation (simulation of an auditorium, reverberation time RT=1.2s) was reproduced based on reflection patterns (room simulation software ODEON, Lyngby Denmark). Within the simulation task the order of test conditions was randomized every five sentences. SRTs were calculated individually for each acoustic scene and averaged to assess the mean everyday life performance SRT.

Results

Mean SRT across all test conditions of the NH group was -9.9 dB SNR. Compared to free-field conditions, the SRT degraded up to 5 dB in reverberation (depending on the acoustic scene). Mean SRT of the CI group was 1.9 dB SNR (without SCAN) and 0.4 dB SNR (with SCAN). Without SCAN, CI users demonstrated decreased speech perception in reverberant conditions comparable with the NH subjects (up to 5.5 dB).

Conclusion

CI users and NH subjects showed similar detrimental effects of reverberation on speech perception. However, the mean SRT of CI users is almost 12 dB worse than the SRT of NH subjects. Automated signal processing based on scene classification is able to improve speech perception of CI users in free-field and reverberant conditions.

32

The effects of SNR driven amplitude compression in hearing aids on output SNR and signal envelope distortion

Christophe Lesimple [cles@bernafon.com], Miquel Sans

Bernafon AG, Bern, Switzerland

Compressing the amplitude of the incoming signal is a standard application of hearing aid amplification. The amount of compression is initially defined by the selected fitting rationale that indicates how much gain should be applied to a speech signal at different input levels. The results from speech intelligibility tests, in controlled environments, indicate that compression can compensate for the loss of audibility for soft speech (Marriage & Moore, 2003; Davies-Venn et al., 2009).

Difficulties appear when noise degrades the speech signal. In this case, the compressor still applies gain based on the input level without distinguishing the signal type. Over-amplification of noise by compressive amplification was found to result in degradation of the output SNR (Naylor & Johannesson, 2009) and speech envelope flattening (Jenstad & Souza, 2007).

A decision block applying corrections to the amplification, based on the SNR estimation, at a phonemic resolution, was designed to address these two negative effects. The principle of SNR driven compression is to estimate if the signal is useful or not, i.e. for noise the effective compression is released to avoid over-amplification. This qualification is not restricted by pre-defined rules for listening environment detection so that it can measure small and fast changes in daily situations.

The effect of different hearing aid signal processing algorithms was measured individually and in different combinations. The output SNR and envelope distortion index were calculated from different spatial distributions of speech and noise sources. When noise is spatially separated from the signal, traditional algorithms, such as directionality, improve the output SNR and the envelope distortion index. However, directional microphone algorithms cannot avoid an output SNR degradation caused by compression at positive input SNRs. SNR driven compression partially compensates for this loss and improves the contrast of the envelope at the output of the hearing aid. The resulting changes in output SNR and the envelope distortion index with different outcome measures will be presented and discussed.

33

Slope of the performance-intensity function and reaction time for speech in different noise types

Jon Øygarden [jon.oygarden@ntnu.no]

Norwegian University of Science and Technology, Trondheim, Norway

Speech recognition was measured in four different noise types for two groups in a project for developing a screening test for the hard of hearing organization in Norway (HLF – Hørselshemmedes Landsforbund). Group 1 consisted of 28 audiology students with normal hearing. Group 2 consisted of 70 persons with varying degrees of hearing loss. Audiometry was also performed for the frequencies of 250, 500, 1000, 2000, 3000, 4000, 6000 and 8000 Hz.

The evaluated noise types were:

1. Stationary speech-shaped noise from the Norwegian Matrix test with high frequencies (HF > 1400 Hz) attenuated by 15 dB.
2. Amplitude-modulated (AM) speech-shaped noise with HF attenuated. The noise was 100% amplitude modulated by a 16-Hz sine tone.
3. Reverse speech babble – four matrix sentences (with silent periods > 200 ms removed) played backwards – with HF attenuated.
4. Stationary speech-shaped noise from the Norwegian Matrix test.

A pilot screening test was developed using Matlab software to measure SRT for words in these four noise types. For each noise type (1-4), 40 test words selected among 10 nouns were used in an adaptive method, where the signal-to-noise ratio was adjusted in ±2 dB steps according to the result of the preceding word. The user responded on a touch screen with symbols and text for each of the nouns. For each participant, the following was measured for each word and noise type: level, selected response, success, time used, estimated threshold.

From these results, a rough estimate of the slope of the performance-intensity function for the different noise types and different hearing thresholds can be obtained. Variations in response time will also be evaluated and presented.

34

Segregation enhancement for hearing impaired listeners using a deep neural networks separation algorithm

Lars Bramsløw [labw@eriksholm.com]

Eriksholm Research Centre

Gaurav Naithani

Tampere University of Technology

Atefeh Hafez

Oticon A/S

Tom Barker

Cirrus Logic Ltd

Niels Pontoppidan

Erikshom Research Centre

Tuomas Virtanen

Tampere University of Technology

Hearing aid users are challenged in listening situations with noise and especially speech-on-speech situations with two or more competing voices. Specifically, the task of segregating two competing voices is very hard, unlike for normal-hearing listeners.

Recently, deep neural network algorithms have shown great potential in tasks like blind source separation of a single-channel (monaural) mixture of multiple voices. The idea is to train the algorithm on relatively short samples of clean speech, thus learning the characteristics of each voice. Once trained for those specific voices, the network can then be applied to mixtures of new speech samples from the same voices.

For this listening task, the benefit of a deep neural network (DNN) based stream segregation enhancement algorithm on hearing-impaired listeners was tested on 15 hearing-impaired listeners. The newly developed Competing Voices Test (Bramsløw et al, 2016) was used, in which pairs of sentences are presented, and the listeners has to repeat a target sentence as cued on a monitor. The cue is a word from the target sentence, presented either before or after playback of the mixed sentences. This competing voices test is based on the Danish HINT test (Nielsen & Dau, 2011; Nilsson et al, 1994).

A mixture of two HINT sentences was separated using DNN and presented to the two ears as individual streams and tested for word score. The results indicate that DNNs have a large potential for improving stream segregation and speech intelligibility in difficult scenarios with two equally important target voices. In such cases, the user will at any time be able to shift attention to the desired target voice.

35

Subjective listening effort: Influence of background noise direction and speaker's gender

Melanie Krueger [m.krueger@hoerzentrum-oldenburg.de], Kirsten C. Wagener, Markus Meis, Michael Schulte

Hörzentrum Oldenburg, Oldenburg, Germany

In today's society, hearing and understanding of speech is essential. In many situations the target speech is masked by some noise which leads to a decrease in intelligibility. The amount of decrease depends e.g. on the direction of the noise source (spatial release of masking) and the voice pitch of the conversation partner. However, interfering noise does not only affect intelligibility but also listening effort. Therefore, we investigated the influence of these factors on the subjectively perceived listening effort.

In addition to speech intelligibility, listening effort is an important outcome measure for comparing hearing aid settings. Using the Adaptive Categorical Listening Effort Scaling method ACALES (Krueger et al., 2017) the individual listening effort was determined for various directions of the noise source. Using this method the individual SNR range from "no effort" to "extreme effort" was determined for 20 young normal hearing subjects. The listening effort measurements were performed with following conditions: S0N0 (speech (S) and noise (N) from 0°), S0N90, S0N135, and S0N180 with a male speaker. Condition S0N0 was also measured with a female speaker (all measurements applying sentences of the male and female Oldenburg sentence test (OLSA), Wagener et al., 1999, Wagener and Brand, 2005, Wagener et al., 2014).

The listening effort decreased with increasing spatial separation between target and noise up to 135°, but increases again at 180°. Similar effects have been reported in the literature on speech intelligibility measurements with OLSA. However, speech intelligibility seems to improve with increasing degrees only up to the maximum at 90°. Subsequently, speech intelligibility decreases again (Beutelmann and Brand, 2006).

In addition, the difference in subjective listening effort was evaluated with a male versus a female speaker, resulting in lower listening effort for the female speaker. In addition, the speech intelligibility with OLSA was determined for a female and male speaker. Reference values show different SRT values for the OLSA with a male speaker (-8.4 dB SNR) and a female speaker (-10 dB SNR, Wagener et al., 2014). This difference seems to be smaller for listening effort.

36

A Model of Concurrent Vowel Identification Without Segregation Predicts Perceptual Errors

Samuel S Smith [samuel.smith@nottingham.ac.uk]
University of Nottingham, UK

Ananthakrishna Chintanpalli
Birla Institute of Technology & Science, India

Michael G Heinz
Purdue University, USA

Christian J Sumner
University of Nottingham, UK

When positioned in a complex auditory environment, individuals with normal hearing are able to identify and concentrate on specific components within that environment, famously termed the cocktail-party phenomenon. There are a multitude of cues which can be used to facilitate auditory stream segregation (e.g., pitch differences, dynamics, onset/offset asynchronies, differences in speech spectral characteristics). Particular attention has been paid to the positive effect that differences in fundamental frequency (F0) between two vowels (steady-state harmonic complexes), presented concurrently, has on their identification (review: Micheyl and Oxenham, 2010).

Computer models exist that predict with some success the improvement in concurrent-vowel identification observed with increasing F0 differences (Meddis & Hewitt, 1992). However, these existing models are poor at predicting listener confusions (Chintanpalli and Heinz, 2013).

Presented is our model of concurrent-vowel identification, which incorporates a naïve Bayesian classifier. This model directly predicts the probabilities of different combinations of two vowels giving rise to an integrated representation of the concurrent-vowel pair presented. This contrasts with previous models, which were deterministic and assumed that a segregation process separated out individual vowel representations based on F0 differences, followed by a comparison with templates of individual vowels. Our model can also incorporate a pitch estimation process, but this is used to restrain the concurrent-vowel pair categories used in classification, and has marginal benefits. Our new 'synthesis' based model was tested for both temporal (autocorrelation-based) and spectral (rate-based) internal representations.

The new model was able to successfully predict confusions with a high degree of accuracy (confusions: R>0.85). In the case where there was no difference in F0 between the vowels, performance was slightly better for the spectral model than the temporal model. However, only when temporal processing was implemented, our model qualitatively replicated the positive effect that differences in F0 have on human concurrent-vowel identification. Overall, our model is much closer to predicting human performance than previous models, and hints at a process that seeks to optimally predict which concurrent-vowel pair led to a corresponding internal representation, rather than to segregate the representation and recognise individual vowels separately.

37

Acoustic analyses of vowel variation for the investigation of perceptual adaptation to speaker properties in channel-vocoded speech: preliminary data.

Olivier Crouzet [olivier.crouzet@univ-nantes.fr]
LLING - Université de Nantes / CNRS, France.

Etienne Gaudrain
CRNL - CNRS / Université Lyon 1, France

Deniz Başkent
UMCG - RUG, The Netherlands

Perceptual analysis of speech requires processing two complementary dimensions: on one side, various acoustic parameters provide access to information concerning voice properties (vocal characteristics, speaker identity, gender, emotional state…) while, on the other side, listeners process sounds into phonological categories (phonemic classes / distinctive features). These two dimensions interact strongly in speech. For instance, when comparing speakers with different vocal tract lengths (VTL), a shorter VTL would be associated with higher "formant frequencies" overall. When comparing different vowels, a single speaker contrasting the vowels /o/~/u/ would lower both the first and second formants by changing the shape of one's own vocal tract. These interactions seem to have a relatively low impact on everyday communication in normal-hearing listeners: it is assumed that some perceptual adaptive mechanisms take place that contribute to the listeners' ability to deal with these interactions. However, these phenomena may constitute a source for the low speech recognition performance that some cochlear-implanted deaf listeners experience in multi-speaker situations. In order to investigate this issue, we are setting-up a series of experiments in order to investigate the influence that adaptive mechanisms to changes in voice properties (e.g. vocal tract length) may exert on phonological classification (e.g. vowel identification) when listening is degraded using channel-vocoded speech.

We plan to replicate a classical experiment (Ladefoged & Broadbent, 1957; see also Sjerps, McQueen & Mitterer, 2013) in order to measure the level of vowel normalization that speakers may reach when processing vocoded speech in contrast to natural speech. In order to address these issues, formant measurements that were issued from a database of natural recordings of Dutch speakers were investigated in order to assess both formant frequency central tendencies and variation. These measurements provide access to information concerning between-vowel average distance and within-vowel individual variation with respect to (1st, 2nd and 3rd) formant frequencies. They also provide us with observations concerning the direction of change from one vowel category to the other for each formant and their relation to the expected impact of differences in vocal-tract length. From these data, a set of 15 Dutch monosyllabic word-pairs were selected and recordings by 3 Dutch speakers were collected. For each vowel contrast (word pairs), an acoustic continuum was generated on the basis of the actual formant frequencies. These results will be described along with a preliminary crowd-sourcing experiment that will provide measurements of psychometric identification curves for each vowel continuum.

Biological inspired MEMS acoustic sensors

Andrew B Reid [andrew.reid@strath.ac.uk], Yansheng Zhang, James FC Windmill
University of Strathclyde

Directional microphones offer a robust method for attenuating background noise, but their development has been incremental, and their benefits modest. We will present our current work on developing novel directional microphones for attenuating background noise using microelectromechanical systems (MEMS) acoustic sensors. The developed MEMS microphone designs are inspired by the hearing mechanisms of two insect species, Ormia ochracea & Achroia grisella, which both have exceptional potential to resolve directional hearing cues. The general sensing concept is based on two mechanically connected membranes in which the two base resonance movement mode shapes, a tilt and bend around the connection, combine to create an apparent increase in resolving directional cues. To replicate this behaviour in an engineering solution we designed a range of silicon MEMS microphones with piezoelectric readout and design considerations looking at a low frequency coverage, a distinct resonance at 8 kHz and a broadband frequency coverage below 10 kHz. The fabrication is done using the cost effective multi-user silicon-on-insulator process PiezoMUMPs, commercially offered by Memscap Inc. The process uses a 10-µm thick single crystal silicon layer for the device structures and a 500-nm thick layer of aluminium nitride as piezoelectric active material, which is used as main acoustic sensing principle. Three current families of devices will be presented which each have a footprint between 1 mm x 2 mm and 1.2 mm x 3 mm. The mechanical and acoustic sensitivity characterisation of the devices will be presented, showing their directional acoustic response and good agreement with analytical estimations and acoustic-structure interaction finite element analysis simulations. Further steps to include the influence of the physical mounting and environment of the MEMS chips on the directional acoustic response of the microphones will also be shown. These MEMS microphones offer novel methods for robust noise attenuation in difficult communication situations.

39

The contribution of salient localizable glimpses on speech intelligibility in a multitalker setting with spatially diffuse signals

Esther Schoenmaker [esther.schoenmaker@uni-oldenburg.de], Steven Van de Par
University of Oldenburg, Germany

Speech intelligibility in a multitalker scene is a challenging task, even for normal-hearing listeners. Spatial separation of the various talkers improves speech intelligibility, leading to a spatial benefit relative to the situation in which all speech signals originate from the same direction.

Typically, in a setting with a small number of talkers, the sparse spectro-temporal character of speech causes interfering speech to only partly mask the target speech. As a result, the local signal-to-noise ratio (SNR) fluctuates strongly within a spectro-temporal representation of simultaneous speech signals.

In a previous experiment (Adv. Exp. Med. Biol. 894, p. 73–81.) spectro-temporal elements of target speech were eliminated from a three-talker mixture when they were lower than a predefined SNR criterion. Measurements of speech intelligibility over a range of such SNR criteria demonstrated that when target speech was reduced to only few, but very salient, spectro-temporal elements („glimpses“) the spatial benefit was sustained and of comparable size to that of intact speech. This suggests that the spatial information contained in high-SNR glimpses is highly important for a proper organization of the auditory scene by the listener.

In a reverberant environment, however, speech signals become interaurally decorrelated, rendering spatial cues less useful as a cue for source segregation. In a new experiment we investigate whether the adverse effect of decorrelated interaural cues can be reduced by delivering useful spatial information to only the most salient glimpses of either the target or the interfering talkers, thereby rendering some of the glimpses localizable.

Our experimental data will be discussed in relation to contributions of monaural and binaural cues to auditory grouping and segregation. In addition, the data will provide insight into the importance of localizability of target versus interfering signals.

40

Data-driven discovery of general mechanisms of cortical processing of natural sounds

Moritz J Boos [moritz.boos@uni-oldenburg.de]

University of Oldenburg, Applied Neurocognitive Psychology Lab

In humans, brain activity can be predicted from a given stimulus representation combining fMRI with voxel-wise encoding models. Yet, the choice of stimulus representation can limit the interpretability of the encoding model. In the auditory domain, an efficient coding of natural sounds a sparsity constraint on their representation - accounts for tuning properties of auditory nerve fibers (Lewicki, 2002). We aim to join these two approaches unsupervised learning and voxel-wise encoding models in fMRI to find general principles underlying the cortical functional organization of auditory processing. For 15 participants from an open 7-Tesla fMRI dataset (Hanke et al., 2014), we predict BOLD activity elicited by an auditory movie. We learn a representation of the Mel-frequency spectrogram of the auditory movie using sparse coding with binary latents (BSC) (Lücke and Eggert, 2010; Henniges et al., 2010). We decompose the encoding model into latent dimensions, using principal component analysis. This reveals three dimensions that generalize across participants. The first dimension is centered on primary auditory cortex. It correlates highly with stimulus energy (r=.78). The other two dimensions are located more lateral, ventral, and anterior along the superior temporal sulcus. While both are sensitive to the varying auditory complexity of the stimulus, the second component relates to the speech signal to noise ratio (SNR) in the auditory movie. Using fMRI activity represented in this three dimensional latent space, we reconstruct ratings of the speech SNR of the stimulus for unseen participants and unseen parts of the movie (r=.8). In conclusion, unsupervised learning and a data-driven decomposition of fMRI activity reveals general mechanisms underlying auditory processing in human temporal cortices.

References

Hanke, Michael, Florian J Baumgartner, Pierre Ibe, Falko R Kaule, Stefan Pollmann, Oliver Speck, Wolf Zinke, and Jörg Stadler (2014). In: Scientific data 1.

Lewicki, Michael S (2002). In: Nature neuroscience 5.4, pp. 356–363.

Lücke, Jörg and Julian Eggert (2010). In: Journal of Machine Learning Research 11, pp. 2855–900.

41

3D printed acoustic metamaterials for small-scale noise control applications

Cecilia Casarini [cecilia.casarini@strath.ac.uk], Ben Tiller, James F.C. Windmill, Joseph C. Jackson

University of Strathclyde

Controlling the propagation of acoustic waves is of great importance to our everyday life, for example to isolate ourselves from industrial or traffic noise and at the same time to be able to enjoy listening to music or to communicate with people around us. Acoustic metamaterials have been studied in the last decade because of their ability to manipulate sound waves in new ways, thanks to their exotic properties such as negative mass density and bulk modulus. By exploiting the resonances in the unit cells that form the structure of the metamaterials it is possible to generate absorption bands where the sound is deeply attenuated. Similar results can be obtained by using phononic crystals, which are materials composed of periodically spaced unit cells, where the wavelength of the frequencies that can be controlled is of the same order of the periodic spacing. However, the properties of acoustic metamaterials are a byproduct of the resonances of their unit cells rather than the distance between them, hence they are subwavelength structures and can enhance acoustic absorption. The objective of the work presented in this poster was to fabricate small-scale acoustic metamaterials that could be included in wearable devices such as headphones or hearing aids and that could attenuate sound in a chosen frequency range. To do so, it was necessary to develop manufacturing techniques to fabricate the unit cells in an accurate and reproducible way. We chose to make use of additive manufacturing technology to fabricate Helmholtz resonators and membranes which are often employed as unit cells at the base of metamaterials. We first 3D printed acoustic metamaterials based on Helmholtz resonators having the shape of soda cans but with a dimension scaled down by a factor of 20 and we were able to obtain absorption bands with a sound transmission loss up to 30 dB. We then modified the dimension of the resonators to “tune” the overtones and achieve multiple absorption bands. Furthermore, we developed a technique to 3D print thin membranes at the base of the resonators to further widen the absorption. These results could contribute to improve the fabrication of metamaterials and hence could lead to applications for noise control such as noise cancelling headphones and smart sensors.

42

Bio-inspired Frequency-adaptive Acoustic System

José Guerreiro [jose.guerreiro@strath.ac.uk], Joseph C. Jackson, James F. C. Windmill

University of Strathclyde

Standard microphones are generally designed with a static and flat frequency response in order to address multiple acoustic applications. However, they may not be flexible or adaptable enough to deal with some requirements. For instance, when operated in noisy environments such devices may be vulnerable to wideband background noise which will require further signal processing techniques to remove it, generally relying on digital processor units. In this work, we consider if microphones could be designed to be sensitive only at selected frequencies of interest, whilst also providing flexibility in order to adapt to different signals of interest and to deal with environmental demands. This research then introduces the fundamentals of a novel concept of signal processing at the sensor level. An acoustic signal processing framework integrating a functional prototype system was engineered to support the concept of a frequency agile sensor. That is a concept where the "transducer becomes part of the signal processing chain" by exploring feedback processes between mechanical and electrical mechanisms that together can enhance peripheral sound processing. This capability is present within a biological acoustic system, namely in the ears of certain moths. That was used as the model of inspiration for a smart acoustic sensor system which provides dynamic adaptation of its frequency response with amplitude and time dependency according to the input signal of interest.

43

Preattentive Processing in the Spatial Unmasking of Speech

Benjamin H. Zobel [bzobel@psych.umass.edu]
University of Massachusetts Amherst, USA

Lisa D. Sanders, Richard L. Freyman
University of Massachusetts Amherst, U.S.

Listeners benefit when target speech and masking speech are spatially separated. Research suggests that under complex listening conditions, when target and masker are perceptually confusable, spatial separation provides a cue that facilitates bottom-up segregation of target and masker, and top-down selective attention to the target. However, it has been difficult to distinguish between the contributions of bottom-up and top-down processes through behavioral measures alone. The present experiment used event-related potentials (ERPs) to examine the role of attention in the spatial unmasking of speech. Vocoded target words were presented with 2-talker vocoded masking babble when target and masker were 1) co-located, and 2) perceptually separated by adding an identical copy of the masker at a separate location that preceded the onset of the co-located masker by 4 ms. When listeners were attending to the sounds in a target detection task, auditory evoked potentials elicited by target words were only observed when target and masker were perceptually separated and masking was released, replicating previous findings (Zobel et al., 2016). Importantly, a similar modulation of ERPs was observed when listeners directed attention away from the auditory modality to engage in a challenging visual 2-back task. These findings provide strong evidence that under complex listening conditions, the perception of spatial separation between target and masker facilitates bottom-up, preattentive processing of target speech.

Speech-in-noise recognition abilities are associated with vocal pitch perception abilities in controls but not in high-functioning autism spectrum disorder

Stefanie Schelinski [schelinski@cbs.mpg.de]
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Katharina von Kriegstein
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; Humboldt University of Berlin, Berlin, Germany

The ability to recognise auditory speech in a noisy environment is critical for successful communication in everyday situations. There is evidence that in autism spectrum disorder (ASD), speech perception is reduced under noisy conditions (Alcantara, Weisblatt, Moore, & Bolton, JChildPsycholPsyc, 2004; Groen et al., JAutismDevDisord, 2009). Currently it is unclear, whether difficulties in speech-in-noise perception are associated with difficulties in perceiving basic acoustic features of voices that are relevant for speech-in-noise perception. A key acoustic feature for speech-in-noise perception is the fundamental frequency which is perceived as vocal pitch (Anderson & Kraus, JAmAcadAudiol, 2010). Here we investigated speech-in-noise recognition abilities and its relation to vocal pitch perception abilities in a group of adults with high-functioning ASD (n = 16) and typically developed individuals (n = 16; matched pairwise on age, gender, and IQ). The ASD group has been previously shown to have difficulties in vocal pitch perception but intact non-vocal pitch perception abilities (Schelinski, Roswandowitz, & von Kriegstein, AutismRes, 2017). In the speech-in-noise recognition test, we investigated the individual thresholds for speech recognition when speech was presented with different levels of speech-shaped noise. The ASD group showed significantly higher thresholds as compared to the control group, i.e. typically developed individuals understood speech in higher noise levels. Within the control group, performance in the speech-in-noise recognition test correlated with performance in vocal pitch, but not non-vocal pitch perception. Within the ASD group, there were no correlations between speech-in-noise recognition and vocal or non-vocal pitch perception abilities. This indicated that in controls better speech-in-noise recognition abilities were associated with better vocal pitch perception, but not in the ASD group. Our results suggest that perceptual impairments, i.e. difficulties in vocal pitch perception, might contribute to speech-in-noise recognition difficulties in ASD. In line with our previous results on vocal emotion recognition this implies that communication difficulties in ASD might not only be based on higher-level cognitive difficulties, but also on impaired basic perceptual processing.

Who are you listening to? Towards a dynamic measure of auditory attention to speech-on-speech

Moïra-Phoebé Huet [moira-phoebe.huet@inserm.fr]

University of Lyon, Lyon, France

Christophe Micheyl

Starkey, Créteil, France

Etienne Gaudrain

University of Groningen, Groningen, The Netherlands

Etienne Parizet

University of Lyon, Lyon, France

Listening to a speech signal is a perceptual task that is ecologically relevant and used in many experimental studies in psychology and neurosciences. During the last 20 years, in particular, a growing number of psychoacoustic or brain imaging studies have used a task of listening to concurrent speech signals in order to investigate, for example, different phenomena of auditory masking, or, brain correlates of selective auditory attention.

In general, in these speech-on-speech tasks, the listener is asked to selectively direct his attention on one of the voices. However, these tasks assume that the listener keeps a constant focus on the target voice, while in real-life situations, listeners' attention can quickly vary from one voice to another. To approach more realistic situations, it is therefore important to have new methods to infer retrospectively if, and at what times, the attention of the listener has turned away from the target voice.

We have therefore developed a paradigm and a set of materials to infer the dynamics of auditory attention over time. After listening to two simultaneous stories — a target and a masker — the participants have to find, among a set of words, those present in the target story. The words were selected according to several criteria such as repetition, their frequency in the language, the recency effect and the primary effect.

In the present study, the masker were uttered by the same talker as the target stories, but the voice parameters (F0 and vocal tract length) were manipulated to parametrically control similarity of the two voices, hence controlling the difficulty of the task.. The difficulty of the task was also manipulated by alternating the modes of the presentation of the stimuli, namely, dichotically and diotically presentations. As the target and masker stimuli are more similar, we expect to observe more responses corresponding to the masker voice, which would reflect more spurious switches of attention to the wrong talker. Preliminary results of this study will be presented.

Effects of global brightness on salience and auditory foreground perception

Francesco Tordini [francesco.tordini@mail.mcgill.ca]
McGill University, CIRMMT

Albert S. Bregman
McGill University, Department of Psychology

Jeremy R. Cooperstock
McGill, Department of Electrical and Computer Engineering

The word salience describes the attention-grabbing quality of a sound. A salient sound stands out in the presence of other competing ones, which become the background of the scene. The concept of salience is ubiquitous in auditory sciences.

However, despite the emphasis placed on the topic, computational models of auditory salience are far from mature. Progress in the field is hampered by the little agreement on behavioral paradigms that truly probe the involuntary, stimulus-driven nature of salience. As a by-product, there is a lack of standard datasets available to the research community. Moreover, current approaches reduce the problem of salience modeling to that of change detection, that is, matching the detection response of a human listener to brief auditory events.

We propose to distinguish between the salience of sound events and that of streams, and we introduce a paradigm to study the latter using repetitive patterns in a competitive, spatial scenario.

We suggest that global descriptors of perceptual features can be used to characterize the streams and predict their salience, hence the perceptual organization of the auditory scene into foreground and background.

However, there are many possible independent features that can be used to describe sounds. Our hypothesis, deemed biologically and computationally reasonable, is that only a few of them cause a sound object to stand out from the scene. Adopting a data-driven approach, we select the features that emerge from the analysis of the data collected using our competitive streaming paradigm and feature-rich natural sounds, such as bird chirps and human voice, and synthetic sounds. The result is a parsimonious interpretation of the rules guiding foreground formation: after loudness, tempo and brightness are the dimensions that have higher priority.

Our data show that, under equal-loudness conditions, patterns with fast tempo and those from certain "preferred" brightness bands tend to emerge from the scene. Moreover, the interaction between tempo and brightness in foreground selection seems to increase with scene complexity or task difficulty.

We propose to use the relations we uncovered as the underpinnings of a computational model of foreground selection, and also as recommendations for sonic information design, particularly that of auditory warning systems.

47

Encoding of Mid-Level Speech Features in MEG Responses

Christoph Daube [c.daube.1@research.gla.ac.uk]
University of Glasgow, Institute of Neuroscience & Psychology

Robin A. A. Ince
University of Glasgow

Joachim Gross
Universität Münster

Studying the neuronal responses to naturalistic speech stimuli offers the perspective of retracing increasingly abstract transformations of the auditory input as it is processed along the auditory pathway. One computational approach to explore hypotheses about such transformations in Magneto- or Electroencephalography (MEEG) data relies on the estimation of multivariate temporal response functions (mTRFs). With this class of regularized linear models, previous studies have demonstrated the relevance of spectrally resolved stimulus energy and linguistically derived phonemic features for the prediction of MEEG signals.

This makes it interesting to investigate potential intermediate acoustic features which could afford the listener such pre-lexical abstraction.

To address this question we recorded the MEG of young healthy human participants (currently N=3) while they listened to a narrative of 1 hour duration. We performed source localisation using an LCMV beamformer and identified regions of interest based on retest-reliable responses to a repeated speech stimulus. In these regions surrounding bilateral Auditory Cortices we estimated mTRFs from bandpass-filtered signals (1-15 Hz) in a 5-fold nested cross-validation framework and compared the predictive power of a number of different feature spaces: Plain stimulus energy, spectrograms, gabor-filtered spectrograms and a set of linguistic articulatory features.

In all participants, we replicated the finding of improved prediction performance when articulatory features were added to cochleagrams. We achieved a comparable increase in prediction performance when we instead used gabor-filtered spectrograms.

Our results demonstrate that mTRFs combining various feature spaces can be estimated reliably for individual grid points from source-localised MEG data. This allowed us to extend previous acoustic models with further mid-level auditory features. Taken together, this is a promising next step to fill the gap between cochlear and linguistic representations of speech stimuli using temporally highly resolved and non-invasive neuroimaging.

How speech statistics limits the number of effective channels in cochlear implants. Implications for sound-coding strategies.

Jacques A Grange [grangeja@cardiff.ac.uk]
Cardiff University

Tarik Siebe, Tim Juergens
Oldenburg University

John F Culling
Cardiff University

Typically, cochlear implant (CI) processors first analyse sound through a bank of evenly spaced filters on a log-frequency scale, before each extracted temporal envelope is used to modulate the excitatory current passed through each implanted electrode. However, factor analysis (FA) of speech shows that the information carried by the temporal modulations of speech is distributed in bands whose width grows non-monotonically with frequency. Given that the spread of excitation (SOE) along the spiral ganglion limits the number of independent channels in CI users to eight or less, optimising transmission of information should be considered in the design of sound-coding strategies. A scree plot derived from factor analysis suggests that six channels should suffice to transmit most of the speech information. This outcome may explain why speech-reception thresholds (SRTs) using the SPIRAL CI vocoder improve markedly more slowly as a function of number of channels beyond 7 simulated channels. A first study tests whether FA-inspired channel bandwidths might improve speech intelligibility. Simulations with normal-hearing listeners employed a pulsatile vocoder (Oldenburg) and adaptively measured SRTs for the following factors: two levels of SOE (200 dB/oct and 8 dB/oct) x two numbers of channels (8 or 11) x three channel definition/allocation strategies. The first strategy (CIS) used evenly-spaced channels and simulated electrodes aligned with the channel centre frequencies. The second and third strategies used FA-selected channels, with simulated electrodes selected from a fixed 22-electrode array; in the second (FAw), selected electrodes were evenly spaced, thereby limiting channel interaction, but leading to spectral warping; in the third (FAu), the selected electrodes were those with a place frequency closest to each of the channel centre frequencies, leading to an almost unwarped electrode map. Simulating SOE significantly elevated thresholds and the number of channels interacted with SOE. Although no effect of strategy was found, spectral warping would be expected to elevate SRTs. However, with 8 channels and SOE typical of CI users, the CIS and FAw strategies led to the two lowest thresholds. A second study that varied the number of activated channels (4, 5, 6, 8, 11, 15 and 22) and simulated typical or no SOE with the FAu strategy, showed that 5 channels sufficed to attain the lowest threshold. Perceptual learning experiments may help elucidate which FA-inspired strategy would be most effective in CI users.

The spatial speech test of real-world listening for assessing binaural hearing

Deborah A Vickers [d.vickers@ucl.ac.uk], Mana Ahnood, Bhavisha Parmar, Jenny Bizley
University College London

The spatial speech test was developed to simultaneously assess the ability of listeners to identify speech and determine the relative location of target sound in the presence of noise (Bizley et al., 2015).

Listeners hear two sequentially presented words from adjacent speakers with a 300 separation and they identify the word and the relative direction of the second word presentation relative to the first. The task is performed in the presence of multiple independent noise sources at an individually determined signal-to-noise ratio.

While this test provided a sensitive assessment of spatial hearing in normal hearing listeners, when bilateral hearing aid users were tested in the same task they were unable to perform the relative-localization aspect despite doing well on a standard clinical localization task measuring the ability to point to a speech source in silence.

Therefore, we sought to adapt the spatial speech test further to make it easier for hearing impaired listeners. To achieve this we maintained the dual-task element of the design, but restricted the noise sources to left or right space. We tested normal hearing listeners (n=11) and bilateral cochlear implant users from 8-80 years of age (n=10). In a subset of listeners we additionally assessed performance with each implant in turn.

The findings showed that the spatial location of the words had a significant effect on relative localization performance for both normal hearing and cochlear implanted listeners. For those tested in the unilateral cochlear implant condition it showed that relative localization performance was not above chance but in the bilateral cochlear implant condition listeners performed significantly better.

For the normal hearing listeners' performance on the word identification aspect of the test was moderated by spatial separation of the words from the noise sources, this was not the case for the bilateral CI users.

Bilateral cochlear implant users were able to perform above chance in both speech perception and relative localization judgements in the presence of noise, although they did not demonstrate spatial release from masking. The test is sufficiently sensitive to detect binaural hearing abilities and demonstrated benefit from the second side implant for relative localization judgements.

Reference

Bizley JK, Elliott N, Wood KC, Vickers DA. (2015) Simultaneous Assessment of Speech Identification and Spatial Discrimination: A Potential Testing Approach for Bilateral Cochlear Implant Users? Trends Hear. 2015; 19 1-11.

How the tongue and lips produce clear speech: CVC words with randomised vowels, transcribed by listeners in normal and noisy conditions

James M Scobbie [jscobbie@qmu.ac.uk], Joan Ma
Queen Margaret University, Edinburgh, Scotland

In noisy conditions, speakers adapt their speech production in a number of ways, which can be usefully studied through acoustic analysis of the speech output or perceptual testing of this output. Noise can be perceptible to the speaker, the speaker and listener, or the listener alone, and in all cases, the speaker may enhance aspects of their speech for the benefit of the listener. Such changes cause a variety of effects such as an expanded vowel space. One hypothesis is that speakers alter their supralaryngeal articulations in order to enhance formant values, to make vowels more perceptibly distinct from each other.

Our focus is on the enhancement of speech production that occurs when the listener's hearing becomes artificially masked by speech babble noise (40dB), as a model of how speech production may change when talking to hearing-impaired listeners.

Our previous pilot work on a single speaker found that tongue and lip articulations were indeed different in a noisy condition. The materials were single /b/+V+/p/ CVC words. They varied only in the vowel, which was one of six monophthongs which were fully randomised over 6 productions. Some vowels like /ʉ/ appeared to change little, and the high front vowels /i/ and /e/ were slightly retracted. Though the initial /b/ and final /p/ were identical every time, in the noisy condition the /b/ was clearly hyper-articulated while the /p/ was not. For this full paper we will analyse 6 speakers of Scottish English to look for general patterns.

We will use Ultrasound Tongue Imaging and a lip-jaw camera synchronised with the acoustics, in order to describe the ways in which different speakers expand their vowel space. Ultrasound provides a mid-sagittal tongue surface image from near the tip of the tongue down to near the root, and is cheap, quiet, and provides dynamic images at a high frame rate. The camera is mounted on a headset worn by the speaker, which also holds the ultrasound probe steady.

We will analyse the difference between the two production conditions by tracing the tongue surface in each vowel and calculating the average position of the target, and comparing differences in vowel shape and location across conditions. In addition, the Bark-transformed area of the F1/F2 acoustic vowel space will be measured.

51

Speech perception under eye-controlled and head-controlled directional microphones in a dynamic 'cocktail party'

Ľuboš Hládek [lubos.hladek@nottingham.ac.uk]
MRC/CSO Institute of Hearing Research - Scottish Section

Bernd Porr
University of Glasgow

Graham Naylor, W. Owen Brimijoin
MRC/CSO Institute of Hearing Research - Scottish Section

Hearing impaired people have difficulties with understanding speech in noisy environments. Despite the fact that hearing aids with directional microphones reduce background noise and thus improve speech intelligibility, the benefit does not always translate to real listening scenarios. One of the reasons is that standard directional microphones amplify sound only in front of the listener, which makes listening to off-axis signals even more difficult than with the omni-directional microphones. In order to address this problem, the directivity of the microphones could be controlled by the eye gaze of the hearing aid user. In the current experiment, hearing impaired and normal hearing subjects listened to listened to single words at a rate of 1 word / 1.5 seconds coming either from 30 degrees to either the left or right, simulating a turn-taking conversation. The task of the participants was to repeat the words out loud. The stimuli were presented through the loudspeaker ring and the directivity pattern of the virtual hearing aids, either head-controlled or eye-controlled, was simulated on-line using head tracking and eye tracking to control the gain of each loudspeaker channel. The stimuli were presented in four conditions, combinations of the two directional steering methods and two acoustic beam widths. Preliminary results indicate a benefit of the eye-controlled technology over the head-controlled technology, however, the benefit was diminished in those subjects who preferred to re-orient using pronounced head movements. These data indicate that the patterns of behavior could be critical for understanding speech benefits in situations like dynamic 'cocktail parties'.

[Work supported by the Oticon Foundation, the Medical Research Council (grants #U135097131 and #MC_UU_00010/4), and the Chief Scientists Office – Scottish Government]

52

The effect of the language proficiency of bilingual adults on the Canadian Digit Triplet Test

Josée Lagacé [josee.lagace@uottawa.ca], Christian Giguère, Véronique Vaillancourt, Suzanne Lteif, Sandrine Pelletier-Laroche
University of Ottawa, Canada

Background

Bilingualism or multilingualism is present in all parts of the world and the Canadian population is no exception to this situation, creating a need for effective clinical tools and guidance in the speech audiometry area. For example, the performance of bilingual or multilingual listeners on a speech test may be lower than the monolingual normative data because of their proficiency in the language of the test being used. It is then difficult to determine if the lower scores are related to the language competencies or indicative of a hearing deficit. The Digit Triplet Test (DTT) was first introduced in Dutch as an automatic self-screening test (Smits et al., 2004). Since then, it has been developed in many different languages. A Canadian English and French version of a digit triplet test (CDTT) has recently been developed (Ellaham et al., 2016) as a measure of speech understanding in noise. Normative data for monolingual adults have been collected for each version of the test. As there are some indications that closed-set speech tests, such as the CDTT, are more effective in evaluating basic speech recognition abilities in noise with bilingual populations than open-set tests, this study aimed at exploring the effect of the language proficiency of bilingual and multilingual adults on the Canadian Digit Triplet Test.

Methods

One hundred (120) adults with normal hearing thresholds were tested with both versions of the CDTT. A questionnaire about the linguistic experience was also completed by all the participants to determine the relative proficiency level in both languages, as well as the language dominance.

Results

Regardless of language proficiency levels, bilingual adults performed similarly on the Canadian English and Canadian French versions of the CDTT, and their performance is at par with that of monolingual individuals (based on previous findings).

Conclusions

Therefore, in a clinical setting, the CDTT can be carried out in the preferred language of the client.

53

Simulating hearing loss in neural networks: Does pre-training on intact speech boost performance on degraded input?

Robert Grimm [Robert.Grimm@uantwerpen.be], Michèle Pettinato

Computational Linguistics & Psycholinguistics Research Center, University of Antwerp, Belgium

We present results from machine learning experiments designed to mimic conditions in human listeners with cochlear implants (CIs). Postlingually deaf (PD) individuals, who receive CIs after a period of normal hearing, often perform better on hearing-related tasks than congenitally deaf (CD) individuals, who are born deaf and receive CIs after an early period of auditory deprivation. We consider two possible reasons for this.

(1) CD individuals might perform worse than PD individuals as a result of early auditory deprivation. Then, after implantation with a CI, the brain may have lost the plasticity necessary for full recovery.

(2) Alternatively, and in contrast to CD individuals, PD individuals might learn to differentiate fine-grained speech structure during their period of normal hearing that is impossible to acquire from CI-delivered signals. This feature structure might then boost hearing performance post-implantation – giving PD individuals an advantage relative to CD individuals.

To evaluate the two possibilities, we train neural networks on (a) normal speech and (b) vo-coded speech that is modified to simulate the input received by people with CIs. We then compare the performance of two networks: a CD network, which is only trained on vo-coded speech; and a PD network, which is pre-trained on normal speech before the training data are vo-coded. We find that the PD network retains sensitivity to fine-grained spectral differences that is absent in the CD network, even though this does not lead to the PD network outperforming the CD network.

However, to transition from intact to vo-coded speech, the PD network only requires minor adjustments to its internal connectivity – which affords rapid adaptation to vo-coded speech. The CD network, in contrast, cannot rely on previous knowledge gained from intact speech and has to start learning from scratch. Thus, if we severely restrict the learning capacity of both networks once exposed to vocoded speech, the PD network reaches peak performance in a fraction of the time it takes for learning to plateau in the CD network.

This suggests that PD individuals outperform CD individuals because the manner in which they process intact speech only requires minor modifications in order to generalize to CI-delivered speech. As a result, they can rapidly adapt to CIs. CD individuals, once implanted, need to develop speech processing capacities from scratch; and unless they are implanted within the first months of life, the brain's reduced plasticity leads to reduced performance relative to PD individuals.