Organization

The Speech in Noise 2014 Workshop is hosted by CNRS - Laboratoire de Mécanique et d'Acoustique, Marseille, France, and will take place in Marseille on January 9th and 10th.

Organising Committee

Sabine Meunier	Laboratoire de Mécanique et d'Acoustique, Marseille, FR
Mark Huckvale	University College London, London, UK
Rolph Houben	Academic Medical Center, Amsterdam, NL
Olivier Macherey	Laboratoire de Mécanique et d'Acoustique, Marseille, FR
Jacques Chatron	Laboratoire de Mécanique et d'Acoustique, Marseille, FR
Sophie Savel	Laboratoire de Mécanique et d'Acoustique, Marseille, FR
Gaston Hilkhuysen	Laboratoire de Mécanique et d'Acoustique, Marseille, FR

The Workshop on the Web

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

Sponsoring

We warmly thank our sponsors: PACA, Genesis, Medel, Neurelec & ANR

Table of Contents

Oral Presentations

O1: Simulating the bimodal excitation spread produced by Cochlear implant bipolar stimulation: Effect on speech intelligibility . . . 2
__Quentin Mesnildrey__ and Olivier Macherey

O2: How CI users can make the best of their implants in SpiN situations: Positioning in a room, head orientation strategy and translational avenues . . . 3
__Jacques A. Grange__ and John F. Culling

O3: Speech recognition in N-talker babble: Patterns of performance with increasing N vary across types of speech material . . . 5
__Tim Green__ and Stuart Rosen

O4: Top-down restoration of speech: Hearing impairment and age can influence it in unexpected ways . . . 6
__Deniz Başkent__

O5: Measuring Listening Effort with Reaction Time to Digits in Noise. . . . 7
Rolph Houben, __Inge Brons__, Maj Haverkate, and Wouter A. Dreschler

O6: Hearing aids in the noisy world: benefits and challenges . . . 8
__Maja Serman__, Sebastian Pape, and Marc Aubreville

O7: Near-End Listening Enhancement for Mobile Phones . . . 9
__Bastian Sauert__

O8: Auditory classification images: How noise can reveal the acoustic cues used in phoneme categorization . . . 10
__Léo Varnet__, Gwendoline Trollé, Willy Serniclaes, Kenneth Knoblauch, Fanny Meunier, and Michel Hoen

O9: Acoustic- and phonological-specific processing of speech perception . . . 11
__Sophie Bouton__ and Anne-Lise Giraud

O10: Beyond speech intelligibility testing: A memory test for assessment of signal processing interventions in ecologically valid listening situations . . . 12
__Thomas Lunner__, Elaine Ng, Mary Rudner, and Jerker Rönnberg

O11: Keynote Lecture The role of temporal fine structure in the perception of speech in background sounds by people with normal and impaired hearing . . . 13
__Brian C. J. Moore__

O12: Enhanced temporal coding in cochlear implants . . . 14
__Raphael Koning__ and Jan Wouters

O13: Perceiving phonetic variation in noise: structure, speaker, accent . . . 15
__Rachel Smith__

Poster Presentations

P1: Does F0-segregation interact with dip listening or spatial unmasking? 2
Thibaud Leclère, Mathieu Lavandier, and Mickael Deroche

P2: Speech understanding in realistic conditions: effects of number and type of interferers, and of head orientation. 3
John F. Culling and Sasha Priddy

P3: Periodicity and aperiodicity in the perception of speech in both steady-state and fluctuating maskers 4
Kurt Steinmetzger and Stuart Rosen

P4: Left temporal alpha-band activity reflects single word intelligibility 5
Robert Becker, Maria Pefkou, Christoph Michel, and Alexis Hervais-Adelman

P5: Use of prosodic information during sentence processing in fluctuating noise 6
Rebecca Carroll and Esther Ruigendijk

P6: Impact of beamforming algorithms on speech perception in cochlear implant users in a moving noise source condition 7
Tobias Weißgerber, Hannah-Lina Grahlmann, Tobias Rader, and Uwe Baumann

P7: Perceptual phase entrainment to the rhythm of speech entails a high-level process 8
Benedikt Zoefel and Rufin VanRullen

P8: Can alpha oscillations in the brain protect speech signals against interfering distractors? . 9
Malte Wöstmann, Antje Strauß, and Jonas Obleser

P9: The role of anticipatory and perseveratory coarticulation cues for speech perception in young and old listeners 10
Rachel Haines, Heather Fortnum, and Antje Heinrich

P10: The effect of retiming speech on masked intelligibility 11
Martin Cooke and Vincent Aubanel

P11: Spatial release from masking for horizontally and vertically distributed sources and the definition of the better ear. 12
Barry Bardsley and John.F. Culling

P12: Simulation data and a model approach for speech perception with electric-acoustic stimulation (EAS) 13
Tobias Rader, Youssef Adel, Uwe Baumann, and Hugo Fastl

P13: Elicitation and analysis of a robust word misperception corpus in Spanish 14
Attila Máté Tóth, Maria Luisa García Lecumberri, and Martin Cooke

P14: Boosting speech intelligibility using spectral reweighting under a constant energy constraint 15
Yan Tang and Martin Cooke

P15: Is there more to saliency than loudness? 16
Francesco Tordini

P16: Intelligibility of spoken and sung sentences in different types of background noise 17
Sarah Knight, Sarah Hawkins, Edward Wickham, Christopher Fox, and Antje Heinrich

P17: Effects of NAL-R on Consonant Vowel Perception in Hearing-Impaired Subjects 18
Christoph Scheidiger, Jont B. Allen, and Torsten Dau

P18: Developing speech recognition materials suitable for non-native speakers 19
Louise Stringer

P19: AAL: Intelligible City for All. Evaluation of ecological intelligibility test for normal hearing and hearing impaired adults in railway station context 20
Yosra Mzah, Neska El Haouij, Nader Mechergui, Amira Ben Jemaa, Meriem Jaïdane, Sonia Larbi, Sylvie Ghalila, Dhafer Malouche, Kevin Lecoz, Yannick Belouard, Cendrine Mercier, Jawad Hajjam, and Alessendra Brescia

P20: Rate discrimination at low pulse rates: Influence of intracochlear stimulation site between normal-hearing and cochlear implant listeners. 21
Pierre Stahl, Olivier Macherey, Sabine Meunier, and Stéphane Roman

P21: Factors limiting vocal-tract-length perception in cochlear-implants 22
Etienne Gaudrain and Deniz Başkent

P22: Perceptual restoration of degraded speech is preserved with advancing age 23
Jefta D. Saija, Elkan G. Akyürek, Tjeerd C. Andringa, and Deniz Başkent

P23: Pulse-spreading harmonic complexes: an advantageous carrier for simulating cochlear implants 24
Olivier Macherey, Quentin Mesnildrey, and Gaston Hilkhuysen

P24: Optimizing a noise reduction system with physical intelligibility metrics 25
Gaston Hilkhuysen and Mark Huckvale

P25: Consonant confusions in frozen and random white noise 26
Johannes Zaar, Søren Jørgensen, and Torsten Dau

P26: The effect of expertise in speech-in-speech and speech-in-noise tasks 27
Clara Suied, Jean-Christophe Bouy, and Guillaume Andéol

Author Index 28

Time table 6$^{\text{th}}$ SpiN Workshop 29

Oral presentations

O1: Simulating the bimodal excitation spread produced by Cochlear implant bipolar stimulation: Effect on speech intelligibility

Quentin Mesnildrey and Olivier Macherey

Laboratoire de Mécanique et d'Acoustique, Marseille, FR

Several studies comparing focused electrode configurations (bipolar or tripolar) with monopolar configuration have failed to find any benefits in terms of speech perception. This study specifically investigates the perceptual effects of bimodal excitation pattern produced by bipolar stimulation, using noise-vocoded CI simulations in which synthesis filters were designed to simulate the spread of activity in auditory neurons as predicted by a simple cochlear model.

Speech intelligibility was tested for normal hearing subjects using the French Matrix Test in the presence of a reversed speech masker. Experiment 1 tested three simulated modes (monopolar "MP", bipolar "BP+1", control "CTRL"), in function of the number of channels. The CTRL mode had synthesis filters identical to the analysis filters. We found a significant interaction between stimulation mode and the number of channels. Performance for MP and BP+1 clearly showed an asymptote at 8 channels whereas subjects continued to improve with CTRL, up to 15 channels. Also, the improvement in scores as a function of the number of channels was smaller for BP+1 than for MP or CTRL. This may be due to the increased channel interactions for BP+1 when each simulated electrode starts to act both as an active and as a return electrode for different channels.

Experiment 2 investigated whether the spectral discontinuity introduced by the bimodal shape may be detrimental to speech intelligibility. BP+1 was re-tested and compared to two new synthesis filters (Continuous "CTN" and Asymmetric "AS"). CTN and AS were identical to BP+1 except that the hole between the two peaks was filled for CTN and one of the peak was removed for AS. CTN and BP+1 yielded equivalent performances suggesting that the spectral discontinuity introduced by the bimodal pattern may not be a problem. Furthermore, performance for AS was better than for BP+1 and CTN. This last observation encourages the use of asymmetric pulse shapes designed to attenuate the bimodal excitation produced by bipolar stimulation.

In a third experiment, we focused on the effect of spacing between electrodes in BP stimulation. With 11 channels and 17 simulated electrodes, three compressed maps were tested with the BP+1 stimulation (*Apical*, *Basal* and *Centered*). Despite the fact that the peaks of the BP excitation pattern are all the more narrow that the electrodes are close, increasing this spacing reduces the number of electrodes which act both as an active and as a return electrode for different channels. To investigate on this effect, we tested two wide BP configurations, *BP+5* and *WideNoise*. In this last condition, overlapping parts of the signals were removed and replaced by a stationary noise spectrally restricted to the same initial frequency range. The BP+1 conditions showed a deleterious effect of mismatch consistent with previous studies. However, subjects performed worse with the BP+5 than with the BP+1 configuration. This result differs from CI users data reported in Pfingst et al, 2001. Two effects can be pointed out as possible explanations. First, the mismatch in our simulations was underestimated compared to CI users. Second, although the BP+5 configuration generates less interactions, the overlapping signals convey information from remote part of the spectrum. The weak correlation between the overlapping envelopes could have been highly deleterious. This assumption was confirmed by the best overall performance obtained with *WideNoise* configuration.

Acknowledgements

Work funded by the ANR (Project DAIMA)

O2: How CI users can make the best of their implants in SpiN situations: Positioning in a room, head orientation strategy and translational avenues

Jacques A. Grange and John F. Culling

Cardiff University, UK
grangeja@Cardiff.ac.uk

Spatial release from masking (SRM) traditionally shows how spatially separating masker from target improves speech in noise (SpiN) intelligibility. With the guidance of a binaural model of SRM (Jelfs et al, 2011), we previously showed that CI users (CIs) can benefit in some configurations from much larger SRM than reported elsewhere for a sub-optimal 90° separation whilst facing the speech. We thereby revealed (Culling et al, 2012) further benefits of bilateral over unilateral implantation (up to 20 dB vs. 6~10 dB previously demonstrated). This preliminary work could translate in helping CIs adopt an optimal position in a noisy setting such as a restaurant. In an effort to study more natural situations we investigated the predicted benefits of freeing the head and specifically those of modest head turns away from the speech direction.

Adult participants were 8 normally hearing listeners (NHs) age-matched to 8 bilateral CIs (BCIs) and 9 unilateral CIs (UCIs). Four spatial configurations (speech at 0°, masker at 0, ± 90 and 180°) and three head orientations (0 and $\pm 30°$) were chosen on the basis of model predictions and expected comfort limits to the gaze angle required to lip-read with a head-turn. The speech was presented in audio or audio-visual (AV) conditions in a sound-deadened room. The model, once fed with impulse responses acquired in the room with a Kemar manikin, predicted a 5~7.5 dB and 4~5 dB benefit of a favourable 30° head turn for NHs and CIs respectively.

We first recorded undirected then directed (i.e. informed), free-head orientation when listeners were presented with a progressively declining speech-to-noise ratio (SNR). Through acquisition of fixed-head audio and AV speech reception thresholds (SRTs) with a bespoke adaptive paradigm making use of SPIN sentences, we then investigated head-orientation improvements with and without lip-reading. Finer measurements of SRTs in audio only were finally acquired with IEEE sentences and a previously employed paradigm.

Whilst our hypothesis was that CIs would make more use of head orientation than NHs, 50% of NHs and just 10% of CIs spontaneously turned their heads. Of those who did, 25% of NHs and 70% of CIs gained improvement in SRM. All performed better in the directed paradigm although CIs made more effective use of the guidance given to them. SRT measurements resulted in NHs and UCIs benefiting from a 30° head turn as predicted. However, BCIs reached only half the predicted benefit. The large summation and squelch levels BCIs benefited from may be of spectral nature and could explain a reduction in BCIs head-turn benefit. Lip-reading yielded a further benefit of 3 dB and 5dB for NHs and CIs respectively, regardless of head orientation, thereby supporting head-orientation as an effective SpiN intelligibility improvement tool for CIs.

The SNR reached in the first experiment reflected the SRM corresponding to the listeners' final head orientation. The three experiments combined demonstrate how CIs could benefit from training to optimise position & head orientation with respect to speech and noise sources in a social setting. This is reinforced by their propensity for facing a speaker. One can easily appreciate how the same tests could lead to similar conclusions for HA users and other hearing impaired listeners.

References

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

Culling, J. F., Jelfs, S., Talbert, A., Grange, J.A. & Backhouse, S. S. (2012). The benefit of bilateral vs. unilateral cochlear implantation to speech intelligibility in noise. *Ear and hearing, 33(6)*, 673-82.

O3: Speech recognition in N-talker babble: Patterns of performance with increasing N vary across types of speech material

Tim Green and Stuart Rosen

Speech, Hearing and Phonetic Sciences, UCL, London, UK
tim.green@ucl.ac.uk

Several factors influence the extent to which a target talker can be understood against a background of other talkers, including differences in voice pitch which can be used to segregate different talkers, the extent to which fluctuations in the level and spectral content of the competing speech allow 'glimpses' of the target speech, and aspects of informational masking, such as the distracting effect of the content of the interfering speech. The balance of these factors can be expected to change as the number of competing talkers increases, and may also vary according to the type of speech material used to assess performance. Here, recognition of target speech from one male and one female speaker was measured in normally-hearing listeners with various numbers of competing talkers (N = 1, 2, 4, 8 or 16; either all male or all female). Four types of speech materials were used: IEEE and BKB sentences, monosyllabic words, and VCVs. Signal-to-noise ratio was fixed for a given combination of target speaker and material. Patterns of performance varied substantially according to the type of speech material and the relationship between the sex of the target speaker and that of the competing talkers. For sentences with same-sex competing talkers there was a precipitous decline in recognition as N increased from 1 to 2, likely reflecting a large amount of informational masking in these conditions. Further increases in N produced either no change or a small increase in performance. For all other combinations of speech material and masker sex, performance continued to decline as N increased beyond 2. For word recognition, performance reached a minimum with N=4, while for VCVs, minimum performance was typically reached with N=8, presumably reflecting the extent to which shorter and shorter glimpses can be useful.

Acknowledgments
Supported by the MRC.

O4: Top-down restoration of speech: Hearing impairment and age can influence it in unexpected ways

Deniz Başkent[1,2]

[1] Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, NL
[2] Research School of Behavioral and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, NL
d.baskent@umcg.nl

In real life communication, interfering sounds commonly obliterate speech signals. To enhance communication, a top-down perceptual restoration mechanism is constantly active, filling in for the inaudible parts of speech. In this process, restoration is achieved through activation of the best lexical candidates, based on the speech features available from the audible parts, as well as using linguistic knowledge, context, and expectations.

Hearing-impaired listeners and users of cochlear implants commonly report difficulty in understanding speech in noise. In our research, we had originally hypothesized that the changes imposed by hearing impairment or hearing devices on the bottom-up speech cues could limit the benefit from the top-down restoration, contributing to this difficulty. Our work with moderately hearing-impaired listeners (Başkent, 2010; Başkent et al., 2010) and acoustic simulations of cochlear implants (Başkent, 2012) strongly supported this hypothesis, and showed no restoration benefit in both situations. Our recent work with actual users of implants revealed a more complex picture, namely, that restoration in implant users is indeed different than that of normal hearing, but is not necessarily reduced in an expected way (Bhargava et al., In revision). Considering both hearing impairment and implant use are strongly correlated with advanced age, we have recently also studied the effect of age alone in a normal-hearing population (Saija et al., 2013). Based on the potential decrease in some cognitive functions due to advanced age, one could expect restoration mechanisms to be also negatively affected. In contrary to this expectation, older people showed a strong benefit from restoration, perhaps relying on long-term linguistic knowledge and vocabulary.

Overall, our research has shown that, indeed changes in sensory or cognitive functioning, for example due to hearing impairment or aging, can affect the top-down restoration mechanisms. While these changes may be contributing to problems of hearing-impaired people in understanding speech in noise, the interactions of hearing impairment, hearing devices, and aging with top-down restoration are observed to be more complex than were initially hypothesized.

Acknowledgments

Support provided by the Netherlands Organization for Scientific Research, VIDI grant 016.096.397, the Rosalind Franklin Fellowship, and the Heinsius Houbolt Foundation. The study is part of t he research program of our department: Healthy Aging and Communication.

References

Başkent, D. (2010). “Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores.” J Acoust Soc Am Letters 128, EL169-EL174.

Başkent, D., Eiler, C. L., and Edwards, B. (2010). “Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss,” Hear Res 260, 54 -62.

Başkent, D. (2012). “Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electricacoustic stimulation,” J Assoc Res Otolaryngol 13, 683 -692.

Bhargava, P., Gaudrain, E., and Başkent, D. (In rev.). “Top-down restoration of speech in cochlear-implant users,” Hear. Res.

Saija, J. D., Akyürek, E. G., Andringa, T. C., and Başkent, D. (2013). “Perceptual restoration of degraded speech is preserved with advancing age,” J. Assoc. Res. Otolaryngol., in press . E-pub: Nov 2013.

O5: Measuring Listening Effort with Reaction Time to Digits in Noise.

Rolph Houben, **Inge Brons**, Maj Haverkate, and Wouter A. Dreschler

Academic Medical Center Amsterdam, Amsterdam, NL
i.brons@amc.uva.nl

Audiological rehabilitation can be beneficial for speech communication even if it does not have a measurable effect on speech intelligibility. For example, single-channel noise reduction is often preferred by hearing-aid wearers, while it does not improve speech intelligibility. We hypothesized that this is caused by an improvement in listening effort. Unfortunately, there is no easy method for the measurement of listening effort. The purpose of the current study is to determine if the measurement of the reaction time to spoken digits in noise can be used to measure listening effort at signal to noise ratios (SNRs) that are high enough to render the speech completely intelligible. We measured reaction times for two tasks. In the first task, participants had to quickly identify the last digit of a triplet ('identification'). In the second task they had to quickly add the first and the last digit ('arithmetic'). Three subsequent experiments will be presented, in which we determined (1) if the reaction time to digits increases for normal-hearing listeners when noise is added to the speech, and if the effect of noise differs for the two tasks, (2) if the reaction time to digits in noise is altered by noise-reduction processing for normal-hearing listeners, and (3) if these noise-reduction results are influenced by hearing loss.

O6: Hearing aids in the noisy world: benefits and challenges

Maja Serman, Sebastian Pape, and Marc Aubreville

Siemens Audiologische Technik GmbH, Erlangen, DE
maja.serman@siemens.com

The auditory world that we are immersed in is full of internal and external noises. Auditory illusions such as phonemic restoration show us that speech perception can be resilient to or even improved by noise in some situations. How noise influences our performance depends on the type of noise and its interaction with the signal, but also on the abilities of the individual to deal with this interaction.

Hearing impairment degrades speech in noise (SiN) understanding; the more severe the hearing loss, the more severe the degradation. In this talk, a short summary of the hearing aid (HA) features that have an impact on SiN understanding will be presented. The talk will also consider how properties of the HA input signal affect signal processing algorithms behaviour, and how well the HA estimation of the sound type (*i.e.* speech, noise, music, etc.) matches the patient's perception of the sound environment.

O7: Near-End Listening Enhancement for Mobile Phones

Bastian Sauert

RWTH Aachen University, DE
sauert@ind.rwth-aachen.de

Mobile telephony is often conducted in the presence of acoustical background noise such as traffic or babble noise. In this situation, the near-end listener perceives a mixture of clean far-end (downlink) speech and environmental noise from the near-end side, which goes along with an increased listening effort and possibly reduced speech intelligibility. As in many cases the noise signal cannot be influenced, the manipulation of the far-end signal is the only way to effectively improve speech intelligibility and to ease listening effort for the near-end listener by digital signal processing. We call this approach *near-end listening enhancement* (NELE).

In this talk, innovative solutions for the problem of near-end listening enhancement are presented. These optimize the intelligibility of the far-end speech in local background noise with respect to the objective criterion *Speech Intelligibility Index* (SII). In contrast to state-of-the-art techniques, the developed methods tackle the problem from the application perspective considering also the requirements and restrictions of realistic scenarios such as in mobile phones. It is of particular importance that the processing adapts dynamically to the sound characteristics of the ambient noise. Hence, an effective intelligibility enhancement is provided in the presence of background noise, while in silence *no* audible modification is applied. The utilized noise tracking algorithm estimates the noise spectrum blindly from the microphone signal, the only access to the acoustical environment. Furthermore, a power limitation in critical bands ensures that the ear of the near-end listener is protected from damage and pain.

In mobile phones, the restrictions of the so-called micro-loudspeakers need to be considered. Especially their maximum thermal load constitutes a major limitation. This leads to an optimization of the SII with the constraint that the total audio power may only be increased up to a maximum power. As a result, significant improvements of speech intelligibility under adverse acoustical conditions are achieved. In the most difficult scenario where an increase of total audio power is not allowed, the word recognition rate improves with the proposed algorithms by up to 22 percentage points.

It is shown, that the developed concepts can also be applied in different devices such as mobile phones, headphones, hands-free conference terminals, car multimedia systems, public address systems, and hearing aids.

O8: Auditory classification images: How noise can reveal the acoustic cues used in phoneme categorization

Léo Varnet[1,6], Gwendoline Trollé[4], Willy Serniclaes[5], Kenneth Knoblauch[2,6], Fanny Meunier[3,6], and Michel Hoen[1,4,6]

[1] Lyon Neuroscience Research Center, CNRS UMR 5292, Auditory Language Processing (ALP) research group, Lyon, FR
[2] Stem Cell and Brain Research Institute, INSERM U 846, Integrative Neuroscience Department, Bron, FR
[3] Laboratoire sur le Langage le Cerveau et la Cognition, CNRS UMR 5304, Auditory Language Processing (ALP) research group, Lyon, FR
[4] INSERM U1028, Lyon Neuroscience Research Center, Brain Dynamics and Cognition Team, Lyon, FR
[5] Université Libre de Bruxelles, UNESCOG, CP191 Bruxelles, BE
[6] Université de Lyon, Université Lyon 1, Lyon, FR
leo.varnet@isc.cnrs.fr

The exact relationship between acoustic/phonetic cues extraction and phoneme categorization still constitutes an open question for research on speech perception. Since the 1950s, speech synthesis and degraded speech have been widely used for the identification of functional phonetic cues; however a major disadvantage of these methods is that they require prior knowledge of the cues being sought. Up to now, there is no turnkey solution for isolating speech cues from natural stimuli.

In this talk, I will present a psychoacoustic imaging method, inspired by recent theoretical developments in visual psychophysics, and allowing experimenters to directly see where humans listen inside natural speech utterances. This "Auditory Classification Image" technique relies on a Generalized Linear Model with smoothness priors to link categorization errors in a speech-in-noise comprehension task with the trial-specific distribution of noise, resulting in a spectrotemporal map of the acoustic cues used to categorize phonemes.

Here I will demonstrate the effectiveness of this method through two examples: 1) a two-alternative forced choice experiment between stimuli 'aba' and 'ada' in noise, and 2) a more complex experiment on phonetic context effects in speech (Mann, 1980).

O9: Acoustic- and phonological-specific processing of speech perception

Sophie Bouton and Anne-Lise Giraud

Auditory Language Group, Department of Neuroscience, University Medical Centre (CMU), Geneva, CH
sophie.bouton@unige.ch

How do we categorize acoustically variable inputs into a discrete phonological code? Perception of spoken syllables relies in the perception of a combination of relevant acoustic cues that match representations. Using a classical procedure of gradually morphed syllables along an acoustic continuum, we disambiguated the neural responses to acoustic processing from those associated with phonological categorization (representation and decision).

Regression analyses revealed temporally and spatially distinct correlates of acoustic, phonological and decisional processing in speech categorical perception. The encoding of each perceptual/decision processing was largely segregated (spatially) speaks to a hierarchical functional neuroanatomy. We find an acoustic vs. phonological dissociation between left PAC (primary auditory cortex) and STS (superior temporal sulcus). Moreover, speech categorical processing varied with the ambiguity of the stimuli. Within regions that were specifically responsive to phonological categories, some were involved when the stimuli where not ambiguous at all (left STS for prototypes) and others only when the categorical decision was difficult (left STG for stimuli close to phonological boundaries). This dissociation signals a functional neuroanatomical hierarchy in the phonological decision process, in which the regions more involved in the decisional steps are close to those where acoustic information is available.

O10: Beyond speech intelligibility testing: A memory test for assessment of signal processing interventions in ecologically valid listening situations

Thomas Lunner[1,2,3], Elaine Ng[1], Mary Rudner[1], and Jerker Rönnberg[1]

[1] Linnaeus Centre HEAD, Swedish Institute for Disability Research, Department of Behavioral Sciences and Learning, Linkping University, SE
[2] Eriksholm Research Centre, Oticon A/S, Snekkersten, DE
[3] of Clinical and Experimental Medicine, Linköping University, SW
TLU@Eriksholm.com

Performance of hearing aid signal processing is often assessed by speech intelligibility in noise tests, such as the HINT, CRM, or SPIN sentences presented in a background of noise or babble. Usually these tests are most sensitive at a signal-to-noise ratio (SNR) below 0 dB. However, in a recent study by Smeds et al. (2012) it was shown that the SNRs in ecological listening situations (e.g. kitchen, babble, and car) were typically well above 0 dB SNR. That is, SNRs where the speech intelligibility in noise tests are insensitive.

Cognitive Spare Capacity (CSC) refers to the residual capacity after successful speech perception. In a recent study by Ng et al. (2010), we defined the residual capacity to be number of words recalled after successful listening to a number of HINT sentences, inspired by Sarampalis et al. (2009).

In a recent test with 26 hearing impaired test subjects we showed that close to 100% correct speech intelligibility in a four talker babble noise required around + 7 dB SNR. At that SNR it was shown that a hearing aid noise reduction scheme improved memory recall by about 10-15%. Thus, this kind of memory recall test is a possible candidate for assessment of hearing aid functionality in ecologically relevant (positive) SNRs.

O11: Keynote Lecture
The role of temporal fine structure in the perception of speech in background sounds by people with normal and impaired hearing

Brian C. J. Moore

Department of Experimental Psychology, University of Cambridge, UK
bcjm@cam.ac.uk

In a normally functioning cochlea, complex broadband signals such as speech are decomposed by the filtering on the basilar membrane (BM) into a series of narrowband signals. The waveform at each place on the BM can be considered as an envelope (ENV) superimposed on a more rapidly oscillating carrier, the temporal fine structure (TFS). A distinction is made between the physical ENV and TFS of the input signal ($\mathrm{ENV_p}$ and $\mathrm{TFS_p}$), the ENV and TFS at a given place on the BM ($\mathrm{ENV_{BM}}$ and $\mathrm{TFS_{BM}}$), and the neural representation of ENV and TFS ($\mathrm{ENV_n}$ and $\mathrm{TFS_n}$).

The importance of ENV and TFS information for speech perception has been explored using various forms of vocoder processing. In such processing, the signal is filtered into several frequency channels, and $\mathrm{ENV_p}$ and $\mathrm{TFS_p}$ are estimated for each channel. The signal in each channel is then manipulated so as to alter either $\mathrm{ENV_p}$ or $\mathrm{TFS_p}$. Finally, the manipulated channel signals are combined and the intelligibility of the processed speech is assessed using normal-hearing or hearing-impaired listeners. There are two problem with such processing. Firstly, stimulus manipulations intended to disrupt $\mathrm{TFS_n}$ while preserving $\mathrm{ENV_n}$ do have disruptive effects on $\mathrm{ENV_n}$, and stimulus manipulations intended to disrupt $\mathrm{ENV_n}$ while preserving $\mathrm{TFS_n}$ do have disruptive effects on $\mathrm{TFS_n}$. Secondly, while vocoder processing can be used to selectively remove the original $\mathrm{TFS_p}$ or the original $\mathrm{ENV_p}$ in some or all channels, $\mathrm{TFS_n}$ and $\mathrm{ENV_n}$ cues are present in the auditory system. Despite these problems, analyses using auditory models suggest that it is possible to process sounds so as to differentially affect $\mathrm{TFS_n}$ and $\mathrm{ENV_n}$ cues. The results of studies using vocoder processing are consistent with the idea that TFS cues are used for speech perception, especially for the perceptual segregation of target speech from background sounds, and that cochlear hearing loss and/or increasing age reduce the ability to use TFS cues. Studies of the correlation between speech reception and sensitivity to TFS, as measured using non-speech stimuli, support these ideas.

O12: Enhanced temporal coding in cochlear implants

Raphael Koning and Jan Wouters

Research group ExpORL, KULeuven, Leuven, BE
Raphael.Koning@med.kuleuven.be

Speech intelligibility can be very good in quiet for cochlear implant (CI) users. In adverse listening conditions, the speech perception rapidly decreases.

Recent studies have confirmed that rapid changes in temporal and spectral content are the most important parts of the speech signal. In this study, a speech enhancement algorithm called enhanced envelope (EE) strategy has been developed that amplifies the onsets of the speech envelope in all frequency bands. The enhanced temporal coding is achieved by introducing additional peak signals at the onsets.

In this study, the EE strategy was evaluated with respect to the advanced combined encoder (ACE) strategy in CI users. The potential of the strategy was investigated in stationary speech shaped noise and with an interfering talker. Additionally, a loudness rating was done to investigate the influence of the enhanced onsets on loudness perception.

All CI users showed an immediate benefit of the EE algorithm in comparison to the ACE reference that they were using in their clinical devices. Overall, SRT improvements around 2 dB were obtained for the stationary speech shaped noise and the interfering masker.

The results suggest that speech intelligibility can be improved in adverse listening conditions by an enhanced temporal coding of the onsets in the speech envelope in the signal processing path of CIs.

O13: Perceiving phonetic variation in noise: structure, speaker, accent

Rachel Smith

University of Glasgow, Glasgow, UK
Rachel.Smith@glasgow.ac.uk

Systematic phonetic variation in the speech signal reflects a number of aspects of linguistic structure and individual and social identity. Knowledge of these patterns has been argued to contribute to speech understanding (e.g. [1], [2]). In this talk I discuss what speech-in-noise studies can tell us about the contribution of phonetic detail to word recognition.

Three experiments tested the perception of various types of phonetic detail under adverse listening conditions, specifically multi-talker cafeteria noise. The first experiment demonstrated that coherent phonetic detail reflecting an utterance's grammatical structure improves intelligibility in noise [3]. The second experiment [4] showed that listeners draw upon their knowledge of speaker-specific and structure-specific detail when listening to speech in noise: specifically, familiarity with the way a particular talker produces phonetic cues to consonants' word affiliation, as in the /s/ of *cat size* vs. *cat's eyes*, confers a small but significant intelligibility advantage. Finally the third experiment [5] investigated the impact of both long-term familiarity with a regional accent, and short-term familiarisation with an individual talker, on intelligibility-in-noise, and found effects of both types of familiarity on intelligibility scores. Ongoing work seeks to relate listeners' performance in noise to specific phonetic differences between the two accents in terms of cues to word boundaries.

Taken together, these studies show that relatively subtle aspects of phonetic detail survive embedding in noise (at least in the type of non-stationary noise used in these studies, which permits glimpsing). They further demonstrate that adverse conditions can reveal a facilitatory role in word recognition for constellations of weak cues that are coherent with the listener's experience of speech.

References

[1] Hawkins, S., and Smith, R. (2001). Polysp: a polysystemic, phonetically-rich approach to speech understanding. *Rivista di Linguistica*, 13, 99-188.

[2] Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31, 373-405.

[3] Baker, R. (2008). The production and perception of morphologically and grammatically conditioned phonetic detail. Unpublished PhD thesis, University of Cambridge.

[4] Smith, R. and Hawkins, S. (2012). Production and perception of speaker-specific phonetic detail at word boundaries. *Journal of Phonetics*, 40, 213-233.

[5] Smith, R., Holmes-Elliott, S., Pettinato, M., & Knight, R.-A. (2013). Cross-accent intelligibility of speech in noise: Long-term familiarity and short-term familiarisation. *The Quarterly Journal of Experimental Psychology*, DOI: 10.1080/17470218.2013.822009.

Poster presentations

P1: Does F0-segregation interact with dip listening or spatial unmasking?

Thibaud Leclère[1], Mathieu Lavandier[1], and Mickael Deroche[2]

[1] Université de Lyon - École Nationale des Travaux Publics de l'État (ENTPE) Laboratoire Génie Civil et Bâtiment (LGCB), Lyon, FR

[2] Johns Hopkins University School of Medicine, Department of Otolaryngology, Baltimore, USA

thibaud.leclere@entpe.fr

Differences in fundamental frequency (F0) and location between a speech target and a masker as well as amplitude modulations in the masker are helpful cues to improve speech intelligibility in cocktail party situations. Each cue has been thoroughly investigated independently in many studies but it remains unclear whether they interact with each other. Experiment 1 examined potential interactions between F0-segregation and dip listening while experiment 2 examined interactions between F0-segregation and spatial unmasking. Speech reception thresholds were measured for a monotonized or an intonated voice against eight types of harmonic complex interferers. In experiment 1, the eight interferers varied in F0 contour (monotonized or intonated), mean F0 (0 or 3 semitones above the target) and broadband temporal envelope (stationary or 1-voice modulated). In experiment 2, interferers varied in F0 contour, mean F0 and spatial location (colocated or separated from the target). Thirty-two listeners participated in each experiment. The results will be presented and discussed.

P2: Speech understanding in realistic conditions: effects of number and type of interferers, and of head orientation.

John F. Culling and Sasha Priddy

School of Psychology, Cardiff University, Cardiff, UK
cullingj@Cardiff.ac.uk

In real life, interfering noise is frequently uninterrupted, multi-source spatially distributed and reverberant. Three experiments measured SRTs against uninterupted interferers based on continuous discourse in different virtual environments. Experiments 1 and 2 compared the effects of speech, reversed speech, speech-shaped noise and speech-modulated speech-shaped noise as a function of the number of interfering sources at a constant overall noise level (controlled at source). Experiment 1 manipulated reverberation using a simulated room, while Experiment 2 used recordings from an acoustic manikin in a real restaurant. Experiment 3 examined the effect of the manikin's head orientation in the presence of eight interferers (speech or speech-shaped noise) in the same restaurant. Experiment 1 found that SRTs were elevated in reverberation and, except in the case of reverberated speech-shaped noise, for larger numbers of interferers. The effect of reverberation decreased with increasing numbers of interferers. Reversed speech produced the lowest and continuous noise the highest SRTs in most cases. The results suggest beneficial effects of masker periodicity and modulation for 1 or 2 interfering sources (SRTs lower for modulated than continuous sources, and lower for single-voice speech and reversed speech sources than for modulated noise). There were detrimental effects of masker intelligibility (lower SRTs with reversed than with forward speech interferers) with 2 interferers, and there was evidence of modulation masking with 8 interferers (lowest SRTs for continuous noise). Experiment 2 found a similar pattern of results to the reverberant case in Experiment 1. Experiment 3 found consistently beneficial effects of head orientation away from the target source. This benefit was driven by improved target-speech level at the ear turned towards the source. As with 8 interferers in Experiments 1 and 2, speech interferers produced slightly higher SRTs than noises.

P3: Periodicity and aperiodicity in the perception of speech in both steady-state and fluctuating maskers

Kurt Steinmetzger and Stuart Rosen

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

Background Hearing-impaired listeners and cochlear implant (CI) users experience great difficulties in speech perception in all types of background noise, and unlike normal-hearing listeners show little benefit from fluctuations in the masker. One popular (if partial) explanation for these difficulties proposes a key role for temporal fine structure (TFS) cues. However, there is controversy over whether TFS has a special role in allowing fluctuating masker benefit or whether its contribution to speech perception is just as important for steady maskers. We investigated the abilities of normal-hearing listeners to perceive speech targets in the background of noise maskers in a variety of conditions mixing presence and absence of periodicity in both target and masker.

Methods Speech Reception Thresholds (SRTs) were measured adaptively for IEEE sentences processed to change their source characteristics (and hence their periodicity). The sentences were combined with four different maskers: speech-shaped noise and harmonic complexes with a dynamically varying fundamental frequency, that were both either steady or fluctuating at 10 Hz. Additionally, we varied the spectral resolution of the target speech over a wide range. In two separate experiments 16 and 12 normal-hearing subjects were tested.

Results It was found that increasing the amount of periodicity in the masker strongly aids speech intelligibility across all levels of spectral resolution, while no such effect was observed for the periodicity of the target speech. Contrary to what has been believed previously, the fluctuating-masker benefit (FMB) is markedly smaller for periodic than for aperiodic maskers, and a substantial FMB required a rather high spectral resolution of the target speech. Furthermore, a natural mix of periodicity and aperiodicity in the target speech was observed to foster the ability to glimpse, no matter if the masker was periodic or aperiodic.

Conclusions Our results show the importance of periodicity in tracking a speech signal through a background noise and suggest that the inability to exploit periodicity in segregating masker and target speech may be an even more important factor in the limitations of hearing-impaired and CI speech perception than the inability to benefit from fluctuations in a masker.

Acknowledgments

This project has been funded with support from the European Commission under Contract FP7-PEOPLE-2011-290000 and the Medical Research Council (Grant Number G1001255).

P4: Left temporal alpha-band activity reflects single word intelligibility

Robert Becker[1], Maria Pefkou[2], Christoph Michel[1], and Alexis Hervais-Adelman[2]

[1] Functional Brain Mapping Lab, University of Geneva, University Medical School, Geneva, CH
[2] Brain and Language Lab, University of Geneva, University Medical School, Geneva, CH
alexis.adelman@unige.ch

Investigations of the electroencephalographic (EEG) correlates of degraded speech perception have often been inconclusive as to whether observed differences in brain responses between conditions result from different acoustic properties of more or less intelligible stimuli or whether they relate to cognitive processes implicated in comprehending challenging stimuli.

In this study we use noise vocoding to spectrally degrade monosyllabic words in order to manipulate their intelligibility. We used spectral rotation to generate incomprehensible control conditions matched in terms of spectral detail. We recorded EEG from 14 volunteers who listened to a series of noise vocoded (NV) and noise-vocoded spectrally-rotated (rNV) words, while they carried out a detection task. We specifically sought components of the EEG response that showed an interaction between spectral rotation and spectral degradation. This reflects aspects of the EEG response that are related to intelligibility of acoustically degraded monosyllabic words, while controlling for spectral detail.

Analyses of event-related potentials showed an interaction effect for a P300-like component at several centro-parietal electrodes. Time-frequency analysis of the EEG signal in the alpha-band revealed a monotonic increase in event-related desynchronization (ERD) for the NV but not the rNV stimuli in the alpha band at a left temporo-central electrode cluster from 420-560ms reflecting a direct relationship between the strength of alpha-band ERD and intelligibility.

By matching NV words with their incomprehensible rNV homologues, we reveal the spatio-temporal pattern of evoked and induced processes involved in degraded speech perception, largely uncontaminated by purely acoustic effects.

P5: Use of prosodic information during sentence processing in fluctuating noise

Rebecca Carroll and Esther Ruigendijk

Exzellenzcluster “Hearing4all”, University of Oldenburg, DE
rebecca.carroll@uni-oldenburg.de

Acoustic cues used for prosodic phrasing (amplitude modulation, pitch contour, rhythmic information) have been shown to be robust information against a stationary speech-shaped noise masker (e.g., Carroll, 2013). Although a noise masker with an amplitude modulated envelope (i.e., stationary noise) has been shown to be beneficial for speech recognition due to 'glimpsing' (e.g., Cooke 2006), preliminary findings suggest that the “prosodic benefit” observed for speech in stationary noise does not necessarily hold to the same degree in a fluctuating noise masker. By manipulating the rhythmic modulation of the noise masker envelope, different cues (rhythm, duration vs. F0 pitch contour) may become more or less reliable and hence useful for stream segregation. Successful processing of sentences which mainly rely on the use of prosodic phrasing for correct interpretation may therefore be differentially affected by different noise maskers (stationary vs. fluctuating vs. no masker). We test two structural conditions at suprathreshold SNRs and near perfect intelligibility. The Closure Positive Shift (CPS; Steinhauer et al., 1999) as an electrophysiological (ERP) correlate of prosodic boundaries is used to determine the impact of these noise maskers on processing the prosodic information in 30 young listeners with normal hearing. Stationary noise is expected to elicit a typical CPS effect comparable to a CPS elicited in silence, whereas rhythmically modulated noise is expected to negatively influence the rhythmic structuring of the speech signal, which may result in cancellation or depletion of the CPS component. The electrophysiological measure will provide a clearer picture not only on the correct interpretation, but also on individual differences due to listening effort. An additional behavioral task determines whether the listener correctly interpreted the sentence. Missing CPS components in one condition may thus be interpreted as a result of misinterpreting the prosodic information.

References

Carroll, R. (2013). *Effects of Syntactic Complexity and Prosody on Sentence Processing in Noise.* Aachen: Shaker.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *JASA, 11,* 1562-1573.

Steinhauer, K., Alter, K., & Friederici, A. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2,* 191-196.

P6: Impact of beamforming algorithms on speech perception in cochlear implant users in a moving noise source condition

Tobias Weißgerber, Hannah-Lina Grahlmann, Tobias Rader, and Uwe Baumann

Audiological Acoustics, ENT Department, Goethe-University Frankfurt am Main, DE
Tobias.Weissgerber@kgu.de

The assessment of speech perception in noise in clinical audiological diagnostics is typically carried out in test setups containing only a few loudspeakers. In order to determine the impact of advanced signal processing algorithms in cochlear implants (CI) or hearing aids on noise suppression or beamforming, a sophisticated sound reproduction system with 128 loudspeakers was realized.

Speech reception thresholds (SRT) in a moving noise source condition were measured in 12 normal hearing participants and 14 CI users (CP810, Cochlear). 7 subjects were bilateral CI users and 7 were bimodal with a hearing aid in the contralateral ear. Speech processor settings were either 'everyday', static beamforming (ZOOM), or adaptive beamforming (BEAM). Speech was always presented from 0°. The moving noise source was generated by means of wave field synthesis. Noise was presented in either two conditions: quasi-continuous (Oldenburg Noise, OLnoise) or modulated (Fastl-Noise).

Average SRT in in the control group was -10 dB SNR. Average SRT in both CI groups was -3 dB (OlNoise) and +10 dB (Fastl-Noise). With the ZOOM setting, a decreased SRT of 4 dB and with the BEAM setting an even more decreased SRT of 7.5 dB compared to the everyday setting was observed on average in both CI groups.

The implementation of adaptive beamforming strategies provides benefit in listening situations with a single moving noise source in the rear. WFS can serve as a sophisticated tool to assess potential benefits of advanced signal processing strategies.

P7: Perceptual phase entrainment to the rhythm of speech entails a high-level process

Benedikt Zoefel and Rufin VanRullen

Centre de Recherche Cerveau et Cognition (CerCo), CNRS, Toulouse, FR
zoefel@cerco.ups-tlse.fr

In speech sound, fluctuations in amplitude and spectral content occur rhythmically. Recent studies show that neural oscillations in auditory cortex entrain to this rhythm of speech, resulting in an alignment of high excitability phases with informative features, and possibly reflecting attentional selection. However, mechanisms of phase entrainment are still unclear, because in a standard speech sequence phonetic information (carried by momentary deviations in spectral content), is generally confounded with overall amplitude modulations: There is no phonetic information when spectral energy is absent; and there are moments of silence or low spectral energy between successive phonemes. Therefore, it is not clear whether brain oscillations truly align to fluctuations in phonetic information (a high-level process), or merely to the rhythmic changes in input sound amplitude (a low-level process).

Here, we disentangled these alternatives by constructing speech/noise stimuli whose spectral energy is comparable at each moment in time. Original speech snippets were recorded by a male speaker reading parts of a novel. Experimental stimuli were constructed by merging original speech with noise - critically, rhythmic fluctuations in spectral content and energy were thus removed. However, some points of the constructed stimuli merely contained noise while others contained real phonemes. Phonetic information still fluctuated rhythmically at ~2-8Hz, providing potential means for oscillatory phase entrainment. We assessed this entrainment by presenting auditory clicks at threshold level at random moments during our speech/noise snippets. Subjects indicated detection of a click by a button press. If there were entrainment to the constructed speech/noise sound, click detection should vary as a function of the original speech envelope (or equivalently, as a function of phonetic information). Indeed, we show that detection of a click depends on the phase of the original speech envelope before and after click onset. Yet by construction, both spectral content and amplitude of the speech/noise sequence were statistically indistinguishable across the different rhythmic phases. Thus, we demonstrate that the auditory threshold can be entrained by the rhythm of speech even when phonetic information is not accompanied by concomitant changes in input sound amplitude. In other words, phase entrainment to speech rhythm entails a high-level process.

P8: Can alpha oscillations in the brain protect speech signals against interfering distractors?

Malte Wöstmann[1,2], Antje Strauß[1], and Jonas Obleser[1]

[1] Max Planck Research Group “Auditory Cognition”, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, DE

[2] International Max Planck Research School on Neuroscience of Communication, Leipzig, DE

woestmann@cbs.mpg.de

Speech processing is demanding if distracting talkers interfere with the to-be-perceived signal. When confusability of target speech and distractors is high, successful understanding requires the protection of target speech against distractor intrusions on a neural level. We presume that cortical alpha (~10 Hz) oscillations associated with functional inhibition of task-irrelevant brain areas significantly support speech understanding by suppressing neural distractor processing. We present data from a recent electroencephalography (EEG) study ($n = 38$), showing a prominent increase in alpha activity when participants listened to target speech while ignoring a distracting talker. When separability of target and distractor was decreased by temporal-fine-structure degradation of acoustic detail, alpha activity increased strongest. Findings suggest that enhanced alpha activity is needed to suppress a distractor that would otherwise interfere with the target signal. To examine auditory distractor suppression directly, we designed a dichotic listening working-memory paradigm. Participants were cued to attend and maintain a stream of four spoken digits on one ear, while ignoring a distracting (same- talker) stream of digits on the other ear. Both streams contained broadband background noise. Participants chose from a visually presented array of digits those presented in the to-be- attended stream. We present initial behavioural data collected with this paradigm. We discuss preliminary results in the context of the hypothesis that relative high alpha activity in brain areas associated with distractor processing facilitates speech perception in complex noise.

P9: The role of anticipatory and perseveratory coarticulation cues for speech perception in young and old listeners

Rachel Haines[1], Heather Fortnum[1], and Antje Heinrich[2]

[1] Nottingham Hearing Biomedical Research Unit, UK
[2] MRC Institute of Hearing Research, Nottingham, UK.
Antje.Heinrich@ihr.mrc.ac.uk

The partial spectral-temporal overlap of speech segments, called coarticulation, is a fundamental property of speech production, and is present in all naturally produced speech. Coarticolatory cues have been shown to be beneficial to speech perception of normally hearing (NH) listeners in quiet, and potentially in background noise. Two types of coarticulatory cues, anticipatory and persevatory, were examined in young NH listeners, older NH adults, and young hearing-impaired listeners. Anticipatory (ANT) cues occur when a subsequent speech segment affects the acoustic representation of the previous segment. Persevatory (PERS) cues occur when the current segment leaves an acoustic trace in the subsequent segment. Stimuli were C_1VC_2 words starting either with a voiced (/b, d, g/) or unvoiced (/t, k, s, ʃ/) consonant. Stimuli were shortened by deleting the VC_2 (ANT) or the C_1 (PERS) portion. These shortened stimuli were then presented either in quiet or in eight-talker babble. For each stimulus the listener had to decide which of two alternative CVC words, displayed on a screen and matched for either C_1 (ANT) or rhyme (PERS), corresponded to the stimulus. Results indicated a differential effect of voicing on coarticulation cues: ANT cues aided word perception irrespective of the voicing of C_1, but PERS cues were only effective for voiced C_1 segments. There was also a differential effect of listener group on coarticulation: while all three listener groups used ANT cues equally well, young HI listeners were less able to use PERS cues compared with young and old NH listeners. This work was funded by a National Institute for Health Research PhD Studentship.

P10: The effect of retiming speech on masked intelligibility

Martin Cooke[1] and Vincent Aubanel[2]

[1] Ikerbasque & Language and Speech Lab, Universidad del Pais Vasco, Vitoria, ES
[2] The MARCS Institute, University of Western Sydney, AU
m.cooke@ikerbasque.org

Speech can be modified in the spectral domain to improve its intelligibility when mixed with noise, producing gains equivalent to up to 5 dB [e.g. 1; see evaluations in 2,3]. We recently proposed an algorithm – GCRetime [4] – which focuses instead on nonlinear durational changes designed to unmask salient speech information in the face of fluctuating maskers. This approach produces gains of around 4 dB in the presence of a competing talker [3].

The current study addressed three questions. First, do spectral and temporal modifications complement each other? Second, do temporal modifications result in speech which is intrinsically more intelligible when taken out of the context of the fluctuating masker which induced the modifications? Finally, to what extent is any benefit of retiming due to mere elongation (slower speech rate)? The new study also tests our earlier findings using speech material in a different language [5].

In the first experiment native listeners identified keywords in sentences masked by a competing talker. Sentences were presented in unmodified form and following spectral, temporal and spectro-plus-temporal manipulation. Significant gains of more than 20 percentage points were observed for both spectral and temporal modifications, replicating [1,4]. However, gains of around 40 percentage points resulted from the combined modification, demonstrating the complementarity of spectral and temporal modifications.

In experiment 2 sentences were presented either unmodified, retimed as in experiment 1, or simply elongated to match the duration of the retimed sentences. Maskers were competing speech, speech modulated noise and speech shaped noise. Results indicate that around half of the benefit of retiming observed in experiment 1 comes from simple elongation of the stimuli. Further, when presented in stationary noise, retimed speech was significantly less intelligible unmodified speech, suggesting that retiming results in harmful manipulations when taken out of the noise-inducing context. We conclude that retiming helps in fluctuating noise in spite of reduced "intrinsic" intelligibility of the modified speech, suggesting that further gains may be possible if harmful artefacts are mitigated.

Finally, one unanticipated outcome was the contrasting effect of mere elongation in the presence of stationary and fluctuating maskers. As in many earlier studies, we found no beneficial effect of a slower speech rate in speech-shaped noise. However, elongation did lead to gains for both modulated maskers. We speculate that while elongation increases the likelihood that speech information will be glimpsed in masker dips, in the stationary noise case elongation simply results in longer glimpses of what is essentially the same speech information.

References

[1] Zorila, T.-C., Kandia, V., & Stylianou, Y. (2012). Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression. Proc. Interspeech, Portland, USA, 635-638.

[2] Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., & Tang, Y. (2013). Evaluating the intelligibility benefit of speech modifications in known noise conditions. Speech Comm., 55, 572-585.

[3] Cooke, M., Mayo, C., & Valentini-Botinhao, C. (2013). Intelligibility-enhancing speech modifications: the Hurricane Challenge. Interspeech, Lyon, France.

[4] Aubanel, V., & Cooke, M. (2013). Information-preserving temporal reallocation of speech in the presence of fluctuating maskers. Interspeech, Lyon, France.

[5] Aubanel, V., Garcia Lecumberri, M.L., & Cooke, M. (submitted). The Sharvard corpus: A phonemically-balanced Spanish sentence resource for audiology. Int. J. Audiology.

P11: Spatial release from masking for horizontally and vertically distributed sources and the definition of the better ear.

Barry Bardsley and John.F. Culling

Psychology Department, Cardiff University, UK
BardsleyB@cardiff.ac.uk

This study reports spatial release from masking measured over headphones. Following Ter-Horst et al. (1993), two speech-shaped noise interferers were displaced symmetrically in virtual space on either side of a speech source. Separations of 45 and 90°horizontally, and of 20 and 40°vertically were used, in both forward facing and side facing conditions. The results were compared with Ter-Horst et al (1993) and the model devised by Lavandier & Culling (2010). Results of the study compare well with the model predictions but differ substantially from the Ter-Horst et al (1993) data, which was collected in a real sound field. As a follow up study we investigated the relative importance of signal-to-ratio and target-speech level at each ear. Speech was presented from 90°. The head-related impulse responses (HRIRs) for the noise were selected independently for each ear, as though the noise originated at either 90°or 115°at each ear. Thresholds depended only on the HRIR used for the ear contralateral to the speech, indicating that signal-to-noise ratio is more important than speech level; listeners used the ear away from the speech.

P12: Simulation data and a model approach for speech perception with electric-acoustic stimulation (EAS)

Tobias Rader[1], Youssef Adel[1], Uwe Baumann[1], and Hugo Fastl[2]

[1] Audiological Acoustics, ENT Department, University Hospital Frankfurt, DE
[2] AG Technische Akustik, MMK, Technische Universitt Mnchen, Munich, DE
youssefadel@me.com

Electric-acoustic stimulation (EAS) denotes simultaneous stimulation of high-frequency hearing by means of a cochlear implant (CI) and of residual low-frequency hearing up to 500 Hz by acoustic stimulation in the same ear. Patients implanted and fitted according to the EAS concept show significantly higher speech intelligibility in complex noise environments compared to bilaterally implanted CI patients.

To investigate the effect of EAS on speech perception in noise we developed a simulation to mimic electric-acoustic stimulation using recordings of the German Oldenburg sentence test (OLSA) and two types of competing noise: (1) pseudo-continuous OL-noise and (2) amplitude-modulated Fastl-noise. Speech perception scores were obtained using a specialized automatic speech recognition (ASR) model to determine characteristic parameters of the synergic EAS results.

Analogous SRT results were found for normal-hearing subjects and the ASR model, although shifted ca. 8 dB SNR apart. The synergic effect of EAS for different cut-off frequencies of the low-pass filtering (simulating residual hearing) was demonstrated and a significant amount of speech information was found for cut-off frequencies $\leq$ 300 Hz. This result could be considered when examining eligibility criteria for EAS.

References

Rader, T., H. Fastl, and U. Baumann (2013): *Speech Perception with Combined Electric-Acoustic Stimulation and Bilateral Cochlear Implants in a Multisource Noise Field.* Ear and Hearing 34(3): 324-332.

P13: Elicitation and analysis of a robust word misperception corpus in Spanish

Attila Máté Tóth[1], Maria Luisa García Lecumberri[1], and Martin Cooke[2,1]

[1] Language and Speech Laboratory, University of the Basque Country, Vitoria, ES
[2] Basque Foundation for Science, Bilbao, ES
toth.attila.mate@gmail.com

Human listeners make mistakes when listening to speech in adverse conditions. In some rare cases these misperceptions are consistent across listeners. These robust misperceptions are of particular interest for two reasons. First, they suggest common strategies listeners use to process speech in adverse conditions. Second, they serve as valuable test stimuli for microscopic intelligibility models, which attempt to predict individual listener responses to each presented stimulus. Previous speech misperception studies [1,2] have focused on anecdotal reports of isolated occurrences in real life, which are difficult to replicate in the absence of the speech signal and the error-inducing context. Other work has elicited misperceptions via speech modifications in the lab and has generally regarded noise as a confounding factor [3]. The current study collected robust, noise-induced Spanish word misperceptions in a laboratory setting. 173 young adults participated in two 1-hour sessions, providing a total of 308157 responses to 53039 different speech in noise tokens. The confusions were collected using an adaptive procedure which focused effort on trials most likely to lead to robust misperceptions. More than 3000 consistent misperceptions were elicited in this way. We present and analyse the confusions based on an extension of the taxonomy introduced in [4] as well as via a novel masker-dependent scheme based on the amount of information recruited from the background which is incorporated into the misperceived word. The corpus will be released as an open resource to the community.

References

[1] A. Cutler and C. Henton, "There's many a slip 'twixt the cup and the lip," in On Speech and Language: Studies for Sieb G. Nooteboom, H. Quené and V . van Heuven, Eds. Netherlands Graduate School of Linguistic, 2004.

[2] K. Tang and A. Nevins, "Naturalistic speech misperception - a computational corpus-based study," in Proceedings of the 43rd Meeting of the North East Linguistic Society, 2012.

[3] A. Cutler and S. Butterfield, "Rhythmic Cues to Speech Segmentation: Evidence from Juncture Misperception," Journal of Memory and Language, vol. 31, pp. 218-236, 1992.

[4] S. Garnes and Z. S. Bond, "A slip of the ear? a snip of the ear? a slip of the year?" in Errors in Linguistic Performance: Slips of the Tongue, Ear , Pen and Hand, V . A.Fromkin, Ed. New York: New York: Academic Press, 1980.

P14: Boosting speech intelligibility using spectral reweighting under a constant energy constraint

Yan Tang[1] and Martin Cooke[2,1]

[1] Language and Speech Lab, Universidad del Pais Vasco, Vitoria, ES
[2] Ikerbasque (Basque Foundation for Science), ES
y.tang@laslab.org

Speech intelligibility in noise can be increased without changing overall RMS energy by reallocating energy to frequency regions which are least masked by noise or which, with a small amplification, produce masking release. Frequency domain reallocation under a constant energy constraint can be realised by spectral weighting. Different combinations of maskers and speech signals might be expected to require different spectral weights for optimal masking release. The current study examines the performance of weights discovered via an optimisation procedure using an intelligibility metric as an objective function.

Spectral weightings were selected by pattern search [1]. A new objective intelligibility model, extended glimpse proportion: xGP [2], was used as the objective function with the candidate weightings as the variable vector. Weightings for six masker types (white, speech-shaped noise, modulated noise, competing speech, lowpass and highpass noise) each presented at two SNRs were computed. While the spectral profiles varied somewhat for the different maskers, in general they showed a pattern of boosting energy above 1 kHz at the expense of energy at lower frequencies.

To evaluate the approach, two listening tests were conducted with native Spanish speakers using a Spanish corpus [3]. In the first experiment, the original speech and optimal spectrally-weighted speech were presented to listeners in the 12 masker conditions. Keyword scores improved by 8 to 55 percentage points for 5 of the 6 maskers, with the lowest improvement for the high-pass noise and small decreases for the white noise masker. In a second experiment, masker and SNR-independent spectral weightings based on the general form of the optimal weightings from experiment 1 were investigated for a subset of maskers (speech-shaped noise, modulated noise and competing speech). These weightings varied in the number of boosted channels (from 5 to 20) in the frequency range 1-7.5 kHz. Intelligibility gains for 20-channel boosting were similar to those observed in experiment 1, while decreasing the number of boosted channels led to reductions in intelligibility.

The two experiments demonstrate that (i) noise- and level-dependent optimal spectral weighting can lead to very substantial intelligibility gains without increasing RMS energy, especially for maskers whose long-term spectrum falls with frequency; (ii) a noise- and level-independent spectral weighting is nearly as effective as those customised for specific maskers. Given the simplicity with which spectral reweighting can be implemented, these findings points to a practical mechanism for intelligibility enhancement in some common noise conditions. Further work is required to find effective boosting strategies for maskers with a uniform or high-pass characteristic, and to investigate alternative optimisation algorithms and objective intelligibility measures that are more sensitive to frequency regions important for speech comprehension.

References

[1] Hooke, R.; Jeeves, T.A. (1961). ""Direct search" solution of numerical and statistical problems". Journal of the Association for Computing Machinery, 8(2): 212-229.

[2] Tang, Y and Cooke, M. (in preparation). A glimpse-based intelligibility metric for plain, enhanced and synthetic speech in additive noise conditions.

[3] Aubanel, V., Garcia Lecumberri, M. L., & Cooke, M. (in revision). The Sharvard corpus: A phonemically-balanced Spanish sentence resource for audiology. Int. J. Audiology.

P15: Is there more to saliency than loudness?

Francesco Tordini[1,2]

[1] McGill University, Montréal, CA
[2] Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), CA
tord@cim.mcgill.ca

The ability of human listeners to hear out a sound event from a complex auditory scene is well known. Salience can be defined as the property of a sound to jump to the foreground with respect to other sounds or background noise. Here we present a behavioural test battery aiming to capture the perceived salience of natural sounds in a binaural setting using two competitive sound streams.

Our results demonstrate that perceived loudness effects, although prominent, cannot completely explain foreground/background selection, raising a question as to the degree of overlap between the definitions of loudness and salience in real world scenarios. We also discuss the agreement between the recent ITU-R BS.1770x/EBU-tech3343 broadcasting standards and our subjective loudness ratings.

P16: Intelligibility of spoken and sung sentences in different types of background noise

Sarah Knight[1], Sarah Hawkins[2], Edward Wickham[2], Christopher Fox[3], and Antje Heinrich[1]

[1] MRC Institute of Hearing Research, Nottingham, UK
[2] University of Cambridge, Cambridge, UK
[3] Brunel University, London, UK
Sarah.Knight@ihr.mrc.ac.uk

Speech perception in noise (SPiN) has been widely investigated, but such investigations have not been extended to the perception of singing, despite the frequency with which sung words are encountered in daily life. We therefore know little about whether or not the varied findings of SPiN research can be applied to the perception of sung texts. The current study sought to address this issue.

There are various ways in which sung and spoken speech differ. First, it is reasonable to assume that singing is regarded by listeners as aesthetic and musical rather than information-bearing. Second, even if sung speech is listened to for information-gathering purposes, it differs from spoken speech acoustically because the speech rhythm in sung speech is usually subjugated to musical rhythm, thus disrupting higher-level temporal information; and the pitching of vowels distorts the spectro-temporal information carried in formants. To better understand how sung speech relates to spoken speech we devised a task that was modelled after traditional SPiN tasks and used sung sentences of high and low semantic predictability that were presented in a variety of background noises. Based on speech research which shows that high semantic predictability is particularly useful when intelligibility is difficult, we might expect semantic context to be particularly important for sung speech; on the other hand, if listeners do not regard singing as information-bearing, then the semantic predictability of the sentence may not be important.

Data were collected outside of the usual lab environment, during six live concert performances by a professional British choir (The Clerks). The stimuli were 36 high- and low-predictability context sentence pairs. They were sung by a single male singer against three types of background noise (/sh/; shifting vowel sounds; spoken babble) at two levels (high; low), performed live by the other choir members. The design was fully crossed across the six concerts. The task was 4AFC: a large screen displayed four options, which included the correct response as well as one phonetically similar, one semantically plausible and one moderately phonetically similar and semantically plausible alternative. For each trial, audience members chose the most likely option with regards to the final word using a handheld device.

Preliminary results show significant main effects of noise level, semantic predictability and noise type. Participants were most accurate in low noise levels and when predictability was high; they performed best in /sh/ noise and worst in babble, with the shifting vowel background in between. There was a significant interaction between noise type and noise level showing that there was a significant effect of level in vowel and babble backgrounds but not in the /sh/ condition. There was also a significant interaction between noise type and predictability showing that, although a significant benefit of predictability was apparent for all noise types, the effect varied in size. The results suggest that despite clear acoustic differences between sung and spoken speech and despite the less-well controlled experimental setting, various findings from traditional SPiN research may be applicable to sung words.

P17: Effects of NAL-R on Consonant Vowel Perception in Hearing-Impaired Subjects

Christoph Scheidiger[1], Jont B. Allen[2], and Torsten Dau[1]

[1] Technical University of Denmark, DK
[2] University of Illinois at Urbana-Champaign, USA
csche@elektro.dtu.dk

Psychoacoustic experiments in different conditions were undertaken. In particular, two *consonant vowel (CV)* identification experiments in masking noise were conducted at various *signal-to-noise ratios (SNRs)* with 16 HI ears. In one of the experiments, the CVs were presented with a uniform gain; in the other experiment, a spectral compensation (i.e. NAL-R) for the individual hearing loss was provided. In both gain conditions, the subjects were instructed to adjust the presentation level to their *most comfortable loudness (MCL)*, which is contrary to the common approach of adjusting the presentation level depending on the *pure tone thresholds (PTTs)* and the *long-term average speech spectrum (LTASS)* ([Zurek and Delhorne(1987)], [Posner and Ventry(1977)]). The data demonstrated that the MCL approach led to consistent responses in all subjects. Based on these results, a more rigorous definition of audibility based on entropy and the [Miller and Nicely(1955)] confusion groups is proposed. Furthermore, the effectiveness of NALR for CV perception was investigated by comparing the confusion matrices of the two experiments. In general, the error and entropy decreased with NALR. The average error decreased from 20.1% ($\sigma = 3.7$) to 16.3% ($\sigma = 2.8$). It was also shown that, with the help of NAL-R, the tested ears became more consistent in their responses for a given token. However, for 15.1% of the *token-ear pairs (TEPs)*, the entropy and error increased with NAL-R.

References

[Miller and Nicely(1955)] George A Miller and Patricia E Nicely. An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, 27(2):338352, 1955.

[Posner and Ventry(1977)] Judd Posner and Ira M Ventry. Relationships between comfortable loudness levels for speech and speech discrimination in sensorineural hearing loss. *Journal of Speech and Hearing Disorders*, 42(3): 370, 1977.

[Zurek and Delhorne(1987)] PM Zurek and LA Delhorne. Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. *The Journal of the Acoustical Society of America*, 82:1548, 1987.

P18: Developing speech recognition materials suitable for non-native speakers

Louise Stringer

Department of Speech, Hearing and Phonetic Sciences, UCL, London, UK
l.stringer.11@ucl.ac.uk

Research into many aspects of speech perception by non-native speakers is growing. However, many speech recognition materials available may not be suitable for this group as they contain uncommon lexical items and rather complex syntactical structures. Other materials available may comprise only a small number of sentences, or a single set of uncontrolled sentences, which limits their usefulness. To address this gap in suitable materials, a novel set of 439 sentence triplets was developed using vocabulary and grammar suitable for intermediate level non-native English speakers. However, sentences are not too simplistic, so are also suitable for native speakers. Each sentence triplet contains related sentences in three different conditions. Type A sentences have high contextual constraint and a highly predictable final keyword (e.g.: The dolphins are swimming in the sea), Type B sentences are weakly constrained with a neutral keyword (e.g.: The children are playing in the sea), and Type C sentences have a high contextual constraint but an anomalous final keyword (The dolphins are swimming in the road). Using three conditions in this way makes the materials more versatile than a single, uncontrolled set of sentences. A greater number of sentences allows more experimental conditions (e.g.: multiple noise masker types), and multiple contextual conditions mean the materials may be used in task types that an uncontrolled single set may not be suitable for. This poster presents the development of this novel material set. Stages described include the selection of vocabulary and grammar used, matching sentences across conditions and assessing the probability of final keywords.

P19: AAL: Intelligible City for All. Evaluation of ecological intelligibility test for normal hearing and hearing impaired adults in railway station context

Yosra Mzah[1], Neska El Haouij[1], Nader Mechergui[1], Amira Ben Jemaa[1], Meriem Jaïdane[1,2,3], Sonia Larbi[1,2,3], Sylvie Ghalila[1], Dhafer Malouche[4], Kevin Lecoz[6], Yannick Belouard[6], Cendrine Mercier[7,5], Jawad Hajjam[5], and Alessendra Brescia[8]

[1] CEA-LinkLab, plateforme commune de recherche et de développement entre CEA et Telnet, TN
[2] Ecole Nationale d'Ingénieur de Tunis (ENIT), TN
[3] Unité de recherche Signaux et Systèmes (U2S), ENIT, TN
[4] Ecole Sup´erieure de la Statistique et de l'Analyse de l'Information (ESSAI), TN
[5] Centre d Expertise National des Technologies de lInformation et de la Communication pour lautonomie (Centich), FR
[6] Centre d'Evaluation et Réeadaptation des troubles de l'Audition (CERTA), FR
[7] University of Nantes, Laboratory CREN, FR
[8] ESCOOP European Social Cooperative - Cooperativa Sociale Europea - sce, IT
mzah.yosra@gmail.com

The project I'CityForAll aims at enhancing the sense of safety and self-confidence of presbyacousic persons whose hearing impairment increases with age. For elderly, this impacts on the intelligibility of vocal messages and their perception of the distance-direction of alarm sounds and of their alarming power.

Two mobility situations are considered: in public confined spaces and in urban spaces. For public confined spaces, the ICT solutions consist in smart loudspeakers for better intelligibility of vocal announcements. For urban mobility, I'CityForAll partners are developing an embedded system in vehicles for a better perception and localization of alarms.

For the intelligibility assessment of public address systems in confined spaces, an in-lab ecologically-oriented protocol was developed, which is derived from the standard Hearing In Noise Test (HINT) used by audiologists. The modifications we introduced to the HINT test concern two aspects. The first one consists in taking into account the reverberation in intelligibility evaluation, coupled to background noise. The second idea is to place the test-subjects in the realistic context of a railway station by using railway station noise and vocal announces as test conditions.

The proposed eco-test has been carried out at Centich[9] with a population of 23 subjects including normal hearing and hearing impaired persons. The statistical analysis of the results shows that:

- We obtain different intelligibility score distributions for different test-conditions but having the same equivalent SNR[10]: test condition with noise and reverberation and test-condition with noise only. Hence, the effect of reverberation on intelligibility is different from that of additive noise

- When the test-subjects have been told they are supposed to be in a railway station at the beginning of the ecological test (reverberation + railway station noise + vocal announce), the distribution of the intelligibility scores are different from those obtained during the standard tests under the same test conditions but with HINT sentences.

[9] Centich: Centre d'Expertise National des Technologies de l'Information et de la Communication pour l'autonomie, France
[10] SNR: Equivalent SNR is a measure that takes into account reverberation and noise

P20: Rate discrimination at low pulse rates: Influence of intracochlear stimulation site between normal-hearing and cochlear implant listeners.

Pierre Stahl[1], Olivier Macherey[1], Sabine Meunier[1], and Stéphane Roman[2]

[1] Laboratoire de Mécanique et dAcoustique, CNRS, Marseille, FR
[2] Dept. of Pediatric Otolaryngology and Neck Surgery, AMU, Marseille, FR
stahl@lma.cnrs-mrs.fr

Pitch appears particularly useful for speech comprehension in complex environment. One possible reason for the poor speech comprehension of cochlear implant (CI) users in the presence of background noise is that their temporal pitch perception remains weaker than in normal hearing (NH). For CI subjects, temporal pitch perception is usually limited to rates below about 300 pulses per second (pps). The present study aims to better understand the limitations of the temporal pitch mechanism in CI users by focusing on rate discrimination at *low* pulse rates.

The focus on low rates was motivated by data obtained in NH listeners showing that frequency difference limens (DLs) for bandpass filtered harmonic complexes were lowest for complex tones presented at the apex (Krumbholz et al., 2000). The fact this was true even when the lowest harmonics were unresolved (rank >12) suggests that temporal cues are more salient when originating from the apex. However, because their stimuli had a constant bandwidth on a Hz scale, it remained unclear whether the better performance observed in low frequency regions was due to the greater portion of neurons being excited or simply to the fact that it was lower.

We present two experiments that measure and compare frequency DLs at four fundamental frequencies (F0s ranging from 20 to 104 Hz) in NH and in CI listeners. First, we extend the results of Krumbholz et al. by comparing DLs for harmonic complexes filtered in two frequency regions (lower cut-offs of 1200 and 3600 Hz) for three different bandwidths (ranging from 250 to 6750 Hz). Second, we measure rate DLs for users of the Med-EL device at two different intracochlear sites. The two sites were chosen to be one of the three most apical and one of the three most basal channels of the device. DLs were measured for 500-ms stimuli in a 3I2AFC task.

For NH listeners, frequency DLs significantly improved as F0 increased from 20 to 104 Hz. Performance was better in the low than in the high frequency region, even when the stimuli had the same bandwidth expressed in percentage of the filters lower cut-off frequency. Although data collection with CI subjects is ongoing, preliminary results also suggest better performance for the apical than for the basal channel.

Acknowledgments

Work funded by the ANR (Project DAIMA)

References

Krumbholz et al. (2000) JASA, 108(3): 1170-80

P21: Factors limiting vocal-tract-length perception in cochlear-implants

Etienne Gaudrain and Deniz Başkent

University Medical Center Groningen / University of Groningen, NL
e.p.c.gaudrain.cam@umcg.nl

Recent studies have highlighted the difficulties that cochlear-implant (CI) users have with perceiving vocal-tract length (VTL) variations across speakers. VTL perception is not only a necessary component of voice gender identification, it is also crucial for the segregation and understanding of concurrent talkers. In this study we aim to identify which aspects of CI stimulation are responsible for poor VTL discrimination performances. We measured just-noticeable-differences (JNDs) for VTL and fundamental frequency (F0) using an adaptive procedure, in normal-hearing participants listening to various vocoders simulating different aspects of electrical stimulation. The vocoders differed in terms of number of channels, carriers and spread of excitation. These measures are also to be compared to similar measures in actual CI users. The JND for F0 increases, i.e. worsens, when the spectral resolution is degraded, but quickly reaches a plateau. In contrast, the JND for VTL steadily increases when spectral resolution decreases. The type of carrier does not affect VTL discrimination. Finally, the electrical spread of excitation seems to be the strongest factor limiting VTL perception in CIs.

P22: Perceptual restoration of degraded speech is preserved with advancing age

Jefta D. Saija[1,2,3,4], Elkan G. Akyürek[2,4], Tjeerd C. Andringa[3,4], and Deniz Başkent[1,4]

[1] Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, NL
[2] Department of Psychology, Experimental Psychology, University of Groningen, NL
[3] Faculty of Mathematics and Natural Sciences, Artificial Intelligence and Cognitive Engineering (ALICE), University of Groningen, NL
[4] Research School of Behavioral and Cognitive Neurosciences, University of Groningen, NL
j.d.saija@umcg.nl

We investigated the age effect on top-down repair of degraded speech using phonemic restoration of interrupted speech. Cognitive skills, like processing speed, memory functioning, and selective attention, can diminish due to age-related changes in the nervous system. Despite this, we show that older adults can successfully compensate for degradations in speech perception. Critically, the older participants of this study were not pre-selected for high performance on cognitive tasks, but only screened for normal-hearing. Perception of meaningful sentences interrupted with silent intervals was measured with and without noise filling the gaps, with young (<26 years, ± 22 years) and older (>62 years, ± 66 years) normal-hearing listeners. The phonemic restoration effect, observed by an increase in intelligibility after adding filler noise, provided a measure of top-down repair of interrupted speech. Additionally, the speech was slowed and speeded by a factor of two to observe effects of cognitive processing speed on restoration.

In this study, older participants showed poorer intelligibility of degraded speech than younger participants, as expected from previous reports of aging effects. However, in conditions that induced top-down restoration, robust compensation in speech perception similar to the younger group was observed. Slowed-down speech gave more time for cognitive processing, resulting in even stronger compensation. Speeded speech was detrimental for intelligibility, especially for the older participants. Even though older adults understand interrupted speech worse than young adults, they can still use top-down processes to extract cues and restore meaningful sentences when noise fills the gaps. Likely, older adults rely on their life-long language experience and rich accrued vocabulary, to effectively restore interrupted speech.

P23: Pulse-spreading harmonic complexes: an advantageous carrier for simulating cochlear implants

Olivier Macherey, Quentin Mesnildrey, and Gaston Hilkhuysen

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

The pulse spreading harmonic complex (PSHC) is a pulsatile broadband signal. We earlier claimed that PSHCs can be optimized such that they show little intrinsic modulations after cochlear filtering. This claim is based on the observation that modulation detection thresholds (MDTs) are lower for broadband PSHCs than for broadband pseudo-random noise.

However, the regular pulses of a PSHC combined with nonlinear cochlear filtering could create audible distortion products. Consequently, the low MDTs for PSHCs might be the result of a detection of modulations on the distortion products, instead of representing little intrinsic modulations. We here consider this alternative explanation.

In a control experiment we found evidence for PSHC generating distortion products, particularly at high presentation levels. But masking these distortion products held little consequences for MDTs, a finding undermining the alternative explanation and favouring the idea that optimized PSHCs have little intrinsic modulations after cochlear filtering. A preliminary intelligibility study using vocoders with sine, noise or PSHC carriers also supports our initial interpretation. Consequently, optimised PSHCs appear interesting carriers for speech vocoders, typically used to simulate cochlear implants in normal hearing listeners.

Acknowledgments
Work funded by the ANR (Project DAIMA).

P24: Optimizing a noise reduction system with physical intelligibility metrics

Gaston Hilkhuysen and Mark Huckvale

University College London, London, UK
hilkhuysen@lma.cnrs-mrs.fr

Most noise reduction systems have various adjustable parameters. Their settings typically hold consequences for speech intelligibility. We previously found that expert listeners were unable to provide parameter values that optimize intelligibility. In this follow-up study we evaluate whether intelligibility metrics, based on the physical properties of the signals involved, are able to identify optimal settings.

Various metrics were evaluated in two noise types combined with sixteen different parameter settings of a commercial noise reduction system. Most metrics predicted that noise reduction would improve intelligibility. Such improvements could not be observed in a listening study. One metric correctly suggested the settings that held no effects on intelligibility.

P25: Consonant confusions in frozen and random white noise

Johannes Zaar, Søren Jørgensen, and Torsten Dau

Centre for Applied Hearing Research, Technical University of Denmark
jzaar@elektro.dtu.dk

This study investigates the effect of deterministic (*"frozen"*) white noise maskers on the recognition and confusion of consonants by normal-hearing listeners. Consonant-vowel combinations (CVs) consisting of 15 Danish consonants combined with the vowel /i/ were used as speech stimuli. Only one recording of each CV was used, all spoken by the same male talker. The CVs were presented in quiet and in white noise at 6 signal-to-noise ratios (SNRs). For each SNR, three noise conditions were defined: *CV & frozen noise "A", CV & frozen noise "B", CV & random noise* (i.e., noise newly generated for every presentation). 8 normal-hearing native Danish listeners participated in the experiment. The results show that (i) the influence of the masking noise on the perception of a given speech token varies considerably across the frozen noise tokens, (ii) the within-listener consistency is larger for frozen noise than for random noise, and (iii) more systematic confusions occur for frozen noise than for random noise, particularly at low SNRs. These observations indicate that different noise tokens with uniform statistical properties can have different effects on the perception of the same speech token. Studies investigating consonant confusions in noise typically use random noise maskers that are merely defined in terms of their overall statistical and spectral characteristics. The results discussed here suggest that the influence of the noise on a token-by-token basis might be underestimated and in fact account for the typically observed variability across listeners.

P26: The effect of expertise in speech-in-speech and speech-in-noise tasks

Clara Suied, Jean-Christophe Bouy, and Guillaume Andéol

Institut de recherche biomédicale des armées, département Action et cognition en situation opérationnelle, Brétigny-sur-Orge, FR
clara.suied@irba.fr

Air Force operators are often faced with complex acoustic environments, including loud background sound and multiple radio communications. These operators self-report large inter-individual differences in their ability to stream simultaneous radio communications, apparently due to their level of expertise (i.e. years of practice). We used a psychophysical task to attempt to quantify the effect of expertise in laboratory-controlled conditions. Speech-in-noise and speech-in-speech intelligibility tasks were performed using the Corpus Response Measurement (CRM) paradigm. The influence of signal-to-noise (or target-to-masker) ratio, as well as the type of mask (same-talker, same-sex, different-sex target and masker voices) were investigated. Moreover, to explore possible neural bases for behavioural performance, we explored the effect of the auditory efferent system; here, the medial olivocochlear (MOC) system, through the recording of oto-acoustic emissions both in the presence and in the absence of broadband noise. This measure was chosen because: the MOC system is thought to facilitate discrimination of speech in noise; differences in the MOC activity have been observed due to (musical) expertise. We compared three groups of participants: experts air force operators, beginner air force operators, and naïve control participants. No differences were found between experts and beginners in the CRM tasks; however, very large differences (up to 50%) in the recognition performances were observed between the naïve control and the expert+beginners groups. With the current dataset, no correlation was found between the behavioral performance and the oto-acoustic emissions measurement. Although we are still collecting data for additional participants, a first important result can be highlighted: air force operators all seem to be "cocktail-party experts". The lack of a difference between the expert and beginner groups may be interpreted in the light of the nature of their real activity, which is much more demanding than a simple intelligibility task. As a next step, measures of cognitive efforts will be considered through the use of dual-task paradigms.

Author Index

Adel, Youssef 13
Akyürek , Elkan G. 23
Allen, Jont B. 18
Andéol, Guillaume 27
Andringa, Tjeerd C. 23
Aubanel, Vincent 11
Aubreville, Marc 8

Başkent, Deniz 6, 22, 23
Bardsley, Barry 12
Baumann, Uwe 7, 13
Becker, Robert 5
Belouard, Yannick 20
Ben Jemaa, Amira 20
Bouton, Sophie 11
Bouy, Jean-Christophe 27
Brescia, Alessendra 20
Brons, Inge 7

Carroll, Rebecca 6
Cooke, Martin 11, 14, 15
Culling, John 3, 12

Dau, Torsten 18, 26
Deroche, Mickael 2
Dreschler, Wouter 7

El Haouij, Neska 20

Fastl, Hugo 13
Fortnum, Heather 10
Fox, Christopher 17

Gaudrain, Etienne 22
Ghalila, Sylvie 20
Giraud, Anne-Lise 11
Grahlmann, Hannah-Lina 7
Grange, Jacques 3
Green, Tim 5

Haines, Rachel 10
Hajjam, Jawad 20
Haverkate, Maj 7
Hawkins, Sarah 17
Heinrich, Antje 10, 17
Hervais-Adelman, Alexis 5
Hilkhuysen, Gaston 24, 25
Hoen, Michel 10
Houben, Rolph 7
Huckvale, Mark 25

Jørgensen, Søren 26
Jaïdane, Meriem 20

Knight, Sarah 17
Knoblauch, Kenneth 10
Koning, Raphael 14

Larbi, Sonia 20
Lavandier, Mathieu 2
Leclère, Thibaud 2
Lecoz, Kevin 20
Lecumberri, Maria Luisa García 14
Louise Stringer, Louise 19
Lunner, Thomas 12

Macherey, Olivier 2, 21, 24
Malouche, Dhafer 20
Mechergui, Nader 20
Mercier, Cendrine 20
Mesnildrey, Quentin 2, 24
Meunier, Fanny 10
Meunier, Sabine 21
Michel, Christoph 5
Moore, Brian 13
Mzah, Yosra 20

Ng, Elaine 12

Obleser, Jonas 9

Pape, Sebastian 8
Pefkou, Maria 5
Priddy, Sasha 3

Rönnberg, Jerker 12
Rader, Tobias 7, 13
Roman, Stéphane 21
Rosen, Stuart 4, 5
Rudner, Mary 12
Ruigendijk, Esther 6

Saija, Jefta D. 23
Sauert, Bastian 9
Scheidiger, Christoph 18
Serman, Maja 8
Serniclaes, Willy 10
Smith, Rachel 15
Stahl, Pierre 21
Steinmetzger, Kurt 4
Strauß, Antje 9
Suied, Clara 27

Tóth, Attila Máté 14
Tang, Yan 15
Tordini, Francesco 16
Trollé, Gwendoline 10

VanRullen, Rufin 8
Varnet, Léo 10

Wöstmann, Malte 9
Weißgerber, Tobias 7
Wickham, Edward 17
Wouters, Jan 14

Zaar, Johannes 26
Zoefel, Benedikt 8

Time table 6^{th} SpiN Workshop, January 2014

	Thursday 9th	Friday 10th
08:00 - 09:00	Welcome Coffee	
09:00 - 10:00	Introduction Quentin Mesnildrey	Poster Session + Coffee
10:00 - 11:00	Jacques Grange Coffee Break	
11:00 - 12:00	Tim Green Deniz Başkent	Sophie Bouton Thomas Lunner
12:00 - 13:00	Lunch	Lunch
13:00 - 14:00	Inge Brons	Keynote Lecture Brian Moore
14:00 - 15:00	Maja Serman Tea Break	Tea Break
15:00 - 16:00	Raphael Koning	Bastian Sauert Rachel Smith
16:00 - 17:00	Léo Varnet Poster Session + Drinks	Closing
17:00 - 18:00		
18:00 - 19:00		