Miller, C. A., Brown, C. J., Abbas, P. J., and Chi, S. L. (**2008**), "The clinical application of potentials evoked from the peripheral auditory system" Hearing Research **242** (1-2), 184-97

Picton, T. W., John, M. S., Dimitrijevic, A., and Purcell, D. W. (**2003**). "Human auditory steady state responses" International Journal of Audiology **42** (2), 177-219

Picton, T.W. (**2011**). *Human auditory evoked potentials*, Plural Publishing, San Diego

Poelmans, H., Luts, H., Vandermosten, M., Boets, B., Ghesquière, P., and Wouters, J. (**2011**). "Auditory steady-state cortical responses indicate deviant phonemic-rate processing in adults with dyslexia" Ear and Hearing, in press

Neurocortical mechanisms of comprehension in degraded speech

JONAS OBLESER[1], ANTJE STRAUß[1] AND ANNA WILSCH[1]

[1] *Max Planck Research Group "Auditory Cognition", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*

Comprehending speech is an astonishing faculty of the human brain, especially so under adverse listening conditions. How and by which neural mechanisms do we cope so well with the fleeting percepts of speech? In addition to "facilitating" influences such as semantic context, listeners also cope with challenging listening situation by fully exploiting their sensory and cognitive resources, e.g. their working memory capacities ("compensation"). We will propose a framework for studying neural signatures of facilitation and compensation, and discuss recent data from functional MRI (fMRI), magneto- and electroencephalography studies (M/EEG; with an emphasis on neural oscillations) that utilize acoustically degraded speech stimuli to study the neural underpinnings of these facilitation and compensation mechanisms in detail.

INTRODUCTION

The so-called "bottom-up" flow of auditory information from the sensory periphery to the central auditory areas in the superior temporal plane (and further on in the cerebral cortex) is not as well understood as comparable processes in the visual domain. Nevertheless, great progress has been made over the last decades. For recent reviews (from a neuroscience perspective) of suggested mechanisms and pathways see e.g. Hackett (2008), Rauschecker and Scott (2009), and Schreiner and Winer (2007).

Our starting point in this paper is the well accepted and intuitively clear assumption that these so-called bottom-up or afferent processes in audition alone cannot explain the astonishing human ability to cope with substantial degradation of the auditory input and still achieve comprehension – this is arguably most relevant in comprehending speech. In principle, however, it applies to any sound of potential behavioural relevance (e.g., we are also able to recognise known music pieces or known voices of music in adverse listening situations).

Two notes of caution: First, for simplicity, the term "top-down" shall be assigned to all these processes that help facilitate or compensate in the process of comprehension in adverse listening. This does not imply that all such top-down processes are assumed to feed back to basic auditory processing levels, as top-down in the strict sense would indicate. In contrast, it has been convincingly argued before that this is not very likely to be the case, and that top-down adjustments to the

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

bottom-up processing of (degraded) sound are likely to occur only over iterative trials or listening situations in a predictive or weight-adjusting manner (e.g., Norris and McQueen, 2008, Davis and Johnsrude, 2007; Obleser and Eisner, 2009). Second, this text will use the terms "degradation", "degraded hearing", and "adverse listening" interchangeably. All data presented were acquired in normal-hearing young listeners and "degradation" is operationalised as various levels of noise-band vocoding (Shannon *et al*., 1995). Noise-vocoding is an effective technique to manipulate the spectral or fine-temporal detail while preserving the amplitude envelope of the speech signal. This renders the signal more or less intelligible in a graded and controlled way, depending on the number of bands used; more bands yield a more intelligible speech signal (cf. Fig. 2). However, we tentatively claim a relevance of the suggested mechanisms for various kinds of acoustic adversity and auditory challenges, ranging from noisy parties and phone lines over mild age-related hearing loss to cochlear implants. Fig. 1 sums up our suggestion of how the wide field of comprehension in adverse listening can be systematised and studied.

Fig. 1: Suggestion for a systematic approach in studying the cognitive neuroscience of (speech) comprehension.

FACILITATION IN DEGRADED SPEECH

Building up expectations from context is arguably a key capability to process speech with the enormous speed as listeners manage to do in everyday life situations. This has been shown in various psycholinguistic experiments. More than half a century ago, Miller *et al.* (1951) found that recognition is faster when words are presented in syntactical contexts rather than in isolation; this is not only because of structural embedding but to a large extent because of the semantic context provided by a sentence (e.g., Kutas and Hillyard, 1980). Here, we would like to outline this facilitating process provided by such semantic context because it is the basis for building up expectations and making predictions from an incoming speech signal.

The ability to integrate words into a context relies on an acquired mental lexicon where different kinds of relationships between words are stored (Levelt, 1989). Frequent co-occurrences of words lead to a probabilistic connection of them in the mental lexicon. Therefore, context can be described in terms of the psycholinguistic framework of regularities and deduction of predictions: A keyword, even when heard degraded and with some uncertainty, opens up "slots" which have to be filled in, and the expectation of that completion is dependent on the probability of a word's typical co-occurrences.

Especially when it comes to degraded speech signals and limited intelligibility, the benefit of predictability becomes apparent. Kalikow *et al.* (1977) developed a battery of 400 sentences, half of them of high and half of them of low predictability, and presented them in a noisy (multi-speaker babble) background. They showed a better comprehension for highly anticipated sentence endings as opposed to less expected completions. More recent studies like Stickney and Assmann (2001) found the same effect using the same sentences but acoustically varied by means of a band-pass filter around different centre frequencies.

To sum up, we use our mental lexicon to build up context and deduce predictions, which will facilitate comprehension – most importantly so under adverse listening conditions.

The acoustic–cognitive interface: Benefits from predictability

As argued above, context is beneficial in developing predictions, and it has been already shown that these are helpful in different kind of noisy environments. One way to simulate the hearing impression of cochlear implant users that is commonly used is noise-band vocoding. Substantial amounts of clinical and neuroimaging data have been gathered using vocoded speech. Shannon *et al.* (2004) have argued in a meta-analysis that performance is mainly dependent on number of bands available and only to a lesser degree on different kinds of stimuli (e.g. words in isolation or in a sentence, melodies, first or second language).

We took into account previous work on predictability and examined possible interactions between higher linguistic–cognitive functions (i.e., profiting from high predictability in a given context) and acoustic degradation. The Kalikow *et al.* (1977) sentences were used to manipulate sentence predictability, and noise vocoding was applied to create different levels of intelligibility with more numbers of bands being more comprehensible.

In the first experiment, subjects listened to 2-, 4-, 8-, 16-band and clear speech. It revealed that at 8-band speech there is a significant difference between semantically high and low probable sentences in accuracy. Subjects performed at ceiling level when context was constraining and only around 50% without any semantic cues. The second experiment was needed to determine whether this finding was specific to 8-band vocoding or not, and could confirm that this intermediate level of degradation seemed to provide the critical amount of spectral information needed to establish top-down processing. Lastly, the third experiment was conducted in order to rule out the presence or absence of clear speech in the experiment on overall comprehension performance (it could have been that switching between degraded and clear speech might have exerted a distracting influence; this was not the case).

Collapsing data from all three behavioural experiments, the conclusion is straight-forward: The greatest comprehension benefit for high predictable sentences is given at an intermediate level of degradation (Fig. 2). The best fit for mean percentage correct as a function of the logarithmic number of bands is a sigmoid curve with steeper slope for sentences with stronger compared to weaker semantic constraints.

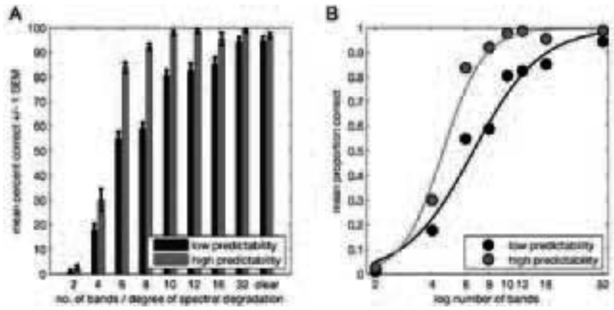

Fig. 2: Aggregated results using data from three behavioral experiments (N=54). A. Mean percentage correct pooled across all three experiments (based on contributions from N=54 subjects; ±1 s.e.m.) for the different amounts of spectral detail, separately for high-predictability (grey bars) and low-predictability sentences (black bars). B. Logistic functions, fitted separately for high-predictability (grey) and low-predictability (black) responses based on the same data (i.e., all noise-vocoding data available from all three experiments).

Functional neuroanatomy of facilitated comprehension

The underlying neural mechanisms of this behaviourally very clear-cut interaction of signal degradation and semantic context are not understood, although the functional neuroanatomy of basic auditory processes, including simple intelligibility variations, has been studied extensively in the last decade (e.g., Binder *et al.*, 1996; Scott *et al.*, 2000; Davis and Johnsrude, 2003).

We addressed the question of how speech comprehension in degradation might be facilitated by contextual cues in a functional magnetic resonance imaging (fMRI) study (Obleser *et al.*, 2007). Again, we were using the Kalikow *et al.* (1977) sentences in the critical 8-band vocoded format, flanked by 2-band and 32-band speech. Activity in both hemispheres in superior temporal sulci and the left inferior frontal gyrus correlated with the amount of spectral detail in the speech signal.

More surprisingly, also higher-order cortical areas remote from auditory cortex proper were found which were associated with high-predictable more than low-predictable speech comprehension under moderately degraded (8-band) signal conditions. Especially the left angular gyrus appeared to be involved in facilitated comprehension. Remarkably, the "functional connectivity" between angular gyrus and a range of other heteromodal (i.e., not strictly auditory) areas in the left hemisphere (prefrontal cortex; inferior frontal cortex; posterior cingulate cortex) was enhanced only for predictable sentences (Fig. 3).

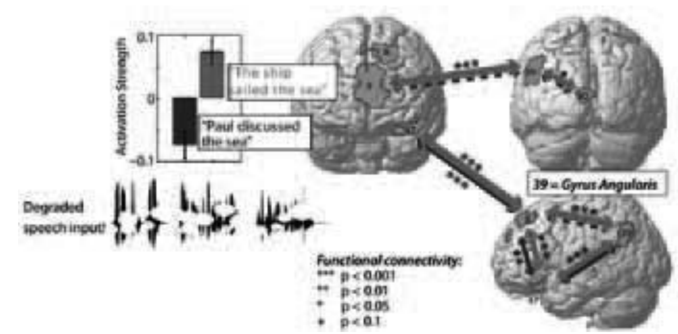

Fig. 3: Overview over the changes in functional connectivity between brain regions for the high-predictability 8-band condition (grey) and the low-predictability 8-band condition (black). Numbers in clusters indicate Brodmann's areas. All arrows shown indicate an increase in the positive correlation for the high-predictability condition, whereas no brain areas are more strongly interlinked for low predictability. $+p < 0.10$; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$. The dashed black lines indicate non-significant correlations.

Electroencephalography (EEG): Time scales of facilitation

Presuming the behavioural and functional neuroimaging data described above, we turned back to the aforementioned concepts of semantic context and their application in EEG studies. Applying this to degraded speech, we hypothesised that the acoustic–semantic integration should happen comparably early and should therefore be reflected in the amplitude of the N400 component (i.e., approximately 400 ms after the onset of a sentence-final keyword, first described by Kutas and Hillyard (1980)). Aydelott *et al.* (2006) had reported a dependency of the N400 on acoustic degradation by varying predictability of sentences in low-pass filtered, disconnected words. Opposed to that, we used naturally connected speech. More importantly, though, we manipulated expectancy not only by the preceding semantic context (e.g., "builds…house" vs. "sees…house") but also by presenting two more or less prototypical sentence endings from the semantic frame (e.g., "house" vs. "mansion"; Strauß, Kotz, & Obleser, in prep.). Again, vocoding was applied; after pre-testing 4-band, 8-band, and clear speech were chosen to supply hardly intelligible, intermediate, and fully intelligible signal quality, respectively. Results show that in clear speech, comprehension is facilitated by high context independent of prototypicality (i.e., the N400 is only sensitive to the semantic frame but not to exact sentence ending). Hardly intelligible signals seem to prevent early facilitation mechanisms altogether. At an intermediate level of signal degradation (8-bands), however, only prototypical completions are facilitating comprehension (i.e., only prototypical sentence endings yield a reduced N400). We take this as indication that top-down facilitation processes do take place in moderately degraded speech, but they are narrowed down in sensitivity to track only the most likely sentence ending.

These data demonstrate how a seemingly high-level semantic process such as the integration effort reflected in the N400 is affected by varying levels of degradation.

In an earlier EEG study, Obleser and Kotz (2011) sticked to a twofold predictability manipulation and varied degradation by using 1-band, 4-band, and 16-band vocoding. The N400 showed the expected sensitivity to signal degradation (see above). Additionally, we also expected facilitation of integration (an enhancement for high-expectancy words) being reflected in a modulation of Gamma-band (γ, > 30 Hz) oscillations. Many studies on the top-down formation of percepts have focussed on this most prominent electrophysiological signature of "Gestalt" formation in the visual domain (for review see Fries, 2009), but also in the auditory, and the speech domain (Hannemann *et al.*, 2007). Enhanced Gamma-band synchrony can surface as power enhancements (compared to selected baseline periods) in the Gamma-band range when analysing the EEG data for non-phase-locked oscillations. Here, we observed 60-80 Hz gamma oscillations around 600 ms after the onset of the sentence's final keyword. They thus followed the N400 and were enhanced for highly expected sentence endings.

Intriguingly, this study also yielded a correlation of the late Gamma effect with the earlier neural response to the sentence onset (N100) at a fairly intelligible signal quality (16-band speech). We have good reasons to associate the N100 with the allocation of compensatory processes, which brings us closer to analysing the relationship of facilitatory and compensatory processes in comprehending degraded speech (see Fig. 5). In the following section, we further elaborate on this relationship.

COMPENSATION IN DEGRADED SPEECH

As the previous section showed, accumulating information from the perceived speech signal can facilitate comprehension of degraded speech. Compensatory top-down mechanisms provided by cognitive functions, somewhat independent of the speech signal properties, support the comprehension process as well. These cognitive functions are mainly working memory and attention – two essential (and non-independent) top-down ingredients in speech comprehension. Thus, we propose that compensation can also be understood as cognitive effort, meaning that the comprehension of degraded speech taxes our cognitive resources.

The information degradation hypothesis (Pichora-Fuller and Singh, 2006) takes the important role of cognitive decline in older people into account. It points out the general importance of cognitive functions during the processing of speech. Furthermore, it emphasises the interaction of sensory and cognitive decline: Older people with peripheral hearing loss are challenged on the sensory level and might lack top-down compensatory resources on a cognitive level.

The interplay of working memory and attention works efficiently for comprehending clear speech. The listener attends to the speech signal, the signal is transferred and stored into working memory, and comprehension is achieved as the speech signal progresses. This is accomplished with hardly any effort in normal hearing people. Both functions (attention and working memory) are subsumed under the construct of

"cognitive control". Cognitive control is understood as supporting relevant processing and monitoring of behaviour such as the inhibition of competing sensory stimuli (Sharp *et al.*, 2006; Eckert *et al.*, 2008). According to Shinn-Cunningham and Best (2008), important features of attention during speech perception are object formation and object selection. That refers to the cognitive effort essential to detect, group, and filter the physical features of the sound. Simultaneously, they point out that in case of signal degradation the attention system is greatly taxed because the perception of the stimulus-inherent features is hindered. That in turn imposes greater cognitive load and leads to the allocation of additional cognitive resources.

Not only the attention system but also working memory is taxed by the comprehension of degraded speech (Pichora-Fuller *et al.*, 1995). Pichora-Fuller and colleagues conducted a working memory task with sentences presented in multi speaker babble parametrically changing the signal-to-noise ratio (S/N). Evidence was found that adverse S/N led to poorer working memory task performance, which they explained with the greater amount of resources being allocated for the processing of auditory information.

Neural signatures of compensation: Empirical findings

We propose that working memory should be best viewed as two distinct mechanisms in speech comprehension: First, a "sensory working memory" deals with buffering and encoding of the perceived speech signal. Second, after the signal is encoded, it is further stored and processed somewhat independent of the signal domain. The latter mechanism is supposedly identical to Baddeley's "phonological store" (Baddeley and Hitch, 1974). In the following, we would like to argue that both mechanisms are taxed to a greater extend under degradation and hence require a high level of compensatory effort.

Regarding the sensory working memory, evidence has been provided for a strong involvement of the planum temporale (PT) during the processing of degraded speech. The PT is a brain area posterior to Heschl's gyrus, which is known to house primary auditory cortex. In general, the PT is engaged in the analysis of complex auditory input. The more entropic and hence less predictable a sound the more active is the PT, better responding to spectro-temporal complexity (Griffiths and Warren, 2002; Overath *et al.*, 2007). Based on these and other results, Obleser and Eisner (2009) argued that the PT plays an important role in the pre-lexical processing of speech but is not specific to verbal information. With a meta-analytical approach Obleser *et al.* (2009) examined the results of ten functional magnetic resonance imaging (fMRI) studies that investigated the processing of un- or hardly intelligible speech. The focus was on the contrast of degraded speech over clear speech. The results led to the conclusion that the activity of the PT rather increases with decreasing comprehensibility of the speech signal, arguably pointing to its increasing importance in the "sensory working memory" aspects in degraded speech.

Evidence on increased effort for neural encoding of degraded speech has also been found on the temporal level of speech processing. Miettinen *et al.* (2011) conducted a magnetoencephalography (MEG) study on the processing of distorted speech signals. In principle, MEG reflects the magnetic field counterpart to the electric potential measured from the scalp in EEG. They could show a stronger auditory evoked field after 100 ms after word onset (N100m) for degraded compared to clear speech. Within the current framework, this early N100m effect can be taken as an early indication of high processing effort in distorted speech.

It is our working hypothesis that the enhanced activity in PT and the more vigorous N100 (see also Obleser & Kotz, 2011) both reflect the allocation of additional neural resources in the processing of degraded speech – a compensatory function at the rather early, sensory level. Fostering the PT–N100 link, it is noteworthy that the PT (together with adjacent Heschl's gyrus) is a likely generator structure of the N100 component (Scherg *et al.*, 1989, cf. Obleser and Eisner, 2009). After the (more or less effortful) neural encoding of the speech signal, information is stored and processed in working memory. This is where cognitive control comes into play. Eckert *et al.* (2008) conducted an fMRI study with younger and older people performing a word recognition task on degraded speech. Older people showed enhanced activity in regions responsible for cognitive control such as the anterior cingulate cortex (ACC) and the midfrontal gyrus (MFG). Since older people often suffer from cognitive decline they need to allocate more cognitive resources in order to understand degraded speech.

Alpha oscillations and their role in speech processing

Besides the Gamma-band results mentioned above, only sparse evidence has been gathered so far on human brain oscillatory responses in speech comprehension. In a recent review, Weisz *et al.* (2011) summarise results from studies that examined changes in the prominent alpha rhythm band. This allows some tentative conclusions on the role of the prominent alpha rhythm (8–13 Hz) in auditory and speech processing:

Recent studies on auditory alpha rest on the theoretical framework by Jensen and Mazaheri (2010). The authors suggest that alpha activity is able to functionally inhibit task-irrelevant regions. Hence, alpha *suppression* can be taken to indicate *less* need for functional inhibition, i.e., active cognitive processing such as accessing the lexical meaning of an utterance. When degradation of the speech signal is added, this typical suppression of alpha is parametrically reduced with increasingly severe degradation (Obleser & Weisz, in revision; Obleser and Weisz, 2010). The alpha modulation as a function of acoustic degradation appears approximately 500 ms after word onset (roughly coinciding with word recognition, i.e., when the word meaning is accessed). Fig. 4 shows the monotonic change in alpha power (significant effect; tested on trials of a single subject, Weisz *et al*., 2011) as the signal quality is varied.

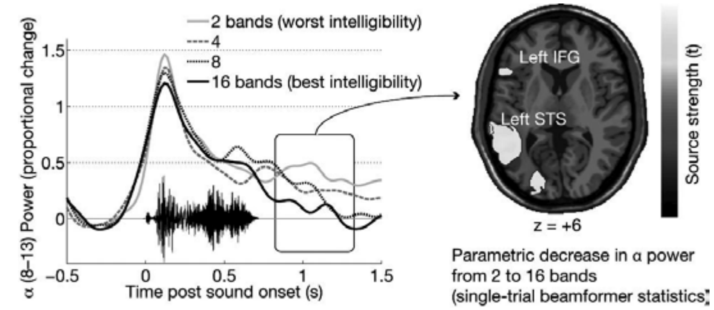

Fig. 4: Single-subject data on late alpha suppression in word comprehension (EEG) as a function of signal degradation. The left panel shows alpha power over time. The right panel shows the source localization of the late alpha-power effect projected onto a standard brain template.

Also following from the functional inhibition framework, active processing is assumed to be reflected in enhanced gamma-band activity. This "trade-off" has been actually demonstrated, showing that the power of fast gamma-band oscillations is coupled to the phase of slower alpha-band oscillations (Jensen and Mazaheri, 2010).

In a similar vein, evidence has been gathered for increasing alpha power in conditions where a need for functional inhibition arises: This effect occurs with increasing memory load during retention of a set of stimuli (Jensen *et al.*, 2002; Leiberg *et al.*, 2006; see Klimesch *et al.*, 2007 for review). Taking these findings into account and knowing that the processing of degraded speech taxes working memory as well as selective attention, an increase in alpha band activity is expected to be found for the processing of more severely degraded speech. This would reflect additional allocated cognitive resources in terms of inhibition of task-irrelevant information; preliminary data summed in Weisz *et al.* (2011) favour this interpretation.

Summing up the initial findings on alpha oscillations in degraded speech, it appears that fluctuations of the alpha frequency band can supply valid information on speech comprehension success. Because of the good signal-to-noise ratio of the alpha band signal, changes in alpha power might be able to provide insight into speech comprehension even on the single-subject or single-trial level and/or in clinical populations, where testing time is often limited.

"BOOTSTRAPPING" THE SPEECH SIGNAL: A TRADE-OFF OF FACILITATION AND COMPENSATION

From a strictly auditory perspective (i.e., ignoring the potential benefit from visual cues for now), a listener in an adverse listening situation must "bootstrap" her way out of this adversity and towards comprehension by accumulating as much information as possible from the degraded signal itself. As we have argued here, facilitation and compensation are two possible neuro–cognitive mechanisms that aid speech comprehension in such adverse listening situations.

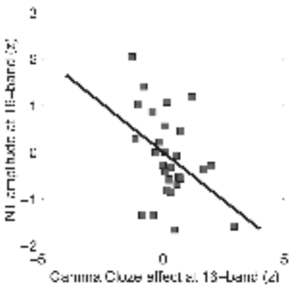

Fig. 5: Negative correlation between individual N100 amplitude (z-normalised) at sentence-onset in response to moderately degraded speech (16 bands) and individuals' Gamma-band power enhancement at the sentence-final keyword (z-normalised, in response to 16-band speech); Pearson's $r = -0.41$, $p < 0.03$.

Of course, this raises questions on the interplay of both mechanisms. Are both strategies applied to the same extent? What if one strategy does not take effect; does the other one step in? Are there individual preferences for one strategy over another? These are future research questions. At least, we have observed some indication for individual differences in the extent to which effortful, "compensatory" strategies and integrative, "facilitatory" strategies are utilised. As mentioned before, Obleser and Kotz (2011) reported an inverse relationship of the amplitude of the N100 component to degraded speech and the late gamma response when integrating speech meaning (Fig. 5). If, as argued above, we take the early N100 to reflect increased effort and later gamma-band activity to reflect ease of semantic integration, this negative correlation would indicate a trade-off between these two strategies: If initial compensation is enhanced, fewer facilitatory processes are engaged. Inversely, other listeners might sacrifice sensory effort ("listening hard") for more top-down facilitated "easy guessing". This needs further investigation.

REFERENCES

Aydelott, J., Dick, F., and Mills, D. L. (**2006**). "Effects of acoustic distortion and semantic context on event-related potentials to spoken words" Psychophysiology, **43**, 454-464.

Baddeley, A. D., and Hitch, G. J. (**1974**). "Working Memory" in *Psychology of learning and motivation*, G. A. Bower (Ed.), Academic Press, New York, 47-90.

Binder, J. R., Frost, J. A., Hammeke, T. A., Rao, S. M., and Cox, R. W. (**1996**). "Function of the left planum temporale in auditory and linguistic processing" Brain, **119**, 1239-1247.

Davis, M. H., and Johnsrude, I. S. (**2003**). "Hierarchical processing in spoken language comprehension" J. Neurosci., **23**, 3423-3431.

Davis, M. H., and Johnsrude, I. S. (**2007**). "Hearing speech sounds: top-down influences on the interface between audition and speech perception" Hearing Res., **229**, 132-147.

Eckert, M. A., Walczak, A., Ahlstrom, J. B., Denslow, S., Horwitz, A., and Dubno, J. R. (**2008**). "Age-related effects on word recognition: reliance on cognitive control systems with structural declines in speech-responsive cortex" J.Assoc. Res. Otol., **9**, 252-259.

Fries, P. (**2009**). "Neuronal gamma-band synchronization as a fundamental process in cortical computation" Ann. Rev. Neurosci., **32**, 209-224.

Griffiths, T. D., and Warren, J. D. (**2002**). "The planum temporale as a computational hub" Trends Neurosci., **25**, 348-353.

Hackett, T. A. (**2008**). "Anatomical organization of the auditory cortex" J. Am. Acad. Audiol., **19**, 774-779.

Hannemann, R., Obleser, J., and Eulitz, C. (**2007**). "Top-down knowledge supports the retrieval of lexical information from degraded speech" Brain Res., **1153**, 134-143.

Jensen, O., Gelfand, J., Kounios, J., and Lisman, J. E. (**2002**). "Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task" Cereb. Cortex, **12**, 877-882.

Jensen, O., and Mazaheri, A. (**2010**). "Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition" Front. Hum. Neurosci., **4**, 1-8.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability" J. Acoust. Soc. Am., **61**, 1337-1351.

Klimesch, W., Sauseng, P., and Hanslmayr, S. (**2007**). "EEG alpha oscillations: The inhibition–timing hypothesis" Brain Res. Rev., **53**, 63-88.

Kutas, M., and Hillyard, S. A. (**1980**). "Reading senseless sentences: brain potentials reflect semantic incongruity" Science, **207**, 203-205.

Leiberg, S., Lutzenberger, W., and Kaiser, J. (**2006**). "Effects of memory load on cortical oscillatory activity during auditory pattern working memory" Brain Res., **1120**, 131-140.

Levelt, W. J. M. (**1989**). "Speaking. From Intention to Articulation" MIT Press, Cambridge, Mass.

Miettinen, I., Alku, P., Salminen, N., May, P. J. C., and Tiitinen, H. (**2011**). "Responsiveness of the human auditory cortex to degraded speech sounds: reduction of amplitude resolution vs. additive noise" Brain Res., **1367**, 298-309.

Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). "The intelligibility of speech as a function of the context of the test materials" J. Exp. Psychol., **41**, 329-335.

Norris, D., and McQueen, J. M. (**2008**). "Shortlist B: a Bayesian model of continuous speech recognition" Psychol. Rev., **115**, 357-395.

Obleser, J., and Eisner, F. (**2009**). "Pre-lexical abstraction of speech in the auditory cortex" Trends Cogn. Sci., **13**, 14-19.

Obleser, J., Eisner, F., Scott, S. K., Kotz, S. A., and Friederici, A. D. (**2009**). "Listen hard: A meta-analysis on brain correlates of failing speech comprehension" Program No. 70.5. 2009 Neuroscience Meeting Planner. Chicago, IL: Society for Neuroscience, 2009. Online.

Obleser, J., and Kotz, S. A. (**2011**). "Multiple brain signatures of integration in the comprehension of degraded speech" NeuroImage, **55**, 713-723.

Obleser, J., and Weisz, N. (**2010**). "Differential influences of spectral and temporal features of speech on human oscillatory brain dynamics" Program No. 837.6. 2010 Neuroscience Meeting Planner. San Diego, CA: Society for Neuroscience, 2010. Online.

Obleser, J., Wise, R. J. S., Alex Dresner, M., and Scott, S. K. (**2007**). "Functional integration across brain regions improves speech perception under adverse listening conditions" The Journal of Neuroscience: J. Neurosci., **27**, 2283-2289.

Overath, T., Cusack, R., Kumar, S., von Kriegstein, K., Warren, J. D., Grube, M., Carlyon, R. P., and Griffiths, T. D. (**2007**). "An information theoretic characterisation of auditory encoding" PLoS Biol., **5**, e288.

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise" J. Acoust. Soc. Am., **97**, 593-608.

Pichora-Fuller, M. K., and Singh, G. (**2006**). "Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation" Trends Amplif., **10**, 29-29.

Rauschecker, J. P., and Scott, S. K. (**2009**). "Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing" Nat. Neurosci., **12**, 718-724.

Scherg, M., Vajsar, J., and Picton, T. W. (**1989**). "A source analysis of the late human auditory evoked potentials" J. Cogn. Neurosci., **1**, 336-355.

Schreiner, C. E., and Winer, J. A. (**2007**). "Auditory cortex mapmaking: principles, projections, and plasticity." Neuron, **56**, 356-365.

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (**2000**). "Identification of a pathway for intelligible speech in the left temporal lobe" Brain., **123**, 2400-2406.

Shannon, R. V., Fu, Q.-J., and Galvin, J. (**2004**). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation" Acta Otol. Suppl., **552**, 50-54.

Sharp, D. J., Scott, S. K., Mehta, M. A., and Wise, R. J. S. (**2006**). "The neural correlates of declining performance with age: evidence for age-related changes in cognitive control" Cereb. Cortex, **16**, 1739-1749.

Shinn-Cunningham, B. G., and Best, V. (**2008**). "Selective attention in normal and impaired hearing" Trends Amplif., **12**, 283-299.

Stickney, G. S., and Assmann, P. F. (**2001**). "Acoustic and linguistic factors in the perception of bandpass-filtered speech" J. Acoust. Soc. Am., **109**, 1157-1165.

Weisz, N., Hartmann, T., Müller, N., and Obleser, J. (**2011**). "Alpha Rhythms in Audition : Cognitive and Clinical Perspectives" Front. Psychol., **2**, 73.

Dynamic and task-dependent encoding of speech and voice in the auditory cortex

MILENE BONTE [1,2] AND ELIA FORMISANO [1,2]

[1] Maastricht Brain Imaging Center and [2]Dept. of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

Speech is at the core of verbal communication and social interaction. It conveys linguistic content and speaker-specific vocal information that listeners exploit for identification. Cortical processing of speech relies on the formation of abstract representations that are invariant to highly variable acoustic input signals and critically depends on behavioral demands. In a series of EEG and fMRI studies we have recently investigated temporal as well as spatial neural coding mechanisms for forming such abstract representations. We focused on categorical and task-dependent neuronal responses to natural speech sounds (vowels /a/, /i/, /u/) spoken by different speakers. Brain activity was measured during passive listening (fMRI, EEG) and during performance of behavioural tasks on vowel or speaker identity (EEG). Our EEG results show that dynamic changes of sound-evoked responses and phase patterns of cortical oscillations in the alpha band (8-12 Hz) closely reflect the abstraction and analysis of the sounds along the task-relevant dimension. Our fMRI results show that spatially distributed activation patterns in early and higher level auditory cortex encode vowel-invariant representations of speaker identity and speaker-invariant representations of vowel identity. Both the transient and task-dependent realignment of neuronal responses (EEG) and the spatially distributed cortical fingerprints (fMRI) provide robust cortical coding mechanisms for forming abstract representations of auditory (speech) signals.

We are suprisingly effcient in understanding who is speaking and what is being said from highly variable speech signals. Furthermore, dependent on the current behavioural goal, we may choose to focus our attention on either speaker identity or speech content and ignore the other dimension. Such adaptive behaviour requires computational mechanisms that enable different (abstract) representations of the same acoustic input.

Cognitive and connectionist models suggest that speech recognition involves the formation of intermediate entities such as phonemes that are invariant to changes in the acoustic input (due to speaker variability, noise or signal distortion) and that can be used for further linguistic processing (McClelland and Elman, 1986; Norris and McQueen, 2008). Similarly, speaker recognition may involve the formation of

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.