Julia Erb: The neural dynamics of perceptual adaptation to degraded speech.
Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences, 2014
(MPI Series in Human Cognitive and Brain Sciences; 159)

The neural dynamics of perceptual adaptation
to degraded speech

Impressum

Max Planck Institute for Human Cognitive and Brain Sciences, 2014

Druck: Sächsisches Druck- und Verlagshaus Direct World, Dresden
Titelbild: © Moritz Ellerich / RaumZeitPiraten / Fabelphonetikum (rhizome schematic) 2014

ISBN 978-3-941504-43-1

The neural dynamics of perceptual adaptation to degraded speech

Der Fakultät für Biowissenschaften, Pharmazie und Psychologie
der Universität Leipzig
eingereichte

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt von
Julia Erb, M.Sc. (Neuroscience)
geboren am 17. September 1985 in Heilbronn

Leipzig, den 14. Februar 2014

Acknowledgements

I would like to thank several people who supported me and decisively contributed to my work during the last three years at the Max Planck Institute.

First, I cordially thank my supervisor Jonas Obleser. Jonas was incredibly supportive in all stages of this thesis, during the conception, analysis and the publication of the experiments. Thank you for your generous encouragement to my development as a scientist.

I am obliged to Erich Schröger and Mark Eckert who agreed to assess my work.

I thank my colleagues and friends from the Auditory Cognition group, Björn Herrmann, Sung-Joo Lim, Mathias Scharinger, Antje Strauß, Anna Wilsch and Malte Wöstmann, for fruitful discussions in Tuesday morning group meetings and Friday evening Cantona sessions – I miss them. In particular, I thank Molly Henry, with whom I discussed many ideas and whose traces can be found throughout this thesis.

I am grateful to Alexandra Ludwig and Michael Fuchs for their prompt acceptance of collaborating on the cochlear implant patient project. Several scientists gave helpful advice and comments at different phases of this dissertation; I particularly thank Frank Eisner, Carolyn McGettigan, Jonathan Peelle and Sonja Kotz.

Many helpers deserve credit for their support in the realisation of this research. In particular, Dunja Kunke and our student assistants tirelessly invited participants and helped with behavioural data acquisition. Without the MRI team, namely André Pampel, Sylvie Neubert, Nicole Pampus, Mandy Jochemko, Anke Kummer and Simone Wipper, acquisition of the imaging data would have been impossible - many thanks for accepting my special scanning wishes. Jöran Lepsien, Lars Meyer and Zoe Schlüter supported me in the analysis of MRI data. I am thankful to Elizabeth Kelly, Armand Morel and Nicola Filippini for proofreading these pages and Kerstin Flake and Steven Kalinke who assured good quality of the graphics and layout of this thesis.

Several people have modified, with their friendship, my geography of Leipzig, and provided me with shelter and support, particularly in the final steps of this thesis. I am thinking of Bruno, Jonathan, Tomas, Roberta, and Christian.

I thank my dear sisters, Stephanie and Christiane, for their unconditional support, and Marlon for changing my universe – you're spring to me, all things to me.

This work is dedicated to my parents, Bärbel and Helmut Erb.

Julia Erb, Leipzig, February 2014

Bibliographic details

Julia Erb, *The neural dynamics of perceptual adaptation to degraded speech*

Fakultät für Biowissenschaften, Pharmazie und Psychologie, Universität Leipzig

Dissertation

201 pages, 339 references, 21 figures, 9 tables

This thesis investigates how the speech perception system copes with challenging listening situations, caused by external degradation of the speech signal, age-related hearing loss, and cochlear implantation. In particular, we aim to isolate predictors of successful adaptation to degraded speech and to understand the underlying neural mechanisms. Five experiments of short-term adaptation to, and processing of, degraded speech were conducted using cognitive and psychoacoustical testing, structural and functional magnetic resonance imaging (fMRI). Speech was degraded using noise-vocoding, which reduces the spectral detail in an auditory signal while leaving the temporal envelope cues intact. The results show that individual differences in thalamic brain morphology and auditory temporal processing skills, as assessed by amplitude modulation rate discrimination, were predictive of rapid adaptation to degraded speech in healthy adults. Importantly, these findings could be replicated in cochlear implant patients whose sensitivity to temporal modulations shortly after implantation could even prospectively predict speech recognition abilities 6 months later. In an fMRI experiment in normal-hearing listeners, the neural plasticity underlying short-term adaptation to degraded speech was found to be rooted in a haemodynamic upregulation of premotor activity and a concurrent down-regulation in a striato-thalamo-cortical circuit. An additional fMRI study on AM rate discrimination revealed a cingulo-opercular network (comprising the anterior insula and anterior cingulate cortex) on which increasingly difficult non-speech discrimination and effortful degraded speech perception jointly relied, most likely reflecting executive (e.g., attentional) processes. In older adults, this network exhibited a generalised upregulation, during perception of both degraded and clear speech, corroborating hypotheses of increased reliance on cognitive control systems in older adults. Lastly, a signal processing scheme which enhances temporal envelope cues in a speech signal was shown to improve comprehension in noise, confirming the importance of temporal envelope cues in speech perception. The results provide novel evidence on the central neural mechanisms of adaptation to adverse listening conditions in listeners of different ages and hearing acuities. These findings are discussed in a speculative framework of adaptive brain mechanisms, integrating evidence across behavioural, cellular and macroanatomical levels.

Contents

"La pensée n'est pas arborescente, et le cerveau n'est pas une matière enracinée ni ramifiée. Ce qu'on appelle à tort « dendrites» n'assurent pas une connexion des neurones dans un tissu continu. La discontinuité des cellules, le rôle des axones, le fonctionnement des synapses, l'existence de micro-fentes synaptiques, le saut de chaque message par-dessus ces fentes, font du cerveau une multiplicité qui baigne, dans son plan de consistance ou dans sa glie, tout un système probabiliste incertain, uncertain nervous system. Beaucoup de gens ont un arbre planté dans la tête, mais le cerveau lui-même est une herbe beaucoup plus qu'un arbre."

- Gilles Deleuze et Félix Guattari (1980), Mille Plateaux,
p. 24, Paris: les Éditions de Minuit

[“Thought is not arborescent, and the brain is not a rooted or ramified matter. What are wrongly called "dendrites" do not assure the connection of neurons in a continuous fabric. The discontinuity between cells, the role of the axons, the functioning of the synapses, the existence of synaptic microfissures, the leap each message makes across these fissures, make the brain a multiplicity immersed in its plane of consistency or neuroglia, a whole uncertain, probabilistic system ("the uncertain nervous system"). Many people have a tree growing in their heads, but the brain itself is much more a grass than a tree.”]

- Translation by Brian Massumi (1987), A Thousand Plateaus,
p. 15, Minneapolis: University of Minnesota Press

Preface

Speech is rarely crystal clear. Yet, listeners are able to rapidly adapt when the acoustic signal is degraded. That is, listeners achieve almost as high speech comprehension from degraded signals as they usually would from clear (i.e., undistorted and unmasked) speech after short periods of exposure to the degraded signal. This dissertation investigates the neural networks that support comprehension when listening conditions become difficult. In particular, we examine adaptive behavioural mechanisms and neural plasticity that allow for rapid adjustment when speech is degraded.

Two sources of speech degradation are investigated. First, speech can be *externally* degraded, due to speaker idiosyncrasies or environmental influences (e.g., in phone lines or due to background noise). Second, on top of such external degradation, declines in the acuity of the peripheral sensory system can *internally* degrade speech. Hearing-impaired or, more drastically, cochlear implant (CI) listeners, whose hearing is restored by means of electric stimulation of the auditory nerve, experience such degradation.

In a first approach, we aim to understand adaptation through listener variability: can individual differences in degraded speech perception reveal the factors that crucially drive adaptation? Both brain structural and behavioural (cognitive and basic auditory abilities) characteristics are examined with respect to their predictive power of adaptation to degraded speech (Experiment I). In parallel, this thesis investigates whether such factors can also explain the variability in speech comprehension success observed in CI patients (Experiment II).

The neural dynamics that support such adaptive mechanisms can only be revealed using neuroimaging methods. Therefore, in a second set of experiments functional imaging is applied in order to trace the neural locus at which adaptation to degraded speech occurs. We first investigate neural speech processing in young, normal-hearing listeners (Experiment III), and subsequently ask how such mechanisms change with age and age-related hearing loss (Experiment VI).

Lastly, we test whether increasing the saliency of acoustic cues in a degraded speech signal, which appear to be crucial in comprehension of degraded speech, can in turn boost intelligibility. The perceptual consequences of manipulating temporal acoustic cues in noise are investigated (Experiment V).

With these three approaches we aim to understand how the perceptual system utilises severely degraded input to interpret speech, which ultimately might enable shaping listening strategies in difficult acoustic conditions.

1 General introduction

Humans are able to recognise speech in a variety of conditions. Although the acoustic speech signal can dramatically vary, listeners are able to robustly understand speech. A word pronounced by speakers with different age, gender, or accent has considerable variation in the acoustic representation (Abercrombie, 1967), but can still be understood. Even within the same speaker, utterances show substantial variability in the acoustic signal, for example due to differences in speech rate (e.g., Summerfield, 1981). On top of across- and within-talker variability, differences in listening environments introduce variation to the speech signal, caused by background noise, competing talkers or reverberation. This robustness of speech recognition in the face of such large variability of the acoustic signal attests a remarkable flexibility to the human speech perception system.

1.1 Perceptual adaptation

Gibson (1969) proposed that the perceptual system undergoes a constant calibration to adapt to the environmental demands, a phenomenon which has been termed “perceptual learning” or “adaptation”. Goldstone (1998) defines perceptual learning as “relatively long-lasting changes to an organism’s perceptual system that improve its ability to respond to its environment and are caused by this environment” (Goldstone, 1998, p. 586). In contrast, more transient short-term perceptual changes occurring immediately when presented with a novel stimulus are often referred to as perceptual adaptation (Helson, 1948). The premise of both definitions is that perceptual learning or adaptation “benefits an organism by tailoring the processes that gather information” (Goldstone, 1998, p.586) and ultimately improves performance. Perceptual learning or adaptation, as it is understood in the present thesis, refers to improvement that is specific to the perceptual judgment. Perceptual learning needs to be clearly distinguished from procedural learning, which refers to an improvement resulting from aspects of task performance not involved in making a perceptual judgment (i.e., habituation with response demands; Hawkey et al., 2004).

1.1.1 Mechanisms of perceptual adaptation

How such perceptual changes occur is not well understood (Goldstone, 1998). A controversial point is the level of information that drives perceptual learning: What are the relative contributions of bottom-up sensory cues and top-down knowledge-based cues?

Behavioural experiments have provided abundant evidence for both mechanisms of perceptual learning. In the visual literature, perceptual adaptation has been observed for unattended, not consciously perceived stimuli (Watanabe et al., 2001), consistent with a bottom-up stimulus-driven account. Further evidence for adaptation occurring early in the sensory system is provided by the finding that training of visual texture segregation is specific to retinal location and eye (and does not transfer to the ipsilateral eye; Karni and Sagi, 1991). In the auditory literature, some have argued that training selective attention to tones changes responses as early as in the cochlea, as indexed by modulation of tone-evoked potentials in the brain stem (Lukas, 1980). Classical conditioning can modify receptive fields in primary auditory cortex of the guinea pig and learning-induced plasticity is highly specific to the trained frequency (Weinberger and Bakin, 1998). In humans, however, training on frequency discrimination appears to generalise (although only partly) to untrained frequencies, (Irvine et al., 2000, Amitay et al., 2005; Delhommeau et al., 2005, for review see Wright and Zhang, 2009), and across ears (Delhommeau et al., 2002; Ari-Even Roth et al., 2003), suggesting that learning occurs at a higher processing stage rather than a sensory encoding stage. Frequency discrimination learning occurs even in the absence of a discriminable stimulus difference (Amitay et al., 2006), but is dependent on the task performed (e.g., improvement occurs following frequency but not intensity discrimination training; Halliday et al., 2011), suggesting that training improves the ability to attend to a task-specific stimulus dimension.

A number of studies have observed perceptual adaptation for speech (for review see Samuel and Kraljic, 2009). Rapid improvement in speech comprehension has been demonstrated when listeners were first confronted with a foreign accent (Bradlow and Bent, 2008; Janse and Adank, 2012). For example, Clarke and Garrett (2004) showed that perceptual adaptation to Spanish- or Chinese-accented sentences occurs already within the first minute of exposure. Similarly, listening to an unfamiliar accent led to fast improvements in comprehension (Adank and Janse, 2010). In the following section, adaptation to degraded speech will be reviewed in more detail.

1.1.2 Speech degradation

In addition to such "natural" alterations of speech, the acoustic speech signal can be artificially degraded in a controlled manner. Artificial speech degradation allows for investigating how the perceptual system reacts when specific acoustic properties are altered. Reductions to both the temporal (Wingfield et al., 1984; Mehler et al., 1993), and the spectral information (Shannon et al., 1995; Rosen et al., 1999; Scott et al., 2000) coded in the clear speech signal have been shown to not disrupt intelligibility severely.

Sentences which are reduced to 40% of the initial duration ("time-compressed speech") are still intelligible (Garvey, 1953). Only a few minutes of exposure leads to substantial improvement in comprehension of time-compressed speech (Voor, 1965; Wingfield et al., 1984; Peelle and Wingfield, 2005; Adank and Devlin, 2010).

Robust speech recognition was also observed for noise-vocoded speech (Shannon et al., 1995). Vocoding was initially developed as a method to reduce information in a speech signal to be transmitted in the telephone line (Dudley, 1939). It removes most of the spectral properties of the auditory signal, while its temporal envelope cues remain largely preserved (Rosen et al., 1999). This is accomplished by dividing the original speech signal into different frequency bands, extracting the amplitude envelope from each band, and reapplying the amplitude envelope to bandpass-filtered noise carriers with matched cut-off frequencies (see Section 2.2). The fewer frequency bands used, the poorer the spectral resolution, and thus the less intelligible the speech signal will be. The resulting speech signal has been described as sounding "like a harsh whisper" (Scott et al., 2000).

A growing number of studies have investigated perceptual learning for vocoded speech. On the one hand, noise-vocoded speech allows for quantification the relative importance of temporal versus spectral cues in the speech signal for comprehension in the intact perceptual system. On the other hand, it serves as a good approximation of how CI patients perceive speech (see Section 1.1.3.) and can therefore be used as CI simulation in normal-hearing listeners.

While fast perceptual learning of vocoded speech has been established by a number of studies (Davis et al., 2005; Hervais-Adelman et al., 2008; Eisner et al., 2010), the mechanisms involved in these perceptual changes are debated. There is evidence consistent with both bottom-up and top-down accounts of learning. (Huyck and Johnsrude, 2012) argue that perceptual learning of vocoded speech is only possible when stimuli are attended, but is absent when auditory or visual distracters are attended, consistent with top-down attentional influences on learning. Lexical access has been shown to facilitate perceptual learning (Davis et al., 2005): Learning was boosted when a degraded sentence was preceded by its clear version such that the identity of the sentence was known ("pop-out" effect), but was absent for nonword sentences, pointing to a higher-level locus for learning.

In contrast, (Hervais-Adelman et al., 2008) observed learning of noise-vocoded nonwords, as well as a transfer effect from trained to untrained words, consistent with a prelexical locus for learning. Further evidence that lexical information is not mandatory for perceptual adaptation comes from cross-language training studies. Training of sine-wave-vocoded speech (Bent et al., 2011) in an unknown foreign language generalised to the untrained native language, depending on the phonological similarities of both languages. Generalisation of noise-vocoded speech learning was also observed among different frequency regions and partly among different carriers (noise, sine-wave or pulse-train carriers; Hervais-Adelman et al., 2011). Under the premise that training can only generalise to untrained conditions if the neural processes affected by training contribute to both conditions, such generalisation patterns can give insights into the neural level at which learning occurs. With this in mind, generalisation across languages, frequencies and different carriers speaks to a level of learning where same amount of abstraction from the acoustic signal is already achieved (Hervais-Adelman et al., 2011). Nevertheless, to trace the neural locus of perceptual learning, neuroimaging techniques can be applied which have provided evidence that processing as early as in the auditory brain stem is changed after perceptual learning of degraded speech. Neuroimaging studies of perceptual learning will be reviewed in Section 1.2.2.

1.1.3 Cochlear implant listeners

Such perceptual adaptation mechanisms are not only important to normal-hearing listeners in degraded listening conditions. They bear particular relevance to listeners with disorders of the auditory system such as sensorineural hearing loss, which endogenously degrades the signal, or, in the most extreme case, CI patients who need to adapt to the coarse input delivered by their hearing aid device (Giraud et al., 2001a; Fallon et al., 2008).

Cochlear implants can restore hearing in deaf patients by means of electric stimulation of the auditory nerve. First proposed in the 1950s by Djourno and Eryies (Eisen, 2003), CIs are today the neural prostheses which are most effective in replacing a sensory organ (Moore and Shannon, 2009). A CI converts sound pressure into electric pulse trains to directly stimulate the auditory nerve, thereby replacing the damaged cochlea. The CI transduced speech signal is considerably distorted and highly artificial however, as the device only provides a limited spectral resolution (Moore and Shannon, 2009). Noise-vocoding (see Section 1.1.2) has been used to simulate cochlear-implant transduced speech in normal-hearing listeners (Shannon et al., 1995). The number of frequency bands used is thought to correspond to the number of electrodes in a CI that stimulate only one specific frequency region of the auditory nerve.

The central auditory system is capable of using the coarse input delivered by the implant, but CI users vary considerably in how well they adapt to their hearing device. Some learn to understand speech even under difficult listening conditions while others are hardly able to use their device. It is still largely unknown which factors drive the adaptation to the CI transduced speech signal. In cochlear-implanted children, demographic factors such as age at implantation, duration of deafness, residual hearing before implantation or the number of electrodes of the CI can only explain about half of the variance in outcome (Blamey et al., 2001; Sarant et al., 2001). Hence, we propose that a combination of cognitive capabilities such as attentional and memory capacities on the one hand, and basic auditory capacities on the other hand better should capture some part of the observed variability.

Degraded speech recognition requires the maintenance of auditory input in working memory for a short period of time, to be available for further phonological analysis. Therefore, working memory has been suggested to be a crucial cognitive factor which influences speech recognition performance in CI users (Cleary et al., 2001). Particularly in paediatric CI users, Pisoni and colleagues provided evidence that working memory scores significantly correlate with word recognition, even after accounting for demographic factors (Cleary et al., 2000; Pisoni and Geers, 2000; Pisoni and Cleary, 2003; Dillon et al., 2004; Fagan et al., 2007; Harris et al., 2013).

On the other hand, basic auditory skills, as far as they can be disentangled from cognitive factors, might be a key to the observed variability. While it is unfeasible to evaluate such auditory capacities prior to implantation (because the patient is mostly fully deafened), basic auditory abilities can be measured post-operatively. One suggestion, how low-level non-speech auditory skills can be usefully assessed just after cochlear implantation will be outlined in the following section.

1.1.4 Temporal envelope cues in speech perception

Both CI transduced and noise-vocoded speech have a very poor spectral resolution. For comprehension, the listener must therefore rely primarily on temporal cues in the speech signal, that is, the temporal intensity modulations in the acoustic signal referred to as "speech envelope". The utilisation of such envelope cues should be easier for listeners with a higher sensitivity to amplitude fluctuations in an auditory signal. One way to evaluate the auditory system's sensitivity to temporal cues is to assess temporal modulation transfer functions (see Chapters 3 and 4 Viemeister, 1979; Wakefield and Viemeister, 1990), where the detectability of modulations in amplitude modulated (AM) sounds is measured as a function of modulation frequency.

Evidence from electrical stimulation experiments suggests that auditory temporal processing is indeed crucial for speech recognition in CI patients. The temporal resolution of CI patients' auditory system at higher frequencies (~50 - 100 Hz) was found to be associated with speech recognition abilities (Shannon, 1992; Fu, 2002; Luo et al., 2008; Chatterjee and Oberzut, 2011). Yet, the most prominent low-frequency modulations in the temporal envelope of speech are near the syllabic rate of approximately 4 Hz (Houtgast and Steeneken, 1985). Therefore, we hypothesised that listeners with higher sensitivity to these low AM frequencies adapt more quickly to vocoded speech or, in the case of CI patients, learn more rapidly to understand CI transduced speech.

As speech envelope cues appear to be necessary and almost sufficient for speech comprehension (Shannon et al., 1995), Apoux et al., (2001) suggested that enhancing the temporal envelope of speech might aid speech comprehension. In normal-hearing listeners, however, expansion of the speech envelope (i.e. increasing the modulation depth of the speech signal) deteriorates speech recognition in quiet (Fu and Shannon, 1998). Nevertheless, some studies suggest that when noise is added, speech envelope expansion aids comprehension (Lorenzi et al., 1999; Apoux et al., 2001). Deeper investigation is needed to distinguish when increased saliency of speech envelope cues is detrimental or beneficial to intelligibility.

1.2 The neural bases of perceptual adaptation

The neural systems that underlie perceptual processing of degraded speech can only be observed using neuroimaging methods. A growing number of studies have investigated brain structure and function mediating learning of speech and non-speech tasks using functional magnetic resonance imaging (MRI) or electrophysiological methods, which will be reviewed below.

1.2.1 Brain morphology and perceptual adaptation

Attempts have been made to assess whether certain brain morphological characteristics are related to perceptual learning. On the one hand, a growing body of literature shows that trained individuals such as taxi drivers (Maguire et al., 2000), phoneticians (Golestani et al., 2011) or musicians (Schlaug et al., 1995; Gaser and Schlaug, 2003; Bermudez and Zatorre, 2005) exhibit differential brain anatomy and there is direct evidence that intensive training alters grey and whiter matter architecture (Draganski et al., 2004; Scholz et al., 2009; Taubert et al., 2010).

Another strand of research comes from the perspective that pre-existing morphological characteristics may predict behaviour: In analogy to certain disorders which are associated with morphological abnormalities, such as stuttering (Foundas et al., 2001; Foundas et al., 2004; Foundas et al., 2013) or dyslexia (Galaburda et al., 1994), brain morphology, even if not pathological, may predict performance. For speech learning, notably Golestani and colleagues have pursued this idea, showing that listeners with increased white matter density in left parietal regions learned faster to discriminate a non-native phonetic contrast (Golestani et al., 2002). Also, the accurate pronunciation of foreign speech sounds was found to be coupled to increased white matter density adjacent to the insular/prefrontal cortex (Golestani and Pallier, 2007). Golestani et al., (2011) further observed that phoneticians show an increased likelihood of multiple transverse gyri in the auditory cortex. Because this was found to be independent of the amount of practice, the authors drew the provocative conclusion that brain structure can determine the profession an individual chooses (Golestani et al., 2011), but which illustrates the limitations of correlative brain structural analyses: Interpreting whether structural differences are shaped by experience or are pre-existing is problematic.

Recently, an auditory categorisation study from our own group demonstrated that the capacity of adaptively directing attention to the most informative dimension (spectral versus durational) of complex acoustic stimuli was predicted by grey matter density in parietal and precentral cortex, whereas categorisation performance was related to morphology of the auditory thalamus (Scharinger et al., 2013). Summing across the reviewed studies, there is evidence that learning and auditory performance are related to brain structural differences, which do not necessarily lie in primary auditory cortex (Foster and Zatorre, 2010). Whether brain structure can more specifically predict learning of degraded speech deserves further investigation.

1.2.2 Brain function and perceptual adaptation

Although behavioural and brain structural studies can inform about the locus of adaptation, the neural dynamics that accompany perceptual adaptation can only be traced online using functional imaging. A small number of experiments have investigated the neural processes mediating perceptual adaptation to degraded speech using fMRI or electrophysiological methods which will be reviewed below.

There is accumulating evidence from electrophysiological studies that perceptual learning of speech is accompanied by changes early in the auditory pathway. The auditory brain stem response, originating in medial superior olive and inferior colliculus, has been found to be altered by perceptual learning (Song et al., 2011; Chandrasekaran et al., 2012; Strait et al., 2012). For example, training of speech perception in noise enhanced the auditory brain stem response to pitch cues and was associated with behavioural improvements that were retained 6 months later (Song et al., 2011).

Further upstream along the auditory pathway, the first-order auditory thalamus (medial geniculate body; MGB; see Figure 1.1), which is the last processing station before the auditory cortex, has been suggested to exert a "gain control" via feedback connections from the cortex, and can thus enhance the detectability of degraded sounds (Bartlett, 2013). Rather than simply transferring auditory information to the cortex, the MGB is proposed to act like a "funhouse mirror" (Bartlett, 2013), that is, to reshape upcoming information. Although the MGB forms part of the ascending auditory system, it has been demonstrated to receive itself massive descending connections from the cerebral cortex, at least in the cat brain (Winer et al., 2001; for review see Winer, 2006). Therefore, an emerging view is that the MGB is not only involved in basic auditory processing, but can be top-down-modulated to process specific features of speech. For example, activity in the MGB has been found to be modulated dependent on the speech task a participant performed (syllable versus speaker recognition; von Kriegstein et al., 2008).

Located just above the auditory thalamus, the pulvinar nuclei of the thalamus (see Figure 1.1) have been shown to be involved in speech processing, as evidenced by direct electrical stimulation. Ojemann and colleagues stimulated the thalamus prior to therapeutic thalamectomy to verify the localisation of ablation. As electrical stimulation of the pulvinar was found to interfere with object naming and increased recall errors during a verbal short-term memory task, the authors attributed a gating function to the pulvinar where the thalamus controls retrieval from memory (Johnson and Ojemann, 2000).

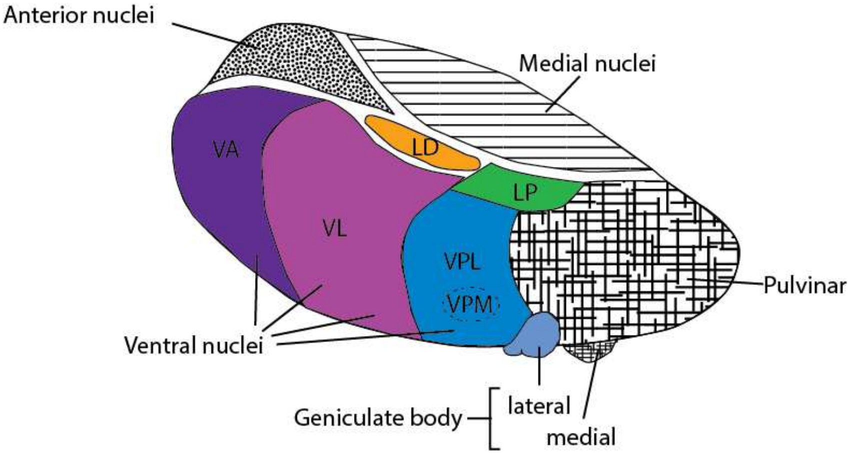


Figure 1.1. Schematic representation of thalamic morphology (left side). The thalamus can roughly be divided into three groups of nuclei, including the anterior, medial and ventral nuclei (VA, VL, VPL, VPM, pulvinar), and the metathalamus, comprising the lateral and medial geniculate bodies (visual and auditory thalamus). VA: ventral anterior nucleus, VL: Ventral lateral anterior nucleus, VPL: Ventral posterior lateral nucleus; VPM: ventral posterior medial nucleus; LD: Lateral dorsal nucleus; LP: Lateral posterior nucleus. Modified from Carpenter (1991).

Higher-level cortical areas have also been suggested to play a crucial role in perceptual adaptation. In particular, premotor cortex activity has repeatedly been associated with perceptual learning of speech. For example, (Adank and Devlin, 2010) conducted an fMRI study of short-term adaptation to time-compressed speech where they observed initially strong activation in the left ventral premotor cortex which decreased as listeners improved in comprehension of time-compressed sentences. Following Liberman's motor theory of speech perception, claiming that speech perception requires a tight coupling of auditory and motor processes (Liberman et al., 1967), the authors concluded that perceptual learning of speech is facilitated by mapping novel acoustic patterns onto articulatory motor plans (Adank and Devlin, 2010). In line with this finding, (Hervais-Adelman et al., 2012) observed increased premotor activity during vocoded speech perception: When perceptual learning was boosted by the successive presentation of a clear, vocoded and clear version of a word (see Section 1.1.2), activity in the premotor cortex increased.

Close to premotor cortex, (Eisner et al., 2010) found inferior frontal gyrus (IFG) activity to correlate with individual learning success in a short-term vocoded speech-learning paradigm with simultaneous written feedback. Following up this hypothesis of a critical role of the IFG in perceptual adaptation, (Sehm et al., 2013) applied transcranial direct current stimulation (tDCS) to the left IFG and angular gyrus during three days of feedback-based training with noise-vocoded minimal pairs. In the most severely degraded condition (two-band vocoding), perceptual learning was facilitated by tDCS over the IFG but not over the intraparietal cortex. This suggests that higher-order language areas significantly contribute to perceptual learning of speech, at least when feedback is provided.

The literature reviewed above is discordant in the point that electrophysiological studies suggest a critical role of subcortical structures in perceptual learning, whereas the majority of functional imaging studies indicate a cortical locus for learning, often placed in premotor cortex (Adank and Devlin, 2010; Hervais-Adelman et al., 2012) or IFG (Eisner et al., 2010; Sehm et al., 2013). The role of subcortical structures in perceptual adaptation might have been disguised due to their susceptibility to movement artefacts (see Section 2.6.1). To follow-up on the possible subcortical involvement in learning, functional imaging studies with a better resolution of subcortical structures are needed.

1.2.3 Neural adaptation in older adults

While normal-hearing young adults have the capacity to rapidly adapt to degraded speech (Samuel and Kraljic, 2009; see above), such perceptual adaptation is not well established in older adults. However, as older adults' sensory acuity commonly declines, they are often confronted with degraded conditions where such adaptation mechanisms become crucial.

A few behavioural studies suggest that rapid perceptual learning of degraded speech is preserved in older adults, showing that older adults are able to adapt to an unfamiliar (Adank and Janse, 2010) or a foreign accent (Gordon-Salant et al., 2010), to time-compressed and noise-vocoded speech (Peelle and Wingfield, 2005). An unresolved question is, however, how hearing loss and cognitive aspects influence adaptation to and neural processing of degraded speech in older listeners.

Neuroimaging studies have frequently observed recruitment of additional neural networks in older compared to young adults (Cabeza et al., 2002; Cabeza et al., 2004; Eckert et al., 2008; Heuninckx et al., 2008; Peelle et al., 2010b). The interpretation of this age-related spread of activation is controversial: On one hand, it has been hypothesised to reflect a loss of specialisation of brain structures (Park et al., 2004), possibly due to a deficit in neurotransmission (Li, 1999). On the other hand, (Cabeza et al., 2002; Cabeza et al., 2004) have argued that recruitment of additional areas serves as compensatory mechanism, by which older adults manage to maintain performance despite degeneration of the sensory cortices (Eckert et al., 2008; Peelle et al., 2010b).

Accordingly, older adults have been hypothesised to increasingly rely on executive, attentional mechanisms for speech comprehension (Pichora-Fuller, 2003). Neuroimaging studies partly corroborate this assumption: Older adults have previously been noted to upregulate cognitive-control-related areas such as the anterior cingulate cortex (ACC) and insula during speech comprehension tasks (Eckert et al., 2008; Harris et al., 2009). Elevated activity in this cingulo-opercular network has been interpreted as an upregulation of error monitoring systems (Sharp et al., 2006). As hearing loss internally degrades speech (Pichora-Fuller, 2003), we hypothesise that these executive systems might additionally be increased or even exhausted in the hearing-impaired. Clearly, more studies are needed to understand how the relative sensory and executive contributions to neural processing of speech changes with age and age-related hearing loss.

1.3 Overview of the present thesis

The reviewed literature has given an overview of behavioural and neuroimaging studies of adaptation to, and neural processing of, degraded speech. From these studies it becomes clear that rapid perceptual adaptation to noise-vocoded speech is a well-established finding, at least in young normal-hearing adults. However, the mechanisms underlying such adaptive processes are less clear. The present thesis is guided by the hypothesis that adaptation to vocoded speech is closely related to processing of the temporal envelope in the non-speech domain, which should be observable both on the behavioural and the neural level. Our research hypotheses will be described below, but for more detailed descriptions the reader is referred to the introduction of each experimental chapter.

The majority of studies have related perceptual learning of degraded speech to top-down influences, showing that working memory and attention are crucial (e.g., Huyck and Johnsrude, 2012; Wild et al., 2012). Here, we pursue an approach from the "bottom-up" perspective, aiming to quantify how much low-level auditory skills contribute to perceptual adaptation to degraded speech. In particular, we test whether the sensitivity of the auditory system to temporal modulations in an acoustic signal can predict how fast a listener adapts to vocoded speech (Chapter 3). Based on previous brain structural studies (Golestani et al., 2002; Golestani and Zatorre, 2004), we hypothesise that brain morphology can explain some part of the interindividual variability observed for adaptation to degraded speech.

Chapter 4 aims to validate these results from normal-hearing adults in CI patients. The experiment intends to prospectively predict speech recognition success from basic auditory non-speech skills in CI patients. So far, in the clinical routine mostly speech tests or pure-tone audiometry are applied, but no measure of auditory temporal processing is usually assessed. Here, we propose that modulation rate discrimination skills can partly characterise a patient's speech comprehension success and would be a valuable addition to the rehabilitation program, to distinguish good and bad users and shape listening strategies accordingly.

Although some neuroimaging studies have found cortical areas to change with perceptual learning (Adank and Devlin, 2010; Eisner et al., 2010), we hypothesise that subcortical structures play a crucial role in the retuning process when a listener is first confronted with a type of degradation which reduces a specific stimulus dimension such as vocoding. Based on auditory brain stem response studies from the Kraus lab (Song et al., 2011; Chandrasekaran et al., 2012; Strait et al., 2012), perceptual learning of speech should involve subcortical structures, which might have been obscured in previous fMRI studies because of their susceptibility to artefacts. Therefore, in Chapter 5, we use a cardiac-gated MRI acquisition procedure (see Section 2.6.1) with increased sensitivity to subcortical structures. The goal of this chapter is two-fold, however: While we firstly aim to identify the neural plasticity that underlies the behaviourally observed adaptation, the second objective is to quantify the convergence of neural (rather than behavioural as in Chapters 3 and 4) mechanisms that subserve processing the temporal envelope for speech and non-speech.

Speech comprehension abilities decline with age and with age-related hearing loss, but it is unclear how this decline expresses in terms of central neural mechanisms. Therefore, Chapter 6 asks both how adaptation to degraded speech is behaviourally altered in older adults, and how neural speech processing changes with age. Based on the degeneration of the auditory cortex commonly observed in older adults with hearing loss (Peelle et al., 2011; Eckert et al., 2012), we anticipate that older adults are forced to rely on secondary structures supporting executive functions (Eckert et al., 2008) when processing of speech becomes difficult.

Regarding the utmost importance of temporal envelope cues in the speech signal for comprehension (see Section 1.1.4), we lastly ask in a behavioural experiment how changes of these envelopes cues affect speech comprehension directly. In addition to noise-vocoding the speech signal, we add two secondary manipulations: We increase (expand) or decrease (compress) the modulation depth of the speech envelope and incrementally add noise to the signal. This allows us to assess the perceptual consequences of altered envelope cues in quiet and in noise. We expect that modification of the envelope cues distorts the signal and thereby decreases intelligibility. Nevertheless, enhanced (expanded) speech envelope cues have been suggested to aid comprehension under difficult listening conditions, in particular when noise is present (Apoux et al., 2004). Such a facilitatory effect of enhancing speech envelope cues would have important implications for hearing aid algorithms.

In sum, the experimental chapters of this thesis tackle the following questions: Can listener variability in perceptual adaptation to degraded speech partly be explained by cognitive, basic auditory capacities and /or brain morphology (Chapter3)? Can individual differences in speech comprehension success between CI patients be predicted by sensitivity of the auditory system to temporal envelope cues (Chapter 4)? What are the neural systems that support adaptation to degraded speech, and which neural networks become engaged when auditory processing of speech and non-speech becomes effortful (Chapter 5)? How do these speech processing networks change with age and age-related hearing loss (Chapter 6)? Can intelligibility of speech in noise be restored when emphasising the temporal envelope cues of the degraded signal (Chapter 7)?

The methods applied to approach these questions will be comprehensively described in Chapter 2. A critical overview of both the behavioural and brain imaging methods is provided to clarify which empirical questions can and cannot be approached with these methods.

2 General methods[1]

This chapter will give an overview of the methods applied in the present thesis, including the stimulus materials presented, the signal processing of the speech materials, the psychoacoustic and brain imaging methods applied.

2.1 Stimulus material: German SPIN sentences

Sentence material used in Experiments I, III and IV was drawn from a German version of the English speech in noise (SPIN) sentences (Kalikow et al., 1977). We developed a German version of the SPIN sentences based on the same criteria on which the English version was based; as in the English version, the German SPIN sentences were controlled for predictability of the final word (high vs. low predictability) which was always a mono- or bisyllabic noun. Homogeneity in sentence length was achieved by restricting all sentences to five to eight words and nine to eleven syllables. According to the "Leipziger Wortschatz" (http://wortschatz.uni-leipzig.de/) all final words occurred with relatively high frequency in the German language (f ranged from 6–16, where frequency f means that the specified word occurs 2^f times more rarely than the reference word "der").

For the present study, only low-predictability sentences were chosen, in order to control and keep low the influence of semantic cues (Obleser et al., 2007). Thus, the listener had to rely primarily on acoustic properties of the sentence to understand it as opposed to high-predictability items where more semantic information would be available. Typical sentences were "Morgen nehmen wir mal wieder Obst" [Tomorrow we will be taking again some fruit; literal translation] or "Gewöhnlich nutzen wir die Räder" [Usually we use the bikes].

---

[1] Parts of this chapter are derived from the articles by Erb et al. (2012, 2013) and Erb & Obleser (2013).

One hundred and thirty low-predictable sentences were recorded by a female speaker of standard German in a sound-damped chamber. The length of the recorded sentences varied between 1623 and 2759 ms. For the short-term adaptation experiment, 100 sentences were used that had proven unanimously low in predictability in a written cloze test (i.e., sentence completion) with ten (different) subjects. A list of these 100 low predictable German SPIN sentences and their high-predictable counterparts are provided in the appendix.

2.2 Signal processing: Noise-vocoding

Artificial degradation can be used to reduce the acoustic information in a speech signal in a controlled manner, whereby the relative contributions of spectro-temporal cues to speech comprehension can be assessed. Here, we used noise-vocoding to degrade speech materials.

Noise-vocoding removes most of the spectral detail coded in the acoustic signal, while its temporal envelope cues remain largely preserved (Rosen et al., 1999). This is achieved by dividing the original signal into frequency bands and extracting the amplitude envelope, that is, the amplitude fluctuations over time, from each analysis band. The amplitude envelopes are then reapplied to bandpass-filtered noise carriers with matched cut-off frequencies (Fig. 2.1). The fewer frequency bands are used, the poorer the spectral resolution of the resulting speech signal will be, meaning that the speech signal is more distorted and therefore more difficult to recognise. If the cut-off frequencies of the analysis bands and the noise bands do not match, a spectral shift (Shannon et al., 1995; Rosen et al., 1999) is introduced which additionally impedes speech comprehension. Noise-vocoding has been used to simulate cochlear-implant (CI) transduced speech in normal-hearing listeners (Shannon et al., 1995). The frequency bands are thought to correspond to the electrodes of the CI which are inserted into the auditory nerve and stimulate one frequency-specific region of the auditory nerve.

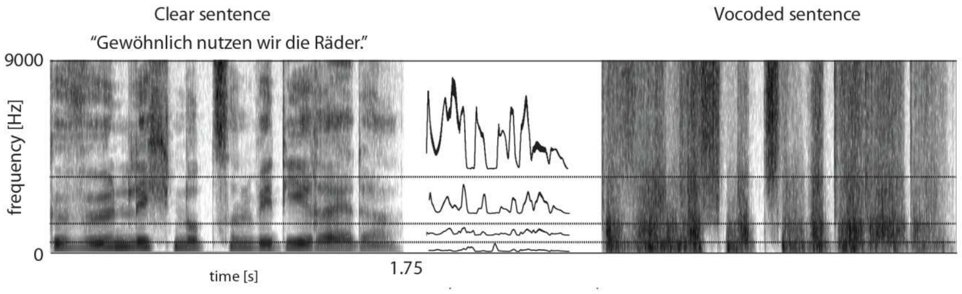


Figure 2.1. Scheme of four-band noise-vocoding. The clear representation of the acoustic signal, in this case the sentence (left) "Gewöhnlich nutzen wir die Räder" [Normally we use the bikes] is divided into four approximately logarithmically spaced adjacent frequency bands spanning 70 to 9000 Hz. The envelope of each frequency band is extracted and reapplied to bandpass-filtered noise carriers with matched cut-off frequencies, thereby effectively smearing the spectral information within a frequency band. The frequency bands are added up again, resulting in a spectrally severely degraded "noise-vocoded" speech signal (left; modified after Davis et al., 2005).

Noise-vocoding was applied to the speech material in MATLAB 7.11. In Experiments I, III and IV, sentences were noise-vocoded using four frequency bands spanning 70 to 9000 Hz. The bands were spaced according to Greenwood's cochlear frequency-position function (Greenwood, 1990). The cut-off frequencies of the analysis bands are specified in Table 2.1. For envelope extraction, we used a 2nd-order, zero-phase Butterworth lowpass filter with a cut-off frequency of 400 Hz. Speech signals were half-wave rectified. Amplitude envelopes were applied to noise carriers with matched cut-off frequencies. Scaling for equal root mean square (RMS) energy was performed channel-wise for each channel envelope. Vocoded and clear versions of the spectrogram of an example sentence are shown in Figure 2.1. In Experiment V, we applied six–band rather than four-band-vocoding and additionally used an envelope modification procedure, which will be described in detail in Section 7.2.2.

Table 2.1. Cut-off and centre frequencies (in Hz) of the four-band noise-vocoded stimuli.

Band	1	2	3	4
Lower cut-off	70	423	1304	3504
Centre frequency	207	764	2156	5634
Upper cut-off	423	1304	3504	9000

2.3 Modelling of learning curves

In order to find a measure for individual perceptual adaptation to vocoded speech over time, we modelled each individual's performance improvement in two different ways: as a power law or as a linear performance increase. It has been claimed that learning follows a power law (Anderson and Schooler, 1991), and power law functions have been used previously to describe short-term perceptual adaptation to time-compressed speech (Peelle et al., 2010a). However, experiments examining perceptual adaptation to vocoded speech have commonly assumed a linear increase in performance over time (Davis et al., 2005).

Here, to test whether a linear or a more complex power law function would better describe the data, both functions were fitted to the individual report scores over time using a least-squares estimation procedure in MATLAB 7.11. The linear trend was described by the function

$$y = a * t + b \tag{1}$$

where y defines performance (proportion words correct), t is time (trial), a is the slope and b the intercept of the fitted function. The power law curve was defined by a function of the shape

$$y = a * t^{m} + b \tag{2}$$

where y defines performance, t is time (trial), m is the learning rate parameter, a the slope and b the intercept of the power law curve. Thus, the power law function has one more parameter than the linear function.

We compared goodness of fit by determining the Bayesian information criterion (BIC; Schwarz, 1978) of the linear and the power law fits within each subject. Note that BIC takes into account also the number of fitted parameters, thereby allowing for a fair comparison between the linear and the power law fit. BIC values were defined as

$$BIC = n * \ln(\sigma_e^2) + \mathrm{k} * \ln(n) \tag{3}$$

where n is the number of observations (n = 100), $\sigma_e{}^2$ is the error variance or sum of squared residuals, k is the number of fitted parameters (k = 2 for the linear fit, k=3 for the power law fit). Smaller

BIC values indicate a better fit (Schwarz, 1978; Priestley, 1981). The BIC increases as a function of σ_e^2 and of k. Models with higher error variance and more fitted parameters are therefore penalised.

2.4 Psychoacoustical methods

Psychoacoustics allow investigating the relationship between physical, acoustic features of a stimulus and perception (Fechner, 1860; Kollmeier et al., 2008). To this end, a stimulus is varied along a physical dimension and a subject's performance is measured as a function of signal level. This allows for the estimation empirical thresholds, that is, the signal level at which a participant is able to detect the presence of a stimulus (detection task) or discriminate the difference in a stimulus (discrimination task), some defined proportion of the time (Macmillan and Creelman, 2005). Here, we used psychoacoustic methods to estimate individual sensitivity to AM rate fluctuations in an acoustic signal.

Psychoacoustic procedures are a combination of response paradigm (which is defined by the stimulus presentation mode, such as the number of intervals presented, and the task) and the method (the measurement strategy; (Marvit et al., 2003)). Two psychophysical methods were applied in the present thesis: The method of constant stimuli, to estimate psychometric curves, and an adaptive staircase method. They were combined with two different response paradigms, a 2-alternative forced choice (2-AFC) AX paradigm and a 3-alternative forced choice (3-AFC) oddball paradigm. Both methods and paradigms will be described below.

2.4.1 Method of constant stimuli

The method of constant stimuli can be used to estimate psychometric functions, which describe sensitivity as a function of stimulus level (Macmillan and Creelman, 2005). Over the experiment, a set of fixed signal levels is presented in a random order and the participant's performance is measured in response to each stimulus.

To build psychometric functions, proportions of correct responses are calculated within a participant for each of the signal levels and averaged over trials. This gives a probability density function of correct responses over levels for each participant. These are then transformed to a cumulative probability density function, to which traditionally a curve of a cumulative normal shape is fitted as an estimate of the psychometric curve (Macmillan and Creelman 2005, Figure 2.2). After z-transformation of this curve, two measures of sensitivity can be derived from such curves, which are expressed in stimulus units: a measure of intercept, which is also called the point of subjective equality (PSE) and a measure of slope, the just noticeable difference (JND) threshold. The PSE is the point where the stimulus is judged on average the same as the standard. The JND is a measure of half the distance between the 25th and 75th percentiles (i.e., the semi-interquartile range, IQR, corresponding to –0.675 and 0.675 after *z*-transformation; Fig. 2.2) of the cumulative probability density function. JND thresholds provide a measure of sensitivity to AM rate, where smaller JND thresholds correspond to a higher sensitivity.

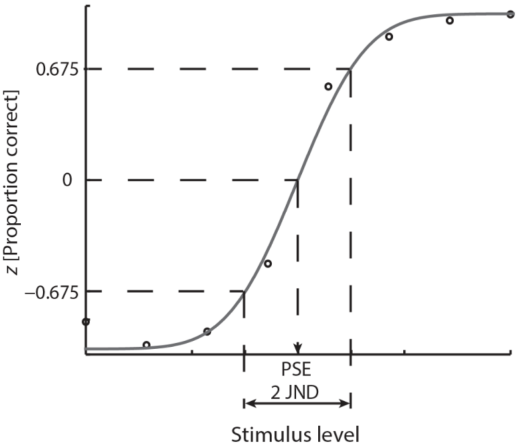


Figure 2.2. Example psychometric function. To estimate psychometric curves, the probability density function of correct responses per signal level is transformed to a cumulative probability density function, to which a curve of a cumulative normal shape is fitted. After z-transformation two measures of sensitivity can be derived the point of subjective equality (PSE) and the just noticeable difference (JND) threshold (modified after Macmillan & Creelman, 2005).

2.4.2 Adaptive methods

A disadvantage of estimating psychometric curves is that it is a time-consuming procedure. In particular, if the approximate psychometric thresholds are unknown in advance of testing (which is particularly difficult in patient populations with a presumably high variability of individual thresholds), data might be collected from points of the psychometric curves that are uninformative of the psychometric threshold. Therefore, Békésy developed adaptive methods (Leek, 2001). Adaptive methods are thought to be more efficient, because they sample only the points of the psychometric curve close to the threshold, but provide less information on the shape of the psychometric curve. Different algorithms are available for adaptive psychophysical testing: While maximum-liklihood methods take into account the complete history of performance (Green, 1993), staircase methods only use performance in the last few trials to decide on the next level to be presented (Levitt, 1971). A third alternative between these two algorithms is the parameter estimation by sequential testing (PEST) approach (Taylor et al., 1983), which adjusts the signal level on every trial. The adaptive staircase method, which was applied in Experiment II, will be described in the following paragraph.

The staircase method uses a computationally simple algorithm. It first presents a stimulus level which is thought to be sufficiently distant from the empirical threshold and therefore easy to detect or discriminate. The signal level (or the level difference between standard and deviant in discrimination tasks) decreases until the participant gives an incorrect response, upon which the signal level (or level difference) increases again, which is referred to as "reversal". The signal level increases, until the participant responds correctly, which in turn triggers another reversal. Thereby, the threshold is approached asymptotically. To estimate the empirical threshold, the levels presented at the ultimate reversals are usually averaged.

Different rules can be defined to decide on the intensity level of the next stimulus of the adaptive track. Commonly, an *n*-up-one-down procedure is applied, where intensity level increases after one incorrect trial and decreases after *n* successive correct trials. Depending on how many correct trials are required to decrease the intensity level, the empirical threshold converges on different points of the psychometric curve: If the level increases after each incorrect and decreases after each correct trial ("One-down-one-up" staircase procedure), the tracking converges on the point of the psychometric function that provides 50% correct responses. A "two-down-one-up" (where intensity level decreases after two successive correct trials) staircase procedure targets the 70% correct point on the psychometric curve, whereas a "three-down-one-up" procedure tracks the threshold that provides 79% correct performance (Levitt, 1971). The step size can be changed from a large one in the beginning, to a smaller one after a fixed amount of reversals, to approach the threshold more quickly. The track will be terminated after a defined number of reversals. For threshold estimation, usually not all trials are used but the stimulus level at the last few reversals is commonly averaged. An example track for a two-down-one-up staircase procedure from Experiment II (see Chapter 4) is shown in Figure 2.3

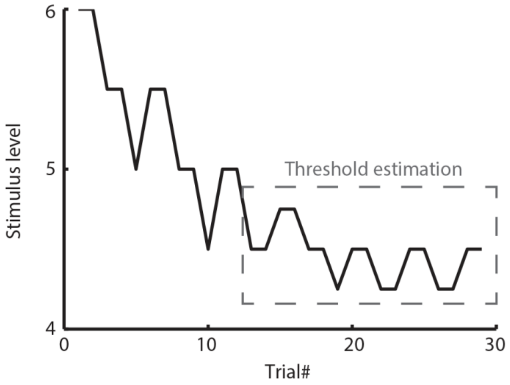


Figure 2.3. Example track for a two-down-one-up staircase procedure (from one participant of Experiment II). Step-size is halved after four reversals, and the procedure is terminated after twelve reversals. For threshold estimation, the stimulus levels at the last eight reversals (marked with a red dotted line) are averaged.

2.4.3 Response paradigms

These psychophysical methods can be combined with different response paradigms which define how many intervals are presented on each trial and which task the participant performs. In an alternative forced choice (AFC) task, the participant is required to select the interval cor-responding to a certain predefined criterion.

For example, in a 2-AFC task, the participant chooses between two options (e.g., “same – different”, “faster – slower”), such that chance performance level is at 50% correct responses. This can take the form of a 2-interval, 2-AFC reminder paradigm, where the standard (of which the signal level is constant) is always presented in the first interval, and the deviant (of which the signal level varies over the experiment) is presented in the second interval. The 2-interval -2AFC reminder paradigm requires identification and verbal labelling of the deviant signal level (e.g., in an AM rate discrimination task, the participant has to indicate, whether the deviant modulation rate was “faster” or “slower”). This paradigm was applied in Experiments III and IV of the present thesis, in combination with the method of constant stimuli.

In the 3-AFC task, the participant chooses between three options, such that chance performance is reduced to 33% correct responses. In the 3-interval, 3-AFC oddball paradigm, a set of three intervals is presented and the participant selects the “odd” interval (the one interval that differs from the other two). Rather than an identification and verbal labelling of the signal level (“faster” or “slower”, see above), this paradigm requires a same-different comparison (e.g., in an AM rate discrimination task, the participant has to indicate, which interval had the different modulation rate). In Experiment I, this paradigm was combined with the method of constant stimuli, and in Experiment II this paradigm was presented as an adaptive staircase procedure.

2.5 Short-term memory tests

Short-term memory is one factor which has been hypothesised to influence degraded speech recognition (e.g., (Rabbitt, 1968), see Chapter 1). Therefore, in the present thesis two widely used short-term memory tests were applied: (1) the digit span (forward and backward; Wechsler, 1987) and (2) a nonword repetition test (based on (Mottier, 1951). Digit span is supposed to measure working memory capacity whereas nonword repetition is thought to test more specifically phonological working memory capacity (Baddeley, 2012). Testing procedures for both measures will be described below.

2.5.1 Auditory forward and backward digit span

The digit span test, which is part of the revised Wechsler memory scale (WMS-R; Wechsler, 1987), is widely used a measure of working memory capacity. Participants listen to series of single digits and are asked to immediately repeat the list of digits in the same order (forward digit span) or in the reverse order (backward digit span). The test has seven levels featuring list lengths starting from three digits increasing by one digit up to nine digits for forward digit span, and list lengths from two to eight digits for backward digit span. Each level comprises two items. The participants' response is marked as correct only if all digits were correctly repeated in the required order. The testing stops when the participant reports none of the items of a level correctly or when all 14 items have been presented. The individual forward and backward digit span measure can either be calculated as the sum of items the participant repeated correctly (the "score" ranging from one to 14; this measure was used in Experiment I) or as the level at which the test was terminated ("level" ranging from three to nine for forward digit span or from two to eight for backward digit span; this measure was used in Experiment IV). Although the level is more commonly used as the individual measure (Wechsler, 1987), in populations with low variability of working memory capacities such as young, healthy adults, it can be advantageous to use the score rather than the level in order to get a more sensitive measure of working memory with higher inter-individual variability.

2.5.2 Auditory nonword repetition test

To specifically test phonological working memory, an extended version of Mottier's test[2] (Mottier, 1951; Welte, 1981) was applied in the current thesis. This test comprises nonwords[3] which each consists of two up to 15 syllables with a consonant-vowel structure, for example, "kapeto". The test has 14 levels, one for each nonword length (counted as the number of syllables) and each level features six items, amounting to a total of 90 nonwords. The first level contains nonwords with a length of two syllables; word length increases by one syllable for each subsequent level.

Participants listen to the nonwords and are asked to repeat back immediately what they heard. Responses are judged online by an experimenter but also recorded for later verification. Only exact repetitions of the nonwords are scored correct. Participants have to report correctly at least four out of six nonwords per level in order to proceed to the next level, otherwise the test is terminated. Nonword repetition test scores are calculated as the total number of correctly repeated items.

2.6 Neuroimaging methods

To study neural mechanisms in humans, few non-invasive methods are available. Electro-encephalography (EEG) was first used in humans in the 1920s, (Berger, 1929) to measure electrical activity from the scalp, which arises from current flows in single neurons. Despite limited spatial resolution, this method has a high temporal resolution in the range of milliseconds. Therefore, it is mainly used in research questions concerning the timing of neural processes. Similarly, magnetoencephalography (MEG) measures the magnetic fields associated with these electric currents with high temporal precision.

More recently, in the 1970s, MRI was first used in diagnostic radiology. Since then, its application in the clinic and research has dramatically increased (Logothetis, 2008). In contrast to EEG, MRI has a poor temporal but high spatial resolution and is therefore used when spatial localisation of neural processes is central for the research question. In the present thesis, both structural and functional MRI

---

[2] The Mottier test, which has also been applied for diagnosing dyslexia, was adapted in-house and kindly supplied by Claudia Männel for testing of healthy adults.

[3] A nonword is a sequence of letters or speech sounds which seems to be a word but is meaningless, i.e., does not belong to the lexicon of a given language.

techniques were applied. The physics of MRI as well as methods to analyse MRI data will be described in the following section.

2.6.1 Principles of structural and functional MRI

The nuclear magnetic resonance phenomenon arises from the magnetic moments of atomic nuclei and their interaction with an external magnetic field (Jezzard et al., 2001). The nuclei of many atoms possess a nuclear moment or spin, that is, they rotate with a certain angular frequency (the Larmor frequency) about an axis.

The spins of protons are randomly oriented, but they align when a strong external magnetic field (referred to as $B_{0;}$ in *z*-direction) is applied. Spins can either align parallel (in the lowest energy state) or anti-parallel (in the higher energy state) to the magnetic field. The imbalance of parallel and antiparallel aligned spins leads to the net magnetisation of the sample.

If a second alternating magnetic field (B_1; a radiofrequency pulse) transverse to the static magnetic field B_0 is applied, the orientation of some spins tilts towards the transverse plane (*x,y* plane). Hence, the magnetisation vector precesses (rotates) about the static magnetic field, which can be detected as resonance signal in a receiver coil placed around the sample. The radiofrequency pulse can tip the magnetisation vectors by different angles, but a signal can only be detected when a net magnetisation exists in the transverse plane (Jezzard et al., 2001).

Echo time. If a 90°-pulse is applied, the spins precess in the transverse plane. Their rotation gets progressively out of phase, however, due to inhomogeneities of the magnetic field (B_0). Therefore, in a spin echo sequence, a second 180° refocusing pulse is applied, which flips the magnetisation about the x-axis, thereby refocusing the phases of the spins and increasing the detectable signal. The time between excitation and data acquisition is termed echo time (TE).

Acquisition time. The period of time required to collect the data is termed acquisition time (TA).

Repetition time. After the system has been perturbed by a radiofrequency pulse, the spins need some amount of time to realign with the magnetic field B_0 and the next radiofrequency pulse can be applied only after a certain delay. This time is referred to as the repetition time (TR).

T1 relaxation. To restore the equilibrium, two relaxation processes take place: T1 and T2 relaxation. T1 relaxation denotes the decay of the longitudinal magnetisation back to its equilibrium and is generally measured in structural MRI. T1 depends on the local molecular motion, which varies in different parts of the brain. For example, cerebro-spinal fluid (CSF) has a slow T1 due to the high amount of freely moving water molecules, whereas white matter has a short T1. Thus, T1 relaxation it can distinguish between different tissues (e.g., CSF and white matter).

T2 relaxation. T2 relaxation refers to the decay of the transverse magnetisation back to its equilibrium. It is generally measured in fMRI. T2 is affected by slow motions of the neighbouring molecules which reduce the coherence of the spins in transverse direction. T2* relaxation also denotes transverse relaxation, but is caused by local variations in the magnetic field. For example, in the brain, the local magnetic susceptibility rapidly changes in regions adjacent to blood vessels, due to the changing levels of paramagnetic deoxyhemoglobin. This effect is exploited in the blood-oxygenation-level-dependent (BOLD) contrast (see below).

Spatial localisation. Localisation of the signal is possible due to small magnetic field gradients that are superimposed on B_0. Because the resonance frequency is proportional to the magnetic field strength, the resonance frequency of the protons differs depending on their localisation in these gradients due to small differences in magnetic field strength. Therefore, the spatial position of the protons along the magnetic field gradients is coded by the differences in resonance frequency (Jezzard et al., 2001).

The haemodynamic response. The BOLD signal relies on the different magnetic properties of oxygenated and deoxygenated haemoglobin (Ogawa et al., 1990b; Ogawa et al., 1990a; Turner et al., 1991). In proximity of an activated neuronal population, the regional cerebral blood flow increases, to supply oxyhaemoglobin. This leads to an increase in of diamagnetic oxyhaemoglobin (of which the consumption is slower than its supply) and a decrease of paramagnetic deoxyhaemoglobin concentration, which in turn results in an increased signal due to fewer spin dephasing (Kim, 2005).

Critically, the BOLD contrast does not measure neuronal activity directly, but captures haemodynamic changes occurring when a population of neurons is firing (Logothetis, 2008). Although the exact physiology related to the haemodynamic response is unknown, the BOLD signal appears to be a consequence of increased energy demands due to synaptic activity (Arthurs and Boniface, 2002). Local field potentials from intracortical recordings have been shown to correlate with the haemodynamic response (Logothetis et al., 2001). The coupling between neuronal activation and increased blood flow depends on various factors, however, and changes with age (D'Esposito et al., 2003) or the use of some drugs, for example (Brown et al., 2003; Liu et al., 2004).

The BOLD response is lagging behind the actual neuronal activation. Depending on the cortical area, it peaks approximately 5s after stimulus presentation. The haemodynamic response in the auditory cortex is relatively fast, peaking approximately 4 s after stimulus onset (Belin et al., 1999). As the time to acquire a volume (acquisition time, TA) is commonly 2 s and can at best decrease to 1 s in high-field MR imaging, the temporal resolution of fMRI is limited (Jezzard et al., 2001).

Sparse sampling. For auditory paradigms, fMRI can be problematic, because the imaging system generates acoustic noise during gradient switching which can amount up to 120 dB SPL in some sequences (Jezzard et al., 2001), which can interfere with auditory stimulation. One way to overcome this problem is to use a sparse-sampling acquisition (Hall et al., 1999). A volume is acquired only every few seconds, meaning that the time between two successively acquired volumes (TR) is longer than TA. Hence, silent periods between scan acquisitions allow for playing stimuli and recording responses in quiet. As the BOLD response requires approximately 5 s to peak, the maximal BOLD response to stimulus presentation can nevertheless be measured the subsequent scan.

Cardiac gating. MR images are highly susceptible to motion artefacts. In particular, heartbeat related movement can cause strong artefacts in subcortical structures. Cardiac gating (von Kriegstein et al., 2008) can partly overcome this issue by synchronising scan acquisition with the heartbeat. Participants' heartbeat is monitored via a pulse oximeter and the scanner waits for the heartbeat to trigger volume acquisition.

2.6.2 Analysis of functional MRI data

Preprocessing. Prior to statistical analyses, the raw functional MR images are generally subjected to a number of preprocessing steps in order to increase the signal-to-noise ratio.

Head motion is a major source of noise in MRI data. Motion correction can be achieved by aligning the time series of volumes with respect to each other using a rigid-body transformation (rotation and translation; Jezzard et al., 2001). Further, differences in brain tissues can cause local inhomogeneities of the magnetic field, in particular close to the air-field sinuses. Those can be corrected for when a fieldmap is available which codes for the strength magnetic field across space. To transform individual data into a standard anatomical space, the functional images are typically first coregistered with the individual anatomical scan. The high-resolution structural scan can then be used to determine the parameters for coregistration with a template which is usually an average of many brains (e.g., Montreal Neurological Institute [MNI] average brain, Collins et al., 1994). Such a normalisation procedure enables the comparison of brain activation patterns across individuals despite the high interindividual variability in brain morphology. Nevertheless, some have argued that high interindividual anatomical variability results in a poor alignment of anatomical features (e.g. locations, shapes and sizes of the gyri and sulci), such that displaying individual activation maps on individual brains would be preferable (e.g., Fedorenko et al., 2012). To compensate for interindividual anatomical variability which cannot be dealt with by normalisation, spatial smoothing is commonly applied as part of the preprocessing pipeline. A Gaussian filter is often used where the kernel width in mm at full-with-at-half-maximum (FWHM) must be specified, for which a value of approximately 8 mm is widely used. However, the optimal size of the smoothing kernel depends on the data and anatomical regions of interest and can only be determined empirically (Hopfinger et al., 2000). Smoothing accounts for the spatial correlation of the fMRI data and increases the signal-to-noise ratio. It also decreases the number of independent resolution elements (“resels”), which have to be taken into account in a correction of multiple comparisons (see below). However, a disadvantage of smoothing is the reduction of spatial precision, which can result in detection of a significant signal in regions where there is none (e.g., white matter).

Statistical analyses. After the preprocessing, images are statistically analysed. The time series of each voxel (i.e., the 3-dimensional pixel of which an MR image is constituted) can be analysed separately in univariate analyses, or the data can be analysed jointly in multivariate analyses, which consider the spatial relationships between voxels (Haynes and Rees, 2006).

A univariate way of analysing the data is the use of a General linear model (GLM; Friston et al., 1994), where a model is fitted to each voxel’s time course. The GLM is defined as

$$y = \beta_1 * x_1 + \beta_2 * x_2 + \cdots + \beta_n * x_n + c + e \tag{4}$$

where y are the actual data, containing the intensity value of a given voxel at each time point (TR). x is the model which codes for a condition (e.g., whether stimulation was on or off). β is the parameter estimate, that is, the value the model must be multiplied by to best fit the data. c is a constant which corresponds to baseline activity and e is the residual error between the model and the data.

Hence, if a certain voxel reacts strongly to a condition modelled in x_1, a large value will be assigned to β_1 in the model-fitting. The parameter estimate is converted into a T-value by dividing it by its standard error for every voxel such that a whole-brain T-statistic map is obtained. To decide, which parts of the brain are activated, a threshold is commonly applied. As approximately 100 000 voxels are compared in a brain, if a conventional threshold of $\alpha < 0.01$ is accepted, the number of false positive will amount to 1000 voxels. Different methods are available to correct for such inflated type-I-error (i.e., accepting voxels as activated when they are actually not) at the whole-brain level, which control for family-wise error rate or which take into account the cluster-extent of thresholded voxels.

2.6.3 Analysis of structural MRI data

Voxel-based morphometry (VBM) is an exploratory whole-brain technique to examine local micro-structural brain tissue properties based on a voxel-wise analysis of brain tissue. To compare individual differences in brain anatomy, spatial alignment must first be achieved between individual scans. The protocol developed by Ashburner and colleagues (Ashburner and Friston, 2000) first segments images into different brain tissues, typically grey matter, white matter and cerebro-spinal fluid (CSF). To this end, each voxel's grey value is assigned a probability of belonging to one of these tissue classes. In order to create a group-specific tissue template (generally of grey and white matter), the diffeomorphic-anatomical-registration-using-exponentiated-lie-algebra (DARTEL) algorithm (Ashburner, 2007) estimates the deformations that best align the segmented individual images onto each other by iteratively registering the images to their average. Six rounds of iterative averaging are typically applied. The deformation estimations are then used to spatially normalise the grey and white matter images to MNI space. Subsequently, as for the functional scans, the images are then normally smoothed (see Section 2.6.2).

The final pre-processed images are modulated, that is, adjusted for the total volume of grey matter identified in the normalisation procedure. Therefore, the voxels of these images represent a measure of regional volume (Ashburner, 2009). Modulated images can be statistically analysed in a multiple regression analysis in the context of the general linear model (see above). However, before statistical assessment, it is recommended to correct for the nonstationary smoothness of the data: areas with a low intersubject variance will have a comparably low spatial smoothness whereas areas with high individual variability will have a high spatial smoothness which increases the probability of detecting significant results. It is possible to correct for such non-stationary smoothness using a non-stationarity correction (Hayasaka et al., 2004).

3 Experiment I: Listener variability in adaptation to degraded speech[4]

3.1 Introduction

Normal-hearing listeners learn to understand noise-vocoded speech surprisingly rapidly when first exposed to this type of degradation, in which the spectral information is severely degraded while temporal envelope cues are largely preserved (see Section 2.2; Davis et al., 2005; Eisner et al., 2010). However, there is considerable variability across listeners in their ability to perceptually adapt to this degraded auditory input. Although there are larger individual differences in speech perception performance in CI users than in normal-hearing adults listening to a CI simulation, normal-hearing listeners still differ considerably in learning to understand vocoded speech (Shannon et al., 2002). Currently it is unclear what drives the adaptation to this degraded speech signal. Demographic factors (e.g., age at implantation, duration of deafness, residual hearing) can only partly explain the variance in speech recognition success (Blamey et al., 2001; Sarant et al., 2001).

Phonological working memory is one cognitive factor that has been linked to the development of speech recognition abilities (Gathercole et al., 1994). According to the model proposed by Baddeley and colleagues, working memory relies on two processes: the “buffer”, which holds memory traces for a few moments, and the “phonological loop”, which can be imagined as a subvocal rehearsal process (Baddeley, 2012). Pisoni and colleagues have investigated the relationship between phonological working memory and individual speech recognition performance in CI users, and found that working memory scores, as measured by digit span, significantly correlated with spoken word recognition

---

[4] This chapter is adapted from the article published in Neuropsychologia by Erb, Henry, Eisner and Obleser (2012).

scores in pediatric CI users, even after statistical partialling out of demographic factors (Pisoni and Cleary, 2003).

In the present study, we intended to examine whether we could replicate these results in normal-hearing adults listening to vocoded speech. Digit span requires the participant to hold online representations of perceived auditory information, taxing the “buffer” in Baddeley’s working memory model. However, as digit span demands memorizing of highly familiar items that have a long-term semantic representation, it may not be a reliable test of phonological working memory (Jacquemot and Scott, 2006), especially if semantic and phonological representations have separable stores as some patient studies suggest (Romani and Martin, 1999). In order to test more specifically phonological working memory, we therefore additionally employed a nonword repetition test in the current study. Degraded speech recognition challenges phonological analysis and requires the maintenance of auditory input in the working memory for a short period of time. Thus, we expected that both tests of working memory, digit span and nonword repetition, should independently explain some of the variability associated with vocoded speech perception.

As noise-vocoded speech has very poor spectral resolution, listeners are forced to rely more on temporal envelope cues for speech recognition. The utilisation of these cues might be easier for listeners with a higher sensitivity to envelope fluctuations in an auditory signal. We therefore hypothesised that the individual capability to adapt to vocoded speech is related to AM rate discrimination thresholds. Amplitude modulation rate discrimination performance provides a measure of the auditory system’s ability to encode temporal information in a waveform’s envelope (Wakefield and Viemeister, 1990). The most prominent low-frequency amplitude modulations in the temporal envelope of speech are near the syllabic rate of approximately 4 Hz (Steeneken and Houtgast, 1980; Houtgast and Steeneken, 1985). As these low AM frequencies are essential for speech perception, we decided to test listeners in their sensitivity to AM rates centred on 4 Hz. In the present study, we deliberately chose to assess sensitivity to AM rate rather than AM depth, which is traditionally used to determine temporal modulation transfer functions (Viemeister, 1979). We assessed thresholds to modulation rate instead, as it captures the temporal aspect of envelope encoding which is emphasised when listening to vocoded speech.

Hence we anticipated that, complementarily to phonological working memory scores, auditory processing skills that are not speech-specific (i.e., AM rate sensitivity) could account for part of the individual differences in perceptual learning of vocoded speech.

A small number of studies have investigated the relationship between brain activation patterns and perceptual adaptation to degraded speech. Notably, Adank and Devlin (2010) conducted an fMRI study of adaptation to time-compressed speech. They observed adaptation-related activity bilaterally in the auditory cortex (posterior superior temporal sulcus; STS) and in the left ventral premotor cortex. The authors concluded that perceptual learning of degraded speech involves mapping novel acoustic patterns onto articulatory motor plans (Adank and Devlin, 2010).

Using a short-term vocoded speech-learning paradigm, Eisner et al. (2010) found that individual performance improvement in vocoded speech perception was correlated with the haemodynamic signal change in the IFG. Additionally they observed a positive correlation between activity in the angular gyrus and individual learning curves over the course of 100 noise-vocoded sentences (but did not have continuous measures of performance throughout the scan period).

Golestani and Zatorre (2004) explored the structural brain correlates of novel speech sound learning. They found that faster phonetic learners had increased white matter density in the parieto-occipital sulcus. Though, to our knowledge, structural correlates of vocoded speech learning have not been investigated so far. Here, we examine whether it is possible to predict perceptual learning of vocoded speech from the brain structure of a listener.

The present study tested normal-hearing adults in a short-term (ca. 20 minutes long) vocoded speech-learning paradigm. A set of tests including digit span, nonword repetition and AM rate discrimination were administered. Additionally, we analysed structural MRI scans using VBM. The aim of the study was to elucidate whether and to what extent cognitive, non-speech auditory, and neuroanatomic measures could predict how well a listener perceptually adapts to vocoded speech.

3.2 Materials and methods

3.2.1 Participants

Eighteen participants (9 females, age range 22–30 years, mean 26.2 years) took part in this study. Participants were recruited from the participant database of the Max Planck Institute for Human Cognitive and Brain Sciences. All were right-handed, monolingual speakers of German with no known hearing or language impairments or neurological problems. They were naïve to noise-vocoded speech and had not participated in a vocoded speech perception experiment before. Participants received financial compensation of € 7, and informed consent was obtained from all participants. Procedures were in accordance with the guidelines of the local ethics committee (University of Leipzig).

3.2.2 Stimuli and experimental design

Vocoded speech perception. Sentence material was drawn from a German version of the English SPIN sentences (Kalikow et al., 1977). Only low-predictability items were chosen. The sentences were spectrally degraded using noise-vocoding. A comprehensive description of the sentence corpus and the noise-vocoding procedure is provided in Section 2.1; the German SPIN sentences are listed in the appendix.

One hundred low-predictable four-band noise-vocoded German SPIN sentences were presented by a personal computer through Sennheiser HD201 headphones using Presentation software. On each trial, participants heard a sentence which was followed by visual presentation of a green light ("go" signal for response) on the screen lasting for 3 seconds. During this time, participants were to respond by repeating the sentence, which was recorded for later off-line scoring. As the experiment was designed to be compatible with a future fMRI sparse sampling experiment, participants heard a recording of scanner noise at the end of each trial. For the same reason, we interspersed the experiment with twenty silent trials lacking any auditory presentation. In sum, the experiment comprised 120 trials lasting 9 seconds each.

Vocoded speech perception performance was evaluated from recordings of participants' responses. Responses were scored as proportions of correctly repeated words per sentence. Scoring comprised all words of a sentence including function words. The marking scheme was liberal such that errors in declension or conjugation were accepted as correct.

Measure of individual adaptation to vocoded speech. To establish that participants learned to understand vocoded speech, that is that their performance improved over the course of the experiment, the experiment was divided post-hoc into two blocks of 50 trials each, which were compared using a paired-sample *t*-test.

In order to find a measure for individual perceptual learning of vocoded speech over time, we modelled each individual's performance improvement as a power law or as a linear performance increase using a least-squares estimation procedure in MATLAB 7.11. We compared goodness of fit by determining the Bayesian information criterion (BIC; Schwarz and Gideon, 1978) of both fits for each participant (for a detailed description of the fitting procedures, see Section 2.3).

Working memory tests. All participants underwent two short-term memory tests: Forward and backward digit span test (Wechsler, 1987; see Section 2.5.1), which is thought to measure short-term memory capacity, and an extended version of Mottier's test (Mottier, 1951; see Section 2.5.2), which measures phonological working memory. For both tests, stimuli had previously been read out and recorded by a female speaker. Participants listened to these recordings over headphones. The individual forward and backward digit span scores were calculated as the sum of all items the participant had repeated correctly ("scores").

Amplitude modulation rate discrimination. Participants were also tested in an AM rate discrimination task. Stimuli were one-second long sinusoidally amplitude-modulated white noises. The standard stimulus was modulated at 4 Hz. Deviant stimuli were modulated at 8 different rates that were linearly spaced between 2 and 6 Hz in steps of 0.57 Hz (hence deviant stimuli were centred around 4 Hz) with a constant modulation depth of $m = 60\%$. The onset phase of the sinusoidal modulation was randomly varied for all stimuli. Stimuli were presented with an inter-stimulus interval of 500 ms.

Participants listened to stimuli on a PC in MATLAB 7.11 via Sennheiser HD-201 headphones at approximately 60 dB SPL. The paradigm was a 3-alternative forced choice (3-AFC) task: In each trial, participants heard two standard stimuli and one deviant sound; the position of the deviant within the trial was randomly varied from trial to trial. Participants' task was to indicate at which position they heard the sound with the different modulation rate (faster or slower). Participants responded via keyboard by pressing the keys "1", "2" or "3" and immediately received feedback after each trial with "right" or "wrong" appearing on the screen. On the whole, every participant listened to 160 trials, such that each deviant modulation rate level was presented 20 times in a random order.

Just-noticeable-difference thresholds in AM discrimination. As a measure of individual sensitivity to AM rate, we derived JND thresholds from psychometric curves as described in Section 2.4.1 (Macmillan and Creelman, 2005). Just-noticeable-difference thresholds provide a measure of sensitivity to changes in AM rate, where smaller JND thresholds correspond to a higher sensitivity. Thresholds are reported as absolute deviation from the standard (4 Hz) in Hz. Due to experimenter error, JNDs for two participants were calculated based on only 60 (rather than 160) trials (i.e., 7 or 8 trials per deviant AM rate).

As for the vocoded speech learning task, in order to test for performance increases over time, data were divided post-hoc into two blocks of 80 trials each and compared using a paired-sample *t*-test. The two previously mentioned participants were excluded from this analysis, because the time scale of their performance was not comparable to the one of the remaining participants (60 versus 160 trials).

3.2.3 Experimental procedure

In order to maximise the comparability of performance between individuals, all participants were tested in the same order as follows: Digit span, nonword repetition, vocoded speech perception, AM discrimination.

3.2.4 MRI data acquisition and analysis

Structural brain scans of all eighteen participants were available through the Institute's brain data base. Scanning was carried out on a 3T Siemens Trio TIM scanner. A T1-weighted MPRAGE sequence was used to acquire 176 sagittal slices, with an acquisition matrix of 256 x 240, yielding a resolution of 1mm^3. Repetition time was 13 ms and echo time was 3.46 ms (n = 15, n being the number of participants) or 3.97 ms (n = 3). For acquisition of the images either a birdcage head coil (n = 3), a 12-channel head coil (n = 7), or a 32-channel head coil (n = 8) was used. The scans were between 1 and 62 months old (mean ± SEM = 20.6 ± 4).

To examine the relationship between brain structure and adaptation to vocoded speech, we used VBM running in SPM8 (Wellcome Department of Imaging Neuroscience, London, UK). As described in Section 2.6.3, we first segmented images into grey matter, white matter and cerebro-spinal fluid (CSF) and created a group-specific grey and white matter template using the DARTEL algorithm (Ashburner, 2007). The images were then smoothed with an isotropic Gaussian kernel of 10 mm FWHM. The final pre-processed images were modulated, that is, adjusted for the total volume of grey matter identified in the normalisation procedure (see Section 2.6.3).

Modulated images were statistically analysed in a multiple regression analysis in the context of the general linear model. We set up separate models where either individual adaptation slopes for learning of vocoded speech or AM discrimination thresholds were entered as covariates of interest. As these covariates of interest were highly correlated, we preferred not to enter them into the same model to avoid that shared variance would lead to uninterpretable results. In order to control for the variance in sex, age, coil of acquisition and time since acquisition (in months) which might confound our results, we added to each model four nuisance variables coding for these factors. In both analyses, we corrected for total grey matter volume (GMV) using global normalisation, because we were not interested in the absolute amount of grey matter, but in how regional GMV compares to elsewhere in the brain. That is, we tested the hypothesis that amount of local GMV relative to total GMV was predictive of the behavioural effects of interest (i.e., adaptation slope or AM discrimination threshold). We used this approach to assure not to confound our results with existing age or sex differences in total GMV (Peelle et al., 2012).

The results were corrected across the whole brain based on cluster extent. Corrections based on cluster extent can be problematic in VBM studies due to the potential nonstationarity of the data, but it is possible to use a non-stationary cluster extent correction (Hayasaka et al., 2004). Therefore we corrected the cluster extent threshold for non-isotropic smoothness using the vbm8-toolbox (developed by Christian Gaser; available at http://dbm.neuro.uni-jena.de/vbm/download/).

In addition to the whole-brain analysis, we also conducted a region of interest (ROI) analysis based on the results by Eisner, et al. (2010). We used the peak coordinate of their correlation between individual learning scores and fMRI signal in the left IFG (MNI-coordinates: -46 26 20) to define an ROI (radius: 3 mm, ROI was built using MarsBar toolbox [Brett et al., 2002] running in SPM8). Results were displayed in MRICRON on the ch2 template in MNI space (Rorden and Brett, 2000).

3.3 Results

3.3.1 Adaptation to vocoded speech

First, we established that participants exhibited performance increases in vocoded speech recognition over the course of the experiment. A paired-sample *t*-test showed that participants' mean report scores in the first half of the experiment were significantly different from the ones in the second half ($t(17) = -7.64$, $p < 0.001$; see Fig. 3.1A). Another index for participants' overall performance improvement over time is the correlation between mean report scores with sentence position within the experiment (“sentence number”), which was also significant (Pearson’s $r = 0.61$, $p < 0.001$, Fig. 3.1B).

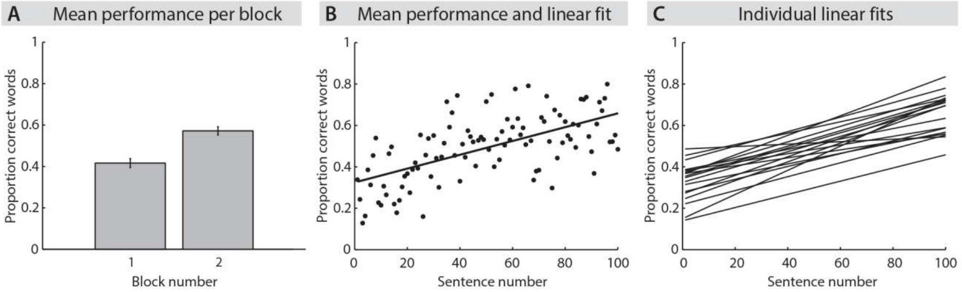


Figure 3.1. Perceptual adaptation to vocoded speech. (**A**) Average performance in the first and second half of the experiment (indicating that learning took place), (**B**) linear fit to the average performance over all trials and (**C**) linear fits to individual speech perception scores, showing individual differences in adaptation to vocoded speech.

To measure individual learning rate, we fitted a linear and a power law curve to each participant’s performance over time (see Section 2.3). To determine which model better described the data, we calculated BICs for the two fits within each participant. A Wilcoxon signed-rank test showed that the BICs of the linear fit (median 239.99, range 223.13–252.73) were significantly smaller than those of the power law fit (median 243.57, range 227.14–261.18; $p < 0.001$). Smaller BIC scores indicate a better fit. Thus, the slope of the linear fit was chosen as an estimate for all individual adaptation rates; in the following analysis we will refer to this simply as “adaptation slope”. Table 3.2 shows the descriptive statistics associated with this perceptual learning measure.

As expected, there was considerable variability in adaptation slopes across listeners; this is easily visible from Figure 3.1C where the linear fits for all participants are shown.

Furthermore, the slopes of the linear fits correlated negatively with their intercepts (Pearson’s $r = -0.69$, $p < 0.01$), indicating that listeners who started out with a good performance improved less over the course of the experiment.

3.3.2 Working memory scores

Forward digit span and backward digit span scores were calculated as the sum of all correctly-repeated items of each test. Similarly, scores for the nonword repetition test were evaluated as the sum of all correctly-repeated items. Descriptive statistics of forward, backward digit span and nonword repetition scores are shown in Table 3.1.

3.3.3 Amplitude modulation rate discrimination

Just-noticeable-difference thresholds for AM discrimination performance were determined as described in Materials and methods. Table 3.1 shows descriptive statistics for JND thresholds. There was no difference in mean performance between the first (mean ± SEM = 0.768 ± 0.018) and the second half (0.767 ± 0.016) of the experiment as shown by a paired samples t-test ($t(15) = 0.08$, $p = 0.95$), indicating that participants' performance did not improve over time in the AM discrimination task.

Table 3.1. Descriptive statistics for test scores across participants.

	Mean	SD	Min	Max	IQR
Adaptation slope	0.0033	0.0012	0.0006	0.0058	0.0012
AM discrimination (JND in Hz)	1.3	0.43	0.66	2.03	0.71
Forward digit span (score)	9	2.4	4	14	4
Backward digit span (score)	8.39	1.97	5	11	3
Nonword repetition test	25.61	3.68	17	30	6

AM: amplitude modulation; JND: just-noticeable-difference threshold; SD: Standard deviation; Min: minimum; Max: Maximum; IQR: Inter-quartile range.

3.3.4 Pearson's correlations

In order to examine whether other test scores could predict adaptation slope, a two-tailed Pearson's correlation was calculated for all pairs of test scores. There was a significant negative correlation between adaptation slope and AM discrimination thresholds, confirming that participants who adapted faster to vocoded speech also performed better in the AM discrimination task. AM discrimination thresholds also correlated negatively with forward as well backward digit span, revealing that better AM discrimination performers had a higher digit span. No significant correlation was found between other pairs of test scores (all correlation coefficients can be found in Table 3.2; a scatterplot of the correlation between AM discrimination thresholds and adaptation slope is shown in Fig. 3.2A).

Table 3.2. Pearson's correlation coefficients for two-tailed correlations between behavioural test scores.

	Adaptation slope	AM discrimination	Forward digit span	Backward digit span	Nonword repetition
Adaptation slope		-0.515*	0.404	0.234	0.145
AM discrimination			-0.727**	-0.526*	-0.262
Forward digit span				0.304	0.417
Backward digit span					-0.213
Nonword repetition					

* $p < 0.05$, ** $p < 0.01$, no indication = non significant.

3.3.5 Partial correlations

To analyse more closely the relationship between adaptation slope, AM discrimination and forward digit span, we calculated partial correlation coefficients between these three variables of interest (Fig. 3.2B). When partialling out forward digit span, the correlation between adaptation slope and AM discrimination threshold dropped to a non-significant trend; AM discrimination and forward digit span remained highly significantly correlated when partialling out adaptation slope.

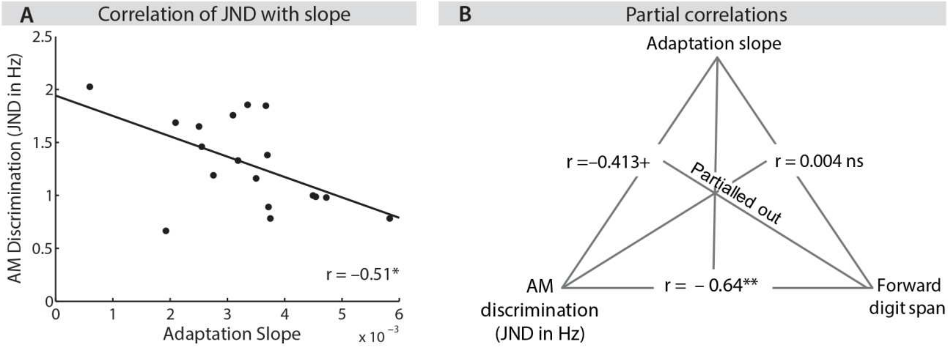


Figure 3.2. Pearson's correlations with adaptation slope. (**A**) Adaptation slope and amplitude modulation (AM) rate discrimination threshold correlate significantly. (**B**) Partial correlation coefficients between adaptation slope, AM discrimination threshold and forward digit span. JND: just-noticeable-difference threshold, ** $p < 0.01$, * $p < 0.05$ level, + $p < 0.1$, ns = non significant.

3.3.6 Voxel-based morphometry results

To identify brain areas predictive of individual adaptation to vocoded speech, multiple regression analysis of the VBM pre-processed modulated grey matter images with adaptation slope as covariate of interest and four nuisance variables coding for sex, age, coil of acquisition and time since acquisition was carried out. Regional volume correlated positively with adaptation slope in the left thalamus (pulvinar), indicating that better learners had increased GMV in the pulvinar. On the other hand, poorer learners had larger GMV in right-lateralised middle frontal gyrus and inferior temporal gyrus, as shown by a negative correlation between regional volume and adaptation slope (see Fig. 3.3 and Table 3.3).

As AM discrimination thresholds were predictive of individual adaptation to vocoded speech at the behavioural level, we wanted to test whether there was an overlap in brain structures predictive of both measures. The whole-brain multiple regression analysis was repeated with AM discrimination thresholds as covariate of interest. We found a positive correlation between AM discrimination thresholds and GMV in the left superior parietal gyrus, meaning that poorer performers had increased GMV. A negative correlation was found in the occipital pole and posterior cingulate gyrus, signifying that better AM discrimination abilities were associated with higher GMV in these regions. There were no overlaps between areas predictive of adaptation slope and AM discrimination rate.

In addition to the whole-brain analysis, we performed an ROI-analysis in the left IFG, where Eisner, et al. (2010) had found a higher fMRI signal during vocoded speech perception in better learners. There was a non-significant trend for a positive correlation between mean GMV in the left IFG (at -46 26 20) and adaptation slope ($z = 1.47$, $p = 0.075$).

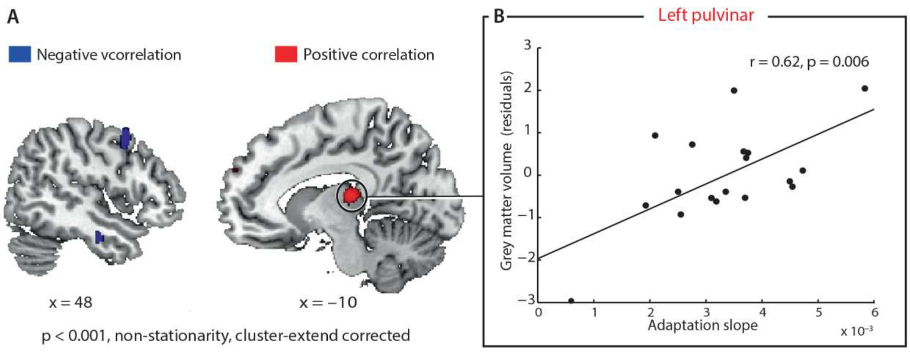


Figure 3.3. Results of voxel-based morphometry analysis. (**A**) Better learners had increased grey matter volume in left thalamus whereas poorer learners had increased GMV in right-lateralised middle frontal and inferior temporal gyrus. (**B**) Scatterplot of the grey matter (studentised) residuals (having removed the effects of all covariates of no interest) and adaptation slope.

Table 3.3. Summary of VBM results.

Location	MNI-Coordinates			*Z*-score	Extent (mm^3)
Adaptation to vocoded speech (slope)					
Positive correlation					
L thalamus (pulvinar)	-12	-24	12	3.56	280
Negative correlation					
R middle frontal gyrus	48	9	51	3.92	152
R inferior temporal gyrus	48	-11	-25	3.68	66
AM discrimination threshold (JND)					
Positive correlation					
L superior parietal gyrus	-15	-47	63	3.71	141
Negative correlation					
L occipital pole	-26	-102	-3	4.07	137
L posterior cingulate gyrus	-12	-44	5	3.47	85

JND: just-noticeable-difference threshold, L: left, R: right.

3.4 Discussion

In the present study, we tested individual short-term adaptation to degraded (i.e., noise-vocoded) speech. We hypothesised that individual differences in adaptation to vocoded speech should be predictable by non-speech auditory, cognitive, and neuroanatomical factors. The main finding of the present study is the predictive potential of AM rate discrimination ability and morphology of the left thalamus in individual perceptual adaptation to degraded speech.

3.4.1 Perceptual adaptation

Perceptual adaptation to degraded speech affords a robust understanding of speech in adverse listening situations. Our results corroborate that perceptual learning of vocoded speech is possible in a short period of time. It has been shown that various other forms of degraded speech or speech that deviates from the norm such as unfamiliar accents (Bradlow and Bent, 2008), speech in noise (Song et al., 2011), and even more artificial forms of degradation such as time-compression (Dupoux and Green, 1997; Pallier et al., 1998) lead to fast perceptual adaptation.

As expected, we found considerable individual variability in perceptual learning of vocoded speech. Nonetheless, all participants were far from ceiling performance after listening to 100 sentences of four-band vocoded speech. The finding that the learning curves in this experiment were better described by a linear than a power law fit reflects the fact that performance changes at a constant rate rather than changing at different rates throughout the learning process (the latter being reflected by a power law curve).

The observed negative correlation between slope and intercept of the linear fits suggests that participants starting better had a more limited scope to improve. This corroborates previous findings that poorer baseline performance is correlated with greater learning in auditory tasks (Amitay et al., 2005) and perceptual learning of speech (Stacey and Summerfield, 2007a).

Despite demonstrations that feedback can greatly enhance the speed of vocoded speech learning (Davis et al., 2005; Hervais-Adelman et al., 2008), we showed in the present study that fast perceptual learning of vocoded speech is still achieved in the absence of feedback.

In contrast to perceptual learning of vocoded speech, no performance improvement over time was observed in the AM discrimination task. This suggests that sensitivity to AM rate is rather a pre-existing, non-improvable feature. Alternatively, exposure to the paradigm during 20 minutes only might not be an appropriately long period to achieve performance improvement. The latter view is supported by a study on AM rate discrimination learning where participants' performance improved after training with 720 trials per day for 6-7 days (Fitzgerald and Wright, 2011).

3.4.2 Working memory in perceptual adaptation

Contrary to our expectations, nonword repetition test scores did not correlate with any of the other measures. This result could reflect a reality, indicating that adaptation to vocoded speech is independent of phonological working memory; or it may simply reflect the inability of the test to capture phonological working memory. The nonword repetition test had initially been designed for diagnostics of dyslexic children (Mottier, 1951); it was adapted in-house for phonological working memory testing in adults by extending the upper limit to a list length of 15 syllables. Yet, this extended version remains to be validated carefully in the normal-hearing adult population.

In contrast to Pisoni and Cleary (2003), we did not find a correlation between forward digit span and vocoded speech learning. Nevertheless, digit span did explain part of the common variance between adaptation slope and AM discrimination threshold as shown by the decrease in correlation strength when partialling out forward digit span (Fig. 3.2B).

The role of working memory in degraded speech perception is an important issue. Rabbitt (1968) suggests with his “effortful hypothesis” that working memory becomes a limiting factor when perceptual effort is required for degraded speech recognition. He observed that masking a word (with noise) makes not only the word itself more difficult to recall, but also the clearly-presented word prior to it. This is proposed to be due to the fact that resources are drawn on to accomplish “perceptual effort” that would otherwise be available for storage in working memory.

Based on this hypothesis, (Piquado et al., 2010) recently proposed a model by which masked words disrupt the short-term memory buffer. Previous work of the same group offered a biologically motivated computational model where, following the idea of Hebbian learning, synaptic strengthening is necessary for the recall of a word. They assume that a degraded stimulus causes weaker neuronal activity (synaptic firing) and therefore leads to fewer synaptic strengthening than a clear stimulus. Thus, recall probability is reduced when the stimulus is degraded (Miller and Wingfield, 2010). The proposed mechanism implies that working memory is a limiting factor in degraded speech perception.

When summing across the results cited here, the following picture emerges: An increased working memory buffer should facilitate adaptation to degraded speech, and it is a reasonable speculation that the relationship between adaptation slope and digit span (here a Pearson’s r of 0.4) would reach significance in a larger sample. But it is also becoming clear that the story of working memory and speech perception is not fully told yet (Jacquemot and Scott, 2006). Moreover, working memory performance itself has been shown to be at least partly predictable by perceptual uncertainty of the sensory input (e.g., Burkholder et al., 2005). Thus, in line with our main finding (see next paragraph), specifically developed auditory short-term memory tests might ultimately prove most predictable of individual adaptation to sensory challenges.

3.4.3 Temporal sensitivity in perceptual adaptation

An important finding of the present experiment is that individual AM rate discrimination thresholds predicted adaptation to vocoded speech. This supports our hypothesis that non-speech auditory skills contribute to perceptual learning of degraded speech.

Clearly, different sources of information can contribute to perceptual learning of degraded speech. Some studies have found evidence for “top-down” influences (i.e., higher-level cognitive processes) on vocoded speech perception, pointing out the importance of lexical information for perceptual learning (Davis et al., 2005; Davis and Johnsrude, 2007) and the facilitatory effect of semantic predictability on speech recognition (Obleser et al., 2007). Other studies have suggested a sublexical mechanism for perceptual learning of vocoded speech. For example, Hervais-Adelman, et al. (2008) showed that training with noise-vocoded nonwords is as effective as with real words for listeners learning to understand vocoded speech. Similarly, (Sebastian-Galles et al., 2000) proposed that adaptation to time-compressed speech operates at a sublexical, phonological level. They found that native speakers of Spanish benefit from training with time-compressed Greek sentences (which is phonologically similar but lexically distant from Spanish; the participants did not understand Greek) as much as from training with Spanish sentences. The authors concluded that adaptation does not depend on comprehension of the adapting sentences, but rather on the phonological properties of the adapting language (Sebastian-Galles et al., 2000). Another study showed that learning can be driven by low-level acoustic cues when listeners adapt to an artificial accent (Idemaru and Holt, 2011). Our findings provide further evidence for the importance of a “bottom-up” acoustic analysis that the auditory system has to perform when confronted with degraded speech (see also Mattys et al., 2009). Initial degraded speech perception studies have underlined that speech waveform envelopes are almost sufficient for speech recognition (Van Tasell et al., 1987; Shannon et al., 1995). Here we show that individual sensitivity to AM is even predictive of degraded speech recognition performance.

Non-speech auditory skills were dependent on short-term working memory, as shown by the correlation with forward digit span. This could partly be due to the design of the AM discrimination task, a 3-AFC task which makes considerable demands on working memory. Thus, listeners who have a poorer auditory short-term memory might perform worse on the 3-AFC task, because they may have more difficulties keeping the three stimuli available for comparison.

We cannot conclude from our results whether the observed relationship between perceptual adaptation and AM discrimination abilities is specific to a modulation rate around 4 Hz, since we did not test other modulation frequencies. There is only cursory evidence that such a relationship is not present for higher modulation rates (i.e., vocoded speech recognition performance and sensitivity to AM depth at 8 Hz; personal communication, Carolyn McGettigan). We thus assume that especially slow amplitude modulations in the syllabic range are predictive of adaptation to degraded speech as they are necessary for accurate speech recognition (Drullman et al., 1994b, a).

3.4.4 Anatomical predictors of perceptual adaptation

We found that individual differences in grey matter volume in the left pulvinar predicted degraded speech learning. There is evidence from electrophysiological studies that the pulvinar is involved in processing speech and short term memory. Johnson and Ojemann (2000) performed electrical stimulations of the thalamus prior to therapeutic thalamectomy in patients with chronic pain, Parkinson's disease and other disorders to verify localisation of ablation. During these procedures, the patient was awake and had to perform language or memory tasks. They observed that electrical stimulation of the dominant pulvinar (i.e., left pulvinar in right-handers) interfered with object naming and increased recall errors during a verbal short-term memory task. The authors attributed a gating function to the pulvinar where the thalamus controls retrieval from memory.

In deaf subjects, the pulvinar appears to play a role in auditory imagery. Lazard et al. (2011) conducted an fMRI study, where post-lingually deaf candidates for cochlear implantation imagined non-speech, non-human sounds and colours (the latter was used as baseline condition). They found increased activity in the right pulvinar in deaf patients compared to controls in the sound imagery versus colour imagery task. Additionally, within the CI candidate group, activity in the right thalamus was positively correlated with duration of hearing loss (Lazard et al., 2011).

In the first-order auditory thalamus, the MGB, von Kriegstein, et al. (2008) found activity to be modulated dependent on specific features of speech processing (von Kriegstein et al., 2008). In this fMRI study, listeners performed either syllable recognition or a speaker/loudness recognition task on the same stimuli. Activity in the MGB was modulated dependent on the task performed.

Furthermore, Peelle, et al. (2011) observed hearing-related changes in neural activity in the thalamus. This fMRI study compared older adults with different hearing abilities (pure-tone audiometric thresholds). Individuals with worse hearing thresholds had less activity in the thalamus when listening to grammatically more complex sentences (Peelle et al., 2011). The authors inferred that the thalamus plays an important role in transferring auditory information to prefrontal and premotor areas and thus is crucial for speech comprehension.

The high interconnectivity of the thalamus with cortical areas could also be an explanation for a role of the pulvinar in adaptation to degraded speech. In mouse studies it has been shown that connectivity between the sensory thalamus (ventral MGB) and auditory cortex is critical for auditory processing (Barkat et al., 2011). A recent animal study on visual perceptual learning stated that learning can be explained by changes in the feed-forward connections between the sensory thalamus and the primary sensory cortex, improving the inference performed by neural circuits (Bejjanki et al., 2011). In non-human primates it has been shown that auditory belt areas receive input from the thalamus, including the pulvinar (Hackett et al., 2007). Our knowledge is of course more limited in humans and is mainly based on diffusion weighted brain imaging data. According to the Oxford Thalamic Connectivity Atlas (http://www.fmrib.ox.ac.uk/connect/; Behrens et al., 2003), which provides probabilistic connectivity values based on diffusion tractography, the peak voxel of our VBM results (at the MNI coordinates −12 −24 12) exhibited a high probability ($p = 0.82$) of being connected to the temporal cortex.

In addition to these positive correlations with perceptual learning of vocoded speech, we found a negative correlation between GMV and adaptation slope in right-lateralised middle frontal gyrus and inferior temporal gyrus (Fig. 3.3A). In the brain morphometry literature, a common explanation of how increased regional brain volumes/thickness can reflect impaired processing capabilities is a lack of normal neuronal pruning during brain development. For example, (Hyde et al., 2007) found increased cortical thickness in the right auditory cortex and IFG in subjects suffering from amusia (i.e., showing impairments in music production and perception). Another recent study found that paediatric patients of pseudohypoacusis, that is, children with an abnormal audiogram despite no known organic causes (normal auditory brain stem responses etc.), have increased GMV in the superior temporal gyrus (STG) as well as the medial frontal gyrus. The authors suggested that insufficient neuronal pruning during development in these areas might cause higher GMV and the associated symptoms (Tomoda et al., 2012). Both studies, although drawn from patient populations, indicate that increased GMV in the temporal cortex (i.e., in a region close to where we found higher GMV in slower vocoded speech learners) is associated with difficulties in auditory processing.

The fact that we did not identify any other regions associated with vocoded-speech learning might be due to the rather small number of participants. Although we did not find increased GMV in the left IFG corresponding to the fMRI results of Eisner et al. (2010) at the whole-brain corrected level of significance, we did observe a trend-level correlation between IFG volume and learning scores in an

ROI-based analysis. Though we do not want to overinterpret this trend, higher brain activity in the IFG in better learners might be paralleled by increased grey matter, corroborating the left IFG's role in perceptual learning.

3.4.5 Clinical implications

The present study adds two important findings to the vocoded speech perception literature: In normal-hearing adults listening to a CI simulation, AM sensitivity as well as thalamic morphology are predictive of perceptual adaptation.

These experiments in normal-hearing listeners could be usefully transferred to CI patients in two different ways: First, we suggest that pre-implantation brain scans can be used to assess anatomical predictors of post-implantation speech comprehension. So far, there have been attempts to prospectively predict CI patients' speech recognition performance from pre-implantation fMRI scans (Lazard et al., 2010), but to our knowledge no studies of brain structural predictors of CI success have been conducted. Second, after cochlear implantation, CI patients could be tested with an AM discrimination paradigm to verify whether their AM sensitivity is also correlated with speech comprehension. This could help clarifying how much of the individual varia-bility in CI patients' speech recognition can be explained by basic auditory processing capabilities.

4 Experiment II: Speech comprehension and temporal processing in cochlear implant patients[5]

4.1 Introduction

Cochlear implants can restore hearing in deaf patients by means of electric stimulation of the auditory nerve (see Section 1.1.3). The CI transduced speech signal is considerably distorted however, as the device only provides a limited spectral resolution. CI users vary considerably in how well they adapt to their device: some learn to understand speech even under difficult listening conditions while others are hardly able to make use of their device. Different factors such as age at implantation, duration and cause of deafness, or the number of implanted electrodes have been proposed to predict speech comprehension success (Blamey et al., 1992). However, they can only partly explain the observed interindividual variability. Therefore, auditory psychophysical tests are needed to characterise and predict a CI patient's speech recognition progress.

In Experiment I (Erb et al., 2012) we showed that in normal-hearing adults listening to a CI simulation, sensitivity to the temporal envelope of a non-speech signal as measured by AM rate discrimination thresholds is predictive of perceptual adaptation. In the present experiment we aim to verify whether this can be transferred to CI patients: Can AM rate discrimination thresholds shortly after implantation prospectively predict a CI patient's speech comprehension success?

There is evidence from electrical stimulation experiments that auditory temporal processing is crucial for speech recognition in CI patients. Fu (2002) found that the temporal resolution of CI patients' auditory system was associated with speech recognition abilities. In CI users with at least six years of experience, AM detection thresholds were assessed by electrical stimulation with steady-state

---

[5] This experiment was designed and performed by Julia Erb, Alexandra Ludwig, Michael Fuchs (Cochlear Implantation Center of the University Clinic, Leipzig), and Jonas Obleser.

pulse trains modulated at 100 Hz. AM sensitivity was significantly correlated with phoneme recognition (of both vowels and consonants). Measuring electric temporal modulation transfer functions (TMTF; Viemeister, 1979, Shannon, 1992) showed that CI patients were most sensitive to modulations between 50 and 100 Hz. At these frequencies, CI users' AM rate discrimination and AM detection thresholds were found to correlate with lexical tone recognition in Mandarin (Luo et al., 2008). Similarly, fundamental frequency (F0) processing in words was correlated with AM discrimination thresholds centered at 50–300 Hz (Chatterjee and Peng, 2008), suggesting that the high-frequency amplitude modulations of speech are important in conveying speech information to CI users. In normal acoustic hearing, cut-off frequencies of the TMTF, which shows low-pass characteristics, are typically lower than in electrical hearing (Bacon and Viemeister, 1985). The higher cut-off frequencies in electric TMTFs are possibly due to differences in dynamic range between acoustic and electrically stimulated listeners (Shannon, 1992).

Electrical stimulation protocols are often only available in specialised centers. Here, we tested whether acoustic AM rate thresholds were equally predictive of speech recognition capabilities in CI patients. Such assessment of acoustic AM rate thresholds could be easily integrated into the postoperative rehabilitation protocol. In contrast to previous electrical stimulation protocols (Shannon, 1992; Luo et al., 2008, Chatterjee and Peng, 2008) and in analogy to Experiment I, we concentrated on low modulation frequencies centered on 4 Hz, which are most important for prosodic and segmental speech information (Houtgast and Steeneken, 1985; Drullman et al., 1994b). To assess patients' speech recognition abilities, we made use of standard German speech (in noise) tests that were assessed as part of the rehabilitation protocol at various time points up to one year after cochlear implantation. We hypothesised that (lower) acoustic AM rate discrimination thresholds could prospectively predict CI patients' (improved) speech reception thresholds and might thus be used as a predictor of individual speech comprehension success.

4.2 Materials and methods

4.2.1 Participants

Sixteen adult CI patients (13 female, 3 male, aged 23–83, mean 53.1 years) participated in this study and were recruited from the CI centre of the University Clinic, Leipzig. They were implanted with one CI containing 12 electrodes (MedEl) or 22 electrodes (Cochlear). All patients were native speakers of German. Etiology and types of deafness were variable, including otitis media, measles, meningitis, sepsis, acute hearing loss, congenital and progressive hearing loss. Duration of deafness before implantation varied from 1–53 years (mean 23.6 years). Five patients had early-onset deafness, whereas eleven patients were late, that is, post-lingually deafened. Only three out of sixteen participants used sign language. All participants gave informed consent. Procedures were approved by, and in accordance with the guidelines of the local ethics committee (University of Leipzig).

4.2.2 Amplitude modulation rate discrimination

The AM rate discrimination testing procedure was slightly modified compared to the protocol used in Experiment I with normal-hearing listeners. In order to adapt to the different and putatively more variable hearing capacities among CI Patients, rather than estimating psychometric curves, we used an adaptive staircase procedure to assess AM rate discrimination thresholds.

Stimuli were sinusoidally amplitude-modulated white noises. As in Experiment I, the standard stimulus was modulated at 4 Hz, whereas deviant stimulus was modulated at different rates between 2 and 6 Hz in steps of 0.5 or 0.25 Hz (for determination of deviant modulation rate on each trial see adaptive tracking procedure below). The onset phase of the sinusoidal modulation was randomly varied for all stimuli and the length of the stimuli also randomly changed between 900 and 1100 ms. Modulation depth was held constant at m = 100% (as opposed to 60% in Experiment I). Stimuli were presented with an inter-stimulus interval of 800 ms (as opposed to 500 ms in Experiment I). All stimuli were peak-normalised with respect to each other.

Experimental procedure. Patients were tested approximately 6 weeks after their CI was switched on for the first time, that is, approximately 10 weeks after cochlear implantation. Testing took place in a sound-proof audiometric cabin. Stimuli were played on a PC in MATLAB 7.11 and were presented in a free-field set-up via an AT900-audiometer (Auritec). The presentation level was set at 30 dB sound pressure level (SPL) above the patient's mean audiometric threshold for band-limited noise at the center frequencies of 500 to 4000 Hz. Thus, the average presentation level was at 67.8 (range 55–79) dB SPL.

Amplitude modulation rate discrimination thresholds were estimated using a three-alternative forced choice (3-AFC) adaptive staircase procedure. On each trial, participants heard two standard stimuli and one deviant sound; the position of the deviant within the trial was randomly varied from trial to trial. The patients' task was to indicate at which position they heard the sound with the different modulation rate. Patients verbally responded with "one", "two" or "three", which the experimenter entered as response, which then started the next trial. The modulation rate difference between standard and deviant increased after each incorrect trial and decreased after two successive correct trials. This "two-down-one-up" staircase procedure converged on the modulation rate difference that provides 70% correct responses. The initial step size was 0.5 Hz and changed to 0.25 Hz after four reversals. The staircase procedure was terminated after twelve reversals. The AM rate threshold was calculated as the average modulation rate of the last eight reversals.

Each participant completed two runs: In the first run, the deviant AM rate starting level was 6 Hz and approached the standard rate of 4 Hz from "above" (i.e., deviant stimuli were modulated at a faster rate than standard stimuli), whereas in the second run, deviant stimulation rate started at 2 Hz, approaching the standard AM rate from "below" (i.e., the deviant was slower than the standard AM rate). Therefore, we obtained two AM rate discrimination thresholds [Hz] from each participant, one approaching 4 Hz from a faster modulation rate ("AM above") and one from a slower modulation rate ("AM below") which we averaged for each participant ("AM mean").

Note that all sixteen participants were tested on the AM threshold above, but only thirteen participants were tested on the AM threshold below due to time limitations. Participants with missing data were excluded from all correlational analyses concerning AM below, such that these analyses always include fewer participants than the AM above correlational analyses and are therefore arguably less reliable. For correlations with AM mean, we used AM above in participants where AM below was missing.

4.2.3 Speech test battery

A battery of speech recognition tests was administered to the CI patients as part of the rehabilitation program a few days up to nine months after switching on the CI for the first time. The Freiburger Number Test (FNT) and the Freiburger Monosyllabic Test (FMT) were applied both in quiet and in noise.

Freiburger Speech Tests. In the Freiburger number test (Hahlbrock, 1953), 20 digits were presented to the patients. In the Freiburger monosyllabic test (Hahlbrock, 1953), 20 monosyllabic German words were presented. Both tests were performed in free field, either in quiet at a presentation level of 65 dB SPL or in steady-state speech-shaped noise, at an SNR of 60 dB. The patients repeated back what they had understood. Speech recognition thresholds were calculated as percent of correctly repeated items. The FNT was assessed a few days, 3 weeks, and 3vmonths after switching on the CI. The FMT was administered 6 weeks, 3 months, 6 months, and 9 months after switching on the CI.

Note that while all sixteen patients had undergone AM rate discrimination testing approximately 6 weeks after implantation, we could not obtain speech reception thresholds for all patients at all time points. Averaged over time points, 12.3 patients participated in the FNT, 7.7 patients participated in the FNT in noise, and 9.5 patients participated in the FMT.

4.3 Results

4.3.1 Perceptual learning of speech

We tested whether speech reception thresholds improved over time using a nonparametric Kruskal-Wallis test. The Kruskal-Wallis test showed no significant effect of time point for speech reception thresholds in the Freiburger monosyllabic test ($\chi^2 = 2.4$, $p = 0.49$; Fig. 4.1A), in the Freiburger number test ($\chi^2 = 1.61$, $p = 0.45$; Fig. 4.1A), and in the Freiburger number test in noise ($\chi^2 = 2.13$, $p = 0.34$, Fig. 4.1A). This failure of finding a significant effect of time in most speech tests is likely due the overall small number of observations ($n = 9.8$, averaged over all speech tests and time points), and a non-negligible number of missing values for different patients at different time points: the number of observations decreased over time from $n = 14$ in the FNT in the 1st week to $n = 7$ in the FMT 9 months afterwards, because some patients had not yet been tested in the later speech tests when writing up the present experiment. Therefore, continued testing is needed to gather more data.

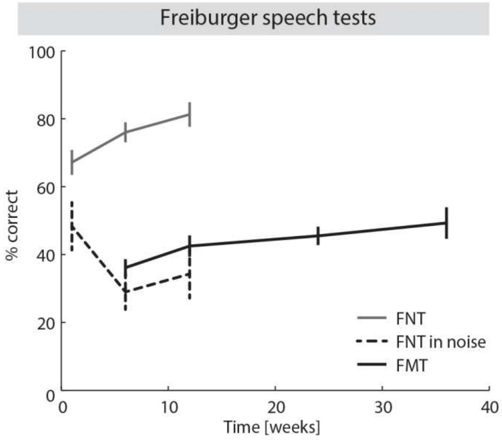


Figure 4.1. Speech reception thresholds as a function of time. Scores in the Freiburger Number Test (FNT) in quiet (grey line) and noise (dotted line), in the Freiburger Monosyllabic Test (FMT; black line) are shown 1 week to 9 months after switching on the CI. Note that the best performance in the FNT in noise is observed in the 1st week due to a clinical sampling artifact: Only patients who had a good performance in the FNT in the 1st week were also tested in the FNT in noise (n = 6), whereas in the 6th week, patients were tested in all speech tests irrespective of performance (n = 10). Error bars show within-subject standard error (Cousineau, 2005).

4.3.2 Amplitude modulation rate discrimination

To test whether AM discrimination thresholds converged when the deviant AM rate approached 4 Hz when starting from 6 Hz or 2 Hz, we correlated AM above and AM below. Both measures were highly correlated, indicating that AM above and below converged (Fig. 4.2A and Table 4.1), justifying the averaging of both measures ("AM mean"). Note, however, that AM above is arguably the more reliable measure, because it has more data points (sixteen data points for AM above as opposed to thirteen data points for AM below).

4.3.3 Spearman's correlations

To test whether sensitivity to non-speech auditory envelopes could prospectively predict speech comprehension success, we tested correlations between AM rate discrimination and speech reception thresholds. A Shapiro-Wilk non-parametric test (Shapiro, 1965) showed that the variables AM below and AM mean were not normally distributed ($p < 0.05$). Therefore, we used Spearman's rank correlations.

Acoustic AM rate discrimination thresholds (AM above) correlated negatively with speech reception thresholds in the Freiburger number test in noise 6 weeks after switching on the CI, indicating that better AM rate discrimination was associated with better number recognition in noise (Fig. 4. 2A).

Table 4.1. Spearman's correlations between AM rate discrimination, demographic factors and speech reception thresholds. Correlations are thresholded at $p < 0.1$.

	AM above	AM below	AM mean	Age	Duration of deafness
AM above	1.00	0.81**	0.96**	-	0.46+
AM below	0.81**	1.00	0.86**	-	-
AM mean	0.96**	0.86	1.00	-	-
Age	-	-	-	1.00	-
Duration of deafness	0.46+	-	-	-	1.00
FNT noise 6 weeks	-0.66*	-	-	-	-0.80*
FMT 3 months	-0.64+	-	-	-	-
FMT 6 months	-0.56+	-	-	-	-
FMT noise 6 months	-	-	-	-	-0.76+

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$; AM: amplitude modulation rate discrimination threshold; FNT: Freiburger Number Test; FMT: Freiburger Monosyllabic Test; time specification denotes time after switching on the CI.

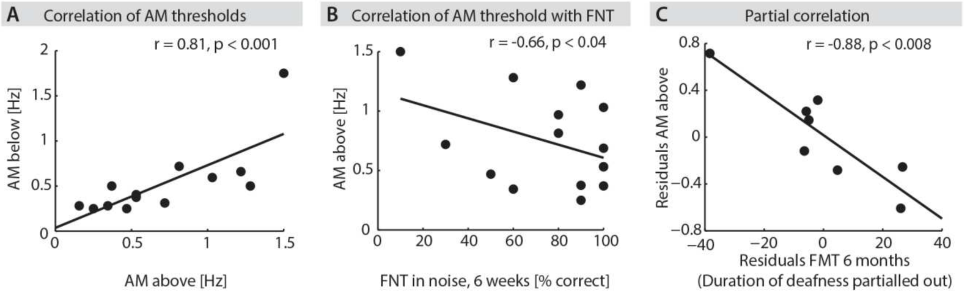


Figure 4.2. Spearman's correlations with AM discrimination thresholds. **(A)** AM rate discrimination threshold (AM above) correlates negatively with AM rate discrimination (AM below), showing that the two measures converge. **(B)** AM rate discrimination threshold (AM above) correlates negatively with the Freiburger number test (FNT) in noise 6 weeks after switching on the implant, showing that better AM discrimination is related to better speech comprehension. **(C)** When controlling for duration of deafness, AM above correlates significantly with the Freiburger monosyllabic test (FMT) after 6 months. Note that the lines show the least-squares regression fits, although correlation coefficients denote Spearman's correlations.

To investigate the predictiveness of AM rate discrimination thresholds compared to demographic factors, we also tested for the correlation of speech reception thresholds with age and duration of deafness. Duration of deafness also correlated significantly with speech reception thresholds in the Freiburger number test in noise 6 weeks after switching on the CI (Table 4.1).

4.3.4 Partial correlations

When partialling out age, AM discrimination thresholds (AM above) and the Freiburger number test in noise (6 weeks after switching on the implant) still correlated significantly (Table 4.2).

When controlling for duration of deafness, the non-significant trend for a correlation between AM discrimination threshold (AM above and AM mean) and the Freiburger monosyllabic test 6 months postoperatively became highly significant (Fig. 4.2C and Table 4.2).

Table 4.2. Partial Spearman's correlations between AM rate discrimination and speech reception thresholds, with duration of deafness or age partialled out. Correlations are thresholded at $p < 0.1$.

	AM above	AM below	AM mean
Age partialled out	-	-	-
FNT noise 6 weeks	-0.74*	-	-
Duration of deafness partialled out			
FNT 3 months	-	-0,91+	-
FMT 3 months	-0,80+	-	-
FMT 6 months	-0.89**	-	-0.80*

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$; AM: amplitude modulation rate discrimination threshold; FNT: Freiburger Number Test; FMT: Freiburger Monosyllabic Test; time specification denotes time after switching on the CI.

4.3.5 Regression analysis

In a final analysis we tested whether speech comprehension 6 months after switching on the implant (in the Freiburger monosyllabic test) could be better predicted by AM rate discrimination thresholds (AM above) compared to duration of deafness and a speech test directly after switching on the implant (Freiburger number test). To this end, all three predictors were included in a general linear model. However, none of the predictors including time of deafness ($t = 0.48$, $p = 0.68$), the Freiburger number test ($t = 2.44$, $p = 0.13$), and AM above ($t = -2.09$, $p = 0.17$) reached significance.

4.4 Discussion

The present study provides preliminary evidence that acoustic AM rate discrimination thresholds centered at 4 Hz are related to speech comprehension in CI patients. Already 6 weeks after switching on the CI, that is, 10 weeks postoperatively, AM rate discrimination thresholds correlated significantly with speech reception thresholds on the same day. When duration of deafness was controlled for, AM rate discrimination thresholds could prospectively predict speech perception performance 6 months after switching on the CI: AM rate discrimination threshold was able to explain ~80% of the variance in the Freiburger monosyllabic test 6 months later. Due to the small sample size and the high variance in various demographic factors, these results should however be interpreted with caution.

4.4.1 Temporal sensitivity correlates with speech recognition

The finding of a correlation between AM rate discrimination thresholds and thresholds in the Freiburger number in noise test on the same day is consistent with a recent observation showing acoustic TMTFs in CI patients being related to speech comprehension. (Won et al., 2011), demonstrated that acoustic modulation detection thresholds averaged over 10 to 300 Hz were significantly correlated both with recognition of disyllabic words in noise and monosyllabic words in quiet. The highest correlations were observed for AM thresholds at 100–150 Hz. A recent study by (Gnansia et al., 2014) showed that the sensitivity to slow temporal modulations is crucial for speech recognition in CI listeners: Modulation detection thresholds at 8 Hz correlated with vowel and consonant identification scores.

Previous experiments assessing electrical TMTFs have shown a similar relationship between AM sensitivity and speech perception in CI users (Cazals et al., 1994, Fu, 2002, Luo et al., 2008; Chatterjee and Oberzut, 2011). Together with previous studies, the present results confirm that temporal modulation information is crucial for CI patients' speech processing.

One important objective of the present study was to investigate whether AM sensitivity shortly after implantation could prospectively predict speech recognition abilities and adaptation to the CI transduced speech signal. In Experiment I, we showed that AM rate discrimination thresholds are predictive of short-term adaptation to a CI simulation in normal-hearing listeners: Improvement during only 20 minutes of listening to noise-vocoded speech correlated with AM rate sensitivity (see Experiment I). Here, we show that AM rate sensitivity can also predict CI users' speech comprehension over a longer time scale of half a year after implantation when accounting for duration of deafness.

4.4.2 Absence of speech recognition improvement

Surprisingly, we failed to find a significant effect of time on postoperative speech reception thresholds. However, this null-finding is likely due to the overall small number of observations and the decrease in number of observations over time, because some patients had not yet been tested in the later speech tests when writing up the present experiment. In analogy to Experiment I, we also fitted individual linear curves to patients' speech reception thresholds over time, to use the slopes as a measure for perceptual learning. However, learning slopes did not correlate with AM discrimination thresholds, possibly due to unreliable fits as a result of missing data points (results not reported). This null-result shows that continued testing is warranted to allow for more reliable conclusions.

4.4.3 Conclusions

The present results suggest that future speech perception in CI users can best be predicted early after implantation when duration of deafness and AM rate discrimination thresholds are available. In contrast to previous studies assessing complete TMTFs which involved extensive testing lasting 2–2.5 hour (Won et al., 2011), the current data indicate that AM discrimination assessment of only 20 minutes duration provides clinically relevant information about CI patients speech comprehension abilities. Such testing could be easily incorporated into the postoperative rehabilitation program. Follow-up testing is needed however, in order to confirm and extend these results and clarify the differential predictiveness of AM rate discrimination thresholds compared to other factors.

5 Experiment III: Neural networks in perceptual adaptation to degraded speech[6]

5.1 Introduction

Humans have the capability to rapidly adapt to degraded or altered speech (see Chapters 1, 3 and 4). The current experiment investigated the short-term neural processes that underlie this adaptation to degraded speech using fMRI.

We simulated CI transduced speech in normal-hearing listeners using noise-vocoding (see Section 2.2; Shannon et al., 1995). Listeners with higher sensitivity to envelope fluctuations in an auditory signal, as measured by AM rate discrimination thresholds, adapt more quickly to vocoded speech (see Experiment I). Therefore, we predicted that the temporal envelope of non-speech sounds (Giraud et al., 2000) and vocoded speech should be processed by shared neural resources.

Rapid perceptual learning of vocoded speech is a well-established finding of behavioural studies (Rosen et al., 1999; Davis et al., 2005; Peelle and Wingfield, 2005; Bent et al., 2011); see Section 1.1.2), but its neural bases are largely unknown. Recently, increased precentral gyrus activity was associated with boosted perceptual learning during the joint presentation of vocoded and clear words (Hervais-Adelman et al., 2012). Further, in a simulated CI rehabilitation program that supplemented presentation of vocoded sentences with simultaneous written feedback, perceptual learning relied on the left IFG (Eisner et al., 2010). However, in everyday situations, listeners rarely receive direct feedback. Therefore, the neural dynamics of feedback-free or self-regulated adaptation are investigated here.

---

[6] This chapter is adapted from the article published in the Journal of Neuroscience by Erb, Henry, Eisner and Obleser (2013).

Subcortical structures likely play a critical role in adaptation to degraded speech, but the specific contributions of distinct structures are uncertain. In Experiment I, we demonstrated that grey matter volume in the left pulvinar thalamus predicted how fast listeners adapted to vocoded speech. However, previous fMRI studies have failed to detect adaptation-related signal changes in subcortical regions, possibly due to these brain areas' susceptibility to MR artefacts. Here, we implemented a cardiac-gated image acquisition procedure, thereby avoiding heartbeat-related artefacts.

The present fMRI study sheds new light on perceptual adaptation to degraded speech with respect to four points: (1) we investigate the convergence of neural mechanisms underlying effortful speech and non-speech processing; (2) we test feedback-free, short-term adaptation; (3) unlike previous perceptual learning studies (Golestani and Zatorre, 2004; Adank and Devlin, 2010; Eisner et al., 2010), we collect word report scores on every trial and are able to model both trial-by-trial fluctuation in comprehension and adaptation to degraded speech; (4) we assess subcortical contributions to adaptation. We show that an "executive" system (Eckert et al., 2009) is recruited when actively coping with difficult speech and non-speech listening conditions. In contrast, activations in the classic "language" network (Scott et al., 2000; Hickok and Poeppel, 2007) are driven by trial-by-trial fluctuations in speech comprehension, not acoustic speech clarity. Finally, we demonstrate that rapid adaptation to degraded speech is accompanied by haemodynamic down-regulation in a cortico-thalamic-striatal network.

5.2 Materials and methods

5.2.1 Participants

Thirty participants (15 females, age range 21–31 years, mean 25.9 years) took part in the study. Participants were recruited from the participant database of the Max Planck Institute for Human Cognitive and Brain Sciences. All were native speakers of German with no known hearing impairments, language or neurological disorders and showed dominant right-handedness according to the Edinburgh inventory (Oldfield, 1971). They were naïve to noise-vocoded speech. Participants received financial compensation of € 16, and gave informed consent. Procedures were approved by the local ethics committee (University of Leipzig).

5.2.2 Stimuli and experimental design

Degraded speech perception. Sentence material was drawn from the German SPIN sentences (see Section2.1), which is controlled for the predictability of the final word (high vs. low predictability). For the present study, only low-predictability sentences were chosen, such that semantic cues were limited and the listener had to rely primarily on acoustic properties of the sentence for comprehension. A complete list of these sentences is shown in the appendix.

The sentences were recorded by a female speaker of standard German in a sound-damped chamber. The length of the recorded sentences varied between 1620 and 2760 ms. Sentences were degraded using four-band noise vocoding as described in Section 2.2.

On each trial, participants first heard a sentence. They were instructed to repeat as much of the sentence as they had understood when a green light appeared on the screen, but to stop talking when the green light disappeared (after 3 s) in order to avoid movement during scan acquisition. Speech production was recorded for later off-line scoring (Eckert et al., 2009; Harris et al., 2009). Responses were scored as proportions of correctly repeated words per sentence ("report scores", Peelle et al., 2013). Scoring took into account all words of a sentence including function words; errors in declension or conjugation were accepted as correct.

Clear speech trials were used as a high-level baseline. Clear speech can be assumed to be fully adapted, and therefore to be processed in a stable way over time. This ensured that no neural adaptation would occur in the baseline condition, whereas another type of artificial speech degradation (e.g., rotated speech) might have led to neural adaptation (even in the absence of behavioural adaptation).

In sum, Experiment IIIa comprised three conditions: (1) four-band vocoded sentences ("degraded speech"; 100 trials in total); (2) clear (non-vocoded) sentences ("clear speech"; 24 trials in total); (3) trials lacking any auditory stimulation ("silent trials"; 20 trials in total). Overall, the experiment comprised 144 trials. Clear speech trials were presented every 5th trial, whereas the silent trials were randomly interspersed (Fig. 5.1C). Sentences were presented to each participant in one of four different orders; presentation order was counterbalanced across participants.

Participants' adaptation curves. As in Experiment I, we modelled individual performance improvement in two different ways: as a power law and as a linear performance increase. To test which function would better describe the data, both curves were fitted to the individual report scores over time using a least-squares estimation procedure in MATLAB 7.11 (for an example fit to the scores averaged over participants, see Fig. 5.2A). We compared goodness of fit by determining the Bayesian information criterion (BIC; Schwarz, 1978) of the each fits within each participant (see Section 2.3).

AM discrimination. Stimuli were one-second long sinusoidally amplitude-modulated white noises. The standard stimulus was modulated at 4 Hz. Deviant stimuli were modulated at 7 different rates that were linearly spaced between 2 and 6 Hz in steps of 0.67 Hz. The middle level was modulated at the same rate as the standard (4 Hz); critically, in this condition participants were unable to distinguish standard and deviant, but still performed the same task. Modulation depth was kept constant at 60%. The onset phase of the sinusoidal modulation was randomly varied for all stimuli, separately for the standard and deviant. Standard and deviant stimuli were presented with an inter-stimulus interval of 500 ms.

The paradigm was a 2-alternative forced choice (2-AFC) task: On each trial, participants first heard the standard stimulus (modulated at 4Hz) followed by one deviant stimulus. After auditory presentation of the sound pair, participants were prompted to respond when a green light appeared on the screen (Fig. 5.1B). Half of the participants were to indicate which sound had a faster modulation rate, while the other halfs task was to indicate which sound had a slower modulation rate in order to counterbalance button presses. Participants responded via a button box in their right hand by pressing the left (for 1st sound) or right key (for 2nd sound).

In sum, there were seven levels of deviant AM rate (comprising 16 trials each); in addition, we interspersed silent trials lacking any auditory stimulation (16 trials). On the whole, every participant listened to 128 trials in a pseudorandom order where trials of the same condition were never presented subsequently.

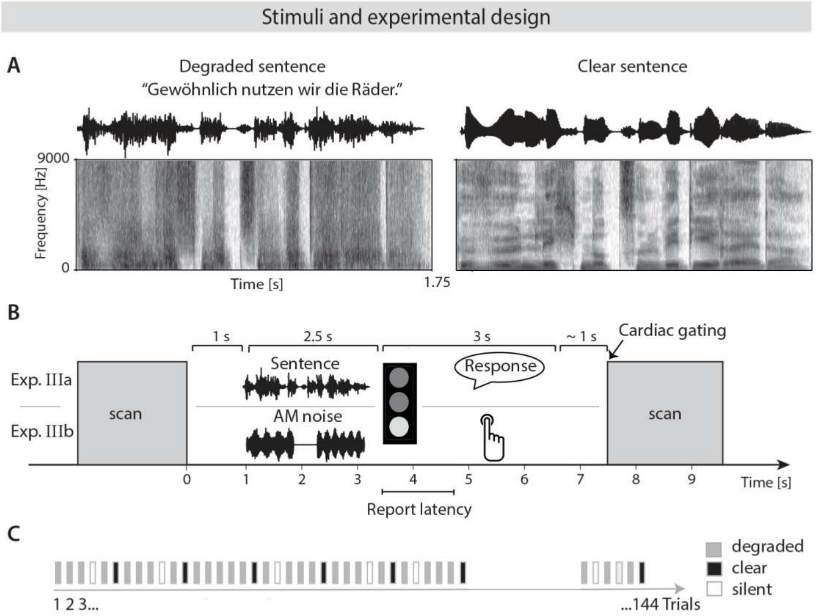


Figure 5.1. Stimuli and experimental design. **(A)** Oscillogram (upper panel) and spectrogram (lower panel) of the sentence "Gewöhnlich nutzen wir die Räder" [Normally we use the bikes] in degraded (left) and clear speech (right). **(B)** Trial structure in Experiments IIIa and IIIb. After acoustic presentation of a sentence or AM stimulus (lasting for approximately 2.5s), participants were prompted to respond when green lights appeared on a screen (lasting for 3s), by repeating the sentence or pressing a button. Subsequent scan acquisition was initiated by cardiac gating. The onset of auditory stimulation preceded the anticipated scan acquisition by a fixed period of 6.5 s, although the actual scan time was variable due to cardiac gating. **(C)** Experiment IIIa comprised 100 degraded speech trials, 24 clear speech trials (inserted after each 5th trial) and 20 randomly interspersed silent trials.

5.2.3 Experimental procedure

In order to maximise the comparability between individuals, all participants were tested in the same order, namely degraded speech perception first (Experiment IIIa), followed by AM rate discrimination (Experiment IIIb). Before participants went into the scanner, they were familiarised with each of the two tasks; they listened to three eight-band vocoded German SPIN sentences as training for Experiment IIIa and three exemplary trials of Experiment IIIb.

In the scanner, to protect participants from hearing damage due to scanner noise, they wore Alpine Musicsafe Pro earplugs, yielding approximately linear 14-dB reduction in sound pressure up to 8 kHz. Auditory stimuli were delivered through MR-Confon headphones using Presentation software. Visual prompts were projected on a screen which participants viewed via a mirror.

Trial timing was identical in Experiment IIIa and 2. Each trial was approximately 9 s long, but actual trial length varied due to cardiac gating (see below). Trials started with a 1 s silent gap, after which participants heard a sentence (Experiment IIIa) or two AM-modulated stimuli (Experiment IIIb) lasting for approximately 2.5 s. Following stimulus presentation (3.5 s into the trial), a green light ("go"-signal for response) was visually presented and lasted for 3 s. After approximately 1 s of silence, scan acquisition with a TR of 2s was triggered using cardiac gating. Thus, the onset of auditory stimulation preceded the anticipated scan acquisition by approximately 6.5 s (Fig. 5.1B).

5.2.4 MRI data acquisition

MRI data were collected on a 3T Siemens Verio scanner. Functional MR images were acquired with a 12-channel head coil using an echo-planar imaging (EPI) sequence [TR ≈ 9000 ms, TA = 2000ms, TE = 30 ms, flip angle = 90°, 3 mm slice thickness, 30 axial slices (ascending), interslice distance = 1 mm, acquisition matrix of 64 x 64, voxel size = 3 x 3 x 3 mm] in two separate runs for Experiment IIIa and IIIb. The acquisition matrix was placed such that the x-axis was in line with the anterior–posterior commissural (AC–PC) line. We used a sparse-sampling procedure, where TR was longer than TA, allowing for silent periods to play stimuli and record responses (Hall et al., 1999).

Additionally, cardiac gating was applied to avoid movement artefacts caused by the heartbeat in subcortical structures (von Kriegstein et al., 2008; see also Section 2.6.1), in which we were especially interested (based on Experiment I). Participants' heartbeat was monitored using an MR compatible pulse oximeter (Siemens) attached to their left ring finger. On each trial, after 9 s had elapsed, the scanner waited for the first heartbeat to trigger volume acquisition. Hence, the actual TR was variable but amounted to 9.45 ± 0.27s (mean ± standard error of the mean, SEM; across all participants).

Following functional imaging, a T1-weighted structural image was acquired with a 32-channel head coil using an MPRAGE sequence [TR = 1300 ms, TE = 3.5ms, flip angle = 10°, 1 mm slice thickness, 176 sagittal slices, acquisition matrix of 256 x 240, voxel size = 1mm^3].

In one participant, we were only able to acquire 136 (as opposed to 144) scans for Experiment IIIa due to technical problems with cardiac gating. A second participant had to be excluded from all analyses concerning Experiment IIIb, because scan acquisition had become desynchronised with stimulus presentation.

5.2.5 Data analysis

Preprocessing. MRI data were analysed in SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK). Preprocessing was carried out separately for Experiment IIIa and IIIb. Structural MRI scans were manually aligned with the coordinate system such that anterior-posterior commissure was in line with the x-axis and the anterior commissure in the origin of the coordinate system. Functional MRI volumes were realigned and unwarped using a fieldmap, coregistered with the structural scan, segmented and normalised to MNI space using the segmentation parameters, and smoothed with an isotropic Gaussian kernel of 8 mm full-width at half-maximum.

MR images were statistically analysed in the context of the GLM. We set up three different models for Experiment IIIa and one model for Experiment IIIb to assess the following effects:

Effects of auditory stimulation. In a basic model for Experiment IIIa, we defined three conditions at the single subject level: degraded speech, clear speech and silent trials. The effect of auditory stimulation was tested by contrasting sound (degraded and clear speech) against silent trials. To avoid overspecification, silent trials were removed from all subsequent analyses.

Effects of speech degradation and trial-by-trial fluctuation with comprehension. In a speech-degradation model, we included two conditions: degraded and clear speech. Additionally, a parametric modulator of the degraded speech trials was defined, representing the behavioural report scores. A regressor of no interest, containing report latencies, was added in order to account for differences in speech production (analysis explained in detail below); this regressor was included in all remaining analyses. To assess effects of stimulus clarity, we contrasted degraded against clear speech trials. To reveal effects of trial-by-trial fluctuations in speech comprehension, we assessed correlations with the regressor representing report scores.

Effects of perceptual adaptation. To model effects of adaptation, we looked for signal changes over time corresponding to participants slow performance increase (“adaptation curves”). However, there are a number of unspecific reasons why BOLD activity could gradually change over time, for example scanner drift or fatigue of the participant. Therefore, we compared the changes over time for vocoded speech to the change in activity seen for the clear speech condition, while taking into account the behaviourally observed adaptation to vocoded speech: Rather than simply testing a condition × time interaction (which would not consider the adaptation curve of a listener), we tested for the three-way interaction for condition × time × behaviour.

To this end, we created two parametric modulators, one for each condition, by multiplying time (i.e., trial number) with the linear adaptation curves. This resulted in a quadratic curve for the degraded speech regressor of which the slope was dependent on the individual adaptation curve (Fig. 5.5, upper panel). However, since there was no perceptual adaptation in the clear speech condition (all sentences were fully comprehended), the "adaptation" curve was flat such that multiplication effectively left the time regressor unchanged, resulting in a linear curve.

Contrasts tested for the difference in these two regressors, ruling out the possibility that a slow change over time is due to slow unspecific signal drifts (which would be present in both conditions; analysis as in (Buchel et al., 1998). Thus, we identified areas where changes over time are more pronounced in the degraded speech condition (where perceptual processing improves) than in the clear speech condition (where perceptual processing remains stable throughout).

Regressor of no interest for report latency. Although the present study was designed to image degraded speech perception, parts of the observed activity may be related to speech production or preparation, because participants overtly repeated back what they had understood starting approximately 3.5 s prior to scan acquisition (see Fig. 5.1B). In particular, participants verbal responses might have been faster for clear relative to degraded speech trials, perhaps leading to partly imaging the BOLD response to speech production, but more so for clear speech trials. Therefore, differences in report latencies might confound the comparison between degraded and clear speech trials. Similarly, adaptation-related signal changes might become confounded, as it is likely that report latencies decrease as participants adapt to degraded speech. To control for this potential confound, we calculated report latency relative to the onset of the visual response cue (see Fig. 5.1B). This measure was included at the first level as one single regressor of no interest in all models concerning Experiment IIIa (described above). For trials where participants did not produce an overt response, the subject-specific mean report latency was entered.

Effects of ΔAM rate. In Experiment IIIb we modelled four conditions; one for each level of AM rate difference between standard and deviant (referred to as "ΔAM rate"): ΔAM rate = 0 Hz, ±0.67 Hz, ±1.33 Hz, and ±2 Hz (note: in the ΔAM rate = 0 Hz condition, the deviant was modulated also at 4 Hz). To assess a linear correlation with ΔAM rate, these conditions were weighted with the contrast vector [−3 −1 1 3] for a positive correlation and [3 1 −1 −3] for a negative correlation with ΔAM rate. Large effects in these contrasts would mean linear scaling with deviance from the standard AM rate. In an additional conjunction analysis (Friston et al., 1999), we tested for the intersection of the effects of ΔAM rate and speech degradation.

All described analyses were whole brain analyses. Regressors were modelled using a finite impulse response (FIR) comprising one bin. A high-pass filter with a cut-off of 1024 s was applied to eliminate low-frequency noise. No correction for serial autocorrelation was necessary because of the long TR in the sparse-sampling acquisition.

Second level statistics were calculated using a one-sample *t*-test. Group inferences are reported at a family-wise error (FWE) corrected voxel-wise threshold of $p < 0.05$, where FWE rate was controlled using random field theory. Only for the adaptation analyses did we use a slightly more lenient threshold of $p<0.001$, where cluster-extent ($k > 20$) was corrected based on a Monte Carlo Simulation (Slotnick et al., 2003). T-statistic maps were transformed to *Z*-statistic maps using spm_t2z.m, and overlaid and displayed on the ch2 template in MNI space included with MRIcron (Rorden and Brett, 2000).

Lateralisation analysis. It is likely that processes of comprehension are left-lateralised (Obleser et al., 2007; Rosen et al., 2011; McGettigan et al., 2012a). We therefore tested for lateralisation of activity related to trial-by-trial fluctuation in speech comprehension. As in the analyses described above, EPI images first were realigned, unwarped and coregistered. The images were then segmented using symmetric grey-matter–, white-matter– and cerebrospinal-fluid–templates, which were created by averaging each original template with itself flipped along the y–axis (Salmond et al., 2000). The segmentation parameters were used to normalise the images to MNI-space, resulting in a normalisation to a symmetric MNI-template (Liegeois et al., 2002) and smoothed at 8 mm FWHM. To conduct a voxel-by-voxel statistical test of laterality, the first-level analyses were performed as described above. Resulting maps were flipped along the y–axis and compared with the original maps in a paired *t*-test (Bozic et al., 2010).

Regions of interest analyses. In order to extract measures of percent signal change in the regions identified by the whole-brain analyses described above, we defined ROIs using the SPM toolbox MarsBar (Brett et al., 2002). Regions of interest were defined as spheres with a radius of 3 mm centred on the identified peak coordinates. Voxels within an ROI were aggregated into a single contrast estimate using the first eigenvariate.

5.3 Results

5.3.1 Behavioural results

In Experiment IIIa, participants reported on average 51.9 ± 1.4% (mean ± SEM) words correctly per degraded sentence. Performance in clear trials was at 99.7± 0.2% correct.

We compared whether a linear or power law (Fig. 5.2A) would better describe the report scores increase over time by calculating Bayes Information Criterion (BIC) for each fit and each participant. The BIC scores for the linear fits (median 242.44, range 217.41–265.11) were smaller than those for power law fits (median 247.98, range 217.93–269.75), as shown by a Wilcoxon signed-rank test ($p < 0.001$), indicating that the linear curve better fit the behavioural data. Thus, we chose the linear fit to describe the participants' improvement ("adaptation curve").

A one-tailed *t*-test on the slopes of the adaptation curves showed that they were significantly greater than zero ($t(29) = 13.18$, $p < 0.001$), indicating that participants adapted to degraded speech, that is that their speech perception performance improved over the course of the experiment (Fig. 5.2A).

In Experiment IIIb, all participants performed well on the discriminable AM stimuli (mean ± SEM = 96.9 ± 0.7%). A paired samples *t*-test showed that there was no improvement from the first half (96.9 ± 0.7%) to the second half (96.7 ± 0.9%) of the experiment ($t(29) = 0.34$, $p = 0.74$).

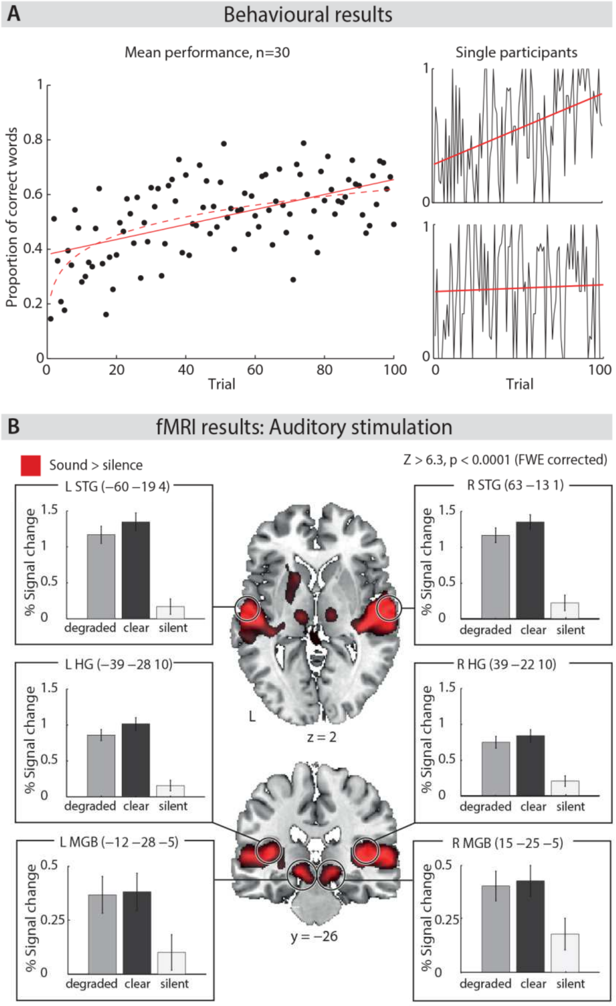


Figure 5.2. Behavioural and fMRI results (effects of auditory stimulation). (A) Behavioural adaptation to degraded speech. Linear (red solid line) and power law fit (dashed line) to mean report scores averaged over participants are shown on the left, and two examples of participants with a steep slope (upper panel) and a shallow slope (lower panel) in their linear fits ("adaptation curves") are displayed on the right. Although there is a general increase in performance, it is clearly visible that report scores strongly fluctuate from trial to trial (right). **(B)** Effects of auditory stimulation in Experiment IIIa. Auditory cortices (Heschls Gyrus, HG; Superior temporal gyrus, STG) and the medial geniculate body (MGB) were active when participants listened to sound (degraded and clear speech) compared to silent trials, confirming that the haemodynamic response to auditory stimulation was captured by the scans. L: left, R: right. Error bars represent SEM.

Controlling for speech-production-related activations. In order to dissociate perceptual and response demands, we estimated report latencies. Report latency in degraded speech trials (977 ± 72 ms, mean ± SEM) was significantly longer than in clear speech trials (662 ± 30 ms; $t(29) = 9.03$, $p < 0.001$). Moreover, report latency decreased over time, based on a *t*-test on the slopes of linear fits to report latencies as a function of trial ($-2.5 \pm 0.26 \times 10^{-3}$, mean ± SEM; $t(29) = 9.94$, $p < 0.001$). Therefore, report latency was regressed out in all fMRI analyses. Importantly, analyses without this regressor of no interest yielded very similar fMRI activation patterns (results not reported). We take this as strong evidence that the observed effects are not driven by speech production, but rather perception.

5.3.2 Functional MRI results

We found extensive activation of the auditory system when contrasting sound against silent trials, in Experiment IIIa (Heschls gyrus, HG; planum temporale; STG; medial geniculate body, MGB; see Fig. 5.2B) as well as in Experiment IIIb (results not shown), confirming that the BOLD response to auditory stimulation was captured by scan acquisition.

Effects of physical speech degradation: degraded versus clear speech. To reveal regions that are modulated by physical speech degradation or clarity, we compared vocoded with clear speech trials. Areas where degraded relative to clear speech yielded an increased BOLD signal included the left supplementary motor area (SMA) and ACC, anterior insula and caudate nucleus bilaterally. In contrast, clear compared to degraded speech yielded higher activations bilaterally in the precentral gyrus spanning the temporal cortices, supramarginal gyrus (SMG), putamen, posterior cingulate cortex and angular gyrus bilaterally (Fig. 5.3A and Table 5.1).

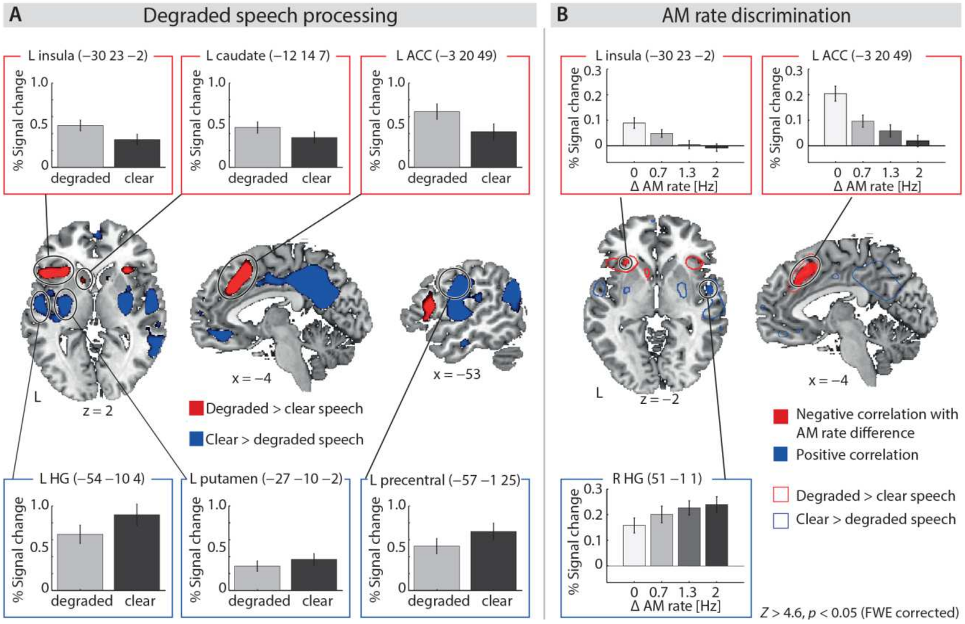


Figure 5.3. Degraded speech processing and AM rate discrimination. (A) Areas sensitive to speech degradation. Degraded relative to clear speech yielded increased activation in the anterior insula, the caudate, supplementary motor area /anterior cingulate cortex (ACC) and inferior frontal gyrus (IFG; in red). Conversely, clear compared to degraded speech revealed increased activity in the temporal cortices, putamen, posterior cingulate, precentral gyrus, supramarginal gyrus (SMG) and angular gyrus. Estimated percent signal change for vocoded, clear and silent trials is shown for selected regions. **(B)** Effects of ΔAM rate in Experiment IIIb and overlaps with Experiment IIIa. As AM rate difference (ΔAM rate) between standard

and deviant increased, activity reduced in the ACC and insula (red). Activity increased with higher ΔAM rate in right Heschls gyrus, left SMG and amygdala (blue). The contours of the Z-statistic maps showing effects of speech degradation are overlaid to illustrate overlaps in activation. A conjunction analysis between the effects of negative correlation with ΔAM rate and degraded > clear speech was significant in the ACC, insula bilaterally and the IFG; conjunction of the inverse contrasts (positive correlation with ΔAM rate ∩ clear > degraded speech) reached significance in the posterior cingulate and SMG bilaterally. L : left, R : right. Error bars represent SEM.

AM rate discrimination and degraded speech processing. The magnitude of the AM rate difference between standard and deviant stimuli (ΔAM rate) correlated positively with activity in right Heschls gyrus, left amygdala and SMG (see Fig. 5.3B and Table 5.1), signifying that activity increased with larger deviance from the standard (and thus easier AM rate discrimination). Conversely, ΔAM rate correlated negatively with activity in the ACC/SMA, left insula and IFG, meaning that the signal increased as modulation rate differences diminished and discrimination became more difficult.

In a series of conjunction analyses, we identified regions commonly involved in processing of degraded speech and of AM stimuli: A conjunction analysis between (1) positive correlations with ΔAM rate and (2) clear > degraded speech was significant in the SMG bilaterally and posterior cingulate. A conjunction of the inverse contrasts (negative correlation with ΔAM rate ∩ degraded > clear speech) yielded a significant cluster in the SMA/ACC, insula bilaterally and left IFG (Table 5.1). There were no commonly activated areas for the "cross-over" conjunction "negative correlation with ΔAM rate ∩ clear > degraded speech" or vice versa.

Trial-by-trial fluctuation with speech comprehension. To reveal areas reflecting trial-by-trial fluctuations in speech comprehension, we tested for correlations with the behavioural report scores. The BOLD signal linearly increased with improved comprehension of degraded speech in bilateral temporal cortices comprising Heschls gyrus, the STS, left IFG, the precentral gyrus bilaterally, the putamen, the thalamus including MGB bilaterally, left angular gyrus, the frontal medial cortex, posterior cingulate, and cerebellum (Fig. 5.4). There were no negative correlations between the fMRI signal and report scores. An additional voxel-by-voxel laterality analysis of this speech comprehension effect (see Methods) revealed that activity in the angular gyrus was significantly left-lateralised (Table 5.1). This is apparent in Fig. 5.4, where angular gyrus activity is seen in the sagittal slice of the left hemisphere only.

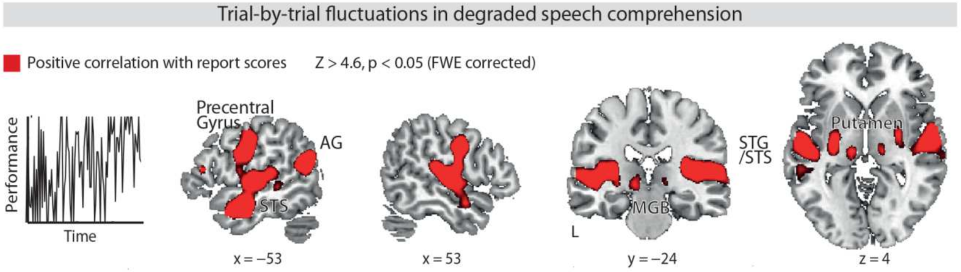


Figure 5.4. Areas varying with speech comprehension. With better speech comprehension on a given trial (graph shown on the left), activity increased in a network comprising the auditory cortices, left IFG, precentral gyrus, MGB, putamen, angular gyrus (AG), and cerebellum. There were no negative correlations with trial-by-trial report scores.

Effects of perceptual adaptation. To identify brain areas which change over time as a function of adaptation, we compared changes over time between degraded and clear speech while taking into account the behaviourally observed adaptation, that is we tested the time × condition × behaviour interaction (Fig. 5.5, upper panel). Thus, we separated changes over time due to adaptation (which only occurred in the degraded speech) from unspecific slow signal changes (which would be present in both conditions and hence be cancelled out in the contrast). As the behavioural change is marked for vocoded speech but virtually absent for clear speech, the BOLD signal is expected to change at different rates (i.e. more change in the vocoded condition because, on top of scanner drift, adaptation-related signal changes are present). Therefore, we modeled a more pronounced increase in the vocoded (quadratic increase) than in the clear speech condition (linear increase). This model separates uninteresting signal drifts from the effects due to the behaviourally observed adaptation.

At a threshold of p(FWE) < 0.05, we found stronger down-regulation over time in degraded relative to clear speech trials as listeners adapted in the ventral anterior thalamic nucleus (Morel, 2007). At the slightly more lenient threshold of $p < 0.001$ (cluster-extent corrected), the caudate nucleus, frontal regions and an occipital cluster spanning from the cerebellum to fusiform gyrus up to the precuneus showed the same effect. Conversely, an activity increase over time in degraded more than clear speech with adaptation was found in the left precentral gyrus and posterior cingulate cortex (Fig. 5.5; Table 5.1).

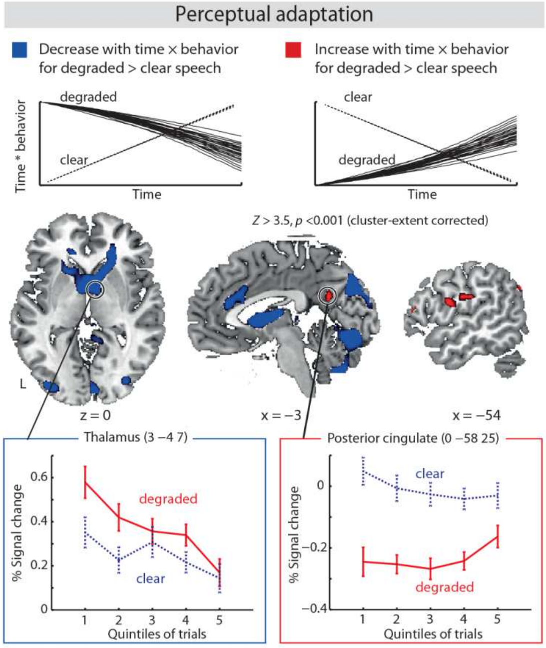


Figure 5.5. Effects of adaptation. To detect changes in activity due to adaptation, we compared the changes over time with performance increase between degraded and clear speech (upper panel). Stronger down-regulation for degraded relative to clear speech was observed in the ventral anterior thalamic nucleus, caudate, frontal, occipital and cerebellar regions (blue). Conversely, activity in the posterior cingulate and left ventral premotor cortex increased over time with adaptation (middle panel). To illustrate changes over time in the thalamus and posterior cingulate, we split up the experiment post-hoc into quintiles for which we separately calculated the contrast estimates for each condition (lower panel).

Table 5.1. Overview of fMRI activation clusters, thresholded at voxel-wise *p*(FWE)<0.05, unless indicated differently.

Location	**MNI-Coordinates**			***Z*-Score**	**Extent (mm^3)**
Degraded > clear speech					
L SMA/ACC	-6	26	40	7.19	3060
L anterior insula	-30	20	-5	7.05	3825
R anterior insula	33	23	-3	5.84	576
L caudate nucleus	-9	11	7	5.18	162
R caudate nucleus	12	17	10	5.45	171
Clear > degraded speech					
R superior frontal gyrus	6	59	28	6.55	7866
R precentral gyrus	54	-7	16	6.5	28647
R putamen	27	-10	-2	7.19	
R MTG	63	-19	-11	6.79	
R angular gyrus	51	-58	22	7.94	
L precentral gyrus	-54	-1	22	7.59	24183
L putamen/ globuspallidus	-27	-10	-2	6.87	
L Heschl’s gyrus	-54	-10	7	7.18	
L angular gyrus	-51	-67	34	7.24	
L MTG	-60	-13	-20	6.69	1044
L posterior cingulate	-6	-37	43	7.47	20700
Positive correlation with ΔAM rate					
R Heschl's gyrus	51	-1	1	5.95	270
L Amygdala	-24	-7	-14	5.37	135
L SMG	-57	-31	19	5.17	279
Negative correlation with ΔAM rate					
L SMA/ACC	-6	17	49	6.19	1143
L anterior insula	-30	23	-2	4.85	54
L IFG, pars opercularis	-48	14	28	5.16	270
Positive correlation with ΔAM rate ∩ clear > degraded speech					
L SMG	-57	-28	22	5.22	477
R SMG	60	-25	19	5.08	261
L posterior cingulate cortex	0	-40	46	5.14	486

(Table 1, continued)

Negative correlation with ΔAM rate ∩ degraded > clear speech					
L SMA/ACC	-6	17	49	7.15	1602
L anterior Insula	-30	23	-2	5.65	324
R anterior Insula	33	23	1	5.25	360
L IFG, pars opercularis	-48	14	25	5.17	315
Positive correlation with trial-by-trial comprehension fluctuations					
L frontal medial cortex	-3	50	-14	6.62	765
L IFG	-51	26	13	5.75	243
L STS	-57	-7	-20	7.47	18279
L precentral gyrus	-54	-10	34	6.58	
L Heschl's gyrus	-57	-13	4	6.67	
L putamen	-30	-16	-2	6.21	
L angular gyrus	-51	-61	19	7.50	
R precentral sulcus	57	-4	31	6.74	11475
R Heschl's gyrus	57	-13	4	6.7	
R putamen	30	-7	-8	6.34	
L thalamus	-12	-19	4	5.56	885
R thalamus	12	-19	4	5.57	306
L posterior cingulate	-3	-52	16	6.22	2376
R cerebellum	15	-61	-23	6.08	3141
L cerebellum	-12	-64	-23	5.73	
Lateralisation of comprehension network					
L angular gyrus	-45	-70	37	5.83	378
Decrease with time × behaviour for degraded > clear speech, *p*< 0.001, *k*> 20					
L superior frontal gyrus	-27	47	31	3.79	666
R MFG	33	47	22	4.89	729
R anteroventral thalamus	3	-4	7	5.07	7929
R caudate	18	20	-5	4.48	
L fusiform gyrus	-24	-61	-14	4.84	19116
Increase with time× behaviour for degraded > clear speech, *p*< 0.001, *k*>20					
L precentral gyrus	-54	-1	19	3.8	189
Posterior Cingulate	0	-58	25	4.05	333

L: left, R: right, ACC: anterior cingulate cortex, IFG: inferior frontal gyrus, MFG: middle frontal gyrus, MTG: middle temporal gyrus, SMA: supplementary motor area, SMG: supramarginal gyrus, STG: superior temporal gyrus, STS: superior temporal sulcus.

5.4 Discussion

The present study was designed to reveal neural systems that support rapid perceptual adaptation to degraded speech and contributes three major novel findings: First, degraded more than clear speech activates an “executive” system which overlaps with the neural substrates of difficult auditory discrimination, namely in the insula and SMA/ACC. Second, BOLD signal in a speech comprehension or “language” network, comprising auditory, premotor cortices and left angular gyrus, depended on speech comprehension rather than physical clarity of the stimulus. Third, the data provide first evidence that self-regulated perceptual adaptation to degraded speech co-occurs with a BOLD down-regulation of subcortical structures.

5.4.1 Executive mechanisms in speech and non-speech processing

Degraded more than clear speech evoked enhanced activity in anterior areas including the insula and the SMA/ACC. This has been observed consistently for difficult comprehension (Giraud et al., 2004; Eckert et al., 2009). Clear speech, in contrast, revealed an expected activity increase in the bilateral temporal cortices (Scott et al., 2000; Davis and Johnsrude, 2003; Giraud et al., 2004; Wild et al., 2012)

When examining to which extent difficult speech and non-speech perception rely on joint neural substrates, a conjunction revealed SMA/ACC and the anterior insula bilaterally. This comparison pertains directly to contributions of bottom-up (i.e., fidelity of AM representations in the ascending auditory pathway) versus top-down mechanisms (e.g., attentional processes) in degraded speech perception. For example, frequency discrimination learning shows partial specificity to the trained frequency (Amitay et al., 2006), consistent with a bottom-up account. In contrast, top-down mechanisms of attention are plausibly involved in perceptual learning (Halliday et al., 2011). Training improves the ability to attend to a task-specific stimulus dimension, and discrimination learning occurs even in the absence of a discriminable stimulus difference (Amitay et al., 2006). Similarly, speech degradation studies have found evidence for bottom-up accounts (Sebastian-Galles et al., 2000; Hervais-Adelman et al., 2008; Idemaru and Holt, 2011) as well as top-down accounts of perceptual learning, in which lexical information aids perceptual adaptation (Davis and Johnsrude, 2003; Davis et al., 2005).

The structures commonly recruited for degraded speech processing and AM discrimination are clearly not specific to auditory envelope processing, but have been suggested by a number of studies to be involved rather in top-down, executive (e.g., attentional) processes (Adank, 2012). Eckert et al. (2009) demonstrated that the insula and SMA/ACC are engaged when tasks become increasingly difficult, independent of modality or task, suggesting that these regions subserve executive processes. Consistently, the anterior insula and SMA showed an enhanced BOLD signal when listeners attended to speech (rather than a distracter), and the speech signal was increasingly degraded (vocoded rather than clear; Wild et al. 2012b). More specifically, these regions are a resource for attention and performance monitoring processes (Dosenbach et al., 2006; Dosenbach et al., 2007; Sadaghiani and D'Esposito, 2012). Thus, we argue that the anterior insula and SMA/ACC fulfill executive processes, and that the recruitment of these executive components is pivotal for a wide range of challenging listening situations.

5.4.2 Revisiting the speech comprehension network

A number of earlier imaging studies manipulated physical stimulus features to vary speech intelligibility and found sensitivity to these manipulations along the superior temporal gyrus and sulcus, often extending into prefrontal and inferior parietal regions (Scott et al., 2000; Davis and Johnsrude, 2003; Zekveld et al., 2006; Obleser et al., 2007; Obleser and Kotz, 2010). In contrast, the present study held physical stimulus properties constant (i.e., four-band vocoding). Therefore, we were able to identify regions where activation varied with actual speech comprehension (i.e., behavioural report scores), independent of acoustic differences.

For a given trial, activity in a perisylvian network was tightly coupled to comprehension. Importantly, these areas (Fig. 5.4) overlapped largely with those activated by clear compared to degraded speech (Fig. 5.3A, blue *Z*-statistic maps). Taken together, this is strong evidence that the observed network supports sensorimotor operations involved in successful comprehension (and, hence, successful repetition) rather than simply indexing sensitivity to physical stimulus characteristics.

Such linguistic processes of comprehension have been proposed to be largely processed in the left hemisphere (McGettigan et al., 2012a; Peelle, 2012). Concordantly, we found evidence for a left-lateralisation of activity in the angular gyrus, a structure which is associated with semantic processing

(Ferstl and von Cramon, 2002); for review see (Price, 2012) and has been suggested to facilitate speech comprehension when signal quality declines (Obleser et al., 2007; Obleser and Kotz, 2010).

5.4.3 Cortical contributions to perceptual adaptation

One important objective here was to identify and track on-line the neural systems supporting adaptation to degraded speech. Depending on adaptation, activity increased over time in cortical areas of the premotor cortex and posterior cingulate and decreased in frontal and occipital areas.

The premotor cortex has been suggested to mediate successful perceptual learning of degraded speech by mapping the unfamiliar auditory signal onto existing articulatory representations of speech sounds. Consistent with this idea, adaptation to time-compressed speech is reflected in a down-regulation of activity in the ventral premotor cortex (Adank and Devlin, 2010). Hervais-Adelman et al. (2012) observed precentral sulcus activation when a vocoded word was paired with its clear representation, a type of feedback which is known to enhance perceptual learning (Davis et al., 2005). Similarly, we note in the present study, although with feedback-free learning, that activity in the premotor cortex increases as a listener adapts to vocoded speech. Thus, perceptual learning might be rooted in sensorimotor integration mediated by speech–productive areas.

Enhanced activity in the posterior cingulate cortex has been noticed by Obleser et al. (2007) when semantic context helped comprehension of degraded speech at an intermediate signal quality (eight-band vocoded speech). This is commensurate with the current data where posterior cingulate activity not only correlated with successful speech comprehension at a given trial (Table 5.1), but increased as a listener adapted to degraded speech. Hence, adaptation-related increase of posterior cingulate activity might directly relate to its facilitating role in degraded speech comprehension.

Although the present experiment did not involve visual manipulations, we found occipital cortex activity (fusiform gyrus, precuneus) to correlate negatively with adaptation. While the fusiform gyrus has been implicated in audiovisual speech perception (Stevenson et al., 2010; Nath and Beauchamp, 2011; McGettigan et al., 2012b), von (von Kriegstein et al., 2003) observed modulation of fusiform gyrus activity during auditory speech perception even in the absence of visual stimuli. Similarly, (Giraud et al., 2001b) have established that visual cortex contributes to auditory speech processing in CI patients but also in normal-hearing listeners, especially when listening to meaningful sentences (Giraud and Truy, 2002). They hypothesised that auditory-to-visual cross-modal interaction contributes to semantic processing. In line with these studies, we speculate that initial recruitment of the visual cortex might help a listener extract meaning when first confronted with a novel form of speech degradation.

A previous study that used a feedback-based vocoded-speech learning paradigm simulating CI rehabilitation programs (Eisner et al., 2010), found the IFG to be involved in successful adaptation. The authors attributed to the IFG a role in the "specific use of simultaneous written feedback to enhance comprehension" (p. 7183), in line the view of the IFG serving as integration site for different sources of information necessary for speech comprehension (Hagoort, 2005; Rauschecker and Scott, 2009). In the present study, we did not observe IFG involvement in adaptation. Likely due to the absence of feedback, listeners might have relied on substantially different neural mechanisms to adapt to degraded speech.

5.4.4 Subcortical contributions to perceptual adaptation

A major novel contribution of the present study was the detection of subcortical involvement in perceptual adaptation. In previous imaging studies, such subcortical contributions might have been obscured due to heartbeat-related artefacts, which we avoided by use of a cardiac-gated scanning protocol.

We identified the ventral anterior nucleus of the thalamus to be functionally involved in adaptation (Fig. 5.5); note that this structure is proximal to but not identical to the pulvinar, where we found structural differences to be predictive of degraded speech adaptation before (Experiment I). Adaptation-related decrease in activity also encompassed the caudate. Although the basal ganglia have primarily been implicated in motor function, there is accumulating evidence that they play an important role in language processing (Lieberman et al., 1992; Fiebach et al., 2003; Kotz, 2006; Kotz et al., 2009). Presumably, they exert their function in language processing through their high connectivity to the cortex (Crosson, 1999): The ventral anterior thalamic nucleus and the caudate form part of the striato-thalamo-cortical loop which is proposed to collect cortical information, funnel and converge it at cortical output areas, thereby reconfiguring cortical activation patterns (Kemp and Powell, 1970; O'Connell et al., 2011). In the context of adaptation to degraded speech, naïve listeners might at first be forced to rely more on information-selection processes supported by the striato-thalamo-cortical loop. Engagement of this pathway is likely to sharpen the cortical representation of a stimulus and ultimately lead to a convergence of the degraded speech signal onto a clear-speech "representation", allowing for enhanced comprehension.

5.4.5 Conclusions

The present work elucidates the central neural mechanisms of rapid adaptation to acoustic speech degradation with respect to three points. First, when listening tasks become increasingly difficult, in the speech as well as the non-speech domain, listeners rely on a common executive network for "effortful listening" (Eckert et al., 2009), involving the SMA/ACC and anterior insula. Second, a perisylvian network subserves speech comprehension and fluctuates with actual comprehension rather than physical stimulus features. Finally, the present data advance the understanding of how a listener adapts to a degraded speech input, demonstrating that rapid adaptation is partly explained by haemodynamic down-regulation in subcortical structures.

6 Experiment IV: Neural networks in older adults' speech comprehension[7]

6.1 Introduction

Speech comprehension can become difficult with age and age-related hearing loss, especially when listening conditions are challenging. Normal-hearing young adults have the capacity to rapidly adapt to degraded speech (Davis et al., 2005; Samuel and Kraljic, 2009; Eisner et al., 2010; see Section 1.1.2). Such short-term perceptual adaptation is not well established in older adults, although it bears particular relevance as older adults are frequently affected by hearing loss. For example, patients with hearing-aids or, more drastically, CI patients need to adapt to an altered and often distorted auditory signal delivered by their device.

In the previous chapter, we have shown in a cohort of young adults that degraded speech processing elicits an increased BOLD response in an "executive" network (Eckert et al., 2009) comprising the anterior insula and ACC. Also, adaptation to degraded speech was shown to be accompanied by haemodynamic down-regulation in a cortico-thalamic-striatal loop (Experiment III). In the current fMRI experiment, we compare these results to a group of older adults with varying degrees of hearing loss to test (1) whether older listeners are able to (behaviourally) adapt to spectrally severely degraded ("noise-vocoded") speech at a rate comparable to young listeners, and (2) how neural processing of degraded speech differs between young and older adults.

---

[7] This chapter has been adapted from the article published in *Frontiers in Systems Neuroscience, Special issue: The effect of hearing loss on neural processing* by Erb and Obleser (2013).

There is evidence that rapid perceptual learning is preserved in older adults (Peelle and Wingfield, 2005; Golomb et al., 2007; Gordon-Salant et al., 2010). For example, older adults are able to quickly adapt to an unfamiliar accent (Janse and Adank, 2012) or a foreign accent (Gordon-Salant et al., 2010). Peelle and Wingfield (2005) showed that older adults adapted to time-compressed and noise-vocoded speech at a similar rate as young adults when starting accuracy was equated.

However, considerable inter-individual variability has been frequently observed in adaptation to vocoded speech (see Chapter 3; Shannon et al., 2002, Eisner et al., 2010). Especially in older adults, the degree of hearing loss and cognitive aspects might substantially impact adaptation to degraded speech. Working memory is one cognitive factor that has been implicated in degraded speech comprehension by a number of studies (Pichora-Fuller et al., 1995; Burkholder et al., 2005; Jacquemot and Scott, 2006; Eisner et al., 2010; Piquado et al., 2010; Obleser et al., 2012). For example, Pisoni and Cleary (2003) observed that working memory scores as measured by digit span significantly predicted speech comprehension in pediatric CI users. In older adults, cognitive factors might be even more closely related to degraded speech recognition (Janse and Adank, 2012), because cognitive decline has been shown to co-occur with sensory decline (Lindenberger and Ghisletta, 2009), which in turn leads to degraded auditory conditions. Thus, we expected working memory capacity in older adults to be related to degraded speech comprehension.

A second factor which heavily affects comprehension is hearing loss. As age-related hearing loss is accompanied by auditory cortex atrophies (Harris et al., 2009; Peelle et al., 2011; Eckert et al., 2012), older adults likely draw on different neural resources for speech comprehension. It is a common observation that older adults recruit additional regions for speech comprehension compared to young adults, although it is uncertain whether this reflects an age-related loss of specialisation of cortical brain regions (Park et al., 2004) or a compensatory mechanism (Cabeza et al., 2002). Peelle et al., (2010b) noted that during processing of syntactically complex sentences, older adults engaged middle frontal regions in addition to a “core sentence-processing network” (comprising MTG and IFG; Peelle et al., 2004; Fiebach et al., 2005) recruited by young adults. The authors interpreted this engagement of additional areas as a compensatory process, whereby the older adults managed to maintain performance despite degeneration of the sensory cortices.

In line with this argumentation, older adults have been hypothesised to engage in more effortful processing during speech comprehension (Pichora-Fuller, 2003). Consistently, Eckert et al. (2008) observed an age-related upregulation of cognitive-control-related frontal areas during an easy word recognition task, while younger adults recruited these areas merely in difficult listening conditions. Harris et al. (2009) further showed that incorrect versus correct word recognition elicited increased activity in the ACC, but more so in older than younger adults, possibly reflecting an age-related upregulation of error monitoring systems (Sharp et al., 2006). Hence, for solving auditory tasks the reliance on cognitive control appears to increase with age.

It is still largely unknown, however, how older adults adapt to degraded speech. In the current fMRI study, we are primarily interested in how short-term adaptation to degraded speech and the involvement of cognitive control networks in speech processing changes with age. Young and older listeners heard and repeated back 100 degraded (four-band-vocoded) sentences as well as a control set of interspersed clear-speech sentences. Thus, we could identify age differences in the neural processes related to both, physical degradation of speech (degraded vs. clear sentences) and trial-by-trial fluctuations in comprehension (covariation of BOLD responses with degraded speech comprehension success).

6.2 Materials and methods

6.2.1 Participants

Sixteen older adults (aged 56–77, mean 67.1 years, 8 female) participated in the study. Their data were analysed jointly with the cohort of 30 young adults (aged 22–31, mean 25.9 years, 15 female) who had participated in Experiment III. Participants were recruited from the participant database of the Max Planck Institute for Human Cognitive and Brain Sciences according to the following criteria: They were native speakers of German; had no language or neurological disorders; showed dominant right-handedness according to the Edinburgh inventory (Oldfield, 1971) and were naïve to noise-vocoded speech. Younger adults self-reported normal hearing (see Section 5.2.1), whereas older adults had normal hearing to moderate hearing loss based on their pure-tone averages (PTA) which were audiometrically assessed (see below). Participants received financial compensation of € 16, and gave informed consent. Procedures were approved by the local ethics committee (University of Leipzig).

Audiometric evaluation. Older adults' pure-tone thresholds were measured at conventional frequencies from 0.25–8 kHz using an Equinox 2.0 AC-440 audiometer (Interacoustics) in a sound-proof chamber. Older participants' PTA (defined as the average thresholds in the listener's better ear at 1, 2 and 4 kHz) indicated normal hearing (< 25dB HL) to moderate hearing loss (40 - 70 dB HL), whereas high-frequency hearing ranged from normal to severe hearing loss (70 – 95 dB HL; audiograms are shown in Fig. 6.1A). Young participants' hearing acuity was not tested but all of them self-reported normal hearing.

Auditory forward and backward digit span. To measure working memory capacity, all participants were tested with a digit span test (WMS-R; Wechsler, 1987). This test is described in detail in Section 2.5.1). For simplicity of the testing procedure, the experimenter read out to the participant lists of single digits between 1 and 9 at a rate of approximately one digit per second, rather than playing recordings (Experiment I). The level at which the test was terminated was taken as the individual forward or backward digit span measure ("level", see Section 2.5.1).

6.2.2 Stimuli and experimental design

Stimuli were German SPIN sentences, which control for the predictability of the final word (Kalikow et al., 1977; see Section2.1 and Appendix). The same low-predictability sentences were used as in Experiments I and III, such that semantic cues were limited and the listener had to rely primarily on acoustic properties of the sentence to understand it. A complete list of these sentences is available in the appendix.

Noise-vocoding was applied to all sentences in MATLAB 7.11 as described in Section 2.2 using four frequency bands spanning 70 to 9000 Hz (filters and frequency bands used are specified in Section 2.2).

As in Experiment III, each trial was approximately 9 s long, but actual trial length varied due to cardiac gating (see Section 2.6.1). Trials started with a 1 s silent gap, after which participants heard a sentence lasting for approximately 2.5 s. Following stimulus presentation (3.5 s into the trial), a green light was visually presented and lasted for 3 s. During this time, participants were to respond by repeating the sentence, but were instructed to stop talking when the green light disappeared in order to avoid movement during scan acquisition. After approximately 1 s of silence, scan acquisition with a TR of 2 s was triggered using cardiac gating (trial timing was identical as in Experiment III, see Fig. 5.1B). Verbal responses were recorded for later off-line scoring (Eckert et al., 2009; Harris et al., 2009) and responses were scored as proportions of correctly repeated words per sentence ("report scores", Peelle et al., 2013). Scoring took into account all words of a sentence (see Section 5.2.2) .

As in Experiment III, clear-speech trials were used as a high-level baseline. Overall, the experiment comprised three conditions: (1) four-band vocoded sentences ("degraded speech"; 100 trials); (2) clear (non-degraded) sentences ("clear speech"; 24 trials); (3) trials lacking any auditory stimulation ("silent trials"; 20 trials), summing up to 144 trials in total. Clear speech trials were presented every 5th trial, whereas silent trials were randomly interspersed. Sentences were presented to each participant in one of four different orders; presentation order was counterbalanced across participants.

Individual adaptation curves. To measure individual learning rate, we modelled learning curves in two different ways: as power law or as linear performance increase (for an example fits to the older group's average performance see Fig. 6.1B). As in Experiments I and III, we compared goodness of fit by determining the Bayesian information criterion (BIC; Schwarz, 1978) of both fits within each participant (see Section 2.3).

6.2.3 Experimental procedure

Before participants went into the scanner, they were familiarised with the task by listening to three eight-band vocoded sentences.

To prevent hearing damage due to scanner noise, participants wore Alpine Musicsafe Pro earplugs while in the bore, yielding approximately linear 14-dB reduction in sound pressure up to 8 kHz. Auditory stimuli were delivered through MR-Confon headphones using Presentation software. Presentation level was adjusted for each participant such that loudness was subjectively comfortable and equal across both ears. This ensured that stimuli were presented at a level well above participants' thresholds in the speech range frequencies such that all participants were able to perceive the sentences. Visual prompts were projected on a screen which participants viewed via a mirror.

6.2.4 MRI data acquisition

Functional MRI data were collected on a 3-T Siemens Verio scanner. Functional MR images were acquired with a 12-channel head coil using an EPI sequence [TR ≈ 9000 ms, TA = 2000ms, TE = 30 ms, flip angle = 90°, 3 mm slice thickness, 30 axial slices (ascending), interslice distance = 1 mm, acquisition matrix of 64 x 64, voxel size = 3 x 3 x 3 mm]. The acquisition matrix was placed such that the x-axis was in line with the anterior–posterior commissure. We used a sparse-sampling procedure, allowing for silent periods to play stimuli and record responses (Hall et al., 1999). As in Experiment III, cardiac gating was applied to avoid movement artefacts caused by the heartbeat in subcortical structures (von Kriegstein et al., 2008; see also Section 2.6.1).

Following functional imaging, young participants received a T1-weighted structural brain scan with a 32-channel head coil using an MPRAGE sequence [TR = 1300 ms, TE = 3.5ms, flip angle = 10°, 1 mm slice thickness, 176 sagittal slices, acquisition matrix of 256 x 240, voxel size = 1mm^3].

Older adults' anatomical scans for registration with the functional images were available through the Institute's brain database. Scanning had been carried out on a 3T Siemens Trio TIM scanner using T1-weighted MPRAGE sequence to acquire 176 sagittal slices, with an acquisition matrix of 256 x 240, yielding a resolution of 1mm^3.

For one older participant the scanner had become desynchronised with the presentation script, such that he had to be excluded from the fMRI data analyses, resulting in a total of 15 older participants included in the analyses. In one young participant, we were only able to acquire 136 (as opposed to 144) scans due to technical problems with cardiac gating.

6.2.5 Data analysis

Note that all behavioural and MRI analyses of single participants closely matched the procedures established previously in Experiment III.

Preprocessing. MRI data were analysed in SPM8. Functional images were realigned and unwarped using a fieldmap, coregistered with the structural scan, segmented and normalised to MNI space using the segmentation parameters, and smoothed with an isotropic Gaussian kernel of 8 mm full-width at half-maximum.

Statistical analyses. MR images were statistically analysed in the context of the GLM. We set up identical models as in Experiment III (see Section 5.2.5). One model assessed effects of speech degradation and effects of trial-by-trial-fluctuations in comprehension. In this model, we included two conditions: degraded and clear speech. To avoid overspecification, silent trials were not modelled in the analyses, but solely used for an initial quality check of the data confirming that sound compared to silent trials yielded large clusters of activity in the auditory cortex. Additionally, a parametric modulator of the degraded speech trials was defined, representing the behavioural report scores. A regressor of no interest, containing report latencies, was added in order to account for differences in speech production (analysis explained in detail in Section 5.2.5, "Regressor of no interest for report latency"). To assess effects of stimulus clarity, we contrasted degraded against clear speech trials. To reveal effects of trial-by-trial fluctuations in speech comprehension, we assessed correlations with the regressor representing report scores. To look for effects of hearing loss, we correlated PTA on the second level with the contrast degraded > clear speech.

All described analyses were whole-brain analyses. Regressors were modelled using a FIR comprising one bin. A high-pass filter with a cut-off of 1024 s was applied to eliminate low-frequency noise. No correction for serial autocorrelation was necessary because of the long TR in the sparse-sampling acquisition.

Second level statistics were calculated using a one-sample *t*-test and group differences were assessed using a two-samples *t*-test. We recognise that comparisons between groups of different sample sizes (here: 15 older adults versus 30 younger adults) are problematic; especially when variance differs between groups, the group with the larger variance should comprise more samples (Samanez-Larkin and D'Esposito, 2008). There is evidence, however, that BOLD signal variability actually decreases in older adults (Garrett et al., 2010). Further, we wanted to avoid discarding data which were already available for the 30 young adults, resulting in a larger sample size of the young compared to the older adults.

We are aware of the problem that haemodynamics likely change with age (e.g., due to vascular changes), such that simple group differences in the BOLD signal could possibly reflect differences in neurovascular coupling rather than actual differences in neural processing. To overcome this issue, we only tested for age group × condition interactions when assessing age effects on neural processing (Samanez-Larkin and D'Esposito, 2008).

Group inferences are reported at a threshold of $p < 0.001$ and a cluster-extent of $k > 20$ to correct for inflated type I error at the whole-brain level, as based on a Monte Carlo Simulation (Slotnick et al., 2003). T-statistic maps were transformed to Z-statistic maps, and overlaid and displayed on the ch2 template in MNI space included with MRIcron (Rorden and Brett, 2000).

Regions of interest analyses. In order to extract measures of percent signal change in the regions identified by the whole-brain analyses described above, we defined ROIs using the SPM toolbox MarsBar (Brett et al., 2002). Regions of interest were defined as spheres with a radius of 3 mm centred on the identified peak coordinates and voxels within an ROI were aggregated using the first eigenvariate.

6.3 Results

6.3.1 Behavioural results

Vocoded speech comprehension. Older adults reported on average 28.0 ± 2.8% (mean ± SEM) words correctly per degraded sentence. Performance in clear trials was at 98.0 ± 1.4% correct. In comparison, young adults' degraded speech recognition was substantially better ($t(44) = 8.23$, $p < 0.001$), with on average 51.9 ± 1.4% words correct per degraded sentence and 99.7 ± 0.2% correct per clear sentence (Fig. 6.1D, left).

Perceptual adaptation. We compared whether a linear or power-law fit would better describe the report scores' increase over time by calculating BIC for each fit and each older participant (Fig. 6.1 B, C). The BIC scores for the linear fits (median 222.09, range 95.21–268.48) were smaller than those for power law fits (median 226.69, range 99.81–273.09), as shown by a Wilcoxon signed-rank test ($p < 0.001$), indicating that the linear curve better fit the behavioural data. In the young participants, we had also shown that linear fits were more adequate than the power law fits to describe participants' learning curves (see Experiment III). Thus, the slope of the linear fit (adaptation slope) was taken as a measure of individual perceptual adaptation to degraded speech.

The BIC compares goodness of fit between different models but does not give an estimate of absolute goodness of fit. Absolute goodness of fit as measured by R^2 in the older adults amounted to $R^2 = 0.061 \pm 0.01$ (mean ± SEM) for the power law fit and $R^2 = 0.072 \pm 0.016$ for the linear fit. R^2 did not differ between the linear and the power law fit, as shown by a Mann-Whitney U test ($p = 0.95$). Note however, that a direct comparison of R^2 between different models does not allow for a fair comparison, as R^2 does not take into account the number of fitted parameters.

In order to test whether goodness of fit differed between age groups, we compared R^2 of the linear fits in the two groups. R^2 in the older adults did not differ from R^2 in the younger adults (0.065 ± 0.009), as shown by a Mann-Whitney U test ($p = 0.92$), indicating that the goodness of fit was comparable in older and younger adults. Although these single-subject R^2 values are relatively small (only approximately 7% of the variance is explained by the fitted model), mean report scores correlated highly with sentence number ($r = 0.57$, $p < 0.001$) when averaging over the sixteen older adults. Davis et al. (2005) have reported similar correlation coefficients for mean report scores with sentence number in their vocoded speech learning study (but did not report single-subject R^2-values).

To confirm that the slopes of the linear fits were a sensible measure of learning, we used another more traditional measure of learning. For each participant, we subtracted the mean performance over the last 20 trials from the mean performance over the first 20 trials (Bent et al., 2009; Eisner et al., 2010). This performance difference (Δ performance) between the beginning and end of the experiment amounted to 0.22 ± 0.03 proportion correct (mean ±SEM) in the older adults and 0.27 ± 0.02 proportion correct in the younger adults. Across age groups, Δ performance was highly correlated with the adaptation slopes ($r = 0.91$; $p < 0.001$).

To test whether older and younger adults differed in their rate of learning, we compared both the Δ performance and adaptation slopes between groups. According to a two-samples *t*-test, there was neither a difference in Δ performance between older and young adults ($t(44) = -1.25$, $p = 0.22$), nor in the adaptation slopes of older ($2.5 \pm 0.39 \times 10^{-3}$, mean ± SEM) and young adults ($2.7 \pm 0.21 \times 10^{-3}$; $t(44) = -0.65$, $p = 0.52$), indicating that both groups were comparable in the rate with which they adapted to degraded speech (Fig. 6.1D, right).

Finally, to exclude the possibility that variability in the adaptation curves (shown in Fig. 6.1C) was a consequence of the counterbalancing of material across participants (in four different presentation orders), we tested whether presentation order had an effect on adaptation slope. A Kruskal-Wallis test was not significant ($\chi^2(15) = 4.34$, $p = 0.23$), indicating that the fact that different listeners received different materials at different time points did not influence adaptation.

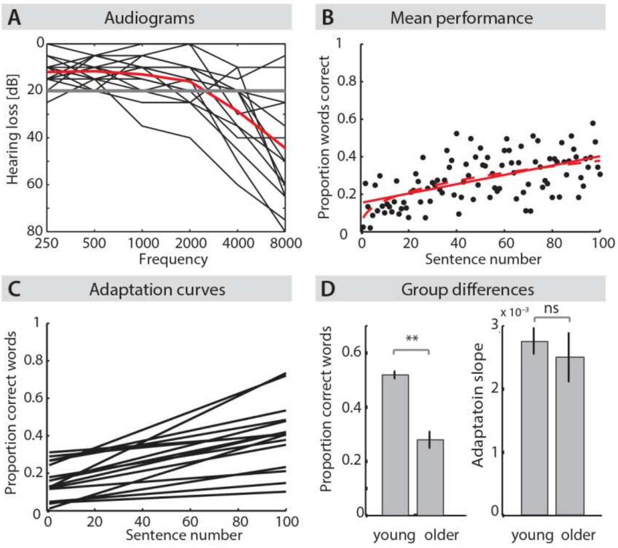


Figure 6.1. Behavioural results. **(A)** Audiograms of older adults. Older participants are affected by varying degrees of sensorineural hearing loss. Hearing loss for each participants' better ear (black line) and mean over all participants (red line) are shown. **(B)** Mean report scores (averaged over all older participants) with a linear (red solid line) and power law fit (red dashed line). **(C)** Older adults' adaptation curves. Linear fits to report scores, showing individual differences in adaptation to degraded speech. The slope of the linear fit was taken as measure for perceptual adaptation to degraded speech (adaptation slope). **(D)** Group comparison of mean speech comprehension and adaptation. Young and older adults differ in mean degraded speech comprehension but not adaptation slopes (**$p < 0.001$, ns = non-significant).

Spearman's correlations. In order to examine whether in older adults, adaptation slope was related to other factors (i.e., age, forward and backward digit span, PTA, mean performance), a two-tailed Spearman's correlation was calculated for all pairs of variables. We adjusted for multiple comparisons by controlling the false discovery rate (Benjamini, 1995), which resulted in a critical $p = 0.019$ at a false discovery rate of $q = 0.05$ ("significant") and $p = 0.039$ at $q = 0.1$ ("non-significant trend").

Within the older adults, adaptation slope correlated positively with backward digit span (Fig. 6.2A and Table 6.1), indicating that listeners with a larger working memory capacity adapted faster to degraded speech. Note that this correlation remained significant when the outlier participant with a slope of 0.006 was removed ($\rho = 0.62$, $p = 0.18$). Pure-tone average and age showed a non-significant trend for a negative correlation with adaptation slope, meaning that older age and hearing loss were associated with slower adaptation rates. Similarly, older age and PTA were negatively related to average speech comprehension performance. Finally, age also correlated significantly with PTA, indicating that older adults had greater hearing loss (Table 6.1).

Table 6.1. Spearman's correlations between behavioural measures.

	Slope	Age	DSF	DSB	PTA	Perform
Slope		–0.54+	0.37	**0.60***	–0.55+	0.45
Age			0.01	–0.18	**0.65***	**–0.67***
DSF				0.25	–0.26	0.10
DSB					–0.44	0.09
PTA						**–0.61***
Perform						

Significant correlations are shown in bold. * significant at $q < 0.05$; + significant at $q < 0.1$; Slope: adaptation slope, DSF: digit span forward, DSB: digit span backward, PTA: pure-tone average, Perform: Mean vocoded speech comprehension.

To analyse more closely the relationship between adaptation slope, age, digit span and hearing loss, we calculated Spearman's partial correlation coefficients between these four variables of interest, with two variables partialled out at a time (Fig. 6.2B). Adaptation slope still correlated significantly with backward digit span, even after partialling out age and hearing loss, indicating that the latter could not explain the relationship between working memory and adaptation. The correlation between PTA and age also remained significant in the partial correlation. On the other hand, the non-significant trend for a negative correlation of adaptation slope with hearing loss and age broke down in the partial correlation (Fig. 6.2B).

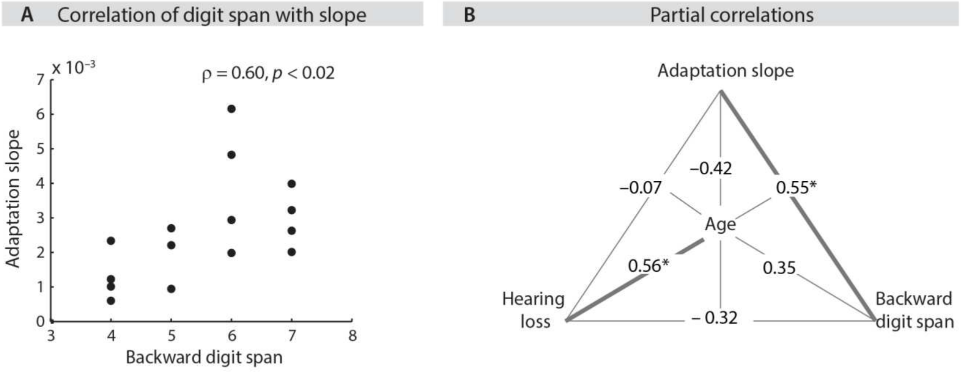


Figure 6.2. Spearman's correlations between behavioural measures. **(A)** Adaptation slope correlated positively with backward digit span. This correlation remained significant when the outlier participant with a slope of 0.006 was removed ($\rho = 0.62$, $p = 0.18$). **(B)** Partial correlations between adaptation slope, age, backward digit span and hearing loss (pure-tone average; PTA). Spearman's correlation coefficients are shown. Note that two measures are partialled out at a time; for example, the correlation between adaptation slope and digit span of $\rho = 0.55$ is controlled for age and hearing loss. Only the adaptation–digit span correlation and the age-hearing loss correlation remain significant in the partial correlation. Significant correlations are shown as bold lines (* $p < 0.05$).

6.3.2 Functional MRI results in older adults

The results reported below refer to the group of older adults exclusively. For the cohort of young adults, activation clusters and coordinates of peak activity are described in detail in Chapter 5.

Degraded speech processing. To reveal regions that are engaged in degraded speech processing, we compared degraded with clear speech trials. At a cluster-extent corrected threshold of $p < 0.001$, degraded compared to clear speech elicited an increased BOLD signal in the left anterior insula. On the other hand, clear compared to degraded speech yielded higher activations bilaterally in the precentral gyrus spanning the temporal cortices, supramarginal gyrus (SMG), right putamen, posterior cingulate cortex and angular gyrus bilaterally (Fig. 6.3A).

Trial-by-trial fluctuations in degraded speech comprehension. To identify areas where the BOLD signal varied with trial-by-trial fluctuations in speech comprehension, we tested for correlations with the behavioural report scores. The haemodynamic response linearly increased with comprehension of degraded speech in bilateral temporal cortices comprising Heschl's gyrus, the middle temporal gyrus, the precentral gyrus bilaterally, the putamen bilaterally, left angular gyrus, and middle frontal gyrus (Fig. 6.3B). Although scans might have been sensitive to both speech perception and production, we controlled for report latency (see Materials and Methods) to model the haemodynamic response to auditory input rather than speech production. Note, however, that including a regressor for report latency might not completely control for production-related activity, such that motor cortical activity (apparent in Figures 6.3A and 6.3B), could be due to more speech production during more intelligible compared to less intelligible trials.

Correlation with hearing loss. We further tested, whether hearing acuity had an influence on neural processing. For the contrast degraded > clear speech, we found a negative correlation with PTA in the right and left anterior insula (Table 6.2 and Fig. 6.3C). The correlation showed the following pattern: Older adults with better hearing acuity had elevated activity for degraded compared to clear speech. Conversely, listeners with greater hearing loss had an increased BOLD signal for clear relative to degraded speech in the anterior insula (Fig. 6.3C, right panel). This Pearson's correlation between insula dynamics and hearing loss was significant even after partialling out age ($r = -0.79$, $p = 0.001$), confirming that the correlation was not driven by age.

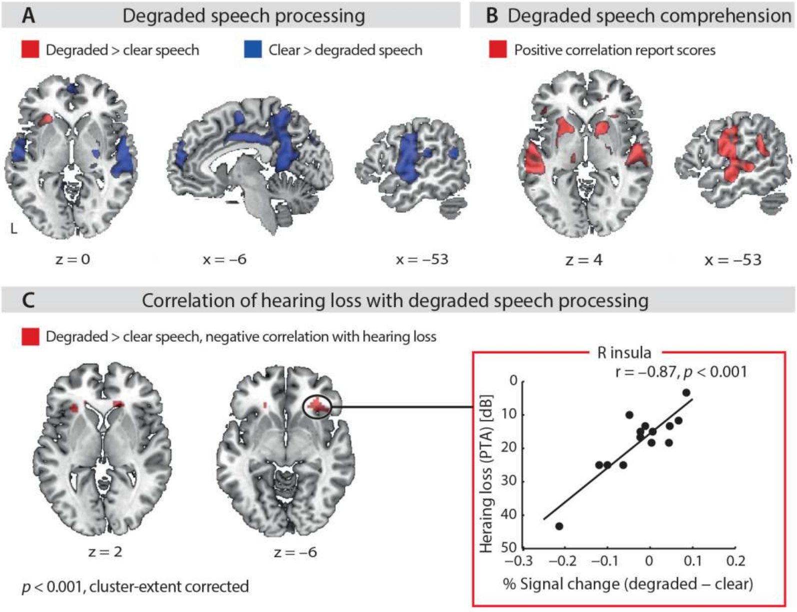


Figure 6.3. Functional MRI results in older adults. (A) Areas sensitive to speech degradation. In older adults, degraded relative to clear speech perception activated the anterior insula, whereas clear more than degraded speech activated a network comprising the precentral gyrus, the temporal cortices, supramarginal gyrus (SMG), right putamen, posterior cingulate cortex and angular gyrus bilaterally. **(B)** Areas varying with speech comprehension. With better speech comprehension on a given trial, activity increased in a network comprising the auditory cortices, precentral gyrus, left angular gyrus (AG), putamen and middle frontal gyrus. There were no negative correlations with trial-by-trial report scores. **(C)** Negative Correlation of hearing loss with activity related to degraded speech processing. Pure-tone average (PTA) correlated negatively with the degraded > clear speech contrast in the right and left anterior insula. The plot of this correlation shows that older adults with better hearing acuity had an increased BOLD signal for degraded compared to clear speech in the right anterior insula. Conversely, listeners affected by more severe hearing loss had elevated activity for clear relative to degraded speech. Note that the correlation remains significant when removing the outlier participant with a PTA of 43 dB ($r = -0.76$, $p = 0.002$) and when partialling out age ($r = -0.79$, $p = 0.001$).

6.3.3 Functional MRI differences between young and older adults

Degraded speech processing. Generally, older and younger adults showed largely overlapping activations during degraded speech processing (see Chapter 5 for the activation clusters in young adults). However, the ACC exhibited a group difference: young adults showed a higher increase in ACC activity when comparing degraded to clear speech than older adults did. This age group × degradation interaction was driven by a reduced dynamic range in older adults, who displayed persistently elevated levels of ACC activity in both conditions (Fig. 6.4A and Table 6.2). Extracting individual percentage signal change values from the ACC region of interest (as identified by the age group × degradation interaction) showed that, across groups, activity dynamics in the ACC were related to performance: Individuals with an overall better degraded speech comprehension showed a higher ACC differentiation for degraded versus clear speech (Fig. 6.4B). However, this correlation was driven by age, because the correlation was not significant in both groups separately (Pearson's correlation in young adults: r = –0.07, p = 0.73; older adults: r = 0.39, p = 0.16) and the correlation broke down when controlling for age (r = 0.15, p = 0.34). Within older adults, however, the correlation reached a non-significant trend when partialling out age (r = 0.46, p = 0.09). However, the age group × mean performance interaction in the ACC haemodynamic signal failed to reach significance ($t(41)$ = 1.5, p = 0.14).

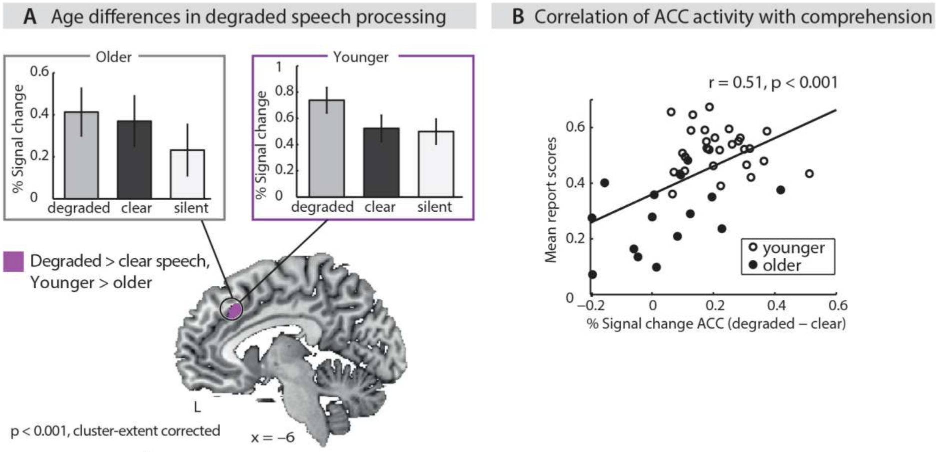


Figure 6.4. Age group × degradation interaction. (A) Young adults showed a greater difference in anterior cingulate cortex (ACC) activity between degraded and clear speech than older adults. This age group × degradation interaction was driven by a reduced dynamic range in older adults, who displayed increased levels of ACC activation during perception of both clear and degraded speech. This is apparent in the bar graphs showing percent signal change for both groups and conditions (upper panel). For comparison, signal change for silent trials is shown, illustrating that older but not younger adults upregulate ACC activity for clear compared to silent trials. **(B)** Correlation of ACC dynamics with performance. A post-hoc ROI analysis showed that across groups, individuals with a better overall degraded speech comprehension (mean report scores) showed a higher differentiation (degraded minus clear speech) in ACC activity.

Trial-by-trial fluctuations in degraded speech comprehension. Following up on the comprehension-dependent fluctuations observed in the cohort of young listeners (Experiment III), we also tested for such fluctuations in the older adults. While observing again substantial overlap between groups, an age group × comprehension interaction, also in prefrontal cortex, was manifest: Older adults' BOLD signal positively correlated with comprehension (i.e., report scores) in the middle frontal gyrus (MFG). Young adults did not show such a correlation in MFG. Young adults, on the other hand, exhibited additional correlations with comprehension in the left fusiform gyrus, right cerebellum and posterior cingulate cortex, where older adults did not show such a correlation (Fig. 6.5 and Table 6.2).

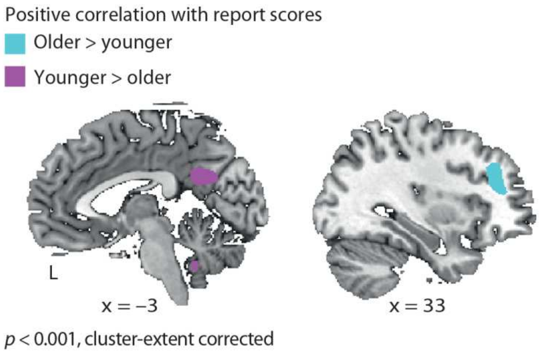


Figure 6.5. Age group × comprehension interaction. In older adults the haemodynamic response in middle frontal gyrus (MFG) covaried with comprehension of degraded speech on a trial-by-trial basis (i.e., report scores); young adults did not exhibit such a correlation. In contrast, young adults showed additional correlations with report scores in the posterior cingulate cortex, left fusiform gyrus and right cerebellum.

Table 6.2. Overview of fMRI activation clusters showing a correlation with hearing loss or significant group × condition interactions, thresholded at $p < 0.001$ (cluster-extent corrected).

Location	**MNI-Coordinates**			***Z*-Score**	**Extent (mm^3)**
Degraded > clear speech, negative correlation with PTA (older adults)					
R anterior insula	33	35	-5	4.38	432
L anterior insula	-30	26	1	4.08	207
Degraded > clear speech, young > older					
L SMA/ACC	-3	26	40	3.58	207
Positive correlation with trial-by-trial comprehension fluctuations, older > young					
R MFG	33	35	34	4.27	754
Positive correlation with trial-by-trial comprehension fluctuations, young > older					
L fusiform gyrus	-30	-43	-11	4.65	855
R cerebellum	12	-49	-35	3.94	477
Posterior cingulate gyrus	0	-58	28	4.49	648

L: left, R: right, ACC: anterior cingulate cortex, MFG: middle frontal gyrus, SMA: supplementary motor area.

6.4 Discussion

The current study intended to compare degraded speech processing between young and older adults to characterise how underlying neural mechanisms change with age. The main results can be summarised as follows: First, although degraded speech comprehension overall appeared deteriorated in older adults, both older and younger adults adapted to degraded speech at the same rate. Within older listeners, better working memory predicted faster adaptation rates and hearing loss predicted worse speech comprehension. Second, hearing loss was related to a distinct activation pattern in the anterior insula during degraded speech processing. Third, young listeners showed an expected modulation of ACC activity depending on task difficulty (i.e., degraded greater than clear speech), whereas older adults displayed elevated levels of ACC activity throughout, consistent with a persistent upregulation in this cognitive-control related area. Within the ACC, a greater dynamic range predicted better speech comprehension. Finally, for correctly comprehended degraded speech trials, older listeners recruited middle frontal regions in addition to a core speech comprehension network which younger listeners relied on, most likely reflecting a compensatory mechanism. We will now discuss these results in more detail.

6.4.1 Perceptual adaptation is comparable in young and older adults

Even though degraded speech comprehension was substantially reduced in older listeners, their ability to gradually adapt to degraded speech over the course of the experiment was preserved, as the slopes of their linear fits (adaptation curves) did not differ from young adults (Fig. 6.1C). Yet, as the R^2-values were relatively low (but cf. Davis et al., 2005, for comparable R^2-values), it remains questionable what no difference between age groups in a parameter of a badly fitting model actually means. Importantly however, we have shown that a second, model-free measure of learning (Δ performance) is strongly correlated with adaptation slope, and does not differ between age groups either. We take this as evidence that adaptation is comparable in young and older adults.

This result is consistent with a finding by Peelle and Wingfield (2005) who showed that older adults' (aged 65–78) perceptual learning of both time-compressed and spectrally shifted noise-vocoded speech was comparable to those of young adults: when older adults' starting accuracy was equated with young adults, both groups' speech comprehension improved at the same rate when listening to 20 degraded sentences. Thus, older and younger adults appear to show equivalent behavioural adaptation to degraded speech. However, this does not warrant that the underlying neural mechanisms are identical.

Older listeners with a better short-term memory adapted faster to degraded speech. The correlation of adaptation with backward but not forward digit span reached significance. The latter demands simple maintenance and repetition. In contrast, backward digit span requires the listener to perform an operation on the items held in the memory buffer (i.e., inverse the order of the digits) and has specific demands on central executive mechanisms (Baddeley, 2012), which are supposedly involved in effortful speech perception. Rabbitt (1968) suggested an effortful hypothesis, according to which working memory becomes a limiting factor when perceptual effort is required in degraded speech comprehension: Masked words disrupt the short-term memory buffer, because resources that would otherwise be available for encoding in short-term memory are diverted for perceptual effort (Piquado et al., 2010). Our result lends further support to the hypothesis that, such cognitive and perceptual abilities become coupled more tightly at an older age (Baltes and Lindenberger, 1997).

6.4.2 Hearing loss deteriorates adaptation and alters insula dynamics

One aim here was to identify how individuals with hearing loss recognise and process degraded speech. Unsurprisingly, a decline in hearing acuity was associated with worse average degraded speech comprehension in older adults. This is best explained by the fact that on top of the exogenous signal degradation (i.e., four-band vocoding), hearing loss endogenously distorts the signal and reduces audibility (Pichora-Fuller and Souza, 2003).

In addition, hearing loss was accompanied by changes in central neural mechanisms. While older adults with worse hearing activated the anterior insula more for clear than degraded speech, better-hearing older listeners had increased anterior insula activity during degraded relative to clear speech perception. Importantly, this activation pattern in the insula could not be explained by age. In Experiment III, we had shown that younger adults rely on the anterior insula in difficult listening. Here, it appears that only better-hearing older adults succeed to recruit the anterior insula in adverse listening conditions, whereas older adults affected by substantial hearing loss show the inverse dynamics. These hearing loss effects were found even though speech was presented at an audible level for each participant, supporting the notion that these are centrally-driven changes.

The insula together with the ACC have commonly been termed the “cingulo-opercular” system (e.g., Harris et al., 2009). Here, an interesting dissociation between insula and ACC dynamics emerges: Anterior insula activation was affected by hearing loss (independent of age), while ACC activity was altered by age (see Section 6.4.3).

There is accumulating evidence that the insula plays a crucial role in top-down, executive processes (Eckert et al., 2009; Menon and Uddin, 2010; Sterzer and Kleinschmidt, 2010; Adank, 2012). For example, Wild et al. (2012) showed that in young listeners, the insula exhibited an enhanced BOLD signal when listeners attended to speech (rather than a distracter), and the speech signal was degraded (vocoded rather than clear). Activation in the anterior insula further scaled with task-difficulty of a temporal non-speech auditory task, dependent on attention (Henry et al., 2013)

The current results are somewhat more complicated, as insula engagement depended on the extent of hearing loss; only those listeners with very mild, age-typical hearing loss (<20 dB HL) exhibit such a pattern, while more hearing-impaired listeners did recruit insular cortex into clear-speech (i.e. already at milder task demands), but less so into degraded-speech processing (higher task demands). Intriguingly, this pattern of insula activation is reminiscent of the Crunch-hypothesis, according to which older adults show a load-dependent inverse u-shaped pattern of activity in cognitive-control-related areas (Reuter-Lorenz and Cappell, 2008).While this finding deserves thorough follow-up experimentation, the insula seems to be crucial for cognitive control (Eckert et al., 2009) and the observed alteration of its haemodynamics might add up to declines of the peripheral auditory system and manifest in the observed deterioration in speech recognition.

6.4.3 Age-related upregulation of ACC activity reflects increased cognitive comtrol

The main objective of the current study was to identify age differences in the neural systems supporting degraded speech processing. We found an age group × degradation interaction in the ACC, where young listeners showed a higher activation difference between degraded and clear speech than older adults. The latter displayed elevated levels of ACC activation during both conditions, indicating a reduced dynamic range in older adults. Thus, the older adults’ BOLD signal in the ACC appears to be less informative and flexible, as it is differentiating less between a degraded and clear speech input. This is consistent with an age-related decrease in the variability of the haemodynamic signal (Garrett et al., 2010, 2011).

As we intended to compare the results in the older group to the ones obtained in young adults (Experiment III), we normalised older adults to a young adult (MNI) template. This can be problematic due to group differences in brain morphology. For example, partial volume effects, that is, sampling both grey and white matter in one voxel, may increase in older adults, who commonly exhibit grey matter atrophy in frontal regions. One solution to overcome this problem is to only test for the interaction (Samanez-Larkin and D'Esposito, 2008). As we found an age group × condition interaction in the ACC, it is unlikely that older adults' smaller dynamic range in ACC activity could be driven by the non-diffeomorphic normalisation of older adults to a young adult (MNI) template.

Age-related functional and structural changes in frontal lobe systems supporting cognitive control have been previously noted (Cabeza et al., 2004; Colcombe et al., 2005; Sharp et al., 2006; Eckert et al., 2008; Harris et al., 2009). For example, Harris et al. (2009) showed that ACC activity increases for incorrect compared to correct word recognition, but more so in older adults. The authors linked this to auditory cortex architecture, showing that ACC recruitment correlates with age-related neurodegeneration of the auditory cortex. Along the same line, Sharp et al. (2006) showed that aging is accompanied by greater cognitive control, as indexed by higher ACC and prefrontal cortex activity during semantic and syllabic decisions on noise-vocoded words. Both Harris et al. (2009) and Sharp et al. (2006) observed that activity in ACC increased with age and was detrimental to performance. Therefore, they interpreted the age-related increase of ACC activation as upregulation of error monitoring systems (Dosenbach et al., 2006; Dosenbach et al., 2007).

In contrast, the present findings show that a higher dynamic range of ACC activity (i.e. the degree to which the ACC became relatively engaged and disengaged in degraded and clear speech, respectively) was associated with better speech comprehension. However, this correlation was best explained by age, as the correlation broke down when correlating groups separately. The following picture emerges: Dynamic range of ACC activity decreases with age which in turn is detrimental to speech comprehension. Rather than playing a compensatory role for deficits due to aging (Cabeza et al., 2002), the observed ACC dynamics might reflect a generalised upregulation of cognitive control with age irrespective of task difficulty (see also the discussion of dedifferentiation versus compensation hypothesis below).

6.4.4 Age-related compensatory prefrontal activity during speech comprehension

For successful speech comprehension, younger adults' activated a perisylvian network (Experiment III), where the BOLD signal was tightly coupled to actual speech comprehension (i.e., behavioural report scores), rather than acoustic properties of the sentences. Older adults additionally showed a correlation with report scores in middle frontal gyrus. Eckert et al. (2008) similarly demonstrated that older adults engage the MFG when words are most intelligible. However, their design varied speech intelligibility by low-pass filtering words such that effects due to acoustic differences could not be disentangled from actual comprehension. In contrast, the current study held physical stimulus properties constant (i.e., four-band vocoding) and thus identified regions where activation varied with actual speech comprehension (i.e., behavioural report scores). Therefore, the present data provide evidence that additional MFG activation observed in older adults is related to comprehension rather than physical speech clarity.

Age-related additional recruitment of middle frontal or lateral prefrontal cortex has been repeatedly observed, for example, during working memory tasks (Cabeza et al., 2002), visual attention (Cabeza et al., 2004), word recognition (Eckert et al., 2008), and for processing of syntactically complex sentences (Peelle et al., 2010b). Two hypotheses were put forward to explain the frequently observed recruitment of additional brain regions by older adults not observed in young adults: The dedifferentiation hypothesis (Baltes and Lindenberger, 1997) interprets the extra activation as difficulties in recruiting specialised neural mechanisms for the relevant task, hence as a loss of neural specialisation (Park et al., 2004). Such a generalised non-functional spread of activation has been attributed to a deficit in neurotransmission with a decrease in signal-to-noise ratio in neural firing (Li, 1999). If this hypothesis is true, engagement of additional regions should not correlate with task performance. On the other hand, the compensation hypothesis suggests that recruitment of additional brain regions plays a compensatory role, for example in counteracting performance decline due to neurodegeneration in domain-specific brain areas (e.g., the auditory cortex; Harris et al., 2009; Peelle et al., 2011; Eckert et al., 2012), and should therefore improve performance (Cabeza et al., 2002; Heuninckx et al., 2008).

In the current study, engagement of the MFG in older adults covaried with report scores, that is, MFG activity did in fact increase with better performance. We take this as evidence for a compensatory mechanism in older adults, whereby the MFG is recruited in addition to the core speech comprehension network (see Experiment III) when speech comprehension succeeds. In sum, our data contribute to, but cannot solve, the ongoing debate of dedifferentiation versus compensation, in that the age-group-differences observed in two prefrontal areas, ACC and MFG, are to be interpreted with opposing conclusions regarding the compensation hypothesis by Cabeza and others (2002, 2004).

6.4.5 Conclusions

Our results show distinct age-related changes of responses in prefrontal cortex. Higher anterior cingulate and middle frontal gyrus activities are found associated with better performance in adverse listening conditions. However, unlike younger adults, older adults do not succeed in selectively modulating ACC activity depending on listening difficulty, but exhibit generalised upregulated levels of ACC activity also in easy listening conditions. In contrast, MFG activity appears to be truly compensatory, as older adults recruit frontal areas in addition to a speech comprehension network when comprehension succeeds. Moreover, more hearing-impaired older adults involve the anterior insula more even in clear speech comprehension. As all three structures have been linked to cognitive control, the results provide further evidence that older adults increasingly rely on cognitive control networks when adapting to challenging listening conditions, at the potential expense of these systems' dynamic range.

7 Experiment V: Speech comprehension after envelope modification in quiet and in noise[8]

7.1 Introduction

Speech envelope cues are necessary and almost sufficient for speech comprehension (Shannon et al., 1995). The idea that speech information is mainly carried by the low-frequency modulations in the signal was proposed by (Houtgast and Steeneken, 1971). They developed the speech transmission index (STI; IEC60268-16, 2003) based on a weighted sum of modulation transfer functions (Steeneken and Houtgast, 1980). The STI allowed for accurate prediction of speech intelligibility, at least under certain conditions, for example in reverberating halls (but cf. Jorgensen and Dau, 2011, for a more robust metric to predict speech intelligibility based on the signal-to-noise ratio in the envelope domain), corroborating the importance of amplitude modulations in speech recognition (Houtgast and Steeneken, 1985).

In line with this hypothesis, we have shown that normal-hearing listeners with higher sensitivity to slow amplitude fluctuations in an auditory signal, as measured by AM rate discrimination thresholds, learn more quickly to understand vocoded speech (Experiment I). In CI listeners, AM rate discrimination thresholds also predicted speech recognition (Experiment II). Similarly, we have demonstrated that processing the temporal envelope of speech and modulation rate discrimination draw on common neural mechanisms (Experiment III). Together, these experiments provide converging evidence for common mechanisms of speech and non-speech envelope processing, confirming the importance of speech envelope cues in comprehension (Houtgast and Steeneken,

---

[8] This experiment was designed by Julia Erb, Molly Henry, and Jonas Obleser.

1985). Therefore, we hypothesised that speech comprehension could be facilitated by enhancing speech envelope cues, that is, rendering them more salient.

The previous experiments of this thesis examined how listeners react when confronted with a spectrally severely degraded speech signal, thereby being forced to rely on the *intact* temporal envelope cues. In the present behavioural experiment, we investigated the perceptual consequences when both the spectral detail and speech envelope cues are manipulated: In addition to noise-vocoding speech, we modified speech envelopes by either expanding (i.e., increasing the modulation depth) or compressing them (i.e., reducing the modulation depth of the signal; Lorenzi et al., 1999).

Such speech envelope modifications have mainly been investigated for the development of hearing aid algorithms which often apply amplitude compression (Fu and Shannon, 2000; Shannon et al., 2004). Both hearing-impaired and CI listeners typically show reduced dynamic ranges and loudness recruitment (i.e., an abnormally-rapid growth in perceived loudness with intensity increase), most probably due to a loss of the fast-acting compression in the damaged cochlea (Moore and Glasberg, 1993; Moore et al., 1995). To correct for loudness recruitment, most CI speech processors use compression, that is, a non-linear (logarithmic) function to map acoustic sound pressure to electrical current (Shannon et al., 2004). In contrast, if intensity modulations are the main carriers of speech information (Steeneken and Houtgast, 1980; Houtgast and Steeneken, 1985; Shannon et al., 1995; Experiments I-III), normal-hearing listeners' speech recognition should actually improve when expanding the speech envelope, that is, enhancing modulation depth (Plomp, 1988).

A few behavioural studies have assessed the effects of envelope expansion on speech recognition in quiet. Freyman and Nerbonne, (1996) expanded the very slow amplitude fluctuations below 20 Hz which led to dramatically poorer consonant recognition. Van Buuren et al. (1999) observed no improvement of intelligibility or sound quality after envelope modification. Short sentences in noise or masked by a competing speaker were best understood and rated most intelligible by both normal-hearing and hearing-impaired listeners when amplitude modulations below 32 Hz were neither expanded nor compressed but rather left unmodified. Similarly, Fu and Shannon (1998) showed that non-linear amplitude mapping deteriorated phoneme recognition in quiet. They extracted the low-frequency modulations of syllables below 160 Hz and applied a power-law transformation of which the exponent varied from compressive (k = 0.3) to expansive (k = 3). Spectral detail was reduced using four-band vocoding. Both normal-hearing listeners and CI patients performed best when loudness was linearly mapped, at least in quiet conditions (Fu and Shannon, 1998).

In contrast, when noise was added and higher cut-off frequencies for envelope extraction were applied, listeners actually benefited from speech envelope expansion. Lorenzi et al. (1999) extracted amplitude fluctuations below 500 Hz of pseudowords in quiet and in noise and raised the envelope to the power of $k = 2$. While consonant recognition in quiet slightly deteriorated, recognition in noise improved after envelope expansion. In a follow-up study, Apoux et al. (2001) combined three different signal-to-noise ratios (SNRs) and three different envelope exponents to identify the level which yielded the best phoneme recognition performance. At all SNRs, envelope expansion was beneficial to performance, with more extreme expansion schemes yielding better performance.

Following up on these findings, Apoux et al. (2004) tested how speech recognition performance was influenced by different cut-off frequencies for envelope extraction, and whether expansion before and after addition of noise was beneficial to speech recognition, in stationary and fluctuating noise. This yielded somewhat contradictory results: They confirmed, on the one hand, a slight benefit when expansion was applied after addition of stationary noise, but only when enhancing very slow amplitude fluctuations (< 16 Hz). On the other hand, amplitude expansion before addition of noise led to no effect or was even detrimental to intelligibility. Also, envelope expansion of speech in fluctuating noise did not yield any benefits.

To clarify and extend these previous and in part contradictory results, we here parametrically varied both SNR and envelope modification. In a behavioural study that extended the parameter space to pilot such a parametric design for potential use in fMRI (e.g., Obleser et al., 2008; Schonwiesner and Zatorre, 2009), we wanted to test whether we could reproduce the interaction of envelope modification and SNR, which has been observed in some studies (Lorenzi et al., 1999).

To ensure that speech recognition depended exclusively on envelope cues, we six-band-vocoded all stimuli (isolated words). In a full factorial design, stimuli were varied along two dimensions: the SNR was parametrically varied (in five levels, plus one noise-free level). The temporal envelopes of amplitude fluctuations (< 400 Hz) were compressed or expanded using a power-law transformation (in five steps from compressive to expansive). Speech comprehension and intelligibility were assessed on every trial. This allowed us to quantify how much listeners profited from envelope expansion at different noise levels.

7.2 Materials and methods

7.2.1 Participants

Participants were recruited from the participant database of the Max Planck Institute for Human Cognitive and Brain Sciences. Ten participants (5 females, age range 23–30) took part in the behavioural study. All were right-handed, monolingual speakers of German with no known language impairments or neurological problems. All participants had self-reported normal-hearing. Participants received financial compensation of € 10.50. Informed consent was obtained from all subjects. Procedures were in accordance with the guidelines of the local ethics committee (University of Leipzig).

7.2.2 Stimuli and experimental design

Stimuli. Every participant listened to 480 words in total that were randomly drawn from a pool of 547 German mono-, bi- or trisyllabic nouns (Kotz et al., 2002). Words were recorded by a female speaker in a sound-proof chamber. The length of the words varied from 284 to 1044 ms (mean 589 ms). These materials were identical to the ones used in studies by Obleser et al. (2008) and Obleser and Weisz (2012).

To degrade words, we used a similar procedure as described in the study by Apoux et al. (2001). First, a white noise masker was added to the audio signal, at 5 different SNRs from 18 to -6 dB in steps of 6 dB. The noise duration was adjusted to match the word duration. In a sixth condition no background noise was present. Subsequently, the summed signals were six-band vocoded and subjected to envelope compression or expansion as follows. The signal was divided into six frequency bands spanning 70 to 9000 Hz that were spaced according to Greenwood's cochlear frequency-position function (Greenwood, 1990). The cut-off frequencies of the six adjacent analysis bands were 70, 268, 633, 1304, 2539, 4813, and 9000 Hz. For extraction of the temporal envelope $E(t)$, we used a second-order, zero-phase Butterworth lowpass filter with a cut-off frequency of 400 Hz (for a detailed description of noise-vocoding see General methods).

The speech envelope $E(t)$ was risen to the power of an exponent k, meaning that it was either compressed ($k < 1$), left unmodified ($k = 1$), or expanded ($k > 1$). In sum, k assumed 5 different logarithmically spaced values: 0.5, 0.707, 1, 1.414, and 2. The modified envelope $E(t)^k$ was then multiplied with band-pass filtered noise carriers whose centre- and cut-off frequencies were matched to the analysis bands. The resulting modulated noises were recombined. Finally, all stimuli were normalised to an equal root-mean square (RMS; for example waveforms of the word "Wiese" [meadow] in all 30 possible conditions see Fig. 7.1A). Signal-processing was carried out in Matlab7.11.

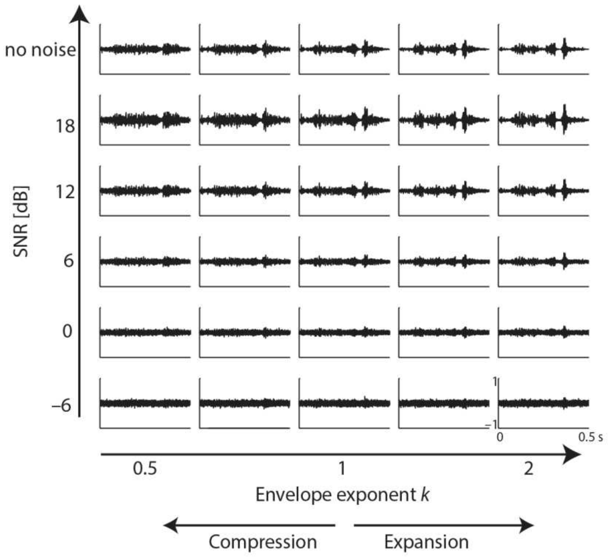


Figure 7.1. Combinations of envelope exponent with SNR. The waveform of the example word "Wiese" [meadow] is shown in all possible combinations of envelope exponent with signal-to-noise ratio (SNR). Both factors were fully crossed in a two-way parametric design.

Task. In order to get a continuous measure of performance, participants rated the intelligibility of auditorily presented words on a scale from one to four and typed what they had understood.

After each auditorily presented word, the question "How intelligible was the word?" appeared on the screen. Participants used the keyboard of the presentation laptop to respond with the keys "1", "2", "3", and "4". The key assignment was counterbalanced across participants, such that half of the participants pressed "1" to indicate that a word was "very intelligible", whereas the other half pressed "4".

Although the intelligibility rating is expected to highly correlate with actual speech comprehension (Davis and Johnsrude, 2003), participants were additionally asked to type what they had actually understood after the rating on each trial. There was no timeout for typing. This second performance measure was added to verify whether actual speech comprehension correlated with the intelligibility rating. Responses were scored as correct or incorrect by comparing the typed string with the presented word (Obleser et al., 2008).

7.2.3 Experimental procedure

Participants were sitting in a sound-dampened chamber in front of a laptop computer. Auditory stimuli were presented through Sennheiser HD 25-SP-II headphones using Presentation software. Presentation level was kept constant across participants at approximately 60 dB SPL. Visual cues appeared on the laptop screen. Participants received a written instruction but did not undergo any training.

7.3 Results

7.3.1 Correlation of intelligibility rating with report scores

The ten listeners who participated in the study both rated intelligibility and typed what they had understood after each presented word (see Section 7.2). To test how closely intelligibility ratings were related to comprehension, we correlated the mean ratings per condition (on a scale from 1–4; Fig. 7.2A) with the mean report scores per condition (percent correct; Fig. 7.2B) for each participant separately. The Spearman's correlation coefficients ($r = 0.89 \pm 0.01$; mean ± SEM averaged over participants) were Fisher-*z*-transformed. A *t*-test on the Fisher-z-transformed correlation coefficients showed that they were significantly different from zero ($t(9) = 23.3$; $p < 0.001$), indicating that intelligibility ratings significantly correlated with actual speech comprehension. The present correlation is arguably lower, but similar to the correlation of $r = 0.99$ between report scores and intelligibility ratings which Davis and Johnsrude (2003) reported. They used a nine-point rating scale and obtained percent correct scores per sentence. Here, we used a coarse four-point rating scale to assure compatibility with functional neuroimaging, and we obtained discrete correct or incorrect responses per word, leading to more coarse resolution of performance and possibly to a decrease in correlation strength.

7.3.2 Effects of envelope exponent and signal-to-noise ratio

The intelligibility ratings showed the expected pattern of an envelope expansion × SNR interaction (Fig. 7.2A): In quiet, participants' performance showed an inverse u-shape as a function of envelope exponent where stimuli were most intelligible at $k = 1$ (i.e., unmodified envelope) and intelligibility decreased when the envelope was compressed ($k < 1$) or expanded ($k > 1$; Fig. 7.2A, top black line). Conversely, when noise was added and the envelope was expanded, subjects actually rated stimuli as more intelligible, indicating that listeners profited from speech envelope expansion when background noise was present. The report scores from the typed responses showed the same pattern of results (Fig. 7.2B), confirming objective benefits of speech intelligibility in noise with expanded envelopes.

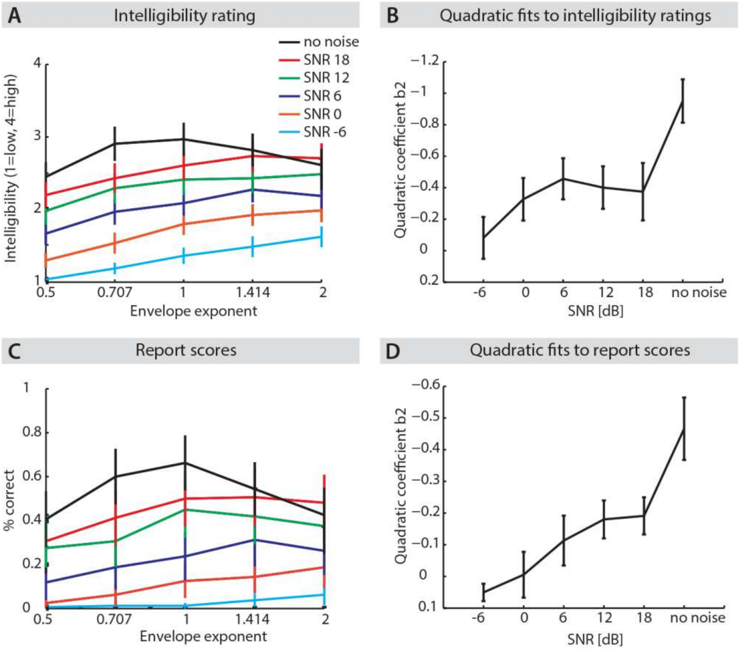
To test whether the described pattern was statistically significant, we averaged intelligibility rating (or correct/ incorrect responses) over trials within each condition and each participant. These mean intelligibility ratings (or mean percentage correct scores) were submitted to a two-way repeated-measures ANOVA with the factors envelope exponent (5 levels) and SNR (6 levels) using PASW Statistics 18.0. All results were Greenhouse Geisser-corrected.

For the intelligibility ratings, we found a significant main effect of envelope exponent ($F(4) = 62.62$, $p = 0.001$,$\eta_p^2 = 0.87$) and of SNR ($F(5) = 145.33$, $p < 0.001$, ,$\eta_p^2 = 0.94$), as well as an interaction of both ($F(20) = 3.4$, $p = 0.008$, ,$\eta_p^2 = 0.27$).

In a non-parametric test, where we computed a 5 × 6 repeated measures ANOVA on the rank-transformed rating data, we obtained the same effects: A main effect of envelope exponent ($F(4) = 68.66$, $p < 0.001$,$\eta_p^2 = 0.88$) and of SNR ($F(5) = 374.84$, $p < 0.001$, $\eta_p^2 = 0.97$), as well as an interaction of both ($F(20) = 2.99$, $p = 0.022$, ,$\eta_p^2 = 0.25$) were significant.

For the report scores, we found the same pattern of results: Both the envelope exponent ($F(4) = 18.96$, $p < 0.001$, $\eta_p^2 = 0.68$) and SNR had a significant effect on report scores ($F(5) = 141.45$, $p < 0.001$, $\eta_p^2 = 0.94$). The envelope expansion × SNR interaction was also significant ($F(20) = 2.89$, $p = 0.022$, ,$\eta_p^2 = 0.24$).

Performing ANOVA-tests on proportional data is problematic, however, as the distributions of these data are not normal. Therefore, we rau-transformed the report scores (Studebaker, 1985) and performed a two-way repeated measures ANOVA on the rau-transformed data. Importantly, rau-transformation did not substantially change our results: We found significant main effect of envelope exponent ($F(4) = 17.98$, $p < 0.001$,$\eta_p^2 = 0.66$) and of SNR ($F(5) = 174.71$, $p < 0.001$, ,$\eta_p^2 = 0.95$), as well as an interaction of both ($F(20) = 2.32$, $p = 0.05$, ,$\eta_p^2 = 0.21$).


Intelligibility rating

Quadratic fits to intelligibility ratings

Report scores

Quadratic fits to report scores

Figure 7.2. Effect of envelope exponent and SNR on speech intelligibility and recognition. (A) Intelligibility rating and **(B)** report scores (from typed responses) show a similar pattern of performance as a function of signal-to-noise ratio (SNR) and envelope condition. Both SNR and envelope exponent have a significant main effect on performance where higher SNR and a higher envelope exponent (i.e., expansion) lead to better performance. Importantly, we observed an envelope expansion × SNR interaction where performance in quiet (upper black line) is best when the envelope is left unchanged ($k = 1$), but in elevated noise levels performance is best when the envelope is expanded ($k = 2$; colored lines). When fitting quadratic curves to **(C)** the ratings or **(D)** report scores as a function of the envelope exponent for each SNR level separately, the quadratic term $b2$ was near zero for the smallest SNR level (SNR = -6 dB) and became increasingly negative with higher SNR, indicating that performance assumed an inverse u-shaped pattern (i.e., the envelope expansion benefit in noise diminished) with increasing SNR.

We wanted to quantify how much performance as a function of envelope exponent changed from quadratic (i.e., inverse-u-shaped) in quiet to linear when noise was added. Therefore, we fitted a quadratic curve of the shape $f(x) = b_2x^2 + b_1x + b_0$ to the ratings as a function of the envelope exponent for each SNR level and each participant separately.

As expected, the quadratic term b_2 was near zero for the smallest SNR level (SNR = -6 dB) and became increasingly negative with increasing SNR level, indicating that the curves assumed an increasingly inverse-u-shaped form with rising SNR level (Fig. 7.2C). This trend was significant: We fitted a linear curve to the quadratic coefficient b_2 as a function of SNR level for each participant separately and tested the slopes against zero using a t-test ($t(9) = -2.5$, $p < 0.05$).

Also, when fitting a quadratic curve to the report scores from the typed responses rather than the ratings as a function of envelope exponent, the quadratic term b_2 showed a significant decrease as a function of SNR level ($t(9) = -7.05$, $p < 0.005$; Fig. 7.2D).

7.4 Discussion

The present study examined the perceptual consequences of expanding or compressing the temporal modulations in speech while incrementally adding noise. Words were six-band-vocoded to ensure that listeners had to rely primarily on temporal envelope cues for comprehension. The results show that expansion and compression of the speech envelope in quiet deteriorates recognition performance. On the contrary, when noise is present, envelope expansion facilitates speech comprehension and can restore intelligibility.

These results replicate and reconcile previous findings by Lorenzi et al. (1999) and Apoux et al. (2001, 2004). Our paradigm introduces two novel aspects compared to previous work by Lorenzi's group: It is the first to combine a parametric variation envelope modification (including both compressive and expansive levels) with a fine-grained parametric variation of signal-to-noise ratio (in five levels), while including a silent condition. Importantly, this allowed us to quantify the extent to which the detrimental effect of envelope expansion on speech recognition in quiet changed into a beneficial effect as noise was incrementally added. A second advantage of our paradigm is that listeners performed a more naturalistic speech recognition task, rather than a 16-AFC task on “logatomes” (Lorenzi et al., 1999; Apoux et al., 2001), such that we were able to show that actual comprehension is affected by envelope modification. Our results will be discussed in the context of previous studies below.

7.4.1 Envelope cues are essential for comprehension of speech in noise

Presence of a steady noise masker expectedly deteriorated word recognition: We found a highly significant main effect of SNR on speech recognition; performance dramatically decreased from an SNR of +18dB to -6 dB. This is consistent with previous studies showing the detrimental effect of noise on recognition of vocoded words, which can partly be explained by reduction in modulation depth due to background noise (Noordhoek and Drullman, 1997).

Envelope expansion of vocoded words in quiet was detrimental to speech recognition (Figures 7.2A and 7.2B, black lines). Such a negative effect of expansion in quiet has been observed previously (Freyman and Nerbonne, 1996; Fu and Shannon, 1998; Lorenzi et al., 1999; van Buuren et al., 1999). Fu and Shannon (1998) pointed out that crude envelope expansion using a power-law transformation distorts the consonant-to-vowel ratio: The vowels, which have generally larger amplitudes, are more amplified than the consonants. Therefore, the consonant-to-vowel ratio decreases in expanded speech, which effectively distorts the signal. If we assume that speech recognition is achieved by some kind of template matching, that is, a comparison of the incoming signal with a template stored in memory, than the template matching is likely to fail with an incoming signal of which modulation depth properties differ drastically from the normal speech signal (Lorenzi et al., 1999). Therefore, a distorted consonant-to-vowel-ratio likely decreases intelligibility (Freyman and Nerbonne, 1996).

In contrast, envelope expansion in noise was found to yield a small but highly significant benefit in terms of correct identification performance and intelligibility ratings (Figures 7.2A and 7.2B), consistent with previous evidence (Lorenzi et al., 1999; Apoux et al., 2001). Yet, earlier studies have only tested a limited number of SNR and envelope combinations. Here, due to the fine-grained variation of SNR and envelope modification, were able to fit quadratic curves to the performance over envelope exponents for each SNR and each participant separately. Importantly, this allowed us to quantify the extent to which a listener was able to profit from envelope expansion in noise.

Lorenzi et al. (1999) presented vocoded speech either in quiet or in noise (at an SNR of 0 dB) with unmodified or expanded (k = 2) envelopes, and found an interaction of both factors. Apoux et al. (2001) combined three SNR levels (-6, 0, 6 dB) with three envelope exponents ($k = 1$; $k = 1.414$; or $k = 2$) and showed small benefits in phoneme identification and reaction times with increasing envelope exponent at all noise levels (but did not observe an interaction). The authors suggested that the beneficial effect of envelope expansion in noise is due to the restoration of modulation depth: The modulation depth of the speech envelope is reduced by the presence of noise (see above), but expanding the speech envelope can compensate for this loss in modulation depth (Apoux et al., 2001).

However, a third study by Lorenzi and colleagues (Apoux et al., 2004) could only partially replicate their previous results. When using the same expansion scheme (with $k = 2$), a small expansion benefit was observed only when low-amplitude fluctuations (< 20 Hz) were expanded (rather than higher modulation frequencies up to ~250 Hz), only in stationary (but not fluctuating) noise, and only for normal-hearing but not hearing-impaired listeners.

The discrepancy between the studies by Apoux et al. (2004, 2001) can be due to different factors. First, the study by Apoux et al. (2001) used noise-vocoded stimuli, just as in the present paradigm, whereas the study by Apoux et al. (2004) did not spectrally degrade stimuli. Hence, listeners could rely on additional spectral cues rather than exclusively temporal envelope cues. Second, the absence of an expansion benefit in fluctuating noise is most likely due to enhancement of not only the signal but also the fluctuating background noise by the expansion algorithm (Apoux et al., 2004).

Together with previous evidence, our finding bears particular relevance for hearing aid algorithms. However, based on the finding that envelope expansion does not improve speech perception in fluctuating noise (Apoux et al., 2004), the "enhanced envelope strategy" might currently be the most promising expansion algorithm (Koning and Wouters, 2012). This scheme expands exclusively the onset of the envelope of a speech signal, such that expansion of the fluctuating background noise can be avoided. The enhanced envelope strategy was shown to not only improve the speech reception thresholds in stationary noise, but also in a competing talker situation. Implementations of such expansion schemes in hearing aid processors might yield improved comprehension in noise, as long as the auditory system's dynamic range is not exceeded (Nejime and Moore, 1997; Lorenzi et al., 1999).

7.4.2 Neural representation of envelope cues

How this expansion benefit is neurally represented remains to be examined: Are there parts of the auditory system which differentially respond to parametric variation of either envelope modification or noise level, and the behaviourally observed interaction of both factors?

For the temporal envelope of non-speech sounds, there is evidence that modulation rates in an acoustic signal are represented hierarchically along the auditory pathway (Giraud et al., 2000; for animal studies see Schreiner and Urbas, 1986; Langner and Schreiner, 1988; Rees and Palmer, 1989). Preferred AM rate has been shown to decrease as one ascends the auditory pathway: While the olivary complex is most responsive to faster modulation rates (>250 Hz), the inferior colliculus (IC) responds to ~30–250 Hz, the auditory thalamus to AM rates at ~16 Hz and the primary auditory cortex is tuned to low AM frequencies at ~8 Hz (Giraud et al., 2000). A preference of human auditory cortex for low modulation rates (< 16 Hz) has been confirmed by several studies (Liegeois-Chauvel et al., 2004;

Schonwiesner and Zatorre, 2009; Overath et al., 2012; see also Experiment III), which matches the AM frequencies that are critical for speech intelligibility. This is consistent with an account whereby each level of the auditory pathway serves as an AM filter (sensu Dau et al., 1997b, a), which enables a decomposition of the temporal modulations along the auditory pathway.

In the present experiment, temporal modulations < 400 Hz were modified such that modulation rate processing at all different levels of the auditory pathway is likely affected. How compression or expansion of these modulations, in quiet and noise, differentially affect neural processing remains to be determined.

7.4.3 Conclusions

The present study demonstrates that expansion of the speech envelope in quiet deteriorates speech recognition just as compression does, whereas envelope expansion in noise facilitates speech comprehension and can restore intelligibility. Such envelope expansion schemes may easily be incorporated into hearing aid algorithms to facilitate speech comprehension in noise. Extending previous work by Lorenzi's group with a fully crossed two-way parametric design now usable for functional neuroimaging experiments (e.g., Obleser et al., 2008), we were able to quantify the extent to which listeners benefited from envelope expansion as the noise level increased.

The previous experiments of this thesis provided evidence for the importance of temporal envelope cues in speech comprehension, showing both in normal-hearing (Experiment I) and CI listeners (Experiment II) a correlation of sensitivity to non-speech temporal envelope cues and degraded speech comprehension. Further, we showed that both speech and non-speech envelope cues are represented in common neural networks, comprising Heschl's gyrus, on the one hand, and the cingulo-opercular network, on the other hand (Experiment III). In this last experiment, we provide direct evidence that enhancing the speech envelope facilitates speech comprehension when noise is present. Envelope cues thus appear to be increasingly important when listening conditions become noisy.

8 General discussion

The current thesis set out to examine the neural systems that support comprehension of, and adaptation to, speech when listening conditions become difficult. In particular, we investigated how normal-hearing listeners, older adults with varying degrees of hearing loss, and CI patients cope with spectral degradation of the signal. To this end, behavioural, brain structural and functional MRI studies were conducted. After summarising the experimental findings, the discussion focuses on the key findings. In particular, it will address the role of subcortical structures in perceptual learning, the cingulo-opercular function in speech perception and sensory processing in general, and the significance of temporal envelope cues for speech comprehension. For more specific discussions of particular findings, we refer the reader to the discussion sections of the respective experimental chapters.

8.1 Summary of the experimental results

Experiment I investigated potential predictors of short-term adaptation to degraded speech, including working memory, basic auditory skills and brain structure. We showed in normal-hearing listeners that left thalamic morphology and sensitivity to modulation rate can predict individual perceptual adaptation to degraded speech. Importantly, the latter was validated in CI patients.

In Experiment II we demonstrated that CI patients' modulation rate sensitivity shortly after implantation did not only correlate with speech recognition abilities on the same day but also 6 months later, provided that duration of deafness was taken into account.

The goal of Experiment III was two-fold: on the one hand, we intended to quantify the convergence of neural mechanisms that subserve modulation rate discrimination and degraded speech comprehension. On the other hand, we aimed to identify the neural systems supporting short-term perceptual adaptation to degraded speech. We report three major findings: first, the degraded speech task revealed an "executive" network (comprising the anterior insula and ACC), parts of which were also activated in the non-speech discrimination task. Second, the haemodynamic signal in a "language" network (Scott et al., 2000), comprising auditory, premotor cortices and left angular gyrus, covaried with speech comprehension rather than physical clarity of the stimulus. Third, in contrast to previous perceptual learning studies, the data provide evidence that perceptual adaptation to degraded speech is not only reflected in an upregulation of cortical activity but accompanied by a down-regulation of subcortical activity. The data highlight differential up– and down-regulation in auditory–language and executive networks, respectively, with important subcortical contributions when successfully adapting to a challenging listening situation.

Experiment IV intended to characterise how such neural speech processing changes with older age and age-related hearing loss. We showed that although degraded speech comprehension appeared deteriorated in older adults, both older and younger adults adapted to degraded speech at the same rate. Amongst older listeners, better working memory predicted faster adaptation rates and more hearing loss predicted worse speech comprehension. Second, hearing loss changed the pattern of activation in the anterior insula during degraded speech processing. In contrast to young listeners, who showed modulation of ACC activity depending on task difficulty (see Experiment III), older adults exhibited elevated levels of ACC activity throughout, consistent with a persistent upregulation in this cognitive-control related area. A reduced dynamic range in ACC was detrimental to degraded speech comprehension. Lastly, for correctly understood degraded speech trials, older listeners recruited middle frontal regions in addition to a core speech comprehension network which younger listeners relied on, most likely reflecting a compensatory mechanism.

Experiment V asked directly how the temporal modulations in speech are processed: Instead of simply six-band-vocoding words, we additionally modified the speech envelope, by either enhancing or decreasing the saliency of speech envelope cues while noise was incrementally added. We show that expansion and compression of the speech envelope in a quiet setting both deteriorate recognition performance; however, when noise is present, envelope expansion does facilitate speech comprehension. Importantly, the parametric variation of envelope modification with a fine-grained parametric variation of signal-to-noise ratio allowed us to quantify the extent to which the detrimental effect of envelope expansion on speech recognition in quiet changed into a beneficial one as noise was incrementally added.

These results will be discussed in the context of previous evidence on cortical and subcortical functions in speech processing below.

8.2 The striato-thalamo-cortical circuit – a distinct role in adaptation to degraded speech?

The role of subcortical structures in speech perception is poorly understood, but two structures are mainly thought to exert a role in speech processing: the basal ganglia and the thalamus. In the neuroimaging experiments of this thesis, convincing evidence was presented for functions of both structures in perceptual learning of degraded speech. In particular, individual differences in the morphology of the left pulvinar thalamus predicted adaptation (Experiment I), while the ventral anterior thalamic nucleus and the caudate were found to be functionally involved in perceptual adaptation (Experiment III).

Before discussing the theoretical implications of these subcortical findings, we will give a brief overview of the anatomy and in particular the connectivity of these subcortical structures. The basal ganglia comprise the globus pallidum and the striatum; the latter can be subdivided into the ventral and the dorsal striatum (caudate nucleus and putamen). While the caudate and putamen are thought of as input nuclei, the pallidum is considered an output nucleus (Parent, 1986). The substantia nigra and the ventral nuclei of the thalamus have strong connections to the basal ganglia and are therefore referred to as "associated structures" (Parent, 1986). The ventral thalamic nuclei form part of the striato-thalamo-cortical loop (Parent and Hazrati, 1995)[9].

The neurotransmitters of the basal ganglia include Glutamate, γ-aminobutyric acid (GABA), and Dopamine. The striatum receives excitatory glutamatergic connections from the cortex and sends inhibitory gabaergic projections to the nuclei of the thalamus. Dopamine, the main transmitter of the substantia nigra, regulates the projections from the substantia nigra to the striatum.

Although most knowledge on the interconnectivity of the basal ganglia, thalamus and cerebral cortex comes from tracing studies in animals (in primate, e.g., Hackett et al., 2007, and non-primate mammals, e.g., Winer et al., 2001), there is no reason to assume that connectivity greatly differs in humans (Winer, 2006). Parent and Hazrati (1995) describe three distinct circuits which project from

---

[9] For a comprehensive description of basal ganglia morphology and connectivity, the reader is referred to Kotz (2006).

the basal ganglia to the cortex: the putamen projects to the motor, premotor cortex and SMA, the caudate nucleus projects to the prefrontal cortex, and the ventral striatum to the ACC. The largest nucleus of the thalamus, the pulvinar, albeit not being part of the striato-thalamo-cortical loop, has direct connections to the STG and has therefore been described as central for auditory processing (Romanski et al., 1997).

Because the basal ganglia and the thalamus are in a nodal position where information is transferred from and to the cortex, they exert multiple functions. Therefore, it has been challenging to specifically define their functions. Generally speaking, these subcortical structures have been proposed to control and select the information flow from and to the cortex (Bar-Gad et al., 2003). Traditionally, the basal ganglia have been linked to motor function (Mink, 1996), but an emerging view is that they also contribute to other, not strictly motoric processes (Middleton and Strick, 2000), such as language processing.

An obvious hypothesis for a role of the basal ganglia in language processing is their contribution to speech production (Crosson, 1999). Disruption of striatal function indeed leads to language production deficits as evidenced by electrical stimulation studies (Van Buren and Ojemann, 1966), and studies of patients with Parkinson's disease (Kotz, 2006). In our paradigm speech production was also involved, to measure comprehension (Experiment III). Still, although speech production increased over time along with adaptation, we argue that the observed down-regulation of striatal and thalamic structures cannot merely be explained by motoric processes, because speech production would be expected to be accompanied by an *increase* rather than a *decrease* of subcortical activity (Mink, 1996). Rather, the discerned subcortical dynamics are likely to reflect perceptual processes, at least to some extent.

In support of this argumentation, more recent studies noted basal ganglia involvement in purely perceptual speech tasks. In particular, syntactical computation was accompanied by activation of the basal ganglia (Moro et al., 2001; Friederici et al., 2003; for review see Kotz et al., 2009). However, such a striatal function in syntactical processing has been attributed to working memory or attentional processes, based on studies in Parkinson's disease patients (Lieberman et al., 1990; Grossman et al., 1991; Lieberman et al., 1992; Grossman et al., 1993).

By now, a relatively well-established function of the basal ganglia is their involvement in learning, in particular, reinforcement learning (Bar-Gad et al., 2003, for review see Graybiel, 2005). The dopaminergic neurons of the substantia nigra project to the striatum, from where projections are sent to the neocortex. This feedforward-feedback connection has been proposed to form the anatomical basis for learning (e.g., motor learning, Joel and Weiner, 1994). Thus, dopaminergic input neurons from the substantia nigra are thought to send a "teaching" signal to the basal ganglia. The basal ganglia in turn "teach" the cortex via thalamo-cortical connections (Graybiel, 2005).

The nature of this dopaminergic teaching signal is controversial. Based on the reward prediction error hypothesis (Montague et al., 1996; Schultz, 2002), phasic dopaminergic activity signals reward prediction errors (Hollerman and Schultz, 1998). Prediction errors are used by reinforcement learning mechanisms to adaptively shape behaviour in order to ultimately maximise reward. More recent hypotheses challenge this view, suggesting that the dopaminergic response is not necessarily related to expected reward but to the novelty or saliency of a stimulus (Redgrave and Gurney, 2006; Laurent, 2008), opening up a new framework for explaining non-reward-related basal ganglia activity. Our finding agrees with the latter view in that a novel stimulus (i.e., vocoded speech) initially activates the striatum, whereas striatal activity decreases as a participant adapts to the stimulus (and its novelty diminishes), even in the absence of feedback or reward.

Perceptual learning studies reporting dopaminergic involvement are rare (see Section 9.1). Convincing evidence for dopaminergic role in the reorganisation of auditory cortical processing comes from an animal study. Bao et al. (2001) demonstrated that electric stimulation of dopaminergic neurons from the ventral tegmental area with a concurrent presentation of a specific tone leads to enhancement of the cortical representation and selectivity of the neural response to the stimulus frequency in auditory cortex (Bao et al., 2001). In humans, one recent non-speech auditory categorisation study without explicit feedback found increased caudate nucleus activity in a condition where auditory sound categories were structured and thus learnable, relative to an unstructured condition (Lim et al., 2013). Functional connectivity between the striatum and STG was enhanced in the learnable condition. The authors interpreted this learning-related modulation of cortico-striatal connectivity as feedback-related signalling from the striatum which reorganises perceptual representations in sensory regions, possibly through dopaminergic influences (Lim et al., 2013). Further, Weis et al. (2013) showed direct evidence for a dopaminergic effect on auditory processing during feedback-based learning: During the learning of the association between frequency-modulated tones and a monetary reward, auditory cortex activity was enhanced. However, this effect was extinguished by the administration of L-DOPA (Weis et al., 2013).

Summing across the studies cited above, the following picture emerges: the basal ganglia and the thalamus (beyond the sensory thalamus) appear to play a greater role in auditory processing than assumed so far (Kotz et al., 2009). Their role in some forms of learning, in particular reinforcement learning, is widely accepted (Graybiel, 2005), and presumably originates from dopaminergic signalling (Shohamy et al., 2008). Our data reconcile both functions and speak to a critical role of thalamic and striatal structures in *perceptual* learning of speech.

Such subcortical involvement in short-term perceptual adaptation to degraded speech has not been detected in previous imaging studies (Adank and Devlin, 2010; Eisner et al., 2010), likely due to artefacts in subcortical regions which we avoided by the use of cardiac gating (see Experiment III). It is reasonable to speculate that subcortical changes precede and underlie the previously observed cortical changes (Pasupathy and Miller, 2005) in premotor cortex (Adank and Devlin, 2010) and IFG (Eisner et al., 2010), in proximity to which we also observed adaptation-related cortical upregulations (Experiment III). In sum, our data allow us to propose that perceptual learning is critically rooted in the unique contribution of the striato-thalamo-cortical circuit. Striato-thalamo-cortical learning systems might generate internal feedback signals, which afford the reorganisation of perceptual representations. Predictions derived from this finding and research avenues that could further test a causal striato-thalamic role in perceptual learning will be outlined in Section 9.1.

8.3 The cingulo-opercular system – executive mechanisms in degraded speech processing?

A concept which has re-gained popularity in the last decade is the idea of “listening effort” (Rabbitt, 1968). It is thought to be an interaction of processing demands and cognitive abilities (Wild et al., 2012). That is, perceiving degraded speech should require more cognitive resources (e.g., attention) than clear speech does. Several recent studies endorsed identifying the neural resources that correspond to listening effort. The cingulo-opercular system was isolated as a good candidate network supporting such executive processes (Eckert et al., 2009; Wild et al., 2012; Vaden et al., 2013).

In support of these findings, we have here repeatedly observed an activity increase in the cingulo-opercular system when auditory processing became difficult: during degraded more than clear speech perception and during AM rate discrimination when stimuli became less discriminable (Experiment III).

The insula most likely exhibits an inverted u-shaped activation function where the haemodynamic response is low during easy tasks (e.g., clear speech perception), increases with higher task demands (e.g., degraded speech perception), and drops again for the most difficult conditions (e.g., quasi unintelligible conditions; Wild et al., 2012). Such an activation profile is compatible with an error-monitoring function of the insula (Sharp et al., 2006). Functional MRI studies from our group support this notion in that the insula was repeatedly shown to scale with task difficulty (Experiment III; Henry et al., 2013; Herrmann et al., 2013b). Moreover, MEG studies with high temporal and spatial resolution demonstrated that neuronal oscillations in the alpha frequency band (8-13 Hz), which can be taken as a marker of increased cognitive load (Obleser et al., 2012), originate primarily from the anterior insula (Wilsch et al., 2014). Elevated cingulo-opercular activity was not only observed during difficult (i.e., physically degraded) speech perception, and when errors were made, but also led to improved speech perception on subsequent trials, possibly recruiting attentional resources following erroneous responses (Vaden et al., 2013). Consistent with this observation, we found that increased ACC activity correlated with overall better speech comprehension performance (Experiment IV).

These regions' profiles were found to be altered with older age and age-related hearing loss. Hearing loss changed the activation pattern in the anterior insula such that activity was elevated during clear rather than degraded speech. We can think of such hearing-loss related changes as of long-term adaptation, where hearing-impaired listeners develop a (mal-) adaptive strategy in order to compensate for their decline in sensory acuity (Cabeza et al., 2002; Cabeza et al., 2004). Recently, a long-term perceptual learning study in normal-hearing listeners who wore a real-time vocoder during two weeks showed upregulated activity of the anterior insula, Heschl's gyrus and the SMA following chronic exposure to vocoded speech (Smalt et al., 2013). Here, none of these regions were observed to be related to short-term adaptation. Instead, we observed activity in the anterior insula to be altered in hearing loss (Experiment IV), a condition which leads to a chronic internal degradation of the auditory signal. Prolonged exposure to speech degradation, due to artificial degradation or hearing loss, appears thus to lead to an upregulation of the cingulo-opercular network, which in some cases can support comprehension (see Figure 6.4; Vaden et al., 2013).

The ACC was generally upregulated in older adults (both during clear and degraded speech perception; Experiment IV). Hence, the older adults' BOLD signal in the ACC appears to be less flexible in response to increased processing demands (i.e., speech degradation), as these systems' dynamic range is exceeded in comparably low task demands (i.e., clear speech perception). Such decreased flexibility is consistent with a previously observed age-related decline in the variability of the haemodynamic signal (Garrett et al., 2010). Garrett et al. (2011) proposes that increased trial-to-trial variability of the haemodynamic response improves performance and decreases reaction times, because relatively more variability enables the neural system to transition flexibly from one state to another. As older brains appear to be less variable than younger brains, that is, exhibit a decreased dynamic range, the capacity of the brain to explore different states is compromised, which in turn appears to be detrimental to performance.

A better understanding of how the deregulation of cingulo-opercular activity affects speech comprehension and auditory processing in general is vital in clinical contexts. In particular, dissociation between the distinct effects of age and hearing loss on neural processing is warranted. For example, hearing-aid- and CI users of different ages and with unique histories of hearing loss likely draw on distinct neural and listening strategies when coping with degraded auditory input (see Section 9.3).

8.4 Temporal envelope cues in speech comprehension

A reappearing theme in this thesis was the close link between processing of non-speech temporal envelope cues and spectrally degraded speech. Modulation rate discrimination abilities were correlated with speech comprehension (Experiments I and II) and relied on similar neural resources as degraded speech processing (Experiment III), confirming the importance of a bottom-up analysis the auditory system performs on the signal (Idemaru and Holt, 2011).

The on-going debate on bottom-up (Norris et al., 2000) versus top-down (McClelland and Elman, 1986) influences on speech perception is beyond the scope of this thesis but will briefly be touched upon here. Its central question is whether during speech processing, "information flow is strictly feedforward [i.e., from the cochlea to auditory cortex and beyond] ("bottom-up") or whether high-level semantic and syntactic computations interact directly with perceptual regions to change their activity, guiding lower-level perceptual processing top-down" (p. 3915, Davis et al., 2011).

Some behavioural studies speak to a top-down account for perceptual learning of speech (Davis et al., 2005; Huyck and Johnsrude, 2012), showing that lexical access facilitates learning (see Section

1.1.2). Neural evidence for top-down mediated perceptual learning has recently been gathered by imaging studies. The IFG, which has been implicated in higher-order language processes (Davis and Johnsrude, 2003; Friederici et al., 2003), was proposed to be the site where knowledge-based cues provided by the feedback are integrated into the computation on the acoustic signal (Eisner et al., 2010). A follow-up study (Sehm et al., 2013) showed that facilitatory tDCS over IFG during training on noise-vocoded minimal word pairs boosted perceptual learning when the acoustic signal was severely degraded. These studies confirm that higher-order language areas significantly contribute to perceptual learning of speech when feedback is provided.

However, in the present thesis, no evidence for a specific IFG involvement in adaptation was found, possibly due to the absence of feedback. Our results suggest that IFG involvement might relate to feedback-based processes but not percetual learning *per se*. Our data impact the debate on the contributions of top-down and bottom-up processes to speech perception, and more specifically, adaptation (e.g. Idemaru and Holt, 2011; Huyck and Johnsrude, 2012) in that they confirm that adaptation is possible in the absence of feedback and when contextual semantic cues are kept low (here, only low-predictable sentences were presented). The observed correlation between AM rate discrimination thresholds and learning further implies that a higher precision in the encoding of the temporal envelope leads to more efficient learning, consistent with important bottom-up contributions to perceptual adaptation. A suggestion of how to test such a causal involvement of temporal envelope encoding in perceptual learning will be outlined in Section 9.6.

In Experiment III we detected elevated Heschl's gyrus activity during AM rate discrimination when standard and deviant AM rate increasingly differed. This was interpreted as increasing activity due to diminishing task difficulty. Nevertheless, such enhanced activity with increasingly distinctive AM rates can also be understood as a release from repetition suppression: the response magnitude of the auditory cortex to a stimulus decreases after repeated presentation of a similar, or in the most extreme case, the same stimulus (Herrmann et al., 2014). Such repetition suppression or "adaptation" effects in the auditory cortex have mostly been investigated using EEG (Naatanen et al., 1988) and are well-documented for frequency-specific adaptation (Herrmann et al., 2013a; Herrmann et al., 2014). Our data are consistent with an account of release from repetition suppression, because the amplitude of the haemodynamic response increased with more distinctive stimuli.

Envelope expansion in noise yielded a small but highly significant benefit for speech recognition (Experiment V). Interestingly, sensorineural hearing loss was observed to alter encoding of the speech envelope. It appears that some listeners with hearing loss show more accurate encoding of the auditory envelope (Anderson et al., 2013), as shown by enhanced AM discrimination thresholds (Fullgrabe et al., 2003). Therefore, an enhancement of speech envelope cues might improve speech recognition performance in the hearing-impaired when fed into hearing aid algorithms (Lorenzi et al., 1999; Apoux et al., 2001). Envelope expansion likely also contributes to an improvement in the "ease of listening", as suggested by the decrease in reaction times to envelope-expanded words by ca. 15 milliseconds (Apoux et al., 2004), possibly freeing cognitive resources for other processes (see Section 9.6).

At which level of the auditory system are such temporal envelope cues are encoded? There is ample evidence that modulation rate is represented hierarchically along the auditory pathway, with lower modulation frequencies being processed higher up in the auditory system (i.e., closer to or in the auditory cortex; Schreiner and Urbas, 1986; Langner and Schreiner, 1988; Rees and Palmer, 1989; Giraud et al., 2000; see also our finding from Experiment III of auditory cortex activity for AM sounds centred on the modulation rate of 4 Hz). The capability of discriminating temporal modulation rates appears to be inherently coupled to the expression of calcium channels at the level of the auditory brainstem (Pirone et al., 2014). Whether a manipulation of modulation *depth* rather than modulation *rate* of speech sounds would similarly be represented in lower structures of the auditory path way or be passed on to the auditory cortex remains to be determined. Our paradigm is suited to resolve the stage at which envelope modification is represented in the auditory system using fMRI. Alternatively, electrophysiological studies could be used to track the neural responses to envelope-modified words with high temporal resolution (Nourski et al., 2009).

9 Implications for future research

This section of the thesis attempts to integrate the findings of this thesis into the larger context of existing evidence on adaptive neural mechanisms. Theoretical implications that follow from the empirical evidence and potential avenues for future research will be discussed. We focus on open questions that were not covered by this thesis and speculate based our observations. In particular, unsolved issues on the functional neuroanatomy of perceptual learning on different scales, from cellular to macroanatomical levels, will be highlighted and experimental set-ups to further elucidate on the brain's plasticity will be proposed. Finally, we will describe a few practical applications derived from the empirical evidence.

9.1 A potential dopaminergic role in perceptual adaptation

Several pieces of evidence were presented for a crucial subcortical involvement in perceptual learning of degraded speech (Experiments I and III). It is reasonable to speculate that the observed learning-related plasticity is mediated by the dopaminergic system, for the following reasons. First**,** dopamine is one major neurotransmitter of the basal ganglia, regulating projections from the substantia nigra to the striatum. Second, there is ample evidence from animal studies that dopamine is a key player in synaptic plasticity (Graybiel, 2005), and is causally involved in different forms of learning, notably in reinforcement learning (Schultz et al., 1997; Bayer and Glimcher, 2005; Steinberg et al., 2013; see Section 8.2). However, there is a gap between the rich literature on the dopaminergic role in higher cognitive functions during feedback-based learning, such as anticipation of reward (Graybiel, 2005) on the one hand, and few studies on the potential dopaminergic functions in *perceptual* processing on the other hand. Therefore, future studies are needed to clarify whether dopaminergic signalling is also a crucially involved in the reorganisation of perception, and could underlie the macroanatomically observed striato-thalamic role in perceptual learning (Experiment III). Different experimental set-ups can be imagined to test this prediction more directly.

The haemodynamic signal is only an indirect measure of neural activity and most likely associated with increased synaptic activity and transmitter release (Arthurs and Boniface, 2002). However, specific neurotransmitters cannot be targeted. In contrast, radioligands-based positron emission tomography (PET) enables the labelling of specific molecular targets, including dopamine (e.g., Cools et al., 2009). Thus PET studies can directly target the dopaminergic system. Different radioligands are available to measure in vivo dopamine release on based on the competition between the ligand and endogenous dopamine ([^{11}C]raclopirde; Salimpoor et al., 2011); to estimate dopamine receptor density (Ikoma et al., 2010); or to assess baseline dopamine synthesis capacity based on the uptake of the ligand ([^{18}F]fluorometsthyrosine, Cools et al., 2009).

Direct online assessment of dopaminergic signalling using PET during short-term perceptual adaptation could clarify whether dopamine critically contributes to self-regulated learning, in the absence of explicit reinforcement. Based on the striatal haemodynamics observed here, a high dopaminergic signal would be expected in the initial learning phase, with dopamine levels decreasing over time as learning reaches a plateau. Further, Cools et al. (2009) demonstrated that reinforcement learning was related to individual differences in baseline striatal dopamine synthesis. Further, PET scanning could directly assess dopamine receptor density as genetic trait (Ikoma et al., 2010), in order to examine a possible relationship between receptor density and learning. Taking a big leap, one could speculate that endogenous dopamine synthesis levels or receptor density partly explain individual differences in perceptual adaptation to challenging listening situations, and possibly account for part of the variance in CI outcome.

Alternatively, the potential dopaminergic role for perceptual learning could be tested by means of a pharmacological intervention, that is, administration of L-DOPA versus dopamine receptor antagonists (e.g., Haloperidol). Moreover, studies in Parkinson's disease patients with a depletion of striatal dopamine could reveal effects of globally altered dopamine concentrations on learning. It is controversial, however, how such a generalised increase or decrease of dopamine levels impacts learning. Animal studies suggest that the absolute quantity of dopamine is not critical for learning, but rather the timing of dopamine release (Shohamy et al., 2008). Hence, a global elevation of dopamine could in fact interfere with learning (Weis et al., 2013) by masking stimulus-induced signalling, thus effectively decreasing the dynamic range of the dopaminergic system. Such impairment of feedback-based reinforcement learning through dopamine medication is supported by studies with Parkinson's patients (Shohamy et al., 2006), but depends on task demands (e.g., dopaminergic medication restores the ability to switch between tasks, but impairs probabilistic learning, Cools et al., 2001). Hence, medical manipulation of global dopamine levels could yield insights into whether the dopaminergic system is involved in the reorganisation of perceptual representations during feedback-free learning.

9.2 Short- and long-term adaptive plasticity

An important issue not addressed by the present thesis is the temporal persistence of adaptation-induced changes. Therefore, we speak of perceptual *adaptation,* rather than *learning,* the latter implying "long-lasting" changes (Goldstone, 1998). Does the striato-thalamo-cortical loop merely mediate short-term perceptual learning, or does it induce long-lasting changes of perceptual representations? If striato-thalamo-cortical connections are the driving system that reorganises perceptual representations, then where and how are these perceptual changes "stored"?

An assumption that we did not explicitly test in the present thesis is that short-term exposure to degraded speech leads to a persistent perceptual reorganisation. Yet, there is empirical evidence in support of this premise showing long-lasting improvements in comprehension several weeks after vocoded speech learning (McGettigan, 2007; Stacey and Summerfield, 2007b), consistent with Goldstone's definition of for perceptual learning (Goldstone, 1998). How such persistent perceptual changes express in terms of central neural mechanisms need to be elucidated in longitudinal studies.

Chronically degraded auditory input might lead to different adaptive mechanisms than short-term exposure to degraded conditions. Long-term learning studies, such as the one by Smalt et al. (2013), where degraded conditions were experimentally induced in normal-hearing adults using a real-time CI simulation over several weeks, can yield insights into how long-term adaptation is afforded by neural plasticity. Further, structural and functional neuroimaging studies in patients suffering from peripheral hearing loss or CI patients provide a unique window on brain plasticity following altered auditory input. In the following sections, different scenarios of chronically degraded auditory input and potential adaptive processes will be outlined.

9.3 Neural plasticity in older age and hearing impairment

The study of hearing-impaired listeners opens up the opportunity to study the (mal-) adaptive mechanisms following long-term deprivation from, or degradation of, auditory input. We have presented evidence for hearing loss-related changes in neural processing corroborating hypotheses of altered executive networks in older adults' speech comprehension (Experiment IV). However, certain aspects of this study may be refined in order to draw more clear-cut conclusions on hearing loss- and age-related neural changes.

In our sample, it was problematic to separate effects of hearing loss from effects of age on neural processing, as both factors were highly correlated. Although within the older group, we were able to robustly show correlations of hearing loss with neural activation patterns, even after controlling for age, comparisons between age groups are confounded by differences in hearing acuity. More carefully controlled experiments with hearing-impaired older participants and age-matched normal-hearing controls, or hearing-impaired young adults respectively, will be able to disentangle the contributions of hearing loss and ageing on neural mechanisms of speech perception. Further, recruiting middle-aged participants rather than studying discrete age groups would allow for scrutinising the gradual neural changes with age in a cross-sectional approach.

The high variance in hearing acuity within the older group might have obscured different aspects of adaptive processes in older age. For example, we did not find a direct neural correlate of short-term adaptation to degraded speech (Experiment IV, results not reported), although we demonstrated that older adults' capacity of perceptual learning is behaviourally preserved. Future studies in a more homogenous older group could clarify how ageing alone might change the neural plasticity underlying short-term adaptation. Identifying neural plasticity in older age is all the more interesting with respect to our dopaminergic hypothesis for adaptation (Section 9.1), because normal ageing has been linked to deficient dopaminergic neuromodulation (Kaasinen et al., 2000; Li et al., 2001; Braver and Barch, 2002).

Dopaminergic deficits have also been implicated in the age-related decrease in functional differentiation and specialisation of the neural response (Li et al., 2001) as evidenced in multiple imaging studies (e.g. Park et al., 2004). We found evidence consistent with both the dedifferentiation (Park et al., 2004) and the compensation hypothesis (Cabeza et al., 2002) of aging: While age-related engagement of the MFG was associated with better performance, and is thus truly compensatory, the persistent upregulation of ACC could not raise older adults' performance to the level of young adults and thus speaks in favour of the dedifferentiation hypothesis. Therefore, our data contribute to, but cannot solve, the ongoing debate of dedifferentiation versus compensation. Whether the observed age-related changes are intrinsically coupled to cortical atrophies (Eckert et al., 2012) needs to be verified in a combination of brain structural and functional MRI studies.

Lastly, the generalised upregulation of the ACC in the older brain was interpreted as an executive mechanism due to higher listening effort. Future studies could evaluate listening effort more directly using intelligibility ratings or reaction times (Desjardins and Doherty, 2013) in addition to word report scores, to verify whether they correlate with cingulo-opercular activity.

9.4 Neural plasticity in cochlear implant patients

To what degree do the adaptive mechanisms observed here in normal-hearing listeners apply to CI patients when adapting to degraded auditory conditions? The finding that both normal-hearing and CI listeners' speech comprehension is strongly related to temporal modulation rate discrimination skills (Experiments I and II) supports similar listening strategies in both groups. Nevertheless, this does not mean that the underlying neural mechanisms are identical.

The work on the neural adaptive processes done so far in CI patients has mostly concentrated on cross-modal plasticity. A visual take-over of the auditory cortex in prelingually deafened patients has been shown to predict a poor cochlear implantation outcome (Lee et al., 2001; Giraud and Lee, 2007; Lee et al., 2007). In support of this view, a recent EEG study demonstrated that visual activation of auditory cortex following cochlear implantation is detrimental to speech recognition (Sandmann et al., 2012). On the other hand, visual cortex activation in response to auditory stimulation has been taken as an adaptive process, because activation amplitudes depended on the time of experience with the implant (Giraud et al., 2001b). In support of this finding, we similarly observed a down-regulation of activity in occipital cortex with adaptation (Experiment III). However, caution must be taken in drawing an analogy between normal-hearing and CI listeners due to the cortical reorganisation induced by long-term deprivation of auditory input (Sandmann et al., 2012).

To our knowledge, no imaging study has linked perceptual adaptation in CI patients to subcortical mechanisms. Based on our finding in normal-hearing adults listening to a CI simulation, striato-thalamo-cortical pathways might be the driving force of such perceptual reorganizations. Hence, our data allow for the speculation that the previously observed cross-modal plasticity in CI patients might be afforded by preceding striato-thalamic activity. Although the available imaging techniques are limited, because CIs are not MR-compatible, PET (Giraud et al., 2001a) could be used to further elucidate the functional neuroanatomy of adaptation to CI transduced speech. A very speculative scenario where adaptive, reorganisational processes are driven by dopaminergic signalling in the striato-thalamo-cortical loop (see Section 9.1) would predict that patients with reduced levels of dopaminergic activity (in the most extreme case, Parkinson's disease patients) are impaired in their ability to adjust to the degraded input.

One important consequence that can unambiguously be deduced from our other key finding of age- and hearing loss-related differences in neural speech processing is that in CI patients, a variety of different factors (e.g., age, history of hearing loss and most likely other factors) will crucially influence neural processes for successful speech comprehension. Thus, it is time to shape CI patients' listening and neural strategies in an individual manner, in the best case on the basis of demographic factors in combination with reliable markers of individual auditory abilities (see also Section 9.6).

9.5 Brain structural plasticity in altered auditory input

Our brain structural analysis examining predictors of adaptation leaves open the question of where inter-subject differences in pulvinar morphology originate from. Two possibilities can be thought of, namely that these individual differences are a pre-existing individual characteristic or are shaped by experience. Whether and to what extent training of degraded speech perception might alter pulvinar morphology, and potentially boost learning, remains to be clarified.

In the last decade, numerous studies have shown that modification of brain structure occurs concurrently with various types of learning (e.g., motor or musical training; Gaser and Schlaug, 2003; Driemeyer et al., 2008; Scholz et al., 2009; Taubert et al., 2010), refuting the idea that the adult brain is a stable, unchanging matter. In contrast to traditional assumptions that brain structural reorganisation only occurs during childhood, these studies provide evidence for a brain that remains plastic beyond childhood and even into older age (Lovden et al., 2010). The cellular mechanisms that mediate the structural plasticity observed on a macroanatomical level are not fully understood. Grey matter changes are unlikely to reflect adult neurogenesis, which is only well-established for the hippocampus (Deng et al., 2009; Tronel et al., 2010), but controversial for the neocortex (Gould et al., 1999; Nowakowski and Hayes, 2000). Rather, grey matter changes involve synaptic, glial and vascular growth (for review, see Zatorre et al., 2012).

Based on MRI studies showing white matter changes in response learning of artificial grammar (Floel et al., 2009) or a foreign language (Schlegel et al., 2012), it is highly plausible that structural changes would be detectable in the adult brain following exposure to degraded speech when using sufficiently sensitive methods (i.e., high-field imaging; Tourdias et al., 2014). Longitudinal MRI studies (e.g., Taubert et al., 2010) of degraded speech learning as well as brain structural investigations in CI patients, as far as feasible, could give insights on how (chronically) degraded auditory input impacts brain structure. Importantly, such studies could show the anatomical locus for the "storage" of learning-induced changes and provide a macroanatomical basis for behaviourally observed long-lasting changes induced by perceptual learning (Stacey and Summerfield, 2007a).

9.6 A causal role for temporal envelope processing in adaptation?

The present thesis refrained from manipulation of learning, but rather observed the neural correlates of self-regulated, feedback-free learning. Previous studies on perceptual learning of speech have attempted to give different forms of feedback, in order to enhance learning by providing top-down knowledge-based cues. For example, written feedback (Eisner et al., 2010) or a clear auditory presentation of the degraded word (Davis et al., 2005; Hervais-Adelman et al., 2012) have been shown to boost learning.

Instead, we show different pieces of evidence that low-level auditory skills are crucial for perceptual learning and comprehension of degraded speech (Experiments I, II, III and V), confirming the importance of a bottom-up influence in perceptual learning (Idemaru and Holt, 2011). A corollary to the observed correlation between AM rate discrimination abilities and adaptation is that one should be able to boost perceptual adaptation by enhancing the sensitivity of the auditory system to temporal non-speech modulations (Amitay et al., 2013). Thus, training and learning in an AM discrimination task might benefit degraded speech comprehension. In Experiment I, we did not find evidence for an improvement in AM rate discrimination over the short period of testing (20 min). However, a more intensive training regimen over several days led to significant improvements in AM discrimination performance (Fitzgerald and Wright, 2011). To speculate, AM discrimination training could enhance the precision of temporal envelope cue encoding in the auditory system, and therefore be a powerful tool to foster perceptual learning of speech.

Some potential clinical applications can be derived from the presented empirical evidence. We gathered preliminary evidence for the predictive power of AM rate discrimination thresholds on future speech perception in CI users, at least when duration of deafness is taken into account (Experiment II). Our results in CI patients confirm that non-speech tests specifically assessing auditory temporal processing would be a valuable feature to add to the post-operative rehabilitation program in addition to the more widespread speech tests. AM discrimination assessment of only 20 minutes duration appears to provide clinically relevant information and could easily be incorporated into the clinical routine. Similarly, assessment of working memory capabilities could be a valuable addition to the clinical test battery, although the results concerning working memory capacities were ambiguous in the present thesis (Experiments I and IV). Using a more comprehensive battery of tests would enable building detailed individual listener profiles, which ultimately might permit the development of individual treatment approaches in order to intervene at specific levels in the auditory processing cascade.

Moreover, we showed that envelope expansion in noise facilitates speech comprehension and can restore intelligibility (Experiment V). Together with previous findings (Lorenzi et al., 1999; Apoux et al., 2001; Apoux et al., 2004), our result bears particular relevance for hearing aid algorithms. Importantly, we extracted the envelope *after* noise was added (rather than before) and found a beneficial effect of envelope expansion. This is reminiscent of what a hearing aid processor, provided with a signal-and-noise mix by the microphone, can achieve without computationally costly and potentially distorting source separation or denoising algorithms. Implementations of such comparably simple expansion schemes in auditory prostheses might yield improved comprehension in noise.

10 Conclusions

The current thesis examined the neural dynamics that support comprehension of speech when listening conditions become difficult. In particular, we showed how normal-hearing listeners, older adults with varying degrees of hearing loss, and CI patients cope with spectral degradation of the signal.

Behavioural evidence showed that AM rate discrimination thresholds predicted adaptation to degraded speech in normal-hearing listeners and in CI patients, implying that AM discrimination thresholds may be used as powerful predictors of CI outcome. A signal processing scheme which enhanced temporal envelope cues of the signal was shown to improve comprehension in noisy environments, confirming the importance of temporal envelope cues in speech perception. Our findings suggest that the precision of temporal envelope cue encoding is decisive for perceptual learning of speech and emphasise the “bottom-up” acoustic analysis the auditory system must perform when confronted with degraded speech.

The neural plasticity underlying short-term perceptual adaptation to degraded speech relied on mainly subcortical contributions. Adaptation was predicted by morphology of the pulvinar, and was accompanied by a haemodynamic down-regulation of striato-thalamic nuclei. The identified structures form part of the striato-thalamo-cortical loop which is proposed to reconfigure cortical activation patterns. Engagement of this pathway is likely to sharpen the cortical representation of the degraded stimulus, potentially through dopaminergic signalling. Our findings contrast with previous evidence for a role of the IFG in feed-back based learning and imply that self-regulated learning is guided by subcortical structures.

Effortful processing of speech and non-speech was accompanied by an increase of cingulo-opercular activity, most likely reflecting executive (e.g., attentional) processes. In older adults, this network exhibited a generalised upregulation, during perception of both degraded and clear speech, corroborating hypotheses of increased reliance on cognitive control systems in older adults. Specifically, we show that the dynamic range of ACC activity decreases with age which in turn is detrimental to speech comprehension. Rather than playing a compensatory role for deficits due to aging, the observed cingulo-opercular dynamics appear to reflect a dysfunctional age-related dedifferentiation of neural networks.

The present thesis advances our understanding of how speech is processed under challenging hearing conditions. The results are of particular relevance with respect to hearing disorders and specifically CI patients for whom adaptive mechanisms permit successful speech comprehension in the absence of a clear signal.

References

Abercrombie D (1967) Elements of General Phonetics. Edinburgh: University Press.

Adank P (2012) The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. Brain and language 122:42-54.

Adank P, Devlin JT (2010) On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. Neuroimage 49:1124-1132.

Adank P, Janse E (2010) Comprehension of a novel accent by young and older listeners. Psychology and aging 25:736-740.

Amitay S, Hawkey DJ, Moore DR (2005) Auditory frequency discrimination learning is affected by stimulus variability. Perception & psychophysics 67:691-698.

Amitay S, Irwin A, Moore DR (2006) Discrimination learning induced by training with identical stimuli. Nature neuroscience 9:1446-1448.

Amitay S, Zhang YX, Jones PR, Moore DR (2013) Perceptual learning: Top to bottom. Vision research.

Anderson JR, Schooler LJ (1991) Reflections of the Environment in Memory. Psychol Sci 2:396-408.

Anderson S, White-Schwoch T, Choi HJ, Kraus N (2013) Training changes processing of speech cues in older adults with hearing loss. Front Syst Neurosci 7:97.

Apoux F, Crouzet O, Lorenzi C (2001) Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners: effects on identification performance and response times. Hear Res 153:123-131.

Apoux F, Tribut N, Debruille X, Lorenzi C (2004) Identification of envelope-expanded sentences in normal-hearing and hearing-impaired listeners. Hear Res 189:13-24.

Ari-Even Roth D, Amir O, Alaluf L, Buchsenspanner S, Kishon-Rabin L (2003) The effect of training on frequency discrimination: generalization to untrained frequencies and to the untrained ear. Journal of basic and clinical physiology and pharmacology 14:137-150.

Arthurs OJ, Boniface S (2002) How well do we understand the neural origins of the fMRI BOLD signal? Trends in neurosciences 25:27-31.

Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95-113.

Ashburner J (2009) Computational anatomy with the SPM software. Magnetic resonance imaging 27:1163-1174.

Ashburner J, Friston KJ (2000) Voxel-based morphometry--the methods. Neuroimage 11:805-821.

Baddeley A (2012) Working memory: theories, models, and controversies. Annual review of psychology 63:1-29.

Bacon SP, Viemeister NF (1985) Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. Audiology : official organ of the International Society of Audiology 24:117-134.

Baltes PB, Lindenberger U (1997) Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? Psychology and aging 12:12-21.

Bao S, Chan VT, Merzenich MM (2001) Cortical remodelling induced by activity of ventral tegmental dopamine neurons. Nature 412:79-83.

Bar-Gad I, Morris G, Bergman H (2003) Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. Progress in neurobiology 71:439-473.

Barkat TR, Polley DB, Hensch TK (2011) A critical period for auditory thalamocortical connectivity. Nature neuroscience 14:1189-1194.

Bartlett EL (2013) The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception. Brain and language 126:29-48.

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129-141.

Behrens TE, Johansen-Berg H, Woolrich MW, Smith SM, Wheeler-Kingshott CA, Boulby PA, Barker GJ, Sillery EL, Sheehan K, Ciccarelli O, Thompson AJ, Brady JM, Matthews PM (2003) Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nature neuroscience 6:750-757.

Bejjanki VR, Beck JM, Lu ZL, Pouget A (2011) Perceptual learning as improved probabilistic inference in early sensory areas. Nature neuroscience 14:642-648.

Belin P, Zatorre RJ, Hoge R, Evans AC, Pike B (1999) Event-related fMRI of the auditory cortex. Neuroimage 10:417-429.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Royal Statistical Society 57, 289 - 300.

Bent T, Buchwald A, Pisoni DB (2009) Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. J Acoust Soc Am 126:2660-2669.

Bent T, Loebach JL, Phillips L, Pisoni DB (2011) Perceptual adaptation to sinewave-vocoded speech across languages. J Exp Psychol Hum Percept Perform 37:1607-1616.

Berger H (1929) Über das Elektrenkephalogramm des Menschen. Archiv für Psychiatrie und Nervenkrankheiten 87:527-570.

Bermudez P, Zatorre RJ (2005) Differences in gray matter between musicians and nonmusicians. Annals of the New York Academy of Sciences 1060:395-399.

Blamey PJ, Pyman BC, Gordon M, Clark GM, Brown AM, Dowell RC, Hollow RD (1992) Factors predicting postoperative sentence scores in postlinguistically deaf adult cochlear implant patients. The Annals of otology, rhinology, and laryngology 101:342-348.

Blamey PJ, Sarant JZ, Paatsch LE, Barry JG, Bow CP, Wales RJ, Wright M, Psarros C, Rattigan K, Tooher R (2001) Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. J Speech Lang Hear Res 44:264-285.

Bozic M, Tyler LK, Ives DT, Randall B, Marslen-Wilson WD (2010) Bihemispheric foundations for human speech comprehension. Proc Natl Acad Sci U S A 107:17439-17444.

Bradlow AR, Bent T (2008) Perceptual adaptation to non-native speech. Cognition 106:707-729.

Braver TS, Barch DM (2002) A theory of cognitive control, aging cognition, and neuromodulation. Neuroscience and biobehavioural reviews 26:809-817.

Brett M, Anton J, Valabregue R, Poline J (2002) Region of interest analysis using an SPM toolbox. Presentation at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan June 2-6, 2002.

Brown GG, Eyler Zorrilla LT, Georgy B, Kindermann SS, Wong EC, Buxton RB (2003) BOLD and perfusion response to finger-thumb apposition after acetazolamide administration: differential relationship to global perfusion. J Cereb Blood Flow Metab 23:829-837.

Buchel C, Morris J, Dolan RJ, Friston KJ (1998) Brain systems mediating aversive conditioning: an event-related fMRI study. Neuron 20:947-957.

Burkholder RA, Pisoni DB, Svirsky MA (2005) Effects of a cochlear implant simulation on immediate memory in normal-hearing adults. International journal of audiology 44:551-558.

Cabeza R, Anderson ND, Locantore JK, McIntosh AR (2002) Aging gracefully: compensatory brain activity in high-performing older adults. Neuroimage 17:1394-1402.

Cabeza R, Daselaar SM, Dolcos F, Prince SE, Budde M, Nyberg L (2004) Task-independent and task-specific age effects on brain activity during working memory, visual attention and episodic retrieval. Cereb Cortex 14:364-375.

Carpenter MB (1991) Core text of neuroanatomy, 4th ed. Edition. Baltimore: Williams & Wilkins.

Cazals Y, Pelizzone M, Saudan O, Boex C (1994) Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants. J Acoust Soc Am 96:2048-2054.

Chandrasekaran B, Kraus N, Wong PC (2012) Human inferior colliculus activity relates to individual differences in spoken language learning. Journal of neurophysiology 107:1325-1336.

Chatterjee M, Oberzut C (2011) Detection and rate discrimination of amplitude modulation in electrical hearing. J Acoust Soc Am 130:1567-1580.

Chatterjee M, Peng SC (2008) Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. Hear Res 235:143-156.

Clarke CM, Garrett MF (2004) Rapid adaptation to foreign-accented English. J Acoust Soc Am 116:3647-3658.

Cleary M, Pisoni DB, Kirk KI (2000) Working Memory Spans as Predictors of Spoken Word Recognition and Receptive Vocabulary in Children with Cochlear Implants. The Volta review 102:259-280.

Cleary M, Pisoni DB, Geers AE (2001) Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. Ear and hearing 22:395-411.

Colcombe SJ, Kramer AF, Erickson KI, Scalf P (2005) The implications of cortical recruitment and brain morphology for individual differences in inhibitory function in aging humans. Psychology and aging 20:363-375.

Collins DL, Neelin P, Peters TM, Evans AC (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. Journal of computer assisted tomography 18:192-205.

Cools R, Barker RA, Sahakian BJ, Robbins TW (2001) Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. Cereb Cortex 11:1136-1143.

Cools R, Frank MJ, Gibbs SE, Miyakawa A, Jagust W, D'Esposito M (2009) Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. J Neurosci 29:1538-1543.

Cousineau D (2005) Confidence intervals in within-subject designs: A simpler solution to Loftus' and Masson's method. Tutorials in Quantitive Methods for Psychology 1:42-45.

Crosson B (1999) Subcortical mechanisms in language: lexical-semantic mechanisms and the thalamus. Brain and cognition 40:414-438.

D'Esposito M, Deouell LY, Gazzaley A (2003) Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. Nature reviews Neuroscience 4:863-872.

Dau T, Kollmeier B, Kohlrausch A (1997a) Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J Acoust Soc Am 102:2892-2905.

Dau T, Kollmeier B, Kohlrausch A (1997b) Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. J Acoust Soc Am 102:2906-2919.

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J Neurosci 23:3423-3431.

Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear Res 229:132-147.

Davis MH, Ford MA, Kherif F, Johnsrude IS (2011) Does semantic context benefit speech understanding through "top-down" processes? Evidence from time-resolved sparse fMRI. J Cogn Neurosci 23:3914-3932.

Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C (2005) Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. J Exp Psychol Gen 134:222-241.

Delhommeau K, Micheyl C, Jouvent R (2005) Generalization of frequency discrimination learning across frequencies and ears: implications for underlying neural mechanisms in humans. J Assoc Res Otolaryngol 6:171-179.

Delhommeau K, Micheyl C, Jouvent R, Collet L (2002) Transfer of learning across durations and ears in auditory frequency discrimination. Perception & psychophysics 64:426-436.

Deng W, Saxe MD, Gallina IS, Gage FH (2009) Adult-born hippocampal dentate granule cells undergoing maturation modulate learning and memory in the brain. J Neurosci 29:13532-13542.

Desjardins JL, Doherty KA (2013) Age-related changes in listening effort for various types of masker noises. Ear and hearing 34:261-272.

Dillon CM, Cleary M, Pisoni DB, Carter AK (2004) Imitation of nonwords by hearing-impaired children with cochlear implants: segmental analyses. Clin Linguist Phon 18:39-55.

Dosenbach NU, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE (2006) A core system for the implementation of task sets. Neuron 50:799-812.

Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, Fox MD, Snyder AZ, Vincent JL, Raichle ME, Schlaggar BL, Petersen SE (2007) Distinct brain networks for adaptive and stable task control in humans. Proc Natl Acad Sci U S A 104:11073-11078.

Draganski B, Gaser C, Busch V, Schuierer G, Bogdahn U, May A (2004) Neuroplasticity: changes in grey matter induced by training. Nature 427:311-312.

Driemeyer J, Boyke J, Gaser C, Buchel C, May A (2008) Changes in gray matter induced by learning--revisited. PLoS One 3:e2669.

Drullman R, Festen JM, Plomp R (1994a) Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95:1053-1064.

Drullman R, Festen JM, Plomp R (1994b) Effect of reducing slow temporal modulations on speech reception. J Acoust Soc Am 95:2670-2680.

Dudley H (1939) The Vocoder. Bell Lab Rec 18:122-126.

Dupoux E, Green K (1997) Perceptual adjustment to highly compressed speech: effects of talker and rate changes. J Exp Psychol Hum Percept Perform 23:914-927.

Eckert MA, Cute SL, Vaden KI, Jr., Kuchinsky SE, Dubno JR (2012) Auditory cortex signs of age-related hearing loss. J Assoc Res Otolaryngol 13:703-713.

Eckert MA, Walczak A, Ahlstrom J, Denslow S, Horwitz A, Dubno JR (2008) Age-related effects on word recognition: reliance on cognitive control systems with structural declines in speech-responsive cortex. J Assoc Res Otolaryngol 9:252-259.

Eckert MA, Menon V, Walczak A, Ahlstrom J, Denslow S, Horwitz A, Dubno JR (2009) At the heart of the ventral attention system: the right anterior insula. Human brain mapping 30:2530-2541.

Eisen MD (2003) Djourno, Eyries, and the first implanted electrical neural stimulator to restore hearing. Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology 24:500-506.

Eisner F, McGettigan C, Faulkner A, Rosen S, Scott SK (2010) Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. J Neurosci 30:7179-7186.

Erb J, Obleser J (2013) Upregulation of cognitive control networks in older adults' speech comprehension. Front Syst Neurosci 7:116.

Erb J, Henry MJ, Eisner F, Obleser J (2012) Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. Neuropsychologia 50:2154-2164.

Erb J, Henry MJ, Eisner F, Obleser J (2013) The brain dynamics of rapid perceptual adaptation to adverse listening conditions. J Neurosci 33:10688-10697.

Fagan MK, Pisoni DB, Horn DL, Dillon CM (2007) Neuropsychological correlates of vocabulary, reading, and working memory in deaf children with cochlear implants. Journal of deaf studies and deaf education 12:461-471.

Fallon JB, Irvine DR, Shepherd RK (2008) Cochlear implants and brain plasticity. Hear Res 238:110-117.

Fechner GT (1860) Elemente der Psychophysik. Leipzig: Breitkopf und Härtel.

Fedorenko E, Duncan J, Kanwisher N (2012) Language-selective and domain-general regions lie side by side within Broca's area. Curr Biol 22:2059-2062.

Ferstl EC, von Cramon DY (2002) What does the frontomedian cortex contribute to language processing: coherence or theory of mind? Neuroimage 17:1599-1612.

Fiebach CJ, Friederici AD, Muller K, von Cramon DY, Hernandez AE (2003) Distinct brain representations for early and late learned words. Neuroimage 19:1627-1637.

Fiebach CJ, Schlesewsky M, Lohmann G, von Cramon DY, Friederici AD (2005) Revisiting the role of Broca's area in sentence processing: syntactic integration versus syntactic working memory. Human brain mapping 24:79-91.

Fitzgerald MB, Wright BA (2011) Perceptual learning and generalization resulting from training on an auditory amplitude-modulation detection task. J Acoust Soc Am 129:898-906.

Floel A, de Vries MH, Scholz J, Breitenstein C, Johansen-Berg H (2009) White matter integrity in the vicinity of Broca's area predicts grammar learning success. Neuroimage 47:1974-1981.

Foster NE, Zatorre RJ (2010) Cortical structure predicts success in performing musical transformation judgments. Neuroimage 53:26-36.

Foundas AL, Mock JR, Cindass R, Jr., Corey DM (2013) Atypical caudate anatomy in children who stutter. Perceptual and motor skills 116:528-543.

Foundas AL, Bollich AM, Corey DM, Hurley M, Heilman KM (2001) Anomalous anatomy of speech-language areas in adults with persistent developmental stuttering. Neurology 57:207-215.

Foundas AL, Bollich AM, Feldman J, Corey DM, Hurley M, Lemen LC, Heilman KM (2004) Aberrant auditory processing and atypical planum temporale in developmental stuttering. Neurology 63:1640-1646.

Freyman RL, Nerbonne GP (1996) Consonant confusions in amplitude-expanded speech. J Speech Hear Res 39:1124-1137.

Friederici AD, Ruschemeyer SA, Hahne A, Fiebach CJ (2003) The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. Cereb Cortex 13:170-177.

Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frackowiak RSJ (1994) Statistical parametric maps in functional imaging: A general linear approach. Human brain mapping 2:189–210.

Friston KJ, Holmes AP, Price CJ, Buchel C, Worsley KJ (1999) Multisubject fMRI studies and conjunction analyses. Neuroimage 10:385-396.

Fu QJ (2002) Temporal processing and speech recognition in cochlear implant users. Neuroreport 13:1635-1639.

Fu QJ, Shannon RV (1998) Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners. J Acoust Soc Am 104:2570-2577.

Fu QJ, Shannon RV (2000) Effects of dynamic range and amplitude mapping on phoneme recognition in Nucleus-22 cochlear implant users. Ear and hearing 21:227-235.

Fullgrabe C, Meyer B, Lorenzi C (2003) Effect of cochlear damage on the detection of complex temporal envelopes. Hear Res 178:35-43.

Galaburda AM, Menard MT, Rosen GD (1994) Evidence for aberrant auditory anatomy in developmental dyslexia. Proc Natl Acad Sci U S A 91:8010-8013.

Garrett DD, Kovacevic N, McIntosh AR, Grady CL (2010) Blood oxygen level-dependent signal variability is more than just noise. J Neurosci 30:4914-4921.

Garrett DD, Kovacevic N, McIntosh AR, Grady CL (2011) The importance of being variable. J Neurosci 31:4496-4503.

Garvey WD (1953) The intelligibility of speeded speech. Journal of experimental psychology 45:102-108.

Gaser C, Schlaug G (2003) Brain structures differ between musicians and non-musicians. J Neurosci 23:9240-9245.

Gathercole SE, Willis CS, Baddeley AD, Emslie H (1994) The Children's Test of Nonword Repetition: a test of phonological working memory. Memory 2:103-127.

Gibson EJ (1969) Principles of perceptual learning and development. New York: Appleton-Century-Crofts.

Giraud AL, Truy E (2002) The contribution of visual areas to speech comprehension: a PET study in cochlear implants patients and normal-hearing subjects. Neuropsychologia 40:1562-1569.

Giraud AL, Lee HJ (2007) Predicting cochlear implant outcome from brain organisation in the deaf. Restor Neurol Neurosci 25:381-390.

Giraud AL, Price CJ, Graham JM, Frackowiak RS (2001a) Functional plasticity of language-related brain areas after cochlear implantation. Brain 124:1307-1316.

Giraud AL, Price CJ, Graham JM, Truy E, Frackowiak RS (2001b) Cross-modal plasticity underpins language recovery after cochlear implantation. Neuron 30:657-663.

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A (2000) Representation of the temporal envelope of sounds in the human brain. Journal of neurophysiology 84:1588-1598.

Giraud AL, Kell C, Thierfelder C, Sterzer P, Russ MO, Preibisch C, Kleinschmidt A (2004) Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. Cereb Cortex 14:247-255.

Gnansia D, Lazard DS, Leger AC, Fugain C, Lancelin D, Meyer B, Lorenzi C (2014) Role of slow temporal modulations in speech identification for cochlear implant users. International journal of audiology 53:48-54.

Goldstone RL (1998) Perceptual learning. Annual review of psychology 49:585-612.

Golestani N, Zatorre RJ (2004) Learning new sounds of speech: reallocation of neural substrates. Neuroimage 21:494-506.

Golestani N, Pallier C (2007) Anatomical correlates of foreign speech sound production. Cereb Cortex 17:929-934.

Golestani N, Paus T, Zatorre RJ (2002) Anatomical correlates of learning novel speech sounds. Neuron 35:997-1010.

Golestani N, Price CJ, Scott SK (2011) Born with an ear for dialects? Structural plasticity in the expert phonetician brain. J Neurosci 31:4213-4220.

Golomb JD, Peelle JE, Wingfield A (2007) Effects of stimulus variability and adult aging on adaptation to time-compressed speech. J Acoust Soc Am 121:1701-1708.

Gordon-Salant S, Yeni-Komshian GH, Fitzgibbons PJ, Schurman J (2010) Short-term adaptation to accented English by younger and older adults. J Acoust Soc Am 128:EL200-204.

Gould E, Reeves AJ, Graziano MS, Gross CG (1999) Neurogenesis in the neocortex of adult primates. Science 286:548-552.

Graybiel AM (2005) The basal ganglia: learning new tricks and loving it. Current opinion in neurobiology 15:638-644.

Green DM (1993) A maximum-likelihood method for estimating thresholds in a yes-no task. J Acoust Soc Am 93:2096-2105.

Greenwood DD (1990) A cochlear frequency-position function for several species--29 years later. J Acoust Soc Am 87:2592-2605.

Grossman M, Carvell S, Gollomp S, Stern MB, Vernon G, Hurtig HI (1991) Sentence comprehension and praxis deficits in Parkinson's disease. Neurology 41:1620-1626.

Grossman M, Carvell S, Gollomp S, Stern MB, Reivich M, Morrison D, Alavi A, Hurtig HI (1993) Cognitive and physiological substrates of impaired sentence processing in Parkinson's disease. J Cogn Neurosci 5:480-498.

Hackett TA, Smiley JF, Ulbert I, Karmos G, Lakatos P, de la Mothe LA, Schroeder CE (2007) Sources of somatosensory input to the caudal belt areas of auditory cortex. Perception 36:1419-1430.

Hagoort P (2005) On Broca, brain, and binding: a new framework. Trends in cognitive sciences 9:416-423.

Hahlbrock KH (1953) [Speech audiometry and new word-tests]. Archiv fur Ohren-, Nasen- und Kehlkopfheilkunde, vereinigt mit Zeitschrift fur Hals-, Nasen- und Ohrenheilkunde 162:394-431.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) "Sparse" temporal sampling in auditory fMRI. Human brain mapping 7:213-223.

Halliday LF, Moore DR, Taylor JL, Amitay S (2011) Dimension-specific attention directs learning and listening on auditory training tasks. Attention, perception & psychophysics 73:1329-1335.

Harris KC, Dubno JR, Keren NI, Ahlstrom JB, Eckert MA (2009) Speech recognition in younger and older adults: a dependency on low-level auditory cortex. J Neurosci 29:6078-6087.

Harris MS, Kronenberger WG, Gao S, Hoen HM, Miyamoto RT, Pisoni DB (2013) Verbal short-term memory development and spoken language outcomes in deaf children with cochlear implants. Ear and hearing 34:179-192.

Hawkey DJ, Amitay S, Moore DR (2004) Early and rapid perceptual learning. Nature neuroscience 7:1055-1056.

Hayasaka S, Phan KL, Liberzon I, Worsley KJ, Nichols TE (2004) Nonstationary cluster-size inference with random field and permutation methods. Neuroimage 22:676-687.

Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. Nature reviews Neuroscience 7:523-534.

Helson H (1948) Adaptation-level as a basis for a quantitative theory of frames of reference. Psychological review 55:297-313.

Henry MJ, Herrmann B, Obleser J (2013) Selective Attention to Temporal Features on Nested Time Scales. Cereb Cortex.

Herrmann B, Henry MJ, Obleser J (2013a) Frequency-specific adaptation in human auditory cortex depends on the spectral variance in the acoustic stimulation. Journal of neurophysiology 109:2086-2096.

Herrmann B, Schlichting N, Obleser J (2014) Dynamic range adaptation to spectral stimulus statistics in human auditory cortex. J Neurosci 34:327-331.

Herrmann B, Henry MJ, Grigutsch M, Scharinger M, Obleser J (2013b) Neural oscillatory dynamics and hemodynamic responses reveal distinct mechanisms underlying perception of time. In: Annual meeting of the society for Neuroscience. San Diego, CA, USA.

Hervais-Adelman A, Davis MH, Johnsrude IS, Carlyon RP (2008) Perceptual learning of noise vocoded words: effects of feedback and lexicality. J Exp Psychol Hum Percept Perform 34:460-474.

Hervais-Adelman A, Carlyon RP, Johnsrude I, Davis MH (2012) Brain regions recruited for the effortful comprehension of noise-vocoded words. Language and Cognitive Processes 27:1145-1166.

Hervais-Adelman AG, Davis MH, Johnsrude IS, Taylor KJ, Carlyon RP (2011) Generalization of perceptual learning of vocoded speech. J Exp Psychol Hum Percept Perform 37:283-295.

Heuninckx S, Wenderoth N, Swinnen SP (2008) Systems neuroplasticity in the aging brain: recruiting additional neural resources for successful motor performance in elderly persons. J Neurosci 28:91-99.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nature reviews Neuroscience 8:393-402.

Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nature neuroscience 1:304-309.

Hopfinger JB, Buchel C, Holmes AP, Friston KJ (2000) A study of analysis parameters that influence the sensitivity of event-related fMRI analyses. Neuroimage 11:326-333.

Houtgast T, Steeneken HJM (1971) Evaluation of speech transmission channels by using artificial signals. Acustica 25:355-367.

Houtgast T, Steeneken HJ (1985) A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. J Acoust Soc Am 77:1069–1077.

Huyck JJ, Johnsrude IS (2012) Rapid perceptual learning of noise-vocoded speech requires attention. J Acoust Soc Am 131:EL236-242.

Hyde KL, Lerch JP, Zatorre RJ, Griffiths TD, Evans AC, Peretz I (2007) Cortical thickness in congenital amusia: when less is better than more. J Neurosci 27:13028-13032.

Idemaru K, Holt LL (2011) Word recognition reflects dimension-based statistical learning. J Exp Psychol Hum Percept Perform 37:1939-1956.

IEC60268-16 (2003) Sound Equipment System - Part 16: Objective rating of speech intelligibility by speech transmission index. . International Electronical Comission, Geneva.

Ikoma Y, Watabe H, Hayashi T, Miyake Y, Teramoto N, Minato K, Iida H (2010) Measurement of density and affinity for dopamine D(2) receptors by a single positron emission tomography scan with multiple injections of [(11)C]raclopride. J Cereb Blood Flow Metab 30:663-673.

Irvine DR, Martin RL, Klimkeit E, Smith R (2000) Specificity of perceptual learning in a frequency discrimination task. J Acoust Soc Am 108:2964-2968.

Jacquemot C, Scott SK (2006) What is the relationship between phonological short-term memory and speech processing? Trends in cognitive sciences 10:480-486.

Janse E, Adank P (2012) Predicting foreign-accent adaptation in older adults. Q J Exp Psychol (Hove) 65:1563-1585.

Jezzard P, Matthews PM, Smith SM (2001) Functional MRI: An Introduction to Methods. Oxford: Oxford University Press.

Joel D, Weiner I (1994) The organization of the basal ganglia-thalamocortical circuits: open interconnected rather than closed segregated. Neuroscience 63:363-379.

Johnson MD, Ojemann GA (2000) The role of the human thalamus in language and memory: evidence from electrophysiological studies. Brain and cognition 42:218-230.

Jorgensen S, Dau T (2011) Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. J Acoust Soc Am 130:1475-1487.

Kaasinen V, Vilkman H, Hietala J, Nagren K, Helenius H, Olsson H, Farde L, Rinne J (2000) Age-related dopamine D2/D3 receptor loss in extrastriatal regions of the human brain. Neurobiology of aging 21:683-688.

Kalikow DN, Stevens KN, Elliott LL (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J Acoust Soc Am 61:1337-1351.

Karni A, Sagi D (1991) Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proc Natl Acad Sci U S A 88:4966-4970.

Kemp JM, Powell TP (1970) The cortico-striate projection in the monkey. Brain 93:525-546.

Kim DS (2005) The cutting edge of fMRI and high-field fMRI. International review of neurobiology 66:147-166.

Kollmeier B, Brand T, Meyer B (2008) Perception of speech and sound. In: Springer handbook of speech processing (Benesty J, Sondhi MM, Huang Y, eds). New York: Springer.

Koning R, Wouters J (2012) The potential of onset enhancement for increased speech intelligibility in auditory prostheses. J Acoust Soc Am 132:2569-2581.

Kotz SA (2006) The role of the basal ganglia in auditory language processing: Evidence from ERP lesion studies and functional neuroimaging. In: The role of the basal ganglia in auditory language processing: Evidence from ERP lesion studies and functional neuroimaging. Leipzig: MPI Series.

Kotz SA, Schwartze M, Schmidt-Kassow M (2009) Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. Cortex 45:982-990.

Kotz SA, Cappa SF, von Cramon DY, Friederici AD (2002) Modulation of the lexical-semantic network by auditory semantic priming: an event-related functional MRI study. Neuroimage 17:1761-1772.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. Journal of neurophysiology 60:1799-1822.

Laurent PA (2008) The emergence of saliency and novelty responses from Reinforcement Learning principles. Neural networks : the official journal of the International Neural Network Society 21:1493-1499.

Lazard DS, Giraud AL, Truy E, Lee HJ (2011) Evolution of non-speech sound memory in postlingual deafness: implications for cochlear implant rehabilitation. Neuropsychologia 49:2475-2482.

Lazard DS, Lee HJ, Gaebler M, Kell CA, Truy E, Giraud AL (2010) Phonological processing in post-lingual deafness and cochlear implant outcome. Neuroimage 49:3443-3451.

Lee DS, Lee JS, Oh SH, Kim SK, Kim JW, Chung JK, Lee MC, Kim CS (2001) Cross-modal plasticity and cochlear implants. Nature 409:149-150.

Lee HJ, Giraud AL, Kang E, Oh SH, Kang H, Kim CS, Lee DS (2007) Cortical activity at rest predicts cochlear implantation outcome. Cereb Cortex 17:909-917.

Leek MR (2001) Adaptive procedures in psychophysical research. Perception & psychophysics 63:1279-1292.

Levitt H (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am 49:Suppl 2:467+.

Li SC, Lindenberger U, Sikstrom S (2001) Aging cognition: from neuromodulation to representation. Trends in cognitive sciences 5:479-486.

Li SC, Lindenberger, U. (1999) Cross-level unification: a computational exploration of the link between deterioration of neurotransmitter systems dedifferentiation of cognitive abilities in old age. In: Cognitive neuroscience of memory (Nilsson LG, Markowitsch, H.J., ed), pp pp 103–146. Seattle: Hogrefe and Huber.

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. Psychological review 74:431-461.

Lieberman P, Friedman J, Feldman LS (1990) Syntax comprehension deficits in Parkinson's disease. The Journal of nervous and mental disease 178:360-365.

Lieberman P, Kako E, Friedman J, Tajchman G, Feldman LS, Jiminez EB (1992) Speech production, syntax comprehension, and cognitive deficits in Parkinson's disease. Brain and language 43:169-189.

Liegeois-Chauvel C, Lorenzi C, Trebuchon A, Regis J, Chauvel P (2004) Temporal envelope processing in the human left and right auditory cortices. Cereb Cortex 14:731-740.

Liegeois F, Connelly A, Salmond CH, Gadian DG, Vargha-Khadem F, Baldeweg T (2002) A direct test for lateralization of language activation using fMRI: comparison with invasive assessments in children with epilepsy. Neuroimage 17:1861-1867.

Lim SJ, Holt LL, Fiez JA (2013) Context-dependent modulation of striatal systems during incidental auditory category learning Poster presentation at the 43rd Annual Meeting of the Society for Neuroscience, San Diego, US.

Lindenberger U, Ghisletta P (2009) Cognitive and sensory declines in old age: gauging the evidence for a common cause. Psychology and aging 24:1-16.

Liu TT, Behzadi Y, Restom K, Uludag K, Lu K, Buracas GT, Dubowitz DJ, Buxton RB (2004) Caffeine alters the temporal dynamics of the visual BOLD response. Neuroimage 23:1402-1413.

Logothetis NK (2008) What we can do and what we cannot do with fMRI. Nature 453:869-878.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150-157.

Lorenzi C, Berthommier F, Apoux F, Bacri N (1999) Effects of envelope expansion on speech recognition. Hear Res 136:131-138.

Lovden M, Bodammer NC, Kuhn S, Kaufmann J, Schutze H, Tempelmann C, Heinze HJ, Duzel E, Schmiedek F, Lindenberger U (2010) Experience-dependent plasticity of white-matter microstructure extends into old age. Neuropsychologia 48:3878-3883.

Lukas JH (1980) Human auditory attention: the olivocochlear bundle may function as a peripheral filter. Psychophysiology 17:444-452.

Luo X, Fu QJ, Wei CG, Cao KL (2008) Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users. Ear and hearing 29:957-970.

Macmillan NA, Creelman CD (2005) Detection Theory: A User's Guide, 2nd Edition. Mahwah, NJ.: Lawrence Erlbaum Associates.

Maguire EA, Gadian DG, Johnsrude IS, Good CD, Ashburner J, Frackowiak RS, Frith CD (2000) Navigation-related structural change in the hippocampi of taxi drivers. Proc Natl Acad Sci U S A 97:4398-4403.

Marvit P, Florentine M, Buus S (2003) A comparison of psychophysical procedures for level-discrimination thresholds. J Acoust Soc Am 113:3348-3361.

Mattys SL, Brooks J, Cooke M (2009) Recognizing speech under a processing load: dissociating energetic from informational factors. Cognitive psychology 59:203-243.

McClelland JL, Elman JL (1986) The TRACE model of speech perception. Cognitive psychology 18:1-86.

McGettigan C (2007) Factors affecting the perception of noise-vocoded speech: stimulus properties and listener variability. Thesis, University College London.

McGettigan C, Evans S, Rosen S, Agnew ZK, Shah P, Scott SK (2012a) An application of univariate and multivariate approaches in FMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. J Cogn Neurosci 24:636-652.

McGettigan C, Faulkner A, Altarelli I, Obleser J, Baverstock H, Scott SK (2012b) Speech comprehension aided by multiple modalities: behavioural and neural interactions. Neuropsychologia 50:762-776.

Mehler J, Sebastian N, Altmann G, Dupoux E, Christophe A, Pallier C (1993) Understanding compressed sentences: the role of rhythm and meaning. Annals of the New York Academy of Sciences 682:272-282.

Menon V, Uddin LQ (2010) Saliency, switching, attention and control: a network model of insula function. Brain structure & function 214:655-667.

Middleton FA, Strick PL (2000) Basal ganglia output and cognition: evidence from anatomical, behavioural, and clinical studies. Brain and cognition 42:183-200.

Miller P, Wingfield A (2010) Distinct effects of perceptual quality on auditory word recognition, memory formation and recall in a neural model of sequential memory. Front Syst Neurosci 4:14.

Mink JW (1996) The basal ganglia: focused selection and inhibition of competing motor programs. Progress in neurobiology 50:381-425.

Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16:1936-1947.

Moore BC, Glasberg BR (1993) Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech. J Acoust Soc Am 94:2050-2062.

Moore BC, Glasberg BR, Vickers DA (1995) Simulation of the effects of loudness recruitment on the intelligibility of speech in noise. British journal of audiology 29:131-143.

Moore DR, Shannon RV (2009) Beyond cochlear implants: awakening the deafened brain. Nature neuroscience 12:686-691.

Morel A (2007) Stereotactic atlas of the human thalamus and basal ganglia. New York: Informa Healthcare.

Moro A, Tettamanti M, Perani D, Donati C, Cappa SF, Fazio F (2001) Syntax and the brain: disentangling grammar by selective anomalies. Neuroimage 13:110-118.

Mottier G (1951) [Examination of the speech of children with reading difficulties]. Folia phoniatrica 3:170-177.

Naatanen R, Sams M, Alho K, Paavilainen P, Reinikainen K, Sokolov EN (1988) Frequency and location specificity of the human vertex N1 wave. Electroencephalography and clinical neurophysiology 69:523-531.

Nath AR, Beauchamp MS (2011) Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. J Neurosci 31:1704-1714.

Nejime Y, Moore BC (1997) Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise. J Acoust Soc Am 102:603-615.

Noordhoek IM, Drullman R (1997) Effect of reducing temporal intensity modulations on sentence intelligibility. J Acoust Soc Am 101:498-502.

Norris D, McQueen JM, Cutler A (2000) Merging information in speech recognition: feedback is never necessary. The Behavioural and brain sciences 23:299-325; discussion 325-270.

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA, 3rd, Brugge JF (2009) Temporal envelope of time-compressed speech represented in the human auditory cortex. J Neurosci 29:15564-15574.

Nowakowski RS, Hayes NL (2000) New neurons: extraordinary evidence or extraordinary conclusion? Science 288:771.

O'Connell MN, Falchier A, McGinnis T, Schroeder CE, Lakatos P (2011) Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. Neuron 69:805-817.

Obleser J, Kotz SA (2010) Expectancy constraints in degraded speech modulate the language comprehension network. Cereb Cortex 20:633-640.

Obleser J, Weisz N (2012) Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cereb Cortex 22:2466-2477.

Obleser J, Eisner F, Kotz SA (2008) Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J Neurosci 28:8116-8123.

Obleser J, Wise RJ, Alex Dresner M, Scott SK (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. J Neurosci 27:2283-2289.

Obleser J, Wostmann M, Hellbernd N, Wilsch A, Maess B (2012) Adverse listening conditions and memory load drive a common alpha oscillatory network. J Neurosci 32:12376-12383.

Ogawa S, Lee TM, Nayak AS, Glynn P (1990a) Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 14:68-78.

Ogawa S, Lee TM, Kay AR, Tank DW (1990b) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci U S A 87:9868-9872.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97-113.

Overath T, Zhang Y, Sanes DH, Poeppel D (2012) Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: fMRI evidence. Journal of neurophysiology 107:2042-2056.

Pallier C, Sebastian-Galles N, Dupoux E, Christophe A, Mehler J (1998) Perceptual adjustment to time-compressed speech: a cross-linguistic study. Memory & cognition 26:844-851.

Parent A (1986) Comparative Neurobiology of the Basal Ganglia. New York: Wiley.

Parent A, Hazrati LN (1995) Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. Brain research Brain research reviews 20:91-127.

Park DC, Polk TA, Park R, Minear M, Savage A, Smith MR (2004) Aging reduces neural specialization in ventral visual cortex. Proc Natl Acad Sci U S A 101:13091-13095.

Pasupathy A, Miller EK (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature 433:873-876.

Peelle JE (2012) The hemispheric lateralization of speech processing depends on what "speech" is: a hierarchical perspective. Frontiers in human neuroscience 6:309.

Peelle JE, Wingfield A (2005) Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. J Exp Psychol Hum Percept Perform 31:1315-1330.

Peelle JE, Cusack R, Henson RN (2012) Adjusting for global effects in voxel-based morphometry: Gray matter decline in normal aging. Neuroimage 60:1503-1516.

Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb Cortex 23:1378-1387.

Peelle JE, Olafsen T, Davis MH, Wingfield A (2010a) Perceptual learning in speech comprehension governed by power law dynamics. Poster presentation at the 40th Annual Meeting of the Society for Neuroscience, San Diego, US November 13-17, 2010.

Peelle JE, Troiani V, Wingfield A, Grossman M (2010b) Neural processing during older adults' comprehension of spoken sentences: age differences in resource allocation and connectivity. Cereb Cortex 20:773-782.

Peelle JE, Troiani V, Grossman M, Wingfield A (2011) Hearing loss in older adults affects neural systems supporting speech comprehension. J Neurosci 31:12638-12643.

Peelle JE, McMillan C, Moore P, Grossman M, Wingfield A (2004) Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: evidence from fMRI. Brain and language 91:315-325.

Pichora-Fuller MK (2003) Cognitive aging and auditory information processing. International journal of audiology 42 Suppl 2:2S26-32.

Pichora-Fuller MK, Souza PE (2003) Effects of aging on auditory processing of speech. International journal of audiology 42 Suppl 2:2S11-16.

Pichora-Fuller MK, Schneider BA, Daneman M (1995) How young and old adults listen to and remember speech in noise. J Acoust Soc Am 97:593-608.

Piquado T, Cousins KA, Wingfield A, Miller P (2010) Effects of degraded sensory input on memory for speech: behavioural data and a test of biologically constrained computational models. Brain Res 1365:48-65.

Pirone A, Kurt S, Zuccotti A, Ruttiger L, Pilz P, Brown DH, Franz C, Schweizer M, Rust MB, Rubsamen R, Friauf E, Knipper M, Engel J (2014) alpha2delta3 Is Essential for Normal Structure and Function of Auditory Nerve Synapses and Is a Novel Candidate for Auditory Processing Disorders. J Neurosci 34:434-445.

Pisoni DB, Geers AE (2000) Working memory in deaf children with cochlear implants: correlations between digit span and measures of spoken language processing. The Annals of otology, rhinology & laryngology Supplement 185:92-93.

Pisoni DB, Cleary M (2003) Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. Ear and hearing 24:106S-120S.

Plomp R (1988) The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. J Acoust Soc Am 83:2322-2327.

Price CJ (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 62:816-847.

Priestley MB (1981) Spectral Analysis and Time Series. London: Academic Press.

Rabbitt PM (1968) Channel-capacity, intelligibility and immediate memory. The Quarterly journal of experimental psychology 20:241-248.

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nature neuroscience 12:718-724.

Redgrave P, Gurney K (2006) The short-latency dopamine signal: a role in discovering novel actions? Nature reviews Neuroscience 7:967-975.

Rees A, Palmer AR (1989) Neuronal responses to amplitude-modulated and pure-tone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise. J Acoust Soc Am 85:1978-1994.

Reuter-Lorenz P, Cappell KA (2008) Neurocognitive Aging and the Compensation Hypothesis. Current Directions in Psychological Science 17:177-182.

Romani C, Martin R (1999) A deficit in the short-term retention of lexical-semantic information: forgetting words but remembering a story. J Exp Psychol Gen 128:56-77.

Romanski LM, Giguere M, Bates JF, Goldman-Rakic PS (1997) Topographic organization of medial pulvinar connections with the prefrontal cortex in the rhesus monkey. The Journal of comparative neurology 379:313-332.

Rorden C, Brett M (2000) Stereotaxic display of brain lesions. Behavioural neurology 12:191-200.

Rosen S, Faulkner A, Wilkinson L (1999) Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. J Acoust Soc Am 106:3629-3636.

Rosen S, Wise RJ, Chadha S, Conway EJ, Scott SK (2011) Hemispheric asymmetries in speech perception: sense, nonsense and modulations. PLoS One 6:e24672.

Sadaghiani S, D'Esposito M (2012) Dissociating brain networks for alertness and selective attention. In: Poster presentation at the 42nd SfN annual meeting, New Orleans, US. New Orleans, USA.

Salimpoor VN, Benovoy M, Larcher K, Dagher A, Zatorre RJ (2011) Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. Nature neuroscience 14:257-262.

Salmond CH, Ashburner J, Vargha-Khadem F, Gadian DG, Friston KJ (2000) Detecting bilateral abnormalities with voxel-based morphometry. Human brain mapping 11:223-232.

Samanez-Larkin GR, D'Esposito M (2008) Group comparisons: imaging the aging brain. Social cognitive and affective neuroscience 3:290-297.

Samuel AG, Kraljic T (2009) Perceptual learning for speech. Attention, perception & psychophysics 71:1207-1218.

Sandmann P, Dillier N, Eichele T, Meyer M, Kegel A, Pascual-Marqui RD, Marcar VL, Jancke L, Debener S (2012) Visual activation of auditory cortex reflects maladaptive plasticity in cochlear implant users. Brain 135:555-568.

Sarant JZ, Blamey PJ, Dowell RC, Clark GM, Gibson WP (2001) Variation in speech perception scores among children with cochlear implants. Ear and hearing 22:18-28.

Scharinger M, Henry MJ, Erb J, Meyer L, Obleser J (2013) Thalamic and parietal brain morphology predicts auditory category learning. Neuropsychologia 53C:75-83.

Schlaug G, Jancke L, Huang Y, Steinmetz H (1995) In vivo evidence of structural brain asymmetry in musicians. Science 267:699-701.

Schlegel AA, Rudelson JJ, Tse PU (2012) White matter structure changes as adults learn a second language. J Cogn Neurosci 24:1664-1670.

Scholz J, Klein MC, Behrens TE, Johansen-Berg H (2009) Training induces changes in white-matter architecture. Nature neuroscience 12:1370-1371.

Schonwiesner M, Zatorre RJ (2009) Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proc Natl Acad Sci U S A 106:14611-14616.

Schreiner CE, Urbas JV (1986) Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF). Hear Res 21:227-241.

Schultz W (2002) Getting formal with dopamine and reward. Neuron 36:241-263.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593-1599.

Schwarz G (1978) Estimating the Dimension of a Model. Annals of statistics 6:461-464.

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123 Pt 12:2400-2406.

Sebastian-Galles N, Dupoux E, Costa A, Mehler J (2000) Adaptation to time-compressed speech: phonological determinants. Perception & psychophysics 62:834-842.

Sehm B, Schnitzler T, Obleser J, Groba A, Ragert P, Villringer A, Obrig H (2013) Facilitation of inferior frontal cortex by transcranial direct current stimulation induces perceptual learning of severely degraded speech. J Neurosci 33:15868-15878.

Shannon RV (1992) Temporal modulation transfer functions in patients with cochlear implants. J Acoust Soc Am 91:2156-2164.

Shannon RV, Galvin JJ, 3rd, Baskent D (2002) Holes in hearing. J Assoc Res Otolaryngol 3:185-199.

Shannon RV, Fu QJ, Galvin J, Friesen L (2004) Speech perception with cochlear implants. In: Cochlear implants: auditory prostheses and electric hearing (Springer handbook of auditory research) (Zeng FG, Popper AN, Fay RR, eds). New York: Springer.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303-304.

Shapiro SSW, M.B. (1965) An analysis of variance test for normality (complete samples). Biometrika 52:591–611.

Sharp DJ, Scott SK, Mehta MA, Wise RJ (2006) The neural correlates of declining performance with age: evidence for age-related changes in cognitive control. Cereb Cortex 16:1739-1749.

Shohamy D, Myers CE, Kalanithi J, Gluck MA (2008) Basal ganglia and dopamine contributions to probabilistic category learning. Neuroscience and biobehavioural reviews 32:219-236.

Shohamy D, Myers CE, Geghman KD, Sage J, Gluck MA (2006) L-dopa impairs learning, but spares generalization, in Parkinson's disease. Neuropsychologia 44:774-784.

Slotnick SD, Moo LR, Segal JB, Hart J, Jr. (2003) Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Brain research Cognitive brain research 17:75-82.

Smalt CJ, Gonzalez-Castillo J, Talavage TM, Pisoni DB, Svirsky MA (2013) Neural correlates of adaptation in freely-moving normal hearing subjects under cochlear implant acoustic simulations. Neuroimage 82:500-509.

Song JH, Skoe E, Banai K, Kraus N (2011) Training to Improve Hearing Speech in Noise: Biological Mechanisms. Cereb Cortex.

Stacey PC, Summerfield AQ (2007a) Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. J Acoust Soc Am 121:2923-2935.

Stacey PC, Summerfield AQ (2007b) Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. J Acoust Soc Am 121:2923-2935.

Steeneken HJ, Houtgast T (1980) A physical method for measuring speech-transmission quality. J Acoust Soc Am 67:318-326.

Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013) A causal link between prediction errors, dopamine neurons and learning. Nature neuroscience 16:966-973.

Sterzer P, Kleinschmidt A (2010) Anterior insula activations in perceptual paradigms: often observed but barely understood. Brain structure & function 214:611-622.

Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW (2010) Neural processing of asynchronous audiovisual speech perception. Neuroimage 49:3308-3318.

Strait DL, Parbery-Clark A, Hittner E, Kraus N (2012) Musical training during early childhood enhances the neural encoding of speech in noise. Brain and language 123:191-201.

Studebaker GA (1985) A "rationalized" arcsine transform. J Speech Hear Res 28:455-462.

Summerfield Q (1981) Articulatory rate and perceptual constancy in phonetic perception. J Exp Psychol Hum Percept Perform 7:1074-1095.

Taubert M, Draganski B, Anwander A, Muller K, Horstmann A, Villringer A, Ragert P (2010) Dynamic properties of human brain structure: learning-related changes in cortical areas and associated fiber connections. J Neurosci 30:11670-11677.

Taylor MM, Forbes SM, Creelman CD (1983) PEST reduces bias in forced choice psychophysics. J Acoust Soc Am 74:1367-1374.

Tomoda A, Kinoshita S, Korenaga Y, Mabe H (2012) Pseudohypacusis in childhood and adolescence is associated with increased gray matter volume in the medial frontal gyrus and superior temporal gyrus. Cortex 48:492-503.

Tourdias T, Saranathan M, Levesque IR, Su J, Rutt BK (2014) Visualization of intra-thalamic nuclei with optimized white-matter-nulled MPRAGE at 7T. Neuroimage 84:534-545.

Tronel S, Fabre A, Charrier V, Oliet SH, Gage FH, Abrous DN (2010) Spatial learning sculpts the dendritic arbor of adult-born hippocampal neurons. Proc Natl Acad Sci U S A 107:7963-7968.

Turner R, Le Bihan D, Moonen CT, Despres D, Frank J (1991) Echo-planar time course MRI of cat brain oxygenation changes. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 22:159-166.

Vaden KI, Jr., Kuchinsky SE, Cute SL, Ahlstrom JB, Dubno JR, Eckert MA (2013) The cingulo-opercular network provides word-recognition benefit. J Neurosci 33:18979-18986.

Van Buren JM, Ojemann GA (1966) The fronto-striatal arrest response in man. Electroencephalography and clinical neurophysiology 21:114-130.

van Buuren RA, Festen JM, Houtgast T (1999) Compression and expansion of the temporal envelope: evaluation of speech intelligibility and sound quality. J Acoust Soc Am 105:2903-2913.

Van Tasell DJ, Soli SD, Kirby VM, Widin GP (1987) Speech waveform envelope cues for consonant recognition. J Acoust Soc Am 82:1152-1161.

Viemeister NF (1979) Temporal modulation transfer functions based upon modulation thresholds. J Acoust Soc Am 66:1364-1380.

von Kriegstein K, Patterson RD, Griffiths TD (2008) Task-dependent modulation of medial geniculate body is behaviourally relevant for speech recognition. Curr Biol 18:1855-1859.

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL (2003) Modulation of neural responses to speech by directing attention to voices or verbal content. Brain research Cognitive brain research 17:48-55.

Voor JB, Miller, J.M. (1965) The effect of practice on the comprehension of worded speech. Speech Monographs 32:452-455.

Wakefield GH, Viemeister NF (1990) Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise. J Acoust Soc Am 88:1367-1373.

Watanabe T, Nanez JE, Sasaki Y (2001) Perceptual learning without perception. Nature 413:844-848.

Wechsler D (1987) Wechsler Memory Scale - Revised Edition. Manual. New York: The Psychological Corporation.

Weinberger NM, Bakin JS (1998) Learning-induced physiological memory in adult primary auditory cortex: receptive fields plasticity, model, and mechanisms. Audiology & neuro-otology 3:145-167.

Weis T, Brechmann A, Puschmann S, Thiel CM (2013) Feedback that confirms reward expectation triggers auditory cortex activity. Journal of neurophysiology 110:1860-1868.

Welte V (1981) [Der Mottier-Test, ein Prüfungsmittel für die Lautdifferenzierungsfähigkeit und die auditive Merkfähigkeit]. Sprache - Stimme - Gehör 5:121-125.

Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS (2012) Effortful listening: the processing of degraded speech depends critically on attention. J Neurosci 32:14010-14021.

Wilsch A, Henry MJ, Herrmann B, Maess B, Obleser J (2014) Alpha Oscillatory Dynamics Index Temporal Expectation Benefits in Working Memory. Cereb Cortex.

Winer JA (2006) Decoding the auditory corticofugal systems. Hear Res 212:1-8.

Winer JA, Diehl JJ, Larue DT (2001) Projections of auditory cortex to the medial geniculate body of the cat. The Journal of comparative neurology 430:27-55.

Wingfield A, Lombardi L, Sokol S (1984) Prosodic features and the intelligibility of accelerated speech: syntactic versus periodic segmentation. J Speech Hear Res 27:128-134.

Won JH, Drennan WR, Nie K, Jameyson EM, Rubinstein JT (2011) Acoustic temporal modulation detection and speech perception in cochlear implant listeners. J Acoust Soc Am 130:376-388.

Wright BA, Zhang Y (2009) A review of the generalization of auditory learning. Philosophical transactions of the Royal Society of London Series B, Biological sciences 364:301-311.

Zatorre RJ, Fields RD, Johansen-Berg H (2012) Plasticity in gray and white: neuroimaging changes in brain structure during learning. Nature neuroscience 15:528-536.

Zekveld AA, Heslenfeld DJ, Festen JM, Schoonhoven R (2006) Top-down and bottom-up processes in speech comprehension. Neuroimage 32:1826-1836.

List of Figures

List of Tables

Abbreviations

ACC	anterior cingulate cortex
AG	angular gyrus
AM	amplitude modulation
ANOVA	analysis of variance
BOLD	blood-oxygen-level dependent
CI	cochlear implant
CSF	cerebro-spinal fluid
DARTEL	diffeomorphic anatomical registration using exponentiated lie algebra
DSB	Backward digit span
DSF	Forward digit span
EEG	electroencephalography
EPI	echo-planar imaging
FDR	false-discovery rate
FIR	finite impulse response
fMRI	functional magnetic resonance imaging
FMT	Freiburger monosyllabic test
FNT	Freiburger number test
FWE	family-wise error
FWHM	full-width-at-half-maximum
GLM	general linear model
GMV	grey matter volume
HG	Heschl’s gyrus
HL	hearing loss
HRF	hemodynamic response function
IFG	inferior frontal gyrus
L	left
MEG	magnetoencephalography
MFG	middle frontal gyrus
MGB	medial geniculate body
MNI	Montreal Neurological Institute

MPRAGE	magnetisation-prepared rapid gradient echo
MRI	magnetic resonance imaging
MTG	middle temporal gyrus
PTA	pure-tone average
R	right
RMS	root mean square
ROI	region of interest
SEM	standard error of the mean
SMA	supplementary motor area
SMG	supramarginal gyrus
SNR	signal-to-noise ratio
SPIN	speech in noise
STG	superior temporal gyrus
STS	superior temporal sulcus
TA	acquisition time
tDCS	transcranial direct current stimulation
TE	echo time
TMTF	Temporal modulation transfer function
TR	repetition time
VBM	voxel-based morphometry

Appendix: The German SPIN sentences

	Low-predictable sentences	**High-predictable sentences**
1	Wir freuen uns sehr über die Laken	Sie bezieht das Bett mit frischen Laken
2	Paul erzählte von dem neuen Netz	Er fing diesen Fisch in seinem Netz
3	Frau Meier spricht über die Küste	Das Boot segelte entlang der Küste
4	Ich möchte das mit der kleinen Blüte	Alle Blumen standen in voller Blüte
5	Peter ärgert sich über die Gläser	Die Biertrinker hoben ihre Gläser
6	Er kümmert sich jetzt nicht um den Pfeil	Ihn traf dort ein vergifteter Pfeil
7	Regelmäßig holt er sich eine Wunde	Am Knie bildet sich Schorf über der Wunde
8	Maria hielt das sehr lange für Weizen	In der Auswahl trennt sich die Spreu vom Weizen
9	Daniel dachte überhaupt nicht an Sport	Boxen ist ein gefährlicher Sport
10	Frau Maier sorgte sich um den Arm	Nehmen Sie das Baby auf den Arm
11	Johannes benötigte schnell den Block	Reiß doch ein paar Zettel aus dem Block
12	Wir betrachteten aufmerksam die Truhe	Der Schatz befindet sich in einer Truhe
13	Kathrin entdeckte vorhin ein Nest	Die Wespen schwärmten aus ihrem Nest
14	Eva und Anton mustern die Scheiben	Schneide den Schinken in dünne Scheiben
15	David hoffte vergeblich auf einen Kuss	Sie gab ihm einen innigen Kuss
16	Ich glaube dir die Sache mit dem Finger	Ich schnitt mir aus Versehen in den Finger
17	Max verschwindet eilig mit dem Wachs	Die Flamme der Kerze schmolz das Wachs
18	Er bittet ihn dauernd um das Schloss	Dieser Schlüssel passt nicht in das Schloss
19	Hans ärgerte sich über den Kratzer	Woher hat dein Auto diesen Kratzer
20	Er erkundigte sich nach einer Wiege	Das Baby schlief in seiner Wiege
21	Herr Koch rätselt über die Dornen	Ein Rosenstrauch hat spitze Dornen
22	Andreas weiß mehr über das Floß	Die schiffbrüchigen Segler bauten ein Floß
23	Der alte Mann registrierte das Tor	Der kleine Junge schießt den Ball ins Tor
24	Er überprüft sehr genau den Unfall	Es gab keine Verletzten bei dem Unfall
25	Unmöglich findet sie auf Anhieb das Deck	Der Matrose schrubbte sorgfältig das Deck
26	Der junge Mann bemerkte nicht die Maus	Sie spielten mit der Polizei Katz' und Maus
27	Frau Krause widmet sich mit Anna dem Spiel	Sie machte gute Miene zum bösen Spiel

28	Sie wussten schon lange von dem Fell	Der große Bär hat ein dickes Fell
29	Ich verkaufe Marie niemals das Sofa	Jens schlief auf dem ausgezogenen Sofa
30	Leider fehlt uns eine passende Dose	Die Plätzchen sind in der silbernen Dose
31	Thomas war gestern wieder nicht im Stall	Klaus geht zum Kühemelken in den Stall
32	Frau Kaiser reagierte nicht auf den Schock	Seit dem Zusammenstoß steht er unter Schock
33	Hanna untersucht penibel die Hüften	Lars stemmte die Hände in die Hüften
34	Anja und ich kümmern uns um den Saft	Zum Frühstück trank er frisch gepressten Saft
35	Er achtete nicht auf seine Faust	Er schlug mich mit der geballten Faust
36	Er fragte mich nach der neuen Mannschaft	Der Trainer qualifiziert seine Mannschaft
37	Julia zeigte Felix einen Schwan	Das Entlein schwamm mit dem weißen Schwan
38	Anton lügt bezüglich seiner Münze	Zur Entscheidung werfen wir eine Münze
39	Jeder braucht unbedingt einen Besen	Das Dienstmädchen fegt das Haus mit dem Besen
40	Sie fragten nach dem hinteren Schuppen	Als Holzlager baute er einen Schuppen
41	Julius fragt nach der neuen Straße	Er parkte den Wagen am Rand der Straße
42	Philipp ignoriert die Fragen der Leute	Sie hörte nicht auf das Geschwätz der Leute
43	Wir hörten schon von der Beute	Der Einbrecher entkam mit der Beute
44	Wieder befassten sie sich mit der Hose	Er machte sich aus Angst in die Hose
45	Zerstreut erwähnte sie nicht mal die Rippe	Gott schuf Eva aus Adams Rippe
46	Herr Wegener observiert die Uhr	Wir hörten das Ticken der alten Uhr
47	Sie waren überrascht vom Applaus	Nach dem Stück ernteten sie viel Applaus
48	Monika überprüfte den Film	Der Fotograf wechselte den Film
49	Stephan lernte alles über den Kiefer	Vom Kaugummikauen schmerzt mein Kiefer
50	Das ist für heute meine erste Suppe	Sie kochte ihm eine herzhafte Suppe
51	Wir unterhalten uns noch über den Preis	Mit dem Beitrag gewann sie den ersten Preis
52	Frau Lehmann ergatterte das letzte Brot	Der Asket lebt nur von Wasser und Brot
53	Wir erinnern uns alle an den Schrei	Vor Schreck entfuhr ihr ein gellender Schrei
54	Niemand besprach genau die Handlung	Er beging eine strafbare Handlung
55	Lena redet nicht gern über ihr Bein	Der Schwimmer bekam einen Krampf im Bein
56	Frau Keller zeigte mir meine Reihe	Unsre Plätze sind in der zweiten Reihe
57	Sie dachte gar nicht mehr an die Höhle	Der Löwe schlief in seiner Höhle
58	Alexandra wirft es in das Feuer	Das gebrannte Kind meidet Feuer
59	Der Junge holt draußen die Falle	Die kleine Maus tappte in die Falle
60	Frau Jung erinnerte an den Tee	Nach dem Aufgiessen zieht der schwarze Tee

61	Mario interessiert sich für das Recht	Der Richter gab dem Ankläger Recht
62	Ein Mann untersuchte gründlich den Ochsen	Vor den Pflug spannte man früher den Ochsen
63	Helena unterschätzte den Tribut	Die Katastrophe fordert ihren Tribut
64	Sie musste nach Hause wegen ihres Rocks	Sie kürzte den Saum ihres neuen Rocks
65	Morgen entfernen sie schließlich den Stamm	Ein alter Baum hat einen dicken Stamm
66	Kathrin versucht es mit den Karten	Laut Spielregeln mischt man nun die Karten
67	Tom freut sich über diese schöne Seide	Das Halstuch besteht aus glänzender Seide
68	Er unterhielt sich lange mit dem Gast	Herr Fischer ist ein gern gesehener Gast
69	Er sieht nicht gut aus ohne Gürtel	In Karate hat er den schwarzen Gürtel
70	Sie stellte sich neben die Ecke	Folgen Sie der Straße um die Ecke
71	Am liebsten fotografiert er die Taube	Als Friedensvogel gilt die weiße Taube
72	Täglich produziert sie ein neues Kleid	Die Braut trägt ein langes weißes Kleid
73	Er erspäht in der Ferne die Flut	Die Gezeiten sind Ebbe und Flut
74	Gewöhnlich nutzen wir die Räder	Ein Auto hat immer vier Räder
75	Die Leute begutachteten die Perlen	Sie fädelt auf die Kette teure Perlen
76	Die Gruppe orientiert sich in der Bucht	Der Hafen liegt in einer stillen Bucht
77	Ich zeige dir noch einmal das Kalb	Eine Kuh säugt nur ihr eigenes Kalb
78	Sven gab ihr letzte Woche den Leim	Zusammengeklebt wird das Holz mit Leim
79	Isabel erwähnte auch die Klinge	Sein Klappmesser hat eine scharfe Klinge
80	Nina hatte keine Ahnung von der Post	Briefmarken bekommst du bei der Post
81	Sein Blick schweifte kurz über das Meer	Der Mann ertrank im stürmischen Meer
82	Ich untersuche den grünen Lack	Vom rostigen Auto blätterte der Lack
83	Heute geht Katharina nicht zum Bach	Sie fischten in einem plätschernden Bach
84	Harald verfolgte den Anfang der Spuren	Der Täter hinterließ sichtbare Spuren
85	Brigitte vertraut nur ihrem Arzt	Fragen Sie Ihren behandelnden Arzt
86	Du verwechselst immer die Löffel	Für Suppen nehmen wir silberne Löffel
87	Mir wird immer schlecht von der Rinde	Wir ritzten ein Herz in die Rinde
88	Gewöhnlich verarbeitet sie kein Mehl	Nun mischt man das Ei mit gesiebtem Mehl
89	Am meisten mag er den roten Ball	Der Mittelfeldspieler köpfte den Ball
90	Den ganzen Tag schon will er keinen Spaß	Austricksen macht ihm diebischen Spaß
91	Zeige mir mal genauer die Bahn	Ich fahre diesesmal mit der Bahn
92	Morgen nehmen wir mal wieder Obst	Sie kaufte Gemüse und frisches Obst
93	Hofffentlich fragt Sascha nach dem Riegel	Der Verbrecher sitzt hinter Schloß und Riegel

94	Frau Schulze wusste von den Flaschen	Bier kauft man in Dosen oder Flaschen
95	Das Mädchen beachtet nicht die Flammen	Aus dem Haus züngelten lodernde Flammen
96	Peter berichtete von dem Regen	Nimm den Schirm mit bei Schnee und Regen
97	Martin erwähnte nicht einmal die Wanzen	In ihrem Bett krabbelten die Wanzen
98	Karin fühlte sich sehr wohl im Wald	Die Arbeiter rodeten den Wald
99	Sie beneidet mich um meine Sauna	Im Bad ist auch eine finnische Sauna
100	Offensichtlich fehlt heute Morgen das Schiff	Der Kapitän lenkte geschickt sein Schiff

Summary

Introduction

Speech is rarely heard under optimal listening conditions. Yet, humans have the capability to rapidly adapt to degraded or altered speech. This challenge is particularly relevant to hearing-impaired listeners, or more drastically, cochlear implant (CI) patients adapting to an extremely distorted auditory input delivered by their hearing device. The thesis presents an investigation of the short-term behavioural and neural processes that underlie successful adaptation to degraded speech.

We simulated CI transduced speech in normal-hearing listeners using noise-vocoding (Shannon et al., 1995), which degrades the spectral detail in an auditory signal, but leaves the temporal envelope cues intact. Normal-hearing listeners learn to understand vocoded speech quickly on the first exposure (Davis et al., 2005; Eisner et al., 2010). However, there is considerable variability across both normal-hearing and CI listeners in their ability to perceptually adapt to the degraded auditory input. Currently, it is unclear which factors drive this adaptation process.

Spectral degradation of the speech signal forces the listener to rely more on temporal envelope cues for comprehension. Hence, we hypothesised that higher sensitivity to slow temporal fluctuations in an auditory signal, as measured by amplitude modulation (AM) rate discrimination thresholds, might be predictive of perceptual adaptation to speech, in normal-hearing and CI listeners (Fu, 2002).

As temporal envelope cues appear to be necessary and almost sufficient for comprehension (Shannon et al, 1995), enhancing these cues has been suggested to aid speech recognition. While envelope expansion deteriorates speech recognition in quiet, some studies suggest that expansion in noise aids comprehension (Apoux et al., 2001). Deeper investigation of this interaction effect may help reveal the mechanisms of speech envelope processing in general.

Although rapid perceptual learning of vocoded speech has been confirmed in many behavioural studies (e.g., Davis et al., 2005), its neural bases are largely unknown. Some electrophysiological studies provide evidence for a subcortical locus of perceptual learning (Song et al., 2011), but functional magnetic imaging (fMRI) studies mainly found adaptation-related cortical changes. For example, feedback-based perceptual learning of vocoded speech was found to rely on the left inferior frontal gyrus (IFG; Eisner et al., 2010). However, as listeners rarely receive direct feedback in everyday situations, the neural dynamics of self-regulated adaptation deserve further research.

Such perceptual adaptation is poorly understood in older adults, although it bears particular relevance for older listeners who are frequently affected by hearing loss. Older adults have been hypothesised to engage in more "effortful" processing during speech comprehension. Eckert et al. (2008) proposed that such increased listening effort is neurally reflected an upregulation of cognitive-control-related frontal areas. For example, errors in a word recognition task elicit increased activity in the anterior cingulate cortex (ACC), but more so in older than younger adults (Harris et al., 2009). Hence, for solving auditory tasks, the reliance on cognitive control appears to increase with age, although it is unclear whether this reflects an age-related loss of specialisation of cortical brain regions or a compensatory mechanism (Cabeza et al., 2002).

In sum, further investigation of the perceptual changes occurring during adaptation to degraded speech is warranted. The present thesis asks whether individual differences in brain structural characteristics, cognitive and basic auditory abilities can predict the extent to which a listener adapts to degraded speech, both in normal-hearing and CI listeners. The neural dynamics underlying such adaptive mechanisms are investigated in young listeners, and compared to older adults with age-related hearing loss. Lastly, we ask whether a signal processing strategy which enhances temporal speech envelope cues can boost intelligibility in quiet and in noise.

Experiments and Results

Five experiments were conducted to tackle the problems outlined above. We applied psychoacoustical testing, structural and functional MRI to study behavioural and neural mechanisms that help a listener adapt to degraded speech.

In a first experiment we aimed to understand adaptation through listener variability. We hypothesised that individual differences in adaptation should be predictable by non-speech auditory, cognitive, or neuroanatomical factors. In a short-term adaptation paradigm, normal-hearing listeners heard and repeated back 100 four-band noise-vocoded sentences. Non-speech auditory skills were assessed using AM rate discrimination, where modulation frequencies were centred on the speech-relevant rate of 4 Hz. Working memory capacities were evaluated, and structural MRI scans were examined for anatomical predictors of perceptual adaptation using voxel-based morphometry. The results show that listeners with smaller AM rate discrimination thresholds learn faster to understand degraded speech.

This ability to adjust to degraded speech is furthermore reflected anatomically in an increased grey matter volume in an area of the left thalamus (pulvinar) that is strongly connected to the auditory and prefrontal cortices. Thus, pulvinar morphology and individual non-speech auditory skills can predict how rapidly a listener adapts to degraded speech.

A second experiment tested whether such factors can also explain the variability in cochlear implantation outcome. Shortly after implantation, we assessed CI recipients' acoustic AM rate discrimination thresholds centred at 4 Hz. We show that thresholds assessed already 10 weeks postoperatively correlate with performance in a standard speech recognition test on the same day. Most importantly, when duration of deafness is controlled for, AM rate discrimination thresholds can prospectively predict performance in a monosyllabic word test 6 months later. Due to the small sample size and the high variance in various demographic factors, continued testing is needed, however.

The goal of the third experiment was two-fold: we intended (1) to quantify the convergence of neural mechanisms that support coping with challenging listening conditions for speech and non-speech, and (2) to identify the neural systems that support short-term perceptual adaptation to degraded speech. In a sparse-sampling, cardiac-gated fMRI acquisition, normal-hearing participants again listened to, and repeated back 100 four-band vocoded sentences. Clear-speech trials were included as baseline. An additional fMRI experiment on AM rate discrimination investigated auditory processing of non-speech slow temporal modulations. The degraded speech task revealed an "executive" network (comprising the anterior insula and ACC, parts of which were also active in the non-speech discrimination task). Trial-by-trial fluctuations in successful comprehension of degraded speech drove the haemodynamic signal change in a perisylvian network considered as classic "language" areas. Finally, we provide first evidence that self-regulated perceptual adaptation to degraded speech co-occurs with a haemodynamic down-regulation of the caudate and ventral anterior thalamic nucleus. These data highlight differential up– and down-regulation in auditory–language and executive networks, respectively, with important subcortical contributions when successfully adapting to a challenging listening situation.

A fourth experiment intended to characterise how such neural speech processing changes with older age and age-related hearing loss. A group of older adults with varying degrees of sensorineural hearing loss participated in an fMRI experiment using an identical short-term adaptation paradigm. Behaviourally, older adults adapted at the same rate as young listeners, although their overall comprehension of degraded speech was lower. Neurally, both older and young adults relied on the left anterior insula for degraded more than clear speech perception. However, anterior insula engagement in older adults was dependent on hearing acuity, such that more hearing-impaired older listeners recruited the insular cortex for clear speech. Young but not older adults relied on the ACC for degraded compared to clear speech. This age group × degradation interaction was driven by a reduced dynamic range in older adults, who displayed elevated levels of ACC activity in both conditions, consistent with a persistent upregulation in cognitive control irrespective of task difficulty. For correct speech comprehension, older adults employed the middle frontal gyrus in addition to a core speech comprehension network on which young adults relied, suggestive of a compensatory mechanism. These results confirm that older adults increasingly recruit cognitive control networks, even under optimal listening conditions, at the expense of these systems' dynamic range.

Given the importance of temporal envelope cues in speech comprehension, we asked in a fifth experiment whether enhancing or diminishing the saliency of envelope cues can impact on speech intelligibility, in quiet and noise. Instead of simply six-band-vocoding words, we additionally modified the speech envelope, by either enhancing ("expanding") or diminishing ("compressing") the modulation depth of the signal while white noise was incrementally added. Importantly, the parametric variation of envelope modification with a fine-grained parametric variation of signal-to-noise ratio allowed us to quantify the extent to which listeners would benefit from envelope expansion as the noise level increased. The results show that envelope expansion in quiet deteriorates speech recognition and intelligibility ratings, just as compression does. On the contrary, when noise is present, listeners show a small but significant speech recognition benefit from expansion, with higher noise levels leading to relatively higher benefits. These results are direct evidence for the importance of envelope cues in speech recognition, especially in the presence of background noise.

Discussion

The current thesis set out to examine the neural systems that support comprehension of, and perceptual adaptation to, speech when listening conditions become difficult. In particular, we investigated how normal-hearing listeners, older adults with varying degrees of hearing loss, and CI patients cope with spectral degradation of the signal.

We were able to isolate individual markers of adaptation to degraded speech: AM rate discrimination thresholds predicted adaptation to degraded speech, supporting our hypothesis that basic non-speech temporal processing skills contribute to perceptual learning of degraded speech. Our findings provide further evidence for the importance of a “bottom-up” acoustic analysis the auditory system has to perform when confronted with degraded speech.

Importantly, these results were replicated in CI patients: Acoustic AM rate discrimination thresholds correlated with CI users’ speech recognition, consistent with previous electrical AM discrimination experiments (Fu, 2002). Most importantly, we show that already early after implantation, AM rate discrimination thresholds can prospectively predict speech recognition six months later, when duration of deafness is accounted for. Together with previous studies, these results confirm that temporal modulation information is crucial for CI patients’ speech processing, and that AM discrimination thresholds might be used as powerful predictors of CI outcome.

Further evidence for the importance of temporal envelope cues for speech perception comes from our finding that expansion of the speech envelope can restore intelligibility when background noise is present (Apoux et al., 2001). Such signal processing schemes could easily be incorporated into hearing aid algorithms to facilitate speech comprehension in noise.

The neural plasticity underlying short-term perceptual adaptation to degraded speech was fostered by important subcortical contributions: Adaptation was predicted by morphology of the pulvinar, and was accompanied by a haemodynamic down-regulation of striato-thalamic nuclei. All identified structures are highly interconnected with the cortex: The pulvinar projects to auditory cortex, and the ventral anterior thalamus and the caudate form part of the striato-thalamo-cortical loop which is proposed to collect cortical information, funnel and converge it at cortical output areas, thereby reconfiguring cortical activation patterns (Kotz et al., 2009). Engagement of this pathway is likely to sharpen the cortical representation of a stimulus and ultimately lead to a convergence of the degraded speech signal onto a clear-speech “representation”, allowing for enhanced comprehension.

Effortful processing of speech and non-speech was accompanied by an increase of insular and ACC activity, most likely reflecting executive (e.g., attentional) processes (Eckert et al., 2009). Interestingly, these regions’ profiles were found to be altered in older age and hearing loss, consistent with previous findings of age-related functional and structural changes in frontal lobe systems supporting cognitive control (Harris et al., 2009). Precisely, we show that the dynamic range of ACC activity decreases with age which in turn is detrimental to speech comprehension. Rather than playing a compensatory role for deficits due to aging (Cabeza et al., 2002), the observed cingulo-opercular dynamics might reflect a generalised upregulation of cognitive control with age irrespective of task difficulty.

Taken together, the present thesis elucidates the central neural mechanisms of adaptation to adverse listening conditions with respect to three major novel points. Listeners rely on a common executive network for “effortful listening”, involving the ACC and anterior insula when listening tasks become increasingly difficult, in the speech as well as the non-speech domain. This network exhibits a generalised upregulation in older adults, indicating an increased reliance on cognitive control systems. Finally, the present data advance the understanding of how a listener adapts to a degraded speech input, demonstrating that rapid adaptation is partly explained by haemodynamic down-regulation in subcortical structures.

Zusammenfassung

Einführung

Gesprochene Sprache wird selten unter optimalen Bedingungen gehört. Jedoch haben Menschen die Fähigkeit sich schnell an ein verzerrtes oder reduziertes akustisches Sprachsignal anzupassen, d.h. Sprache auch bei schlechter Signalqualität verstehen zu lernen. Diese Herausforderung ist besonders relevant für hörgeschädigte Menschen, oder im drastischsten Fall, Cochlea-Implantat (CI)-Patienten, die sich an ein extrem reduziertes Sprachsignal anpassen müssen, das vom Implantat übertragen wird. Die vorliegende Arbeit untersucht Verhaltens- und neuronale Prozesse, die dieser Anpassung an verschlechterte Hörbedingungen zugrunde liegen.

CI-transduzierte Sprache wurde bei Normalhörenden mithilfe von vocodierter Sprache simuliert (Shannon et al., 1995), in der die spektrale Information des akustischen Signals reduziert wird, aber der zeitlich variierende Pegel, d.h. die Schwankungen der zeitlichen Einhüllenden des Signals, intakt bleibt. Bei der ersten Exposition lernen Normalhörende schnell, vocodierte Sprache zu verstehen (Davis et al., 2005; Eisner et al., 2010). Eine beträchtliche interindividuelle Variabilität in der Fähigkeit, sich diesem reduzierten Sprachsignal anzupassen, wurde jedoch bei Normalhörenden und CI-Patienten beobachtet. Derzeit ist unklar, welche Faktoren diesen Anpassungsprozess maßgeblich fördern.

Die Reduktion des spektralen Details im Sprachsignal zwingt den Hörer sich für das Sprachverständnis mehr auf die zeitliche Einhüllende zu verlassen. Eine höhere Sensitivität für zeitliche Modulationen in einem akustischen Signal, wie sie mithilfe von Amplitudenmodulations- (AM)-Ratendiskrimination gemessen werden kann, könnte daher prädiktiv sein für die perzeptuelle Anpassung an verschlechtere Sprachqualität, sowohl bei Normalhörenden als auch bei CI-Patienten (Fu, 2002).

Da die zeitlichen Modulationen im Sprachsignal für das Verständnis notwendig und fast ausreichend sind (Shannon et al., 1995), könnte eine Übertreibung dieser Fluktuationen die Spracherkennung unterstützen. Jedoch wurde gezeigt, dass eine Übertreibung der zeitlichen Einhüllenden bei Stille das Sprachverständnis verschlechtert, während einige Studien darauf hindeuten, dass die Übertreibung der Einhüllenden bei Hintergrundrauschen das Verständnis unterstützt (Apoux et al., 2001). Eine intensivere Untersuchung dieses Interaktionseffekts könnte helfen die Mechanismen der Verarbeitung der Hüllkurve des Sprachsignals aufzuzeigen.

Obwohl schnelles perzeptuelles Lernen von vocodierter Sprache in vielen Verhaltensstudien bestätigt wurde (z.B. Davis et al., 2005), sind dessen neurale Grundlagen weitgehend unbekannt. Einige elektrophysiologische Studien liefern Belege für eine subkortikale Steuerung des Lernprozesses (Song et al., 2011). Jedoch zeigen auf funktioneller Magnetresonanztomographie (fMRT) basierende Lernstudien vor allem kortikale Veränderungen während des Anpassungsprozesses. So wurde beispielsweise gezeigt, dass Feedback-basiertes perzeptuelles Lernen mit Aktivität im linken inferioren frontalen Gyrus einhergeht (IFG; Eisner et al., 2010). Da Zuhörer in Alltagssituationen jedoch selten ein direktes Feedback erhalten, bedarf es weiterer Forschung zur neuronalen Dynamik der selbstregulierten Anpassung.

Solche Anpassungsprozesse sind bei älteren Menschen wenig erforscht, obwohl perzeptuelles Lernen für sie von besonderer Bedeutung ist, da sie häufig von Hörverlust betroffen sind. Laut einer Hypothese zur Sprachverarbeitung bei älteren Erwachsenen müssen Ältere erhöhte Anstrengungen für das Verstehen von Sprache aufwenden. Solche erhöhten Höranstrengungen sollen sich neuronal in einer vermehrten Aktivität in frontalen Hirnregionen, die an der kognitiven Kontrolle beteiligt sind, widerspiegeln (Eckert et al. 2008). Dementsprechend führen Fehler in einer Worterkennungsaufgabe zu erhöhter Aktivität im anterioren cingulären Kortex (ACC), jedoch mehr bei älteren als jüngeren Erwachsenen (Harris et al., 2009). Daher scheint für die Lösung von Höraufgaben der Einbezug von kognitiver Kontrolle mit dem Alter zuzunehmen. Es ist jedoch unklar, ob dies einen altersbedingten Verlust der Spezialisierung von kortikalen Gehirnregionen oder einen Kompensationsmechanismus (Cabeza et al., 2002) widerspiegelt.

Infolgedessen ist eine weitere Untersuchung der Anpassung an verschlechterte Sprachqualität nötig. Die vorliegende Arbeit befasst sich zunächst damit, ob individuelle Unterschiede in Merkmalen der Hirnstruktur, kognitiven und auditiven Fähigkeiten das Ausmaß vorhersagen können, mit dem ein Zuhörer sich an verschlechterte akustische Bedingungen anpassen kann. Dies wird sowohl bei Normalhörenden als auch bei CI-Patienten untersucht. Die zugrundeliegende neurale Dynamik solcher adaptiver Mechanismen wird ferner bei Jüngeren untersucht, um sie dann mit älteren Erwachsenen mit verschiedenen Graden altersbedingter Schwerhörigkeit zu vergleichen. Schließlich wird geprüft, ob eine Signalverarbeitungsstrategie, die die zeitlichen Modulationen im Sprachsignal verstärkt, das Verständnis bei Stille und bei Hintergrundrauschen erleichtern kann.

Experimente und Ergebnisse

Die vorliegende Dissertation umfasst fünf Experimente, um die beschriebenen Fragestellungen anzugehen. Psychoakustische Tests, strukturelle und funktionelle MRT wurden angewandt, um Verhaltens- und neuronale Plastizität zu charakterisieren, die einem Zuhörer helfen, sich an verschlechterte Sprachqualität anzupassen.

Ziel des ersten Experiments war es, die interindividuelle Variabilität bei der Anpassung an verschlechterte Sprachqualität zu verstehen. Unsere Hypothese war, dass individuelle Unterschiede in nichtsprachlichen auditiven, kognitiven oder neuroanatomischen Faktoren prädiktiv sein sollten für das perzeptuelle Lernen von Sprache. In einem Kurzzeitlernexperiment hörten normalhörende Probanden 100 vierband-vocodierte Sätze und wiederholten was sie verstanden hatten. Nichtsprachliche Hörfähigkeiten wurden mithilfe von Modulationsratendiskrimination bewertet, wobei Modulationsfrequenzen um die sprachrelevante Rate von 4 Hz zentriert waren. Gedächtnisfähigkeiten wurden ausgewertet und strukturelle MRT-Aufnahmen wurden mithilfe von voxel-basierter Morphometrie auf anatomische Prädiktoren für perzeptuelles Lernen hin untersucht. Die Ergebnisse zeigen, dass Probanden mit einem kleineren Schwellenwert in der AM-Diskrimination schneller lernen verzerrte Sprache zu verstehen. Diese Fähigkeit, sich verschlechterten Hörbedingungen anzupassen, ist ferner anatomisch reflektiert in einem erhöhten Volumen der grauen Substanz in einem Bereich des linken Thalamus (Pulvinar), der stark mit dem auditiven und präfrontalen Kortex verbunden ist. So können die Morphologie des Pulvinar und nichtsprachliche auditive Fähigkeiten vorhersagen, wie schnell ein Hörer sich an ein reduziertes Sprachsignal anpasst.

Ein zweites Experiment untersuchte, ob solche Faktoren auch die Variabilität des Erfolgs von Cochlea-Implantationen erklären können. Kurz nach der Implantation wurden Schwellenwerte von CI-Trägern in einer akustischen AM-Ratendiskrimination um 4 Hz getestet. Die Ergebnisse zeigen, dass Schwellenwerte bereits 10 Wochen postoperativ mit der Leistung in einem Standard-Sprachverständlichkeitstest zur gleichen Zeit korrelieren. Wenn die Dauer der Taubheit kontrolliert wird, können Modulationsdiskriminationsschwellen sogar prospektiv das Sprachverständnis im Freiburger Wörtertest 6 Monate später vorhersagen. Aufgrund der geringen Stichprobengröße und der hohen Varianz verschiedener demographischer Faktoren ist jedoch eine Fortsetzung der Tests erforderlich.

Ziel des dritten Experiments war es, erstens die Konvergenz der neuronalen Mechanismen zu untersuchen, die helfen, schwierige sprachliche und nichtsprachliche Hörsituationen zu bewältigen, und zweitens die neuronalen Systeme zu identifizieren, die eine schnelle Anpassung an verzerrte Sprache unterstützen. In einem *Sparse-Sampling*, pulsgesteuerten fMRT-Experiment, wurden normalhörenden Versuchspersonen wieder 100 vierband-vocodierte Sätze präsentiert, von denen die Probanden wiederholten, was sie verstanden hatten. Klar gesprochene Sätze wurden als Baseline präsentiert. Ein weiteres fMRT-Experiment zur AM-Ratendiskrimination untersuchte die auditive Verarbeitung von langsamen zeitlichen Modulationen. Die Wahrnehmung von verzerrter Sprache verglichen mit klarer Sprache aktivierte ein „exekutives“ Netzwerk (bestehend aus der anterioren Insula und ACC; Eckert et al., 2009), das teilweise auch aktiviert wurde wenn die Schwierigkeit der nichtsprachlichen Modulationsdiskriminationsaufgabe zunahm. Schwankungen im erfolgreichen Verständnis von reduzierter Sprache korrelierten mit der Hämodynamik in einem perisylvischen Netzwerk, das als klassisches "Sprachnetzwerk" betrachtet wird. Schließlich erfolgte selbstregulierte Anpassung an schlechte Sprachqualität über eine hämodynamische Herunterregulierung von subkortikalen Strukturen (Nucleus caudatus und ventraler anteriorer Thalamuskern) und eine gleichzeitige Hochregulierung der Aktivität von kortikalen Arealen. Diese Daten zeigen die differentielle Hoch- und Herunterregulierung der Aktivität in auditiven, sprachlichen, exekutiven Netzwerken und wichtigen subkortikalen Verbindungen während der erfolgreichen Anpassung an schwierige Hörsituationen.

Ein viertes Experiment charakterisierte Veränderungen neuronaler Sprachverarbeitung im höheren Alter und bei altersbedingtem Hörverlust. Eine Gruppe von älteren Erwachsenen mit unterschiedlichen Ausprägungen an Schwerhörigkeit nahm an einem fMRT-Experiment zur Kurzzeitanpassung an verschlechterte Sprachqualität teil. Ältere lernten gleich schnell wie junge Probanden die verzerrte Sprache zu verstehen, obwohl ihr Verständnis der verzerrten Sprache im Mittel niedriger war. Auf neuronaler Ebene rekrutierten sowohl ältere als auch junge Erwachsene die linke anteriore Insula für die Wahrnehmung von reduzierter Sprache verglichen mit klarer Sprache. Allerdings war die Rekrutierung der anterioren Insula bei älteren Erwachsenen abhängig vom Hörverlust, sodass stark hörgeschädigte Ältere die Insula mehr für klare Sprache als für reduzierte Sprache aktivierten. Jüngere nicht aber ältere Erwachsene zeigten eine erhöhte Aktivierung des ACC für die verzerrte Bedingung verglichen mit der klaren Bedingung. Diese Interaktion von Altersgruppe und Bedingung wurde hervorgerufen durch eine verringerte Dynamik bei älteren Erwachsenen, die erhöhte Werte von

ACC-Aktivität für beide Bedingungen (für ein klares und reduziertes Sprachsignal) zeigten. Diese Beobachtung ist konsistent mit einer anhaltenden Hochregulation in der kognitiven Kontrolle unabhängig von der Aufgabenschwierigkeit. Für ein erfolgreiches Sprachverständnis benutzten ältere Erwachsene den mittleren frontalen Gyrus zusätzlich zu einem perisylvischen Sprachnetzwerk, das junge Erwachsene aktivierten, was auf einen Kompensationsmechanismus hindeutet. Diese Ergebnisse bestätigen, dass ältere Erwachsene selbst unter optimalen Hörbedingungen zunehmend Netzwerke für die kognitive Kontrolle rekrutieren, auf Kosten der Dynamik dieser Systeme.

Angesichts der zentralen Bedeutung der zeitlichen Einhüllenden für das Sprachverständnis, untersuchten wir in einem fünften Experiment, wie sich eine Übertreibung oder Verringerung der Ausgeprägtheit der Einhüllenden auf die Sprachverständlichkeit auswirken, in Ruhe und im Hintergrundrauschen. Statt Worte einfach zu vocodieren, wurde zusätzlich die Einhüllende modifiziert, entweder durch Erhöhen ("Expansion") oder Vermindern ("Kompression") der Modulationstiefe des Signals. Außerdem wurde weißes Rauschen in verschiedenen Signal-zu-Rausch-Verhältnissen hinzugefügt. Die parametrische Variation der Modifikation der Einhüllenden, kombiniert mit einer feinkörnigen parametrischen Variation des Signal-zu-Rausch-Verhältnisses, erlaubte uns zu quantifizieren, inwieweit Probanden von einer Expansion der Einhüllenden profitieren wenn sich der Rauschpegel erhöht. Die Ergebnisse zeigen, dass sowohl die Kompression als auch die Expansion der Einhüllenden in Ruhe die Sprachverständlichkeit verschlechtert. Jedoch zeigen Probanden bei Expansion der Einhüllenden im Hintergrundrauschen eine kleine, aber signifikante Verbesserung des Sprachverständnisses. Diese Ergebnisse sind direkte Evidenz für die Bedeutung der Sprachhüllkurve für das Verständnis, insbesondere in Gegenwart von Hintergrundrauschen.

Diskussion

Die vorliegende Dissertation untersuchte die neuronalen Systeme, die Sprachverständnis und Anpassung unter schwierigen Hörbedingungen erleichtern. Insbesondere wurde erforscht, wie eine spektrale Reduzierung des Signals von Normalhörenden, älteren Erwachsenen mit unterschiedlichem Grad der Schwerhörigkeit und CI-Patienten bewältigt wird.

Wir konnten individuelle Marker der Anpassung an verschlechterte Sprachqualität isolieren: Modulationsdiskriminationsschwellen waren prädiktiv für eine Anpassung an schlechte Hörbedingungen. Dies unterstützt die Hypothese, dass nichtsprachliche auditive Fähigkeiten dazu beitragen, Sprachwahrnehmung unter schlechten akustischen Bedingungen zu lernen. Unsere Ergebnisse sind ein weiterer Beleg für die Bedeutung der *bottom-up* akustischen Analyse, die das auditorische System zu erfüllen hat, wenn sie mit schlechter Sprachqualität konfrontiert ist (siehe auch Mattys et al., 2009) und dass perzeptuelles Lernen durch „Low-Level“ akustische Signale maßgeblich angetrieben wird (Idemaru und Holt, 2011).

Diese Ergebnisse konnten bei CI-Patienten repliziert werden: Akustische AM-Ratendiskriminationsschwellen korrelierten mit dem Sprachverständnis, in Übereinstimmung mit früheren Experimenten, die elektrische AM-Diskrimination getestet hatten (z.B. Fu, 2002). Hier zeigen wir außerdem, dass schon früh nach der Cochlea-Implantation Modulationsdiskriminationsschwellen sogar prospektiv das Sprachverständnis vorhersagen können, wenn die Dauer der Taubheit berücksichtigt wird. Zusammen mit früheren Studien bestätigen diese Ergebnisse, dass zeitliche Modulationen im Sprachsignal entscheidend sind für die Sprachverarbeitung bei CI-Patienten, und dass AM-Diskriminationsschwellen als starke Prädiktoren für den CI-Erfolg verwendet werden können.

Weitere Belege für die Bedeutung der zeitlichen Einhüllenden in der Sprachwahrnehmung kommen von unserer Beobachtung, dass die Expansion der Hüllkurve die Sprachverständlichkeit wiederherstellen kann, wenn Hintergrundgeräusche vorhanden sind (Apoux et al., 2001). Solche Signalverarbeitungsschemata könnten leicht in Hörgerätealgorithmen integriert werden, um das Sprachverstehen bei Hintergrundrauschen zu erleichtern.

Die der Anpassung an schlechte Sprachqualität zugrunde liegende neuronale Plastizität wird hauptsächlich durch subkortikale Strukturen gefördert: Individuelle Anpassung hängt mit der Morphologie des Pulvinar zusammen und wird ferner von einer hämodynamischen Herunterregulation des Nucleus caudatus und des anteroventralen Thalamus begleitet. Alle identifizierten Strukturen sind sehr eng mit der Großhirnrinde verbunden: Das Pulvinar projiziert zum auditiven Kortex (Kotz et al., 2009). Der ventrale anteriore Thalamus sowie der Nucleus caudatus sind Teil der striato-thalamo-kortikalen Schleife, die kortikale Informationen sammelt, bündelt und in kortikalen Zielarealen konvergiert, wodurch kortikale Aktivierungsmuster neu konfiguriert werden. Die Einbeziehung dieser Schleife könnte dazu beitragen, die kortikale Repräsentation eines Stimulus zu schärfen und so zu einer Konvergenz des verzerrten Sprachsignals in eine klare "Repräsentation" führen, und letztlich das Sprachverständnis verbessern.

Angestrengte Verarbeitung von sprachlichen und nichtsprachlichen auditiven Stimuli wurde durch einen Anstieg der Aktivität in der Insula und im ACC begleitet, der wahrscheinlich exekutive (z.B.

Aufmerksamkeits-) Prozesse reflektiert (Eckert et al., 2009). Interessanterweise waren die Aktivierungsprofile in diesen Arealen im Alter und bei Hörverlust verändert. Diese Ergebnisse stehen im Einklang mit früheren Beobachtungen zu altersbedingten funktionellen und strukturellen Veränderungen im Frontallappen, der mit kognitiver Kontrolle assoziiert ist (Cabeza et al., 2002, Eckert et al., 2008; Harris et al., 2009). Hier zeigen wir, dass der dynamische Bereich der ACC-Aktivität mit dem Alter abnimmt, was wiederum für das Sprachverständnis nachteilig ist. Anstatt eine kompensatorische Rolle für Defizite durch Alterung einzunehmen (Cabeza et al., 2002), könnte die beobachtete Dynamik des ACC eine generalisierte Hochregulation der kognitiven Kontrolle unabhängig von der Aufgabenschwierigkeit mit zunehmendem Alter widerspiegeln.

Zusammengefasst charakterisiert die vorliegende Arbeit die zentralen neuronalen Mechanismen der Anpassung an ungünstige Hörbedingungen in Bezug auf drei neue Aspekte: Wenn Höraufgaben schwieriger werden, sowohl in der sprachlichen als auch der nichtsprachlichen Domäne, verlassen sich Zuhörer auf ein gemeinsames exekutives Netzwerk (Eckert et al., 2009), das den ACC und die anteriore Insula einschließt. Die Aktivität dieses Netzwerks zeigt eine generalisierte Hochregulation bei älteren Erwachsenen, die sich vermehrt auf kognitive Kontrollsysteme stützen. Schließlich helfen die Daten zu verstehen, wie ein Hörer sich an eine schlechte Signalqualität anpasst. Denn sie zeigen, dass eine schnelle Anpassung teils durch eine hämodynamische Herunterregulation in subkortikalen Strukturen erklärt werden kann.

Curriculum vitae

Julia Erb

Born in1985 in Heilbronn, Germany

Education

12/2010 – 2014 (exp.)	Doctoral Studies (Psychology) *Max Planck Institute Leipzig, Germany*
09/2008 – 07/2010	Master of Science (Integrative Biology and Physiology, Speciality in Neuroscience), Grade: 14 (good) *University Paris VI, France*
10/2005 – 07/2008	Bachelor of Science (Biomedicine), Grade: 1.7 (good) *University of Würzburg, Germany*
09/1996 – 06/2005	Abitur, Grade: 1.3 (very good) *Elly-Heuss-Knapp-Gymnasium Heilbronn, Germany*

Professional Experience

12/2010 – 12/1213	*Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig* PhD thesis (Advisor: Jonas Obleser and Sonja Kotz): The neural dynamics of adaptation to degraded speech. Functional MRI and behavioural experiments. Analysis using Matlab, SPM, Freesurfer
01/2010 – 09/2010	*FMRIB Center, Department of Clinical Neurology, University of Oxford, UK* MSc thesis and research assistant (Advisor: Kate Watkins): Abnormal brain activation patterns in developmental stuttering during listening and covert reading. Analysis of fMRI data using FSL
04/2009 – 06/2009	*Laboratoire de Neurobiologie, École Normale Supérieure, Paris* Research internship (Advisor: Stéphane Supplisson): Trans-synaptic retrograde tracing of neuronal pathways using the non-toxic fragment C of the tetanus toxin. Primary cell culture, transfection, fluorescence microscopy, patch clamp
01/2007 – 07/2008	*Rudolf-Virchow Center, Würzburg*

BSc thesis and research assistant (Advisor: Heike Hermanns): Characterisation of the Oncostatin M – mediated signal transduction. Cell culture, western blot, flow cytometry

Awards and Honours

12/2010–01/2014	PhD Fellowship of the Max Planck Society
10/2012	DAAD Travel Grant to attend Neuroscience 2012
11/2010	Brain Travel Grant to attend Neuroscience 2010
09/2009 – 06/2010	DAAD Fellowship for Master Studies at the Universities Paris VI and Oxford

Publications

Erb J & Obleser J. Upregulation of cognitive control networks in older adults' speech comprehension (2013). *Frontiers in Systems Neuroscience, Special issue: The effect of hearing loss on neural processing*,7(116), 1-13.

Scharinger M, Henry MJ, **Erb J,** Meyer L, & Obleser J. Thalamic and parietal brain morphology predicts auditory category learning (2013). *Neuropsychologia*, 53C, 75-83.

Erb J, Henry MJ, Eisner F, & Obleser J (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *The Journal of Neuroscience*, 33(26), 10688-10697.

Erb J, Henry MJ, Eisner F, & Obleser J (2012). Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychologia*, 50(9), 2154-2164.

Talks

Erb J. How does the listening brain adapt to challenging acoustic conditions? *Institute Colloquium at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig,* December 2013.

Selbständigkeitserklärung gemäß § 8(2) der Promotionsordnung

Hiermit versichere ich, dass die vorliegende Arbeit ohne unzulässige Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde, und dass die aus fremden Quellen direkt oder indirekt übernommenen Gedanken in der Arbeit als solche kenntlich gemacht worden sind. Ich versichere, dass die vorliegende Arbeit in gleicher oder in ähnlicher Form keiner anderen wissenschaftlichen Einrichtung zum Zwecke einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt und auch veröffentlicht wurde. Es haben keine früheren erfolglosen Promotionsversuche stattgefunden.

Leipzig, den 14. Februar 2014

Julia Erb

MPI Series in Human Cognitive and Brain Sciences:

1 Anja Hahne
Charakteristika syntaktischer und semantischer Prozesse bei der auditiven Sprachverarbeitung: Evidenz aus ereigniskorrelierten Potentialstudien

2 Ricarda Schubotz
Erinnern kurzer Zeitdauern: Behaviorale und neurophysiologische Korrelate einer Arbeitsgedächtnisfunktion

3 Volker Bosch
Das Halten von Information im Arbeitsgedächtnis: Dissoziationen langsamer corticaler Potentiale

4 Jorge Jovicich
An investigation of the use of Gradient- and Spin-Echo (GRASE) imaging for functional MRI of the human brain

5 Rosemary C. Dymond
Spatial Specificity and Temporal Accuracy in Functional Magnetic Resonance Investigations

6 Stefan Zysset
Eine experimentalpsychologische Studie zu Gedächtnisabrufprozessen unter Verwendung der funktionellen Magnetresonanztomographie

7 Ulrich Hartmann
Ein mechanisches Finite-Elemente-Modell des menschlichen Kopfes

8 Bertram Opitz
Funktionelle Neuroanatomie der Verarbeitung einfacher und komplexer akustischer Reize: Integration haemodynamischer und elektrophysiologischer Maße

9 Gisela Müller-Plath
Formale Modellierung visueller Suchstrategien mit Anwendungen bei der Lokalisation von Hirnfunktionen und in der Diagnostik von Aufmerksamkeitsstörungen

10 Thomas Jacobsen
Characteristics of processing morphological structural and inherent case in language comprehension

11 Stefan Kölsch
Brain and Music
A contribution to the investigation of central auditory processing with a new electrophysiological approach

12 Stefan Frisch
Verb-Argument-Struktur, Kasus und thematische Interpretation beim Sprachverstehen

13 Markus Ullsperger
The role of retrieval inhibition in directed forgetting – an event-related brain potential analysis

14 Martin Koch
Measurement of the Self-Diffusion Tensor of Water in the Human Brain

15 Axel Hutt
Methoden zur Untersuchung der Dynamik raumzeitlicher Signale

16 Frithjof Kruggel
Detektion und Quantifizierung von Hirnaktivität mit der funktionellen Magnetresonanztomographie

17 Anja Dove
Lokalisierung an internen Kontrollprozessen beteiligter Hirngebiete mithilfe des Aufgabenwechselparadigmas und der ereigniskorrelierten funktionellen Magnetresonanztomographie

18 Karsten Steinhauer
Hirnphysiologische Korrelate prosodischer Satzverarbeitung bei gesprochener und geschriebener Sprache

19 Silke Urban
Verbinformationen im Satzverstehen

20 Katja Werheid
Implizites Sequenzlernen bei Morbus Parkinson

21 Doreen Nessler
Is it Memory or Illusion? Electrophysiological Characteristics of True and False Recognition

22 Christoph Herrmann
Die Bedeutung von 40-Hz-Oszillationen für kognitive Prozesse

23 Christian Fiebach
Working Memory and Syntax during Sentence Processing.
A neurocognitive investigation with event-related brain potentials and functional magnetic resonance imaging

24 Grit Hein
Lokalisation von Doppelaufgabendefiziten bei gesunden älteren Personen und neurologischen Patienten

25 Monica de Filippis
Die visuelle Verarbeitung unbeachteter Wörter. Ein elektrophysiologischer Ansatz

26 Ulrich Müller
Die katecholaminerge Modulation präfrontaler kognitiver Funktionen beim Menschen

27 Kristina Uhl
Kontrollfunktion des Arbeitsgedächtnisses über interferierende Information

28 Ina Bornkessel
The Argument Dependency Model: A Neurocognitive Approach to Incremental Interpretation

29 Sonja Lattner
Neurophysiologische Untersuchungen zur auditorischen Verarbeitung von Stimminformationen

30 Christin Grünewald
Die Rolle motorischer Schemata bei der Objektrepräsentation: Untersuchungen mit funktioneller Magnetresonanztomographie

31 Annett Schirmer
Emotional Speech Perception: Electrophysiological Insights into the Processing of Emotional Prosody and Word Valence in Men and Women

32 André J. Szameitat
Die Funktionalität des lateral-präfrontalen Cortex für die Verarbeitung von Doppelaufgaben

33 Susanne Wagner
Verbales Arbeitsgedächtnis und die Verarbeitung ambiger Wörter in Wort- und Satzkontexten

34 Sophie Manthey
Hirn und Handlung: Untersuchung der Handlungsrepräsentation im ventralen prämotorischen Cortex mit Hilfe der funktionellen Magnet-Resonanz-Tomographie

35 Stefan Heim
Towards a Common Neural Network Model of Language Production and Comprehension: fMRI Evidence for the Processing of Phonological and Syntactic Information in Single Words

36 Claudia Friedrich
Prosody and spoken word recognition: Behavioral and ERP correlates

37 Ulrike Lex
Sprachlateralisierung bei Rechts- und Linkshändern mit funktioneller Magnetresonanztomographie

38 Thomas Arnold
Computergestützte Befundung klinischer Elektroenzephalogramme

39 Carsten H. Wolters
Influence of Tissue Conductivity Inhomogeneity and Anisotropy on EEG/MEG based Source Localization in the Human Brain

40 Ansgar Hantsch
Fisch oder Karpfen? Lexikale Aktivierung von Benennungsalternative bei der Objektbenennung

41 Peggy Bungert
Zentralnervöse Verarbeitung akustischer Informationen
Signalidentifikation, Signallateralisation und zeitgebundene Informationsverarbeitung bei Patienten mit erworbenen Hirnschädigungen

42 Daniel Senkowski
Neuronal correlates of selective attention: An investigation of electrophysiological brain responses in the EEG and MEG

43 Gert Wollny
Analysis of Changes in Temporal Series of Medical Images

44 Angelika Wolf
Sprachverstehen mit Cochlea-Implantat: EKP-Studien mit postlingual ertaubten erwachsenen CI-Trägern

45 Kirsten G. Volz
Brain correlates of uncertain decisions: Types and degrees of uncertainty

46 Hagen Huttner
Magnetresonanztomographische Untersuchungen über die anatomische Variabilität des Frontallappens des menschlichen Großhirns

47 Dirk Köster
Morphology and Spoken Word Comprehension: Electrophysiological Investigations of Internal Compound Structure

48 Claudia A. Hruska
Einflüsse kontextueller und prosodischer Informationen in der auditorischen Satzverarbeitung: Untersuchungen mit ereigniskorrelierten Hirnpotentialen

49 Hannes Ruge
Eine Analyse des raum-zeitlichen Musters neuronaler Aktivierung im Aufgabenwechselparadigma zur Untersuchung handlungssteuernder Prozesse

50 Ricarda I. Schubotz
Human premotor cortex: Beyond motor performance

51 Clemens von Zerssen
Bewusstes Erinnern und falsches Wiedererkennen: Eine funktionelle MRT Studie neuroanatomischer Gedächtniskorrelate

52 Christiane Weber
Rhythm is gonna get you.
Electrophysiological markers of rhythmic processing in infants with and without risk for Specific Language Impairment (SLI)

53 Marc Schönwiesner
Functional Mapping of Basic Acoustic Parameters in the Human Central Auditory System

54 Katja Fiehler
Temporospatial characteristics of error correction

55 Britta Stolterfoht
Processing Word Order Variations and Ellipses: The Interplay of Syntax and Information Structure during Sentence Comprehension

56 Claudia Danielmeier
Neuronale Grundlagen der Interferenz zwischen Handlung und visueller Wahrnehmung

57 Margret Hund-Georgiadis
Die Organisation von Sprache und ihre Reorganisation bei ausgewählten, neurologischen Erkrankungen gemessen mit funktioneller Magnetresonanztomographie – Einflüsse von Händigkeit, Läsion, Performanz und Perfusion

58 Jutta L. Mueller
Mechanisms of auditory sentence comprehension in first and second language: An electrophysiological miniature grammar study

59 Franziska Biedermann
Auditorische Diskriminationsleistungen nach unilateralen Läsionen im Di- und Telenzephalon

60 Shirley-Ann Rüschemeyer
The Processing of Lexical Semantic and Syntactic Information in Spoken Sentences: Neuroimaging and Behavioral Studies of Native and Non-Native Speakers

61 Kerstin Leuckefeld
The Development of Argument Processing Mechanisms in German.
An Electrophysiological Investigation with School-Aged Children and Adults

62 Axel Christian Kühn
Bestimmung der Lateralisierung von Sprachprozessen unter besondere Berücksichtigung des temporalen Cortex, gemessen mit fMRT

63 Ann Pannekamp
Prosodische Informationsverarbeitung bei normalsprachlichem und deviantem Satzmaterial: Untersuchungen mit ereigniskorrelierten Hirnpotentialen

64 Jan Derrfuß
Functional specialization in the lateral frontal cortex: The role of the inferior frontal junction in cognitive control

65 Andrea Mona Philipp
The cognitive representation of tasks – Exploring the role of response modalities using the task-switching paradigm

66 Ulrike Toepel
Contrastive Topic and Focus Information in Discourse – Prosodic Realisation and Electrophysiological Brain Correlates

67 Karsten Müller
Die Anwendung von Spektral- und Waveletanalyse zur Untersuchung der Dynamik von BOLD-Zeitreihen verschiedener Hirnareale

68 Sonja A.Kotz
The role of the basal ganglia in auditory language processing: Evidence from ERP lesion studies and functional neuroimaging

69 Sonja Rossi
The role of proficiency in syntactic second language processing: Evidence from event-related brain potentials in German and Italian

70 Birte U. Forstmann
Behavioral and neural correlates of endogenous control processes in task switching

71 Silke Paulmann
Electrophysiological Evidence on the Processing of Emotional Prosody: Insights from Healthy and Patient Populations

72 Matthias L. Schroeter
Enlightening the Brain – Optical Imaging in Cognitive Neuroscience

73 Julia Reinholz
Interhemispheric interaction in object- and word-related visual areas

74 Evelyn C. Ferstl
The Functional Neuroanatomy of Text Comprehension

75 Miriam Gade
Aufgabeninhibition als Mechanismus der Konfliktreduktion zwischen Aufgabenrepräsentationen

76 Juliane Hofmann
Phonological, Morphological, and Semantic Aspects of Grammatical Gender Processing in German

77 Petra Augurzky
Attaching Relative Clauses in German – The Role of Implicit and Explicit Prosody in Sentence Processing

78 Uta Wolfensteller
Habituelle und arbiträre sensomotorische Verknüpfungen im lateralen prämotorischen Kortex des Menschen

79 Päivi Sivonen
Event-related brain activation in speech perception: From sensory to cognitive processes

80 Yun Nan
Music phrase structure perception: the neural basis, the effects of acculturation and musical training

81 Katrin Schulze
Neural Correlates of Working Memory for Verbal and Tonal Stimuli in Nonmusicians and Musicians With and Without Absolute Pitch

82 Korinna Eckstein
Interaktion von Syntax und Prosodie beim Sprachverstehen: Untersuchungen anhand ereigniskorrelierter Hirnpotentiale

83 Florian Th. Siebörger
Funktionelle Neuroanatomie des Textverstehens: Kohärenzbildung bei Witzen und anderen ungewöhnlichen Texten

84 Diana Böttger
Aktivität im Gamma-Frequenzbereich des EEG: Einfluss demographischer Faktoren und kognitiver Korrelate

85 Jörg Bahlmann
Neural correlates of the processing of linear and hierarchical artificial grammar rules: Electrophysiological and neuroimaging studies

86 Jan Zwickel
Specific Interference Effects Between Temporally Overlapping Action and Perception

87 Markus Ullsperger
Functional Neuroanatomy of Performance Monitoring: fMRI, ERP, and Patient Studies

88 Susanne Dietrich
Vom Brüllen zum Wort – MRT-Studien zur kognitiven Verarbeitung emotionaler Vokalisationen

89 Maren Schmidt-Kassow
What's Beat got to do with ist? The Influence of Meter on Syntactic Processing: ERP Evidence from Healthy and Patient populations

90 Monika Lück
Die Verarbeitung morphologisch komplexer Wörter bei Kindern im Schulalter: Neurophysiologische Korrelate der Entwicklung

91 Diana P. Szameitat
Perzeption und akustische Eigenschaften von Emotionen in menschlichem Lachen

92 Beate Sabisch
Mechanisms of auditory sentence comprehension in children with specific language impairment and children with developmental dyslexia: A neurophysiological investigation

93 Regine Oberecker
Grammatikverarbeitung im Kindesalter: EKP-Studien zum auditorischen Satzverstehen

94 Şükrü Barış Demiral
Incremental Argument Interpretation in Turkish Sentence Comprehension

95 Henning Holle
The Comprehension of Co-Speech Iconic Gestures: Behavioral, Electrophysiological and Neuroimaging Studies

96 Marcel Braß
Das inferior frontale Kreuzungsareal und seine Rolle bei der kognitiven Kontrolle unseres Verhaltens

97 Anna S. Hasting
Syntax in a blink: Early and automatic processing of syntactic rules as revealed by event-related brain potentials

98 Sebastian Jentschke
Neural Correlates of Processing Syntax in Music and Language – Influences of Development, Musical Training and Language Impairment

99 Amelie Mahlstedt
The Acquisition of Case marking Information as a Cue to Argument Interpretation in German
An Electrophysiological Investigation with Pre-school Children

100 Nikolaus Steinbeis
Investigating the meaning of music using EEG and fMRI

101 Tilmann A. Klein
Learning from errors: Genetic evidence for a central role of dopamine in human performance monitoring

102 Franziska Maria Korb
Die funktionelle Spezialisierung des lateralen präfrontalen Cortex: Untersuchungen mittels funktioneller Magnetresonanztomographie

103 Sonja Fleischhauer
Neuronale Verarbeitung emotionaler Prosodie und Syntax: die Rolle des verbalen Arbeitsgedächtnisses

104 Friederike Sophie Haupt
The component mapping problem: An investigation of grammatical function reanalysis in differing experimental contexts using eventrelated brain potentials

105 Jens Brauer
Functional development and structural maturation in the brain's neural network underlying language comprehension

106 Philipp Kanske
Exploring executive attention in emotion: ERP and fMRI evidence

107 Julia Grieser Painter
Music, meaning, and a semantic space for musical sounds

108 Daniela Sammler
The Neuroanatomical Overlap of Syntax Processing in Music and Language - Evidence from Lesion and Intracranial ERP Studies

109 Norbert Zmyj
Selective Imitation in One-Year-Olds: How a Model's Characteristics Influence Imitation

110 Thomas Fritz
Emotion investigated with music of variable valence – neurophysiology and cultural influence

111 Stefanie Regel
The comprehension of figurative language: Electrophysiological evidence on the processing of irony

112 Miriam Beisert
Transformation Rules in Tool Use

113 Veronika Krieghoff
Neural correlates of Intentional Actions

114 Andreja Bubić
Violation of expectations in sequence processing

115 Claudia Männel
Prosodic processing during language acquisition: Electrophysiological studies on intonational phrase processing

116 Konstanze Albrecht
Brain correlates of cognitive processes underlying intertemporal choice for self and other

117 Katrin Sakreida
Nicht-motorische Funktionen des prämotorischen Kortex:
Patientenstudien und funktionelle Bildgebung

118 Susann Wolff
The interplay of free word order and pro-drop in incremental sentence processing: Neurophysiological evidence from Japanese

119 Tim Raettig
The Cortical Infrastructure of Language Processing: Evidence from Functional and Anatomical Neuroimaging

120 Maria Golde
Premotor cortex contributions to abstract and action-related relational processing

121 Daniel S. Margulies
Resting-State Functional Connectivity fMRI: A new approach for assessing functional neuroanatomy in humans with applications to neuroanatomical, developmental and clinical questions

122 Franziska Süß
The interplay between attention and syntactic processes in the adult and developing brain: ERP evidences

123 Stefan Bode
From stimuli to motor responses: Decoding rules and decision mechanisms in the human brain

124 Christiane Diefenbach
Interactions between sentence comprehension and concurrent action:
The role of movement effects and timing

125 Moritz M. Daum
Mechanismen der frühkindlichen Entwicklung des Handlungsverständnisses

126 Jürgen Dukart
Contribution of FDG-PET and MRI to improve Understanding, Detection and Differentiation of Dementia

127 Kamal Kumar Choudhary
Incremental Argument Interpretation in a Split Ergative Language:
Neurophysiological Evidence from Hindi

128 Peggy Sparenberg
Filling the Gap: Temporal and Motor Aspects of the Mental Simulation of Occluded Actions

129 Luming Wang
The Influence of Animacy and Context on Word Order Processing: Neurophysiological Evidence from Mandarin Chinese

130 Barbara Ettrich
Beeinträchtigung frontomedianer Funktionen bei Schädel-Hirn-Trauma

131 Sandra Dietrich
Coordination of Unimanual Continuous Movements with External Events

132 R. Muralikrishnan
An Electrophysiological Investigation Of Tamil Dative-Subject Constructions

133 Christian Obermeier
Exploring the significance of task, timing and background noise on gesture-speech integration

134 Björn Herrmann
Grammar and perception: Dissociation of early auditory processes in the brain

135 Eugenia Solano-Castiella
In vivo anatomical segmentation of the human amygdala and parcellation of emotional processing

136 Marco Taubert
Plastizität im sensomotorischen System – Lerninduzierte Veränderungen in der Struktur und Funktion des menschlichen Gehirns

137 Patricia Garrido Vásquez
Emotion Processing in Parkinson's Disease:
The Role of Motor Symptom Asymmetry

138 Michael Schwartze
Adaptation to temporal structure

139 Christine S. Schipke
Processing Mechanisms of Argument Structure and Case-marking in Child Development: Neural Correlates and Behavioral Evidence

140 Sarah Jessen
Emotion Perception in the Multisensory Brain

141 Jane Neumann
Beyond activation detection: Advancing computational techniques for the analysis of functional MRI data

142 Franziska Knolle
Knowing what's next: The role of the cerebellum in generating predictions

143 Michael Skeide
Syntax and semantics networks in the developing brain

144 Sarah M. E. Gierhan
Brain networks for language
Anatomy and functional roles of neural pathways supporting language comprehension and repetition

145 Lars Meyer
The Working Memory of Argument-Verb Dependencies
Spatiotemporal Brain Dynamics during Sentence Processing

146 Benjamin Stahl
Treatment of Non-Fluent Aphasia through
Melody, Rhythm and Formulaic Language

147 Kathrin Rothermich
The rhythm's gonna get you: ERP and fMRI evidence on the interaction of metric and semantic processing

148 Julia Merrill
Song and Speech Perception – Evidence from fMRI, Lesion Studies and Musical Disorder

149 Klaus-Martin Krönke
Learning by Doing?
Gesture-Based Word-Learning and its Neural Correlates in Healthy Volunteers and Patients with Residual Aphasia

150 Lisa Joana Knoll
When the hedgehog kisses the frog
A functional and structural investigation of syntactic processing in the developing brain

151 Nadine Diersch
Action prediction in the aging mind

152 Thomas Dolk
A Referential Coding Account for the Social Simon Effect

153 Mareike Bacha-Trams
Neurotransmitter receptor distribution in Broca's area and the posterior superior temporal gyrus

154 Andrea Michaela Walter
The role of goal representations in action control

155 Anne Keitel
Action perception in development: The role of experience

156 Iris Nikola Knierim
Rules don't come easy: Investigating feedback-based learning of phonotactic rules in language.

157 Jan Schreiber
Plausibility Tracking: A method to evaluate anatomical connectivity and microstructural properties along fiber pathways

158 Katja Macher
Die Beteiligung des Cerebellums am verbalen Arbeitsgedächtnis