MAX PLANCK INSTITUTE FOR HUMAN COGNITIVE AND BRAIN SCIENCES LEIPZIG

Antje Strauß: Neural oscillatory dynamics of spoken word recognition. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences, 2015 (MPI Series in Human Cognitive and Brain Sciences; 163)

Neural oscillatory dynamics of spoken word recognition

Impressum

Max Planck Institute for Human Cognitive and Brain Sciences, 2015

Druck: Sächsisches Druck- und Verlagshaus Direct World, Dresden
Titelbild: ©Antje Strauß, 2015

ISBN 978-3-941504-47-9

Neural oscillatory dynamics of spoken word recognition

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 Magistra Artium, Antje Strauß

geboren am 08. Juni 1985 in Blankenburg / Harz

Leipzig, den 01. Oktober 2014

Bibliographic details

Antje Strauß
Neural oscillatory dynamics of spoken word recognition
Fakultät für Biowissenschaften, Pharmazie und Psychologie
Universität Leipzig
Dissertation
163 pages, 359 references, 22 figures

This thesis investigated slow oscillatory signatures of spoken word recognition. In particular, we aimed to dissociate alpha (∼10 Hz) and theta (∼4 Hz) band oscillations to understand the underlying neural mechanisms of lexico-semantic processing. Three experiments were conducted while recording the electroencephalogram (EEG): i) an auditory lexical decision task in quiet, ii) an auditory lexical decision task in white noise, and iii) an intelligibility rating of cloze probability sentences in different level of noise-vocoding (spectrally degraded speech). The results show that alpha oscillations play a role during spoken word recognition in three possible ways: First, induced alpha power scaled with lexicality, that is, with the difficulty to map the phonological representation onto meaning. Post-lexical alpha power was suppressed for words indicating processing of lexico-semantic information. In turn, alpha power was enhanced for pseudowords indicating the inhibition of lexico-semantic processing. Second, induced alpha power was found to be enhanced at the beginning of words embedded in noise compared to clear speech in line with the presumed inhibitory function of alpha. We propose a framework to further assess the role of alpha in selectively inhibiting task-irrelevant auditory objects. Third, pre-stimulus alpha phase was found to modulate lexical decision accuracy in noise. We interpreted this finding to reflect selective inhibition in the sense that stimuli coinciding with the excitatory phase were more likely to be thoroughly processed than when coinciding with the inhibitory phase and were thus ultimately judged correctly. Furthermore, we were able to associate theta oscillations with lexico-semantic processing. First, induced theta power was found to be post-lexically enhanced selectively for ambiguous pseudowords that differed only in one vowel from their real-word neighbours. We interpreted this finding in terms of ambiguity resolution of the response conflict induced by their proximity to real words. We suggest that phonemic information needed to be “replayed” in order to re-compare it with long-term memory representations and thus to resolve ambiguity. Second, in high cloze probability sentences theta power was found to be enhanced just before the onset of the sentence-final word, thus indicating the anticipatory activation of lexico-semantics in long-term memory. The results provide novel evidence on the temporal mechanisms in spoken word recognition. These findings are discussed with regard to their implications of the nonlinearity of speech processing and the reassessment of event-related potentials.

Acknowledgements

First of all, I am much obliged to my supervisor Jonas Obleser who converted me to a natural scientist. I owe him most of my knowledge about signal processing, data analysis and the art of typography. He founded the incredible "Auditory Cognition" group with inspiring and challenging methodological discussions that became almost a family.

Mathias Scharinger sitting to my left and Molly Henry sitting to my right became my scientific foster parents. I would like to thank them for discussing crazy brain measures, playing word games, listening to Bach and Metal simultaneously and their moral support throughout the time.
I would like to thank Malte Wöstmann, Anna Wilsch, Björn Herrmann, Julia Erb, Alex Brandmeyer and Sung-Joo Lim for the constant critical exchange preferably with a cup of espresso or coke in their hand.
I am grateful to Dunja Kunke and a crew of student assistants amongst which were Sergej Schwigon, David Stoppel, Christina Otto, Christoph Daube, Steven Kalinke and Leo Waschke who helped acquiring and preprocessing the data. They created a wonderful working environment.

Furthermore, I would like to thank Sonja Kotz for her initializing ideas and encouragements at different stages of the PhD period. I appreciated a lot the discussions with Hellmuth Obrig about the implications of my results for aphasic patients. Finally, I thank Jörg Jescheniak for accepting and assessing my work.

This work is dedicated to my loving grandfathers Wolfgang Witt and Otto Strauß. It was not granted to both of them to be there to experience the completion of the dissertation with me.

Contents

Was das Gehör betreffe, so schreibe, und zwar nur auf das Oberflächlichste, soll Konrad zum Baurat gesagt haben, sagt Wieser, entweder ein Arzt, was gänzlich falsch sei, oder ein Philosoph darüber, was gänzlich falsch sei. Schreibe ein Arzt über das Gehör, sei das völlig wertlos. Schreibe ein Philosoph darüber, sei das auch völlig wertlos. Man darf nicht nur Arzt und man darf nicht nur Philosoph sein, wenn man sich eine Sache wie das Gehör vornimmt und an sie herangehe. Dazu müsse man auch Mathematiker und Physiker und also ein vollkommener Naturwissenschaftler und dazu auch noch Prophet und Künstler sein und das alles in höchstem Maße.

[Konrad is supposed to have said to the inspector [...] that it is usually either a philosopher or a doctor who writes about the human ear. Neither is adequately prepared for the task and in either case they only treat the phenomenon of hearing in the most superficial manner. If a doctor writes about hearing it is entirely worthless. If a philosopher writes about hearing it is equally worthless. When dealing with such a thing as the human ear, one must be more than a doctor, more than a philosopher. One must be a mathematician, and a physicist, a well-rounded scientist in fact. Nor is that enough either, as one must be something of a prophet and an artist, too—and not just of the common kind.]

Thomas Bernhard. *Das Kalkwerk.*
[Loosely translated by A.S.]

1 General Introduction

Understanding deficient speech is a challenge for each listener in everyday life. Noise caused by traffic and construction sites or by interfering talkers such as a group of toddlers on the playground impose problematic hearing situations. Speech might be also internally degraded because of age-related hearing-loss or signal distortions induced by hearing aids or cochlea implants. Besides these acoustic limitations, speech may be effortful to process because a speaker is lisping or mispronouncing words. The ubiquitous issue of listening under adverse conditions has consequently been examined by all kinds of scientific branches. Psychologists, for example, have investigated the question of whether cognitive processing capacities are deployed and whether additional attention is allocated to deal with these perceptual challenges. Linguists, furthermore, have asked what kind of speech information enables the listener to compensate for the sparse perceptual evidence and, for example, how knowledge about the semantic context can support the comprehension of upcoming words. The current thesis concerns the interface of both psychological and linguistic perspectives and aims at answering the outlined questions in a neuroscientific framework by asking about neural temporal dynamics of speech processing under adverse conditions.

Speech processing means ultimately that meaning is derived from acoustic-phonetic input that unfolds over time. A spoken word, in particular, is supposed to be processed in analogy to reading in a left to right fashion. That is, as the word unfolds in time more and more linguistic information is accumulated until the word is recognized and semantics can be mapped onto the phonological representation. Word recognition can be achieved as soon as the word becomes uniquely different from all other possible words (the so-called word recognition point; Marslen-Wilson, 1987). For example, the recognition point of *banana* occurs at the second */a/* because at this point *banana* is the only possible word candidate that remains. That means most multisyllabic words can be recognized before the complete word has been heard. The time point of word recognition can even be shifted to an earlier position within the word by embedding the word into sentence context (Miller et al., 1951; Grosjean, 1980).

In contrast, if noise is introduced to the acoustic signal, word recognition might be delayed and additional cognitive efforts are required in order to achieve semantic mapping. One problem is the increased confusion of segmental information, i.e. vowels and consonants, in noise (Phatak et al., 2008) which necessitates top-down compensatory processes like

attentional efforts (Rönnberg et al., 2013). Also, word recognition in noise can be improved when words are embedded in predictive sentence contexts (Kalikow et al., 1977).
The current thesis investigates spoken word recognition in ideal and adverse listening conditions by means of electroencephalography. In particular, it asks about the underlying neural temporal dynamics in case semantic mapping is more effortful. The focus lies on determining signatures of slow neural oscillations and thus extends current knowledge gained by analysing event-related potentials. The following sections provide an overview of current models of spoken word recognition. Then, compensatory strategies to deal with spoken word recognition in noise are outlined. Subsequently, neural oscillations and their putative role in speech processing are introduced. Finally, the general hypotheses for the current thesis are derived.

1.1 Spoken word recognition and its cognitive efforts

1.1.1 Psycholinguistic models of spoken word recognition

In spoken word recognition, the basic problem is that an auditory signal unfolding in time needs to be processed such that phonemic evidence needs to be accumulated and mapped onto a representation in long-term memory, i.e. the mental lexicon. The mental lexicon contains all known words of a language together with information about their pronunciation, semantic and syntagmatic relationships (for discussion about its organisation see for example Elman, 2004). Classical ideas about spoken word recognition assume three steps from the acoustic-phonetic analysis to arrive at semantic mapping, namely phonetic identification, lexical selection, and finally integration (Marslen-Wilson, 1987). First, lexical processes are initialized by identifying first phonemes at word onset. Second, more input is received, matching lexical entries can be pre-selected. Lexical search can be for more and more refined. Third, lexical access is accomplished by integrating lexical information and by mapping semantic information onto the phonological representation.
One of the most influential models called COHORT implements word recognition as a purely bottom-up driven process, i.e. in analogy to reading from left to right (Marslen-Wilson and Tyler, 1980; for discussion see Norris et al., 2000). Word onsets pre-activate a cohort of possible words and as the signal unfolds and more phonemic information is available, fewer entries of the cohort match until only one of them is left over. Unfortunately in this model, word recognition fails as soon as a wrong phoneme occurs (worst case already at word onset) as the cohort would be immediately empty; it does not allow any feedback loop which would inform the segmental level about lexical knowledge. This implementation is contradictory to experimental results showing that participants believe to perceive phonemes that had actually been masked (Warren, 1970; Samuel and Ressler, 1986; Sivonen et al., 2006) or that mispronunciations might stay undetected (Cole et al., 1978) because of overriding lexical knowledge.

A first attempt to account for this lack has been offered by Trace (McClelland and Elman, 1986) where the identity of a phoneme varies as a function of lexical context, forward as well as backward (for criticism see Grossberg and Kazerounian, 2011). The influence of contextual information, however, has been further developed to arrive at more precise predictions about the accuracy of spoken word recognition. Hence, there are models available which consider the beneficial effect of higher word frequency (Howes, 1954), the interaction between neighbourhood density and frequency (Goldinger et al., 1989; Cluff and Luce, 1990; Newman et al., 1997), and confusion matrices for vowels and consonants (Miller and Nicely, 1955; Ladefoged, 2005; Phatak and Allen, 2007) to appropriately weight lexical activation (for example, NAM: Luce and Pisoni, 1998, and Shortlist B: Norris and McQueen, 2008). These models are able to predict word recognition accuracy for words in ideal and adverse listening conditions (for a comprehensive review of these models see Jusezyk and Luce, 2002).

In the current thesis, lexical access is studied first by comparing real words and pseudowords. Pseudowords closely resemble real words but do not have a representation in the mental lexicon, i.e. they have no meaning. The resemblance, though, triggers some initial lexical search so that by comparing real words and pseudowords successful and failed semantic mapping can be investigated. Second, the facilitation of lexical access by preceding semantic context is studied as it reveals how strongly context and target word are associated with each other. The robustness to dissociate words and pseudowords on the one hand and the robustness to predict words from context on the other hand will be tested by introducing background noise and by degrading the spectral information of the speech signal itself. Therefore in the following section, adverse listening conditions and required cognitive mechanisms to achieve spoken word recognition in noise will be introduced.

1.1.2 Recognition of spoken word in noise

Adding noise to the speech signal increases the confusability among segmental information such as consonants and vowels (Felty, 2007; Phatak et al., 2008). Thus, in order to overcome confusability, compensatory processes are needed to enable word recognition. For example, working memory as a short-term storage with limited capacities (for a review see Awh et al., 2006) can be used for temporary compensation. A recent model by Rönnberg et al. (2013) suggests that as soon as a mismatch emerges between what can be encoded from the acoustic signal and what is represented in the listener's mental lexicon additional working memory resources are used to on-line disambiguate confusing speech signals. It has been suggested that listeners with higher working memory capacity experience less listening effort under adverse listening conditions (Pichora–Fuller and Singh, 2006; Rudner et al., 2012). This might be due to the fact that more resources can be engaged in reducing confusability and thus listening effort.

Traditionally, the capacity of working memory has been determined by the number of items (e.g., words) that can be stored. Recently, a new concept has emerged that ties the capacity also to the encoding precision of each item (Ma et al., 2014). Crucial for the current thesis, if speech is degraded, confusability is high and stimulus encoding cannot be precise. Therefore, more working memory resources are needed to increase encoding precision. The subprocess of working memory which is dedicated to the short-term storage of phonemically coded information is usually referred to as the *phonological loop* (Baddeley and Hitch, 1974; Baddeley, 2012).

Encoding of degraded speech can be improved (thus reducing working memory load) if attention is allocated in order to enhance the task-relevant signal and to suppress the task-irrelevant noise (Broadbent, 1958; for review see Driver, 2001). The top-down increase of the signal-to-noise ratio is defined as the attentional gain (Ling and Carrasco, 2006). People with higher working memory capacities have been found to also more effectively allocate attention linking the concept attention closely to working memory resources (for discussion see Awh et al., 2006). In the current thesis, attentional processes during encoding and retrieval of words will be of main interest.

The psychological frameworks of Rönnberg et al. (2013) and Baddeley (2012) constitute important bridges between psycholinguistic modelling of spoken word recognition as reported in the previous section and neuropsychological examinations that will be described in more detail in the next section. For example, attempts to find the neural basis of the phonological loop have helped to describe functions of cortical regions (Paulesu et al., 1993). And the other way round, neuropsychological advances can inform these psychological frameworks and modify their conception. The continuing search for a single underlying piece of cortex to subserve the function of the phonological loop has failed up to now so that the concept might need to be reconsidered (Buchsbaum and D'Esposito, 2008).

One assumption of the current thesis is that phonological loop and attention—both especially beneficial in adverse listening conditions—will be reflected in slow neural oscillations. Oscillatory mechanisms indicate dynamic synchronization of brain areas in certain frequency bands, thus temporarily enabling or inhibiting information processing. These assumed neural mechanisms will be laid out in the following section.

1.2 Spoken word recognition and its neural basis

In cognitive neuroscience, word recognition in the sense of meaning retrieval first had been investigated in the visual domain and by means of electroencephalography (EEG; for a detailed description of the method please see Section 2.3). The most prominent neural correlate of lexico-semantic processing had been found when participants read sentences which were completed either by congruent or incongruent words. Semantically incongruent words elicited a more negative amplitude peaking around 400 ms after word onset in comparison to congruent words (Kutas and Hillyard, 1980). This seminal study triggered

30 years of experimental work investigating the so-called N400 component (for review see Kutas and Federmeier, 2011; Van Petten and Luka, 2012).

Besides semantically incongruent sentence contexts, the N400 has also been found to be sensitive to segmental manipulations. This has been shown by using the lexical decision paradigm which is supposed to tap into lexico-semantic mapping comparable to the context manipulation. In this experimental setting, participants hear words or word-like sounds and are asked to respond whether what they just heard was a word or not. Word-like stimuli or pseudowords are words with some phonotactically legal, segmental (or phonetic) alterations. Compared to words or phonotactically illegal nonwords, pseudowords elicit larger N400 magnitudes (i.e., absolute amplitudes; e.g., Bentin, 1987; for review see Kutas and Van Petten, 1994). This is in line with the common interpretation of the N400 as a marker of neural processing effort of semantic mapping. Since pseudowords are phonotactically legal, lexical search is induced but mapping onto lexico-semantic representations in long-term memory is difficult, thus neural processing effort is increased. In sentences, however, neural processing effort is increased because the context is incongruent with the sentence-final word. This led to the view that congruent context facilitates lexico-semantic mapping and therefore reduces the N400 response. In contrast, incongruent context increases semantic integration effort and thus increases the N400 magnitude.

Although the underlying neural effort of processing phonotactically legal pseudowords compared to processing words with preceding incongruent contexts might be fundamentally different, still both are reflected in an increased N400 magnitude. Some authors would argue that the astonishing invariance in latency reflects in both cases the initial access to long-term memory independent of word recognition which happens only at a later stage (as described in Section 1.1.1; Kutas and Federmeier, 2011). Here, another perspective will be introduced that emerged only recently which tries to explain linguistic processes based on neural oscillations (Ghitza, 2011; Giraud and Poeppel, 2012; for a discussion about the relationship between neural oscillations and event-related potentials like the N400 component see Section 2.3). The importance of induced neural oscillations for cognitive functioning has been underestimated so far and has often been disregarded as the noise in the EEG signal. Hence, although the N400 appears to be consistent, neural oscillatory patterns might differ in both experimental settings and thus might reveal different involved cognitive functions.

In principle, oscillatory accounts on speech processing assume that the temporal structure of the input signal, e.g. a spoken sentence, is coupled to the frequency of the neural oscillation applied for processing this information. From the neuronal perspective, it is known that neuronal populations oscillate intrinsically at their preferred frequencies (Buzsáki and Draguhn, 2004) and because of their resonating characteristics, neurons "select" sensory input based on their preferred frequency range (Schroeder and Lakatos, 2009). This has been suggested to lead to a rhythmic sampling of linguistic information (for a review see

Ding and Simon, 2014). Sampling (also often referred to as chunking) evolves because neural oscillations reflect fluctuations in cortical excitability so that if linguistic information coincides with the excitable phase it is more thoroughly processed than if it coincides with the inhibitory phase. These ideas will be now discussed in more detail.

The correspondence between naturally occurring frequency bands in brain oscillations and the rhythms in speech has been modeled computationally, for example, by Ghitza (2011) (an earlier version of the model can be found in Ghitza and Greenberg, 2009). Including some experimental evidence, he argues that delta oscillations ($\sim$ 1 Hz) sample words or prosodic phrases whose physical duration is greater than a second, theta ($\sim$ 4 Hz) samples syllables with durations about 250 ms, beta ($\sim$ 15 Hz) samples phonemes, and gamma ($>$ 30 Hz) samples phonetic features. Poeppel (2003) suggests that sampling frequencies might be asymmetric in the left and right hemisphere of the brain (the so called *asymmetric sampling in time* (AST)-hypothesis). Although speech processing activates primary auditory cortex bilaterally, there might be different temporal integration windows in higher association areas of the left and right auditory cortices. Based on initial neurophysiological evidence, he elaborates that the left might sample rapid changes in the gamma range whereas the right integrates over longer time windows in the theta range.

One must not forget that there are also neurophysiological reasons why neuron populations would oscillate faster or slower. According to the *communication through coherence* view (Fries, 2005; Tiesinga and Sejnowski, 2010; Akam and Kullmann, 2012), phase-locked oscillations in the same frequency band indicate information exchange between affected neurons. On the one hand, the frequency range depends on the size of the neuron populations that communicate with each other: the bigger the size of the population, the slower the oscillation frequency (Buzsáki and Draguhn, 2004). On the other hand, the frequency range depends on the distance between two communicating neuronal populations: the further apart, the slower the oscillation frequency (Buzsáki and Draguhn, 2004). In the case of speech processing, both reasonings, type of speech information and neurophysiological constraints, converge because binding of phonetic features affects certainly fewer neuron populations (may be constricted to primary auditory cortex) than semantic integration of several words in a sentence.

Another multimodal approach to the chunking idea emphasizes that the auditory cortices might settle on preferred frequencies accommodated to articulation-conditioned speech rhythms. Specifically, syllables are not only characterized by rhythmic acoustic amplitude fluctuations but also by cycling mouth openings. That means that the articulatory motor system generates output that is optimal for the central auditory system to process (Giraud and Poeppel, 2012). This sets the stage for interesting evolutionary reasonings for the correspondence between brain and speech rhythms which are beyond the scope of the current thesis.

Beyond the acoustic analysis in auditory cortex, slow neural oscillations have been shown to play a role in higher cognitive functions as well. Most important for the purpose of the current thesis as outlined in the previous sections are attention and long-term (or semantic) memory when retrieving lexico-semantic information. Two frequency bands are associated with these functions, namely alpha (8–12 Hz) and theta (3–7 Hz) oscillations, respectively. In the following, both frequency bands will be characterized in detail which finally leads to the specific hypotheses of the current thesis.

1.2.1 Alpha oscillations and attention[1]

Neural oscillations in the alpha frequency range (∼10 Hz) are the most dominant signals measurable in the human magneto- and electroencephalogram (M/EEG), going back to their first description by Hans Berger (Berger, 1931). The earliest observations of the alpha rhythm revealed that its amplitude is enhanced in humans who are awake but not actively engaged in any task. This finding led initially to the view that high alpha power might simply reflect the default state of brain inactivity or "cortical idling" (for a review, see Pfurtscheller et al., 1996).

Only within the last two decades, the functional significance of alpha oscillations has been recognized and furthermore its ubiquitous role across sensory modalities (visual: for review see Mathewson et al., 2011; sensorimotor: e.g., Haegens et al., 2012; auditory: e.g., Hartmann et al., 2012) and cognitive tasks (working memory: e.g., Jensen et al., 2002; attention: for a review see Klimesch, 2012; decision making: e.g., Cohen et al., 2009). One unifying mechanism suggested for alpha rhythms across modalities and brain areas is that it provides a neural means to functionally inhibit the processing of currently task-irrelevant or task-detrimental information (Jensen and Mazaheri, 2010; Foxe and Snyder, 2011). The functional inhibition hypothesis has received neurophysiological support. For example, both alpha power (i.e., squared amplitude) and alpha phase modulate neuronal spike rate (Haegens et al., 2011) and thus can directly affect the efficiency of neural information flow. In future work beyond the scope of the current thesis, the alpha network needs to be further characterized by its phase–amplitude coupling to gamma oscillations (Jensen et al., 2012) and its role in top-down control as implemented in different cortical layers (Buffalo et al., 2011; Spaak et al., 2012) or in thalamico-cortical communication (Strauss et al., 2010; Roux et al., 2013).

Despite the abundance of studies on the role of alpha activity for visual selective inhibition, there are currently few studies that directly examine the role of alpha activity in the auditory modality. Recently, a series of studies found modulations in alpha power in a variety of auditory tasks prompted by degraded spectral detail (Obleser and Weisz, 2012), missing temporal expectations (Wilsch et al., 2014), working memory load (Obleser et al., 2012; Leiberg et al., 2006), or syntactic complexity (Meyer et al., 2013). Together, these

[1]This section is adapted from parts of the article published by Strauß, Wöstmann, and Obleser (2014). *Front Hum Neurosci 8*, 350.

findings provide good evidence that alpha oscillatory power can be a reliable indicator of auditory cognitive load (see also Luo et al., 2005; Kaiser et al., 2007).
For the current interest in spoken word recognition, alpha oscillations therefore might be for one an important neural mechanism of inhibiting task-irrelevant noise by increasing alpha power. On the other hand, lexico-semantic processing might be indexed by suppressed alpha power, that is enabled neural information flow. In sum, alpha activity might be a neural means to implement attention reflecting what is task-relevant and what is task-irrelevant.

1.2.2 Theta oscillations and semantic memory

Neural oscillations in the theta frequency range (~4 Hz) have been first described in the context of animal studies where they have been observed as the dominant rhythm of the hippocampus (Jung and Kornmüller, 1938; Green and Arduini, 1954). Up to today it is not clear how the hippocampal theta and the cortical theta rhythm observed in the human EEG are related to each other (Cantero et al., 2003; Lisman and Jensen, 2013).
However, hippocampal theta oscillations have been reliably shown to be associated with memory encoding and retrieval in animals (for review see Düzel et al., 2010; Fell and Axmacher, 2011). In humans, depth recordings suggest that theta oscillations are involved in mediating the functional coupling of medial temporal lobe and prefrontal cortex in order to subserve memory functions (for review see Johnson and Knight, 2015). For example, one potential mechanism underlying working memory might be a periodic reactivation of maintained information in theta-timed oscillatory cycles (Fuentemilla et al., 2010). For the current interest in spoken word recognition, this makes theta oscillations a putative neural means to implement the phonological loop sketched in the previous section (Roux and Uhlhaas, 2014).
Another issue is the functional overlap between information retrieval from long-term memory and from semantic memory (Ralph, 2014) suggesting to find theta oscillations in semantic manipulations as well. Indeed, the few studies that investigated slow neural oscillations in language processing found theta power to be enhanced, for example, in case semantic knowledge had been violated in a sentence context (Hagoort et al., 2004). Interestingly, theta enhancement has been found over temporal areas if words described auditory contents and over occipital areas if words described visual contents in line with the idea of sensory-specific semantic memory retrieval (Bastiaansen et al., 2008). These results suggest that theta oscillations could play an important role in both manipulations used in the current thesis, that is when lexico-semantics are more or less predictable from context and when comparing words with meaningless pseudowords.

1.3 General Hypotheses

The previous literature review introduced initial ideas about the relationship between slow neural oscillations and spoken word recognition. First, oscillations might be important for the acoustic analysis of the incoming speech signal. Oscillations might chunk speech into smaller units by temporally aligning peaks of neural excitability with the most informative acoustic cues. Hence, effects in the slightly faster alpha frequency range might be observed if vowels had been manipulated and effects in the slightly slower theta frequency range might be observed if lexical semantics had been manipulated. Second, oscillations might dynamically build neural assemblies by synchronizing in one slow frequency band depending on the task-relevant cognitive function. For example, enabled lexico-semantic processing might be reflected by reduced alpha power whereas accessing long-term memory to semantically integrate words in a sentences might be reflected by effects in the theta frequency range.

Thus, the current thesis aims at determining slow oscillatory signatures of spoken word recognition. In particular, experimental work will tackle the role of oscillations for understanding how lexico-semantic access is achieved if the auditory signal is ambiguous or degraded. To this end, different methodological approaches are applied. Emphasis will be laid on, first, the functional dissociation of alpha and theta oscillations during spoken word recognition, and second, the differential signatures of oscillatory power and phase (or phase-locking) in spoken word recognition especially in effortful listening situations. A third interest lies in the relationship between traditionally analyzed event-related potentials and the oscillatory patterns to reconsider N400 interpretations.

Because the signatures of slow neural oscillations in spoken word recognition are unclear, Chapter 3 first addresses the question how alpha and theta oscillations contribute to lexical access. This problem is approached by using the classical lexical decision task (Marslen-Wilson, 1980) comparing words and word-like pseudowords. We asked whether slow neural oscillations can dissociate lexical integration and ambiguity resolution during lexical access. In particular, we hypothesized to observe alpha power suppression reflecting enabled lexical integration for real words and to observe theta enhancement for pseudowords reflecting periodic reactivation of the word-like phonological patterns to resolve ambiguity. Oscillatory patterns will be related to commonly analyzed event-related potentials. Especially, the interpretation of the N400 as a marker of effortful lexico-semantic processing will be reassessed.

In the next steps, the auditory signal will be degraded by, first, adding white noise to the speech signal and, second, by noise-vocoding the speech signal (thus reducing its spectral content) to increase confusability and task difficulty. The motivation is twofold: On the practical level, experimental results from degraded speech provide insights that can be transferred to special populations (depending on the type of noise e.g., elderly people or cochlea implant patients). On the experimental level, degrading the speech signal allows the controlled lowering of word recognition accuracy so that within participants correlations of brain and behaviour are enabled which otherwise would be impossible due to ceiling effects. Robust versus more vulnerable neural processes can be distinguished.

As a note of caution, adding white noise to the speech signal might trigger additional processes or alter linguistic processes which might not have been induced in quiet listening conditions. This possibility is discussed as a preface to Chapter 4. The short excursion reviews oscillatory mechanisms to accomplish speech recognition in noise and develops the importance of selective inhibition to suppress irrelevant information (Driver, 2001). The comparison of speech in quiet and in noise is also an interesting case to point out functional differences between induced power and phase-locked oscillations.

Chapter 4 paves the way to the analyses of neural phase in Chapter 5. While isolating speech from noise backgrounds might be implemented on the one hand as selective inhibition of the task-irrelevant noise, it might be on the other hand implemented as enhancement of the task-relevant information, e.g. by allocating attention. In Chapter 5, we test the hypothesis whether the selection of a speech stimulus is reflected by neural phase. This has been shown only for low-level perceptual objects such that stimuli coinciding with the excitable neural phase are more likely to be perceived than when coinciding with the inhibitory neural phase (Lakatos et al., 2005; Henry and Obleser, 2012). Here, we ask whether neural phase effects are also crucial for higher cognitive functions such as spoken word recognition. To answer this question, the lexical decision task is repeated with stimuli embedded in white noise such that word recognition accuracy is reduced to be 70 % correct. Neural phase is analyzed in the alpha and theta frequencies. Results will give first insights about the generalizability of rhythmic sensory selection (Schroeder and Lakatos, 2009) and the idea of chunking linguistic information (Giraud and Poeppel, 2012).

Besides inhibitory and sensory selection, Chapter 6 finally aims to clarify the benefits of semantic context facilitating top-down mechanism to improve word recognition. On the one hand, expectations will be gradually reduced by manipulating the *cloze probability* of sentence final words (Taylor, 1953; Kalikow et al., 1977; Bloom and Fischler, 1980). On the other hand, the severity of speech degradation will be progressively enhanced in order to uncover the interaction of adverse listening conditions with semantic facilitation. Traditional event-related potentials will be extended by analyses of slow oscillations. Again, N400 interpretations will be reassessed and will again lead to a functional dissociation of induced and phase-locked oscillations.

In sum, this thesis extends current knowledge about the neural temporal dynamics of spoken word recognition by analysing not only commonly applied event-related potentials but also by looking at slow neural oscillations. The results will have important implications for clinical populations such as aphasics and cochlea implant patients and will also enhance the knowledge about the neuropsychological mechanisms during spoken word recognition.

2 General Methods

2.1 The auditory lexical decision task

Research on lexical access has used a variety of different experimental paradigms, one of the most frequent ones being the auditory lexical decison task (first usage by Marslen-Wilson, 1980). In the auditory lexical decision task, participants are asked to judge as quickly as possible whether a just heard sound was a known word or not ("Yes"/"No"; for a critique of the task's reliability see Diependaele et al., 2012). The experimental paradigm is assumed to tap into lexico-semantic processing. Therefore, it is the method of choice to tackle the current research questions.

The auditory lexical decision task has well-known advantages and pitfalls (summarized in Goldinger, 1996). It provides the possibility to contrast processes of word acceptance, non-word rejection (for a modelling approach to distinguish the two see Dufau et al., 2012), and ambiguity resolution, as studied here in Chapter 3. Also, behavioural responses are gathered on every single trial allowing data analysis with signal-detection methods (Macmillan and Creelman, 2005; see Chapter 5).

One major disadvantage of the auditory lexical decision task is the ecological validity. In everyday life one never has to decide on the lexicality of speech. One rather naturally attempts to assign meaning to what has been heard. Therefore, conclusions about the natural process of lexical access should be treated with caution. As the next paragraph describes in detail, the current design accounts for this by using word-like pseudowords allowing to map the "nonword" response partly onto the more ecologically valid concept of "mispronounced word".

Because reaction times in the lexical decision task have been found to be mainly explained by word frequency (Balota and Chumbley, 1984; Keuleers et al., 2012), in the current set of experiments word frequency has been controlled in order to focus on effects of lexical semantics.

When interpreting lexical decision data, it needs to be considered that performing a decision task presumably affects perceptual processes. Two opposing modelling approaches of the relation between word recognition and lexical decision task are available: One model assumes that lexical processing occurs first, so that in a second step decisions can be made depending on whether word recognition was successful or not (Ratcliff et al., 2004).

According to the other model, perceptual and decisional processes might be completely integrated (Norris, 2009). Interestingly, both approaches reach the same accuracy in predicting reaction times. Thus, the question about the actual relationship between lexical processing and decisional processes still needs to be answered by neurolinguistic endeavours. Implications of the current data concerning this matter will be discussed in Chapter 3 and in Chapter 5.

Stimulus material: words and pseudowords. Stimuli used in Chapters 3 to 5 were adapted from a previous study by Raettig and Kotz (2008) and refined as described below to match requirements for EEG studies and to fit the purpose of the current research questions. From 60 tri-syllabic, concrete German nouns (e.g., /banane/, engl. [banana]; condition is labeled as 'real') two types of pseudowords were derived.
First, 'ambiguous' pseudowords were derived by manipulating the core vowel of the second syllable (e.g., /banene/). Therefore, vowels of the real word conditions were exchanged amongst each other as far as possible, simultaneously considering i) equal exchange probability (i.e., replace /a/ as often with /e/ as with any other vowel), and ii) considering that the cohort is empty at the onset of the manipulated vowel (see Section 1.1.1). By keeping the third syllable intact, the original word remains the only neighbouring real word (e.g., /banane/ is the only real word neighbour of /banene/).
Second, 'opaque' pseudowords were derived by scrambling the syllables across words while keeping the position-in-word fixed (e.g., /ba·poss·ner/ consists of the first syllable of /ba·na·na/, the second syllable of /a·pos·tel/, engl. [apostle], and the third syllable of /rab·bi·ner/, engl. [rabbi]). This way, the overall stress patterns and vowel qualities could be retained (except for 7 items) so that, for example, reduced vowels in third syllables would not be changed. Furthermore, 60 abstract tri-syllabic real words (e.g., /botanik/) were used as fillers to ensure a balanced word-pseudoword ratio. The complete set of words and pseudowords can be found in the Appendix 7.6.

Psycholinguistic considerations. In contrast to Raettig and Kotz (2008), ambiguous pseudowords were manipulated only on the second (not the third) syllable to ensure precise timing necessary in EEG and to experimentally dissect the seemingly sequential process of spoken word recognition according to the following rationale: Auditory word recognition depends heavily on the beginning of the word, that is, the initial syllable (Taft and Forster, 1976; Marslen-Wilson and Zwitserlood, 1989). As studies using word fragment (i.e., first syllable) priming showed (e.g., (Marslen-Wilson and Zwitserlood, 1989; Friedrich et al., 2009; Scharinger and Felder, 2011)), a cohort becomes pre-activated and lexical candidates are isolated. By choosing tri-syllabic German nouns, we were able to use the first syllable to build up initial lexical context, which is identical in all conditions (e.g., /ba·/ would amongst other pre-activate /ba·nane/, engl. [banana]).

At the second syllable, we then introduced some variations by either following a potentially pre-activated real word trace (e.g., /ba·na·/), by exchanging the core vowel (e.g., /ba·ne·/), or by replacing the entire second syllable with a random one (e.g., /ba·poss·/). That way, we perturbed further cohort activation (Taft and Hambly, 1986), i.e. the linear accumulation of lexical evidence towards word identification, in two degrees of severity.
The third and final syllable, however, completed either a clear pseudoword by continuing the wordness violation (e.g., /ba·poss·ner/) or it created an ambiguous case by continuing the initially expected word despite the local manipulation on the second syllable (e.g., /ba·ne·ne/). Thus, word identification should be perturbed but remain possible. By presenting an ending commensurate with the cohort prediction pre-activated at the first syllable, we hypothesized to observe two valid neural strategies of the listener (see Chapter 3). One strategy would be to ignore the local prediction error prompted by the second syllable and to emphasize the global congruence (first and third syllable, as well as suprasegmental features such as prosody) with the overall most likely lexical candidate—comparable to the perception of a slight mispronunciation. The second strategy would be to resolve the ambiguity prompted by the lexical decision task in order to accomplish the task accurately.
In Chapters 4 and 5, we explore lexical decisions in noise. Compensatory processes should change because the manipulated vowel is more easily confused with the original vowel so that word recognition accuracy depends on the successful increase of the signal-to-noise ratio by attention allocation. Also, performance in noise might particularly depend on lexical stress patterns, as stress has been found to be preferably used for segmenting speech in noise (Mattys, 2004).

Comparison of vowels. In Chapter 5, real words and ambiguous pseudowords are compared. These conditions differ by the core vowel of the second syllable only. Vowel identification depends primarily on formant information. Therefore, a post-hoc comparison between formants of 'real' and 'ambiguous' vowels is conducted as summarized in Figure 2.1 in order to reveal the occurance of any systematic vowel shifts. The Euclidian distance between vowels over three formants is calculated as follows:

$$\Delta f = \sqrt{\sum\nolimits_{f=1}^{3} (real_f - am_f)^2}$$

where $real_f$ denotes formants of the real word vowel and am_f the ones of the ambiguous counterpart. Formant distances varied between 58.8 Hz and 1995.1 Hz but two thirds of the vowel pair distances were below 1000 Hz (Fig. 2.1B). Greater distance leads to less confusability (Felty, 2007).
Systematic vowel shifts were assessed by using a bootstrapping procedure because formants and formant distances were not normally distributed (see Fig. 2.1B). First and second formants are the most informative dimensions for vowel identification (Ladefoged,

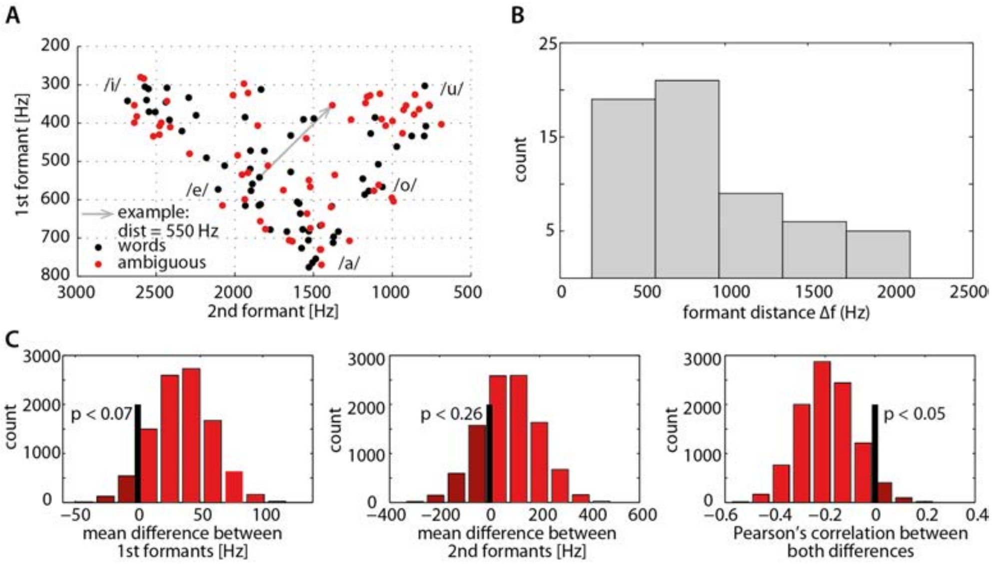

Figure 2.1: Features of the word and pseudword corpus. A. Vowel space of the second-syllable vowels. Black dots mark real word formants and red dots the ambiguous pseudoword formants. The grey arrow depicts an example shift from /elefant/ to /elufant/. **B.** Histogram of Euclidian distances of vowels. $\frac{2}{3}$ of all distances, e.g. /e/–/u/, are below 1 kHz. **C.** Bootstrapping of formant distances between real and ambiguous vowels and their correlation. Ideally, the difference between conditions should be zero which is marked by the thick black line.

2005) which is why this analysis focuses on those two. Sixty differences between real and ambiguous vowel formants were randomly drawn 10.000 times with replacement to generate distributions in Figure 2.1C. Positive values indicate higher formants for real words and negative values higher formants for the ambiguous counterpart. The correlation between formant distances were calculated 10.000 times as well.

Unfortunately, some systematicity in the vowel shifts were disclosed. There is a tendency of downward shifting the first formant from real to ambiguous stimuli (Fig. 2.1C left panel; $p < 0.07$). No significant shift of the second formant between conditions occurred (Fig. 2.1C middle panel; $p < 0.26$). Differences of first and second formants are negatively correlated (Fig. 2.1C right panel; $\rho_{mean} = -0.19, p < 0.05$) indicating that the larger the first formant distance the smaller the second formant distance. This result was driven by the /u/-sounds in the ambiguous condition which were less variable and clustered at lower first formant frequencies.

Although vowel shifts cannot explain the results reported in the following experimental Chapters, future studies might want to control for these in order to reduce performance variability introduced by varying trial-to-trial difficulty in discriminating words and pseudowords.

2.2 Adaptive tracking procedures

Word-pseudoword confusion, and vowel confusion in particular, can be enhanced when presenting speech in noise (Felty, 2007). The increase in perceptual uncertainty allows us to study compensatory neural mechanisms of effortful listening on the one hand and neural signatures of successful versus failing lexical access on the other hand. In order to account for the large inter-subject variability in hearing, psychoacoustic measures are inevitable. Adaptive procedures are handy because an individual signal-to-noise ratio (SNR) can be estimated without time-consumingly collecting data to model a listener's entire psychometric function.

Psychometric functions describe the relationship between performance increase and stimulus level increase as a sigmoid function (Macmillan and Creelman, 2005; see Fig. 2.2A). The point of subjective equality corresponds to the stimulus level at which both response options (e.g., in lexical decisions "Yes"- and "No"-responses) are equally probable and thus accuracy is about 50 % correct. In the current thesis, individual thresholds to discriminate word and pseudoword vowels were of interest at which participants performed about 70.7 % correct. This allows to analyse the data with signal detection methods because a sufficient amount of incorrect trials will be gathered. At the same time, subjective experiences of listening frustration is kept limited. Most importantly, though, compensatory neural mechanisms are best observed at an intermediate level of difficulty. That is at performance levels above chance, i.e. above 50 % accuracy. 70.7 % accuracy is yielded by a "two-down-one-up" staircase procedure (Levitt, 1971). The procedure is adaptive such that the correctness of the response in one trial determines the SNR of the subsequent one. Here, 70.7 % (but not 50 %) accuracy were targeted which necessitates the following algorithm: If the responses of two subsequent trials were correct, the SNR of the next

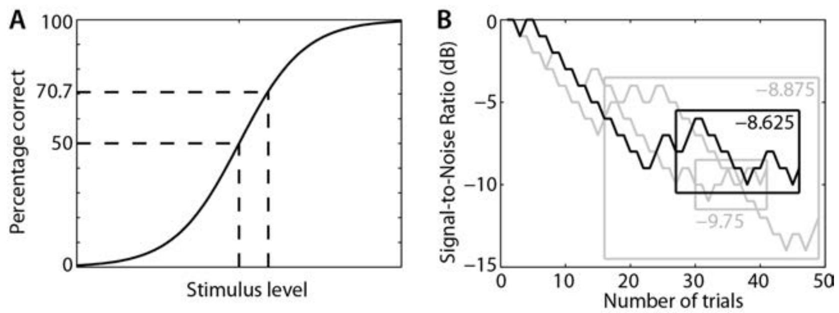

Figure 2.2: A. Psychometric function. Dashed lines exemplify how the one-down-one-up staircase procedure would sample the stimulus level to reach 50 % accuracy and how the two-down-one-up staircase procedure would sample the threshold for 70.7 % correct performance. **B. Example of the two-down-one-up staircase procedure.** Three threshold estimates of one subject (from Chapter 5) are shown. Boxes frame the last 8 reversals, respectively, which were averaged to determine the SNR threshold.

trial decreases so that intelligibility gets worse. As soon as one response is incorrect, the SNR increases so that the next trial is easier. Figure 2.2B illustrates three adaptive tracking procedures to estimate the threshold for one participant from the dataset reported in Chapter 5.

According to Levitt (1971), three parameters are of interest to successfully use adaptive procedures: the initial SNR of the first trial, the step size between two trials and the amount of trials selected to calculating final empirical threshold estimate. First, the initial SNR needs to be set without prior knowledge. If the initial SNR is too far or too close to the assumed empirical threshold, the adaptive procedure becomes inefficient. Second, greater step sizes between two trials are efficient in the beginning of the adaptive procedure to rapidly advance to the empirical threshold. In vicinity of the threshold, smaller step sizes allow a more precise sampling. Third, in order to estimate a reliable threshold, only trials at later stages during the adaptive procedure should be considered, i.e. after several reversals, which closely fluctuate around the empirical threshold. Reversals are defined as the turning points whenever correct responses change to incorrect responses (or vice versa). Here, the duration of the adaptive tracking procedure was set to 12 reversals, but empirical thresholds were determined by averaging across the last eight reversals only, thus discarding the first four reversals.

In order to match the adaptive tracking procedure with the auditory lexical decision task as closely as possible in order to yield transferable thresholds, participants performed a discrimination task (instead of a frequently used detection task) during the tracking. To this end, the second syllables were extracted from the real words and their ambiguous-pseudoword counterparts, including the critical vowel manipulation. Syllable pairs were placed successively (the second syllable starts 500 ms after the onset of the first syllable) in a stream of white noise. Syllable pairs might consist of either two times the real-word syllable or the real-word syllable followed by its ambiguous counterpart. Participants had to indicate whether the second vowel in each pair was the "same" or "different" from the first one. The sound pressure level (SPL) of the white noise was fixed and the SPL of the syllables adapted across trials.

Because of the variable formant distances (see Fig. 2.1B), some single trials might be already more difficult at relatively high SNR than others. Therefore, threshold estimation was repeated three times (as depicted in Fig. 2.2B). These three thresholds were averaged to set the final SNR for the subsequent auditory lexical decision task.

2.3 Electroencephalography

In the current thesis, the electroencephalogram (EEG) recorded from the human scalp was used in all experiments to study auditory word recognition. The excellent temporal resolution of EEG in the range of milliseconds (Speckmann and Elger, 2005) is the decisive advantage over functional magnetic resonance imaging (fMRI) where the temporal resolution is several seconds. Compared to the magnetencephalogram (MEG) in turn, EEG is easier for clinical application in terms of execution and costs. In the following section, the neurophysiological basis of EEG is described followed by the basics of EEG data preprocessing and analysis. Beginning with the traditional approach of studying event-related potentials (ERPs) during lexico-semantic processes, the computationally more advanced procedure of time–frequency power and phase analysis is outlined and source localization techniques are introduced.

2.3.1 The neurophysiological basis of EEG

Recording the EEG from the human scalp quite directly measures neural activity. Voltage fluctuations on the scalp are thought to be a consequence of post-synaptic potentials of cortical pyramidal neurons. Complementary to MEG, EEG captures mostly radially oriented cortical neuron populations but less so the tangentially oriented ones (Lutzenberger et al., 1985). Excitatory post-synaptic potentials at the apical dendrites, for example, generate electronegativity such that a current flows from the nonexcited and electropositive cell soma to the dendrites (Pizzagalli, 2007). Synchronized firing of neuronal populations can reflect corticocortical or thalamocortical information exchange (discussed below).
The greatest limitation of EEG is its spatial resolution. First, because of volume conduction between electrical sources and scalp electrodes, the neuronal signal is captured by neighbouring channels. Second, electrodes measure the sum of huge cortical cell assemblies not only in terms of spread but also in terms of depth, i.e. thalamic sources. But time–frequency analysis of EEG data allows to some extend (and arguably in a more sophisticated manner than ERPs) analogy inference between results from intracranial recordings, for example, from electrocorticography in humans or single-cell recordings in animal studies.

2.3.2 Preprocessing and artefact rejection

Because of the high sensitivity to electric activity, EEG is prone to artefacts arising from line voltage as well as any muscular activity among which are most prominently eye movements or heart beat (for an overview, see Blume et al., 2002). Several automatized techniques are at hand for artefact detection and removal. Commonly, frequency-based filters are applied first, like high-pass filters to even out slow frequency drifts or notch filters to eliminate the line voltage-specific frequency. Also, artefacts can be identified by their characteristic topographical distributions such as the bipolar frontal activity indicating eye movements (Debener et al., 2010). The independent component analysis (ICA; Jung et al., 2000), for example, is a technique of blind source separation which relies on spatio-temporal characteristics of the signal, and is thus useful to extract (and maybe reject) statistically independent source signals from the mixture present in raw EEG recordings. After the identification of independent components and their selective rejection, remaining components will be backprojected to re-gain the now artefact-free signal mixture at electrodes. The greatest advantage of this procedure is the recovery of trials that otherwise would have been needed to be rejected completely. This is highly relevant in valuable patient data which are often noisier than data from young healthy adults but also benefits the present experiments. This is because analyses of neural phase are heavily dependent on number of trials and in particular the bifurcation index used here becomes more reliable with more trials as will be shown in Section 5.6.

2.3.3 Event-related potentials: advantages and limitations

Studying linguistic processes in the brain has traditionally focussed on event-related potentials (ERP) defined as an average over multiple trials time-locked to stimulus onset (Picton et al., 2000; Luck, 2005). Averaging has been thought to diminish the "noise", i.e. ongoing oscillations, in the EEG signal. Indeed, averaging enhances phase-locked responses and therefore mainly reflects synchronized post-synaptic potentials (see Fig. 2.3 left). However as elaborated below, additional information can be yielded by means of time–frequency analysis distinguishing amplitude and phase in different frequency bands.

2.3.4 Time–frequency analysis: evoked versus induced activity

Time–frequency analysis decomposes the EEG signal into different frequency bands allowing the frequency-specific investigation of amplitude effects detached (to a certain extend) from phase influences (Makeig et al., 2004). As exemplified in Figure 2.3 (right column), single trials are Fourier transformed and averaged per frequency yielding a time–frequency representation that contains not only evoked but also induced oscillations (Tallon-Baudry and Bertrand, 1999). In Fig. 2.3, the red blob in the left column, i.e. the evoked activity, is also represented in the right column amongst other (induced) activity. Technically, ERPs can be understood as a mixture of evoked, induced, and instantaneous oscillations although the definite relationship between ERPs and oscillations is still a matter of debate (Mazaheri and Jensen, 2006; Min et al., 2007; Hanslmayr et al., 2007; Klimesch et al., 2007). Evoked activity, that is what is emphasized by ERPs, is phase-locked to the stimulus onset and consistent across stimulus repetitions. Induced activity, in contrast, is not strictly phase- but time-locked and thus correlates with the experimental condition. Instantaneous (sometimes also called spontaneous) activity is uncorrelated with the stimulation (Herrmann et al., 2005).

Fourier transformation by using wavelets. Different methods to achieve the frequency decomposition of EEG signals are available. In the current thesis, Morlet wavelets have

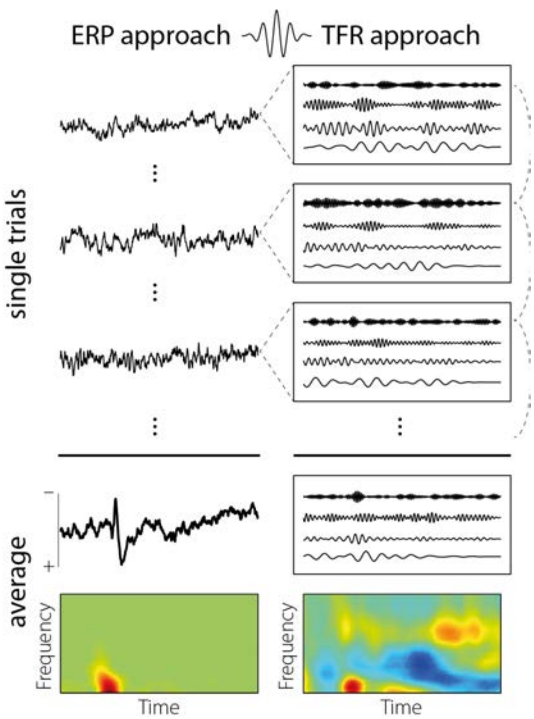

Figure 2.3: Schematic extraction of event-related potentials (ERPs) and time–frequency representations. Left column: single trials recorded in EEG after preprocessing. Their average under the thick line represents the ERP. Below the ERP, its representation in time–frequency space shows high synchronization in lower frequency bands right after stimulus onset. Right column: Fourier transform of every single trial. Fourier transformation is achieved by using Morlet wavelets, an example wavelet is depicted between column titles. If averaging is done over each frequency separately, a mixture of evoked and induced oscillations is gained. See text for details.

been used for the Fourier transformation (Tallon-Baudry et al., 1997). They are complex functions consisting of Gaussian shaped sinusoidal oscillations (an example is schematically depicted between column titles in Fig. 2.3). The real part of the wavelet function represents a sinusoidal oscillation within a certain frequency band and the imaginary part yields, like a Hilbert transform, a 90° phase-shifted signal (Herrmann et al., 2005). The wavelet function is sliding across the EEG signal and convolved with it at each time point ($\pm$ the window width). This way, sinusoidal EEG activity is detected. The number of sinusoidal cycles in the wavelet function should be frequency-specific since one cycle in lower frequencies will cover longer time-windows than in higher frequencies. If the number of cycles is kept constant across frequencies, time-resolution will be worse in lower than in higher frequencies. At the same time, including more time points in lower frequencies leads to higher frequency resolution than in higher frequencies. To account for this trade-off, fewer cycles should be considered for transformation of lower than of higher frequency components.

The resulting complex Fourier transform $F(f,t)$ at frequency f and time t has the form $F(f,t) = x+yi$, where x represents the real part and y the imaginary part (see Fig. 2.4A). Magnitude (or amplitude in EEG data), also refered to as complex modulus or absolute value of the complex number, is thus defined as $|F(f,t)| = \sqrt{x^2+y^2}$ according to the rules of Pythagoras. Power, in turn, which is more often analyzed in neural oscillation literature (see also Chapter 3 and 4, and Section 6.5), is defined as the squared magnitude: $Power(f,t) = |F(f,t)|^2$. The angle ϕ is implicitely given by complex numbers and can be derived by calculating $\phi = arctan\frac{y}{x}$. Finally, inter-trial phase coherence (ITPC; used in Chapter 5 and Section 6.5) is defined as

$$ITPC(f,t) = \left| \frac{1}{N} \sum_{n=1}^{N} \frac{F_n(f,t)}{|F_n(f,t)|} \right|$$

(Lachaux et al., 1999, 2002)

where N is the total number of trials and $F_n(f,t)$ is the Fourier transform of the nth trial. Essentially in the formula, Fourier data is normalized to unit length via division by its magnitude. The sum of normalized Fourier data is then divided by the number of trials. Thus, the absolute part of this complex mean corresponds to the resultant vector length across trials (Berens, 2009).

Although mathematically clearly distinct measures, power and phase may not be independent in EEG data. In empirical data, power may be seen as the envelope of the EEG signal in a specific frequency band and phase as fast fluctuating power. For instance, a sinusoidal oscillation as depicted in Figure 2.4B and C shows with progressing phase from 0° to 90° (or 0 to $\pi/2$ rad) simultaneously a change in magnitude (or amplitude for that matter) indicated by the arrows. Although the absolute magnitude is independent of phase, high magnitude across trials and high inter-trial phase-locking often accompany each other. It has been thus argued that phase effects might just be a more sensitive measure of stimulus-locked activity and power and phase should not be seen as complementary measures (Ding and Simon, 2013).

2.3.5 Source localization

When aiming at estimating underlying sources from M/EEG data, the main problem is that there are infinite source solutions to an EEG scalp topography, which is refered to as the inverse problem (Helmholtz, 1853). In order to localize sources as in Chapter 3, a forward solution and the inverse solution needs to be computed.

The forward model calculates for each source grid point the resulting scalp topography considering volume conduction. Therefore first, a source model is needed which could

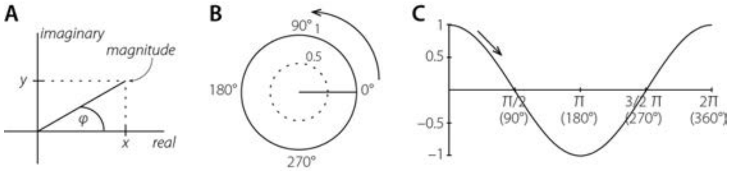

Figure 2.4: **A.** Complex numbers. **B.** Unit cycle. **C.** Cosine function.

be a standard template MRI if, as in the case of the current thesis, no individual MRI is available. Second, a head model extrcated from the MRI anatomical scan is needed, that is a realistically shaped three-layer boundary elements model (BEM) of the brain (Oostenveld et al., 2003) containing information about skin, skull and brain surface. MRI and individual EEG electrode locations need to be co-registered. Then, the so-called lead field can be calculated for each source grid point (in the current thesis with a 1 cm resolution). The lead field contains the forward solution for each source grid point, that means information about how each source grid point is projected onto the surface, i.e. onto each electrode.

After calculating the unique solution of the forward model, the inverse model is estimated. Here, a beamforming technique using DICS (Dynamic Imaging of Coherent Sources; Gross et al., 2001) was applied which estimates sources in the frequency domain, in contrast to other beamforming approaches that estimate sources in the time domain such as the Linearly Constrained Minimum Variance (LCMV; Van Veen et al., 1997). First, the EEG signal at each electrode is Fourier transformed by using multitaper based on discrete prolate spheroidal sequences (DPSS, also Slepian sequences; Slepian, 1978; Mitra and Pesaran, 1999; Jarvis and Mitra, 2001). Multiple tapers improve the spectral precision of power estimates. Second, the cross-spectral density matrix (CSD) is calculated by the cross-correlation of two complex signals (Welch, 1967). In DICS, the two submitted signals are all possible electrode combinations (Gross et al., 2001). Because time information is lost in the CSD, only time windows and frequencies of interest are considered. Again, there is a trade-off between longer time windows allowing better frequency resolution and wider frequency ranges improving time resolution. For example when aiming at beamforming 10 Hz-alpha power, a 700 ms time window subsumes 7 alpha cycles and allows a frequency resolution of $1/0.7$ s $= 1.4$ Hz. The more time points and frequency bins are given the better the final source estimation will be.

The last step uses the forward model, i.e. the individual lead fields, and the CSD to compute an adaptive spatial filter for each source grid point. If a common cross-spectral density for baseline and conditions (in a time and frequency window of interest) had been calculated, single-trials of each conditions can be projected by using this so-called common filter into source space. Thus, single-trial power and subsequent statistical contrasts can be estimated.

3 Alpha and theta power dissociate in spoken word recognition[2]

3.1 Introduction

Accumulating evidence shows that speech comprehension is more completely described by not only looking at evoked but also induced components of the electrophysiological brain response (Giraud and Poeppel, 2012). Besides research concerning the phase (for review see Peelle and Davis, 2012), also power changes of transient slow oscillations have been found to determine language processes (Hald et al., 2006; Bastiaansen et al., 2008; Obleser et al., 2012; Meyer et al., 2013). A functional differentiation between alpha (~10 Hz) and theta oscillations (~4 Hz), even though previously put forward (Klimesch, 1999; Roux and Uhlhaas, 2014; for current debate in audition see Weisz et al., 2011), remains to be shown for speech processing (e.g. an open issue in Obleser et al., 2012; Tavabi et al., 2011).
Generally, alpha oscillations are the predominant rhythm in ongoing neuronal communication and therefore observable in diverse cognitive functions such as auditory processing (sometimes labeled "tau"; Lehtelä et al., 1997; Tavabi et al., 2011; Hartmann et al., 2012), attention (Klimesch, 2012), working memory (e.g., Meyer et al., 2013; Obleser and Weisz, 2012; Wilsch et al., 2014), or decision making (Cohen et al., 2009). A tentative theoretical account on the role of alpha oscillatory activity has only been put forward recently (Jensen and Mazaheri, 2010; Klimesch et al., 2007; Klimesch, 2012): Functional inhibition. In fact, most of the above-cited data are compatible with increased needs for inhibition of concurrent, task-irrelevant, or task-detrimental neural activity. Direct evidence for alpha-mediated inhibition of local neural activity, as expressed in spiking (Haegens et al., 2011) or gamma-band activity (Roux et al., 2013; Spaak et al., 2012), has been provided.
First evidence has shown that greater alpha suppression post-stimulus is associated with more effective language processing: alpha oscillations in response to single words were found to be suppressed as a function of intelligibility of acoustically degraded words (Obleser et al., 2012). This is in line with the inhibitional account meaning that alpha power remains high when the language processing network is inhibited, the crucial mechanism for the present study.

[2]This chapter is adapted from the published article by Strauß, Kotz, Scharinger, and Obleser (2014). *NeuroImage 97*, 387-395.

In contrast to functional inhibition across a range of general cognitive functions plausibly associated with alpha, theta oscillations in human EEG have been related more consistently to episodic memory (e.g., Hanslmayr et al., 2009), sequencing of memory content (e.g., Lisman and Jensen, 2013; Roux and Uhlhaas, 2014), and matching of new information with memory content (e.g., Klimesch, 1999). Moreover, neural periodic reactivation of information held in human short-term memory has been directly related to theta-timed oscillatory cycles (Fuentemilla et al., 2010). Such "replay"of sensory evidence in order to arrive at accurate lexical decisions might be decisive in the present design, especially when input is somewhat ambiguous as outlined below.

Interestingly, theta power enhancement has been observed in a series of language- or speech-specific effects. For example, semantic violations more than world knowledge violations drive theta enhancement during sentence processing (Hagoort et al., 2004; Hald et al., 2006); also, the retrieval of lexico-semantic information (Bastiaansen et al., 2008) and the increasing intelligibility of acoustically degraded words (Obleser et al., 2012) lead to theta enhancement. In the latter study, the alpha suppression reported above was directly proportional to theta enhancement. These results tie theta enhancements in language paradigms to the neural re-analysis of difficult-to-interpret stimulus materials.

In the present study, we want to dissociate neural oscillatory dynamics in the alpha and theta frequency bands in order to link them to segregable functions in spoken word recognition. As a control, however, we also extracted event-related potentials (ERPs) because its N400 component in particular has proven to be a robust index of 'wordness' (Chwilla et al., 1995; Desroches et al., 2009; Friedrich et al., 2009; Laszlo et al., 2012; for review see Friederici, 1997; Van Petten and Luka, 2012). Larger N400 amplitudes, elicited by unexpected (Kutas and Hillyard, 1980; Connolly and Phillips, 1994; Strauß et al., 2013), infrequent words (Rugg, 1990; Van Petten and Kutas, 1990; Dufour et al., 2013), or pseudowords (Friedrich et al., 2006), compared to high-probable or high-frequent real words, have mostly been associated with increased neural processing effort in matching the input signal to items in the mental lexicon. We aim at elucidating this matching process by investigating alpha and theta activity that are framed in terms of inhibition and replay.

We designed an auditory lexical decision task where a word–pseudoword continuum would induce a stepwise reduction in lexical accessibility ('wordness'). Additionally, ambiguous stimuli would evoke a task-dependent conflict (task: "Is it a word?" (yes/no)) and call for re-evaluation of the auditory input. First, we hypothesize that a 'wordness' effect should be observable in the alpha band, with less alpha power when auditory input approximates real words held in the mental lexicon. This effect should be prominent in brain areas associated with lexical processes (e.g., left middle temporal gyrus; Kotz et al., 2002; Minicucci et al., 2013) and would characterize alpha as a signature of enabling lexical integration. Second, we hypothesize that the power of theta oscillations with their ascribed functionality in memory and lexico-semantics would vary with the need for resolving ambiguity.

Altogether, our focus on dissociable slow neural oscillations and their corresponding functional roles during spoken word recognition allows us to contribute to long-standing debates on whether recognition is best conceived as serial, feed-forward mechanisms (Norris et al., 2000) or as parallel, interacting processes (McClelland and Elman, 1986; Marslen-Wilson, 1987). Importantly, time–frequency analyses of on-going EEG activity are ideally suited to extract potentially parallel cognitive processes.

3.2 Methods

3.2.1 Participants

Twenty participants (10 female, 10 male; 25.6 ±2.0 years, M±SD) took part in an auditory electroencephalography (EEG) experiment. All of them were native speakers of German, right-handed, with normal hearing abilities, and reported no history of neurological or language-related problems. They gave their informed consent and received financial compensation for their participation. All procedures were approved of by the ethics committee of the University of Leipzig.

3.2.2 Stimuli

Adapted from Raettig and Kotz (2008), stimuli were 60 three-syllabic, concrete German nouns (termed 'real', e.g., 'Banane' [banana]). For the 'ambiguous' condition, we exchanged the core vowel of the second syllable (e.g., 'Banene'). Finally for the 'pseudoword' condition, we scrambled syllables across words (concrete and abstract, see below), while keeping their position-in-word fixed (e.g., 'Bapossner'). Note that there was a fourth condition with 60 three-syllabic, abstract German nouns not relevant for the current analyses which was necessary to maintain an equal ratio of words and pseudowords. These were considered as fillers and not analyzed further. Previous studies used word-like stimuli in order to investigate lexicality effects on phoneme discrimination (Connine and Clifton, 1987; Frauenfelder et al., 1990; Wurm and Samuel, 1997). An important difference to these studies is that we created a distribution of formant distances between real word vowels and their pseudoword equivalents. For illustration purposes, these difference can be quantified by calculating the Euclidian distance of the first three formants for each vowel pair (Obleser et al., 2003): Distances ranged from 200 Hz (/ɛ/ → /ɪ/, Geselle → Gesille) to 2100 Hz (/oː/ → /iː/, Kommode → Kommide). The majority (approximately one third) of vowel pairs were 600 to 1000 Hz apart from each other (/ə/ → /ɔ/, Batterie → Battorie). Therefore, exchanging a vowel here means that stimuli were lexically but not phonetically ambiguous which calls for ambiguity resolution processes on a decisional rather than a perceptual level (for discussion see Norris et al., 2000). However, we show with this acoustic analysis that lexical ambiguity necessarily corresponds to variance in acoustic input.

Importantly, we controlled for equal ratio of stress patterns across conditions, because in unstressed syllables formant distance decreases, which raises perceptual confusions and task difficulty. The substitution of the vowel marked the deviation point to any existing German word but at the same time did not violate German phonotactic rules. The same holds true for clear pseudowords even though deviation points were not as exactly timed as in the ambiguous condition and alternated between the first and second phoneme of the second syllable. Please note that ambiguous stimuli had only one real word neighbor whereas clear pseudowords might have evoked several real word associations.
All words and pseudowords were spoken by a trained female speaker and digitized at 44.1 kHz. Post-editing included down-sampling to 22.05 kHz, cutting at zero crossings closest to articulation on- and offsets, and RMS normalization. In sum, the experimental corpus consisted of 240 stimuli with a mean length of 754.2 ms $\pm$ 83.5 ms (M$\pm$SD).

3.2.3 Experimental procedure

In an electrically shielded and sound-proof EEG cabin, participants were instructed to listen carefully to the words or word-like stimuli and to perform a lexical decision task.
Figure 3.1A shows the detailed trial timing. After each stimulus, a delayed prompt indicated that a response should be given via button press ("Yes"/"No") to answer whether or not a German word had been heard. The response delay was introduced in order to gain longer trial periods free of exogenous components (due to the visual prompt) or arte-facts (i.e., button press), which are required for a clean time–frequency estimation and source localisation of oscillatory activity. The button assignment (left/right) was counterbalanced across participants such that 10 participants used their left and the other 10 their right index finger for the 'Yes' response. Accuracy scores (percentage correct) and reaction times were acquired. Subsequently, in order to better control for eye-related EEG activity, an eye symbol marked the time period during which participants could blink. Duration of blink break and onset of the next stimulus were jittered to avoid a contingent negative variation. Prior to the experiment there was a short familiarization phase. It consisted of 10 trials taken from Raettig and Kotz (2008) which had similar manipulations but were not used in the present experiment. Then each participant listened to all 240 stimuli. Listeners paused at their own discretion after blocks of 60 trials. The overall duration of the experimental procedure was about 30 min.
Each participant obtained an individually pseudo-randomized stimulus sequence. Note that the order of occurrence for a given ambiguous pseudoword (e.g., 'Banene') and its real word complement (e.g., 'Banane') was counterbalanced across participants in order to control for facilitated word recognition due to ordering effects. As a constraint to pseudo-randomization, their sequential distance was kept maximal (i.e., ~120 other items in between).

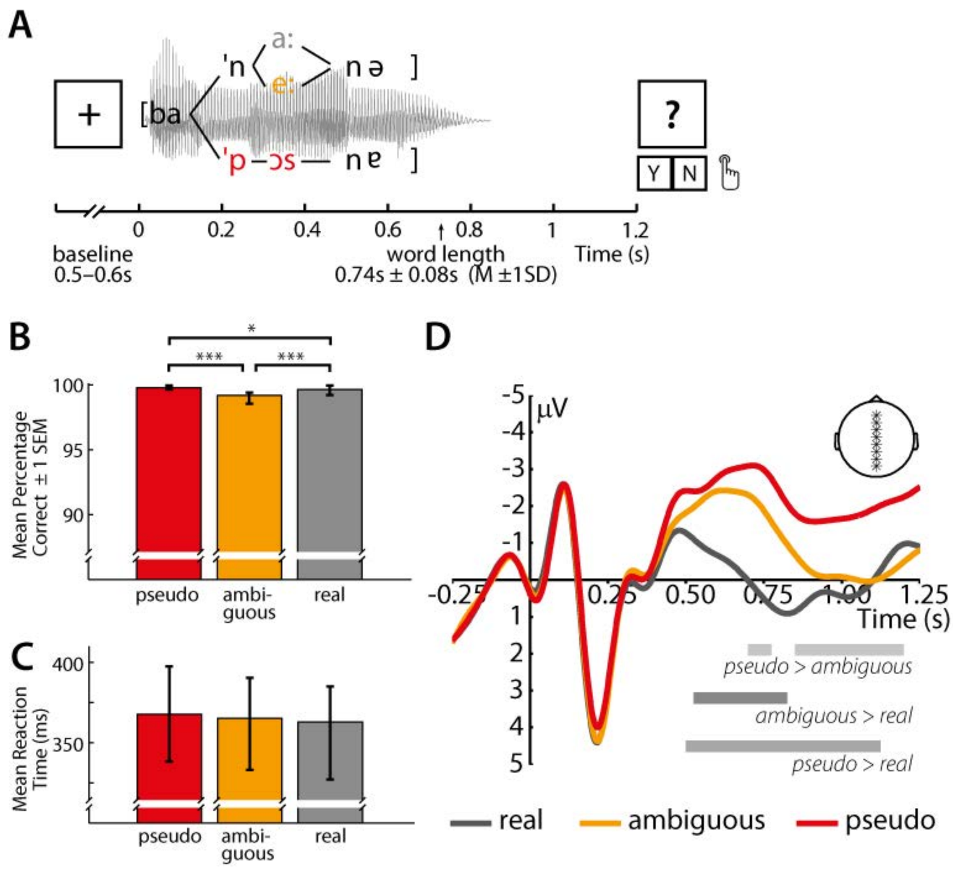

Figure 3.1: Study design and behavioural measures. **A.** Stimulus design and schematic time course of one trial. Stimuli were tri-syllabic German nouns ('real'), 'ambiguous' pseudowords (one vowel exchanged), and clear 'pseudowords' (scrambled syllables across items). **B.** Mean percentage correct $\pm$ 1 SEM (between-subjects standard error of the mean). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ **C.** Mean reaction times relative to the prompt, $\pm$ 1 SEM. **D.** Grandaverage of ERPs over midline electrodes. Grey shaded bars indicate statistical differences.

3.2.4 Electroencephalogram acquisition

The electroencephalogram (EEG) was recorded from 64 Ag–AgCl electrodes positioned according to the extended 10–20 standard system on an elastic cap with a ground electrode mounted on the sternum (Oostenveld and Praamstra, 2001). The electrooculogram (EOG) was acquired at a horizontal (left and right eye corner) and a vertical (above and below left eye) line. All impedances were kept below 5 kΩ. Signals were referenced against the left mastoid and digitized online with a sampling rate of 500 Hz, with a pass-band of DC to 140 Hz. Individual electrode positions were determined after EEG recording with the Polhemus FASTRAK electromagnetic motion tracker (Polhemus, Colchester, VT, USA) for more precise source reconstructions.

3.2.5 Data analysis: event-related potentials

Data pre-processing and analysis was done offline by using the open source Fieldtrip toolbox for Matlab^{TM}, which is developed at the F.C. Donders Centre for Cognitive Neuroimaging in Nijmengen, Netherlands (Oostenveld et al., 2011). Data were re-referenced to linked mastoids and band-pass filtered from 0.1 Hz to 100 Hz. To reject systematic artefacts, independent component analysis was applied and components were rejected according to the 'bad component' definition by Debener et al. (2010). Remaining artefacts were removed when the EOG channels exceeded ± 60 μV for frequencies between 0.3 and 30 Hz, which led to whole trial exclusion (3.6 $\pm$5.3 trials per participant). Resulting clean data were used for subsequent analyses.
To extract event-related potentials (ERPs), epochs were low-pass filtered using a 6th order Butterworth filter at 15 Hz, baseline-corrected (baseline – 0.2 to 0 s), and then averaged over trials per condition. As in previous studies (Strauß et al., 2013; Obleser and Kotz, 2011), auditory evoked potentials were considered to be strongest over midline electrodes (FPz, AFz, Fz, FCz, Cz, CPz, Pz, POz, Oz), which were defined as a region of interest (ROI) for the ERP analysis, best capturing the dynamics of the N400 component. On the ERP amplitudes, we performed a time series analysis (49 consecutive steps of 50 ms width, windows overlap by 25 ms thereby covering a time range from 0 to 1.25 s) using repeated measures ANOVA with the factor of wordness (pseudo, ambiguous, real). We assessed p values with Greenhouse–Geisser-corrected degrees of freedom. If p values survived false discovery rate (FDR) correction for multiple comparisons (i.e., time windows), post-hoc t tests for pairwise comparisons were performed within these time windows.

3.2.6 Data analysis: time–frequency representations

In order to obtain time–frequency representations (TFRs), clean data were re-referenced to average reference. This is important for comparability with source analysis since the forward model needs a common average reference as well. For power estimates of non-phase-locked oscillations, Morlet wavelets were used on single trial data in 20-ms steps from – 700 to 2100 ms with a frequency-specific window width (linearly increasing from 2 to 12 cycles for frequencies logarithmically-spaced from 3 to 30 Hz). Single trials were subsequently baseline-corrected (against the mean of a – 500 to 0 ms pre-stimulus window of all trials) and submitted to a multi-level or "random effects" statistics approach (for application to time–frequency data see e.g., Obleser and Weisz, 2012; Henry and Obleser, 2012).
On the first or individual level, massed independent samples regression coefficient t tests with condition as dependent variable and contrast weights as independent variable (chosen correspondently to our effects of interest, see below) were calculated. Uncorrected regression t values and *betas* were obtained for all time–frequency bins. According to our

hypotheses, our effects of interest were a 'wordness' effect, namely a linear trend [pseudo > ambiguous > real], but also a stimulus-specific or 'ambiguity' effect [ambiguous > (pseudo, real)].

On the second or group level, the *betas* were tested against zero in a one tailed dependent sample t test. A Monte-Carlo non-parametrical permutation method (1000 randomisations) as implemented in the Fieldtrip toolbox estimated type I-error controlled cluster significance probabilities ($\alpha < 0.05$).

To evaluate the influence of baseline correction, we repeated first and second level statistics on absolute power estimates (skipping single trial baseline correction).

3.2.7 Source localisation of time–frequency effects

Source localisation for resulting clusters followed the Fieldtrip protocol on source reconstruction using beamformer techniques (e.g., Medendorp et al., 2007; Haegens et al., 2010; Obleser et al., 2012; Obleser and Weisz, 2012). In short, an adaptive spatial filter (DICS—Dynamic Imaging of Coherent Sources; Gross et al., 2001) based on the cross-spectral density matrix was built by estimating the single trial fast Fourier transformation of time windows and smoothed frequencies of interest (TOI and FOI) using a set of Slepian Tapers (Mitra and Pesaran, 1999). TOI and FOI were determined according to cluster results in sensor space but computational considerations were also taken into account (more time and frequency smoothing allows better spatial estimation): For theta, estimates were centred around 4.5 Hz (±2.5 Hz smoothing) and covered a 700 ms time window from 500 to 1200 ms, thus, three theta cycles and three tapers were used. For alpha (10 Hz ±2 Hz smoothing), a 700 ms time window was defined centred around 1000 ms, which covers approximately seven alpha cycles and results in two tapers.

For source localisation, the individual EEG electrode locations were first co-registered to the surface of a standard MRI template (by applying rigid-body transformations using the ft_electroderealign function). By co-registering to this template, a realistically shaped three-layer boundary elements model (BEM) provided by the Fieldtrip toolbox (Oostenveld et al., 2003) based on the same template was used. We were then able to calculate individualised forward models (i.e., lead fields) based on individual electrode positions and a standard head model for a grid with 1 cm resolution. Using the cross-spectral density matrices and the individual lead fields, a spatial filter was constructed for each grid point, and the spatial distribution of power was estimated for each condition in each subject. A common filter was constructed from all baseline and post-trigger segments (i.e., based on the cross-spectral density matrices of the combined conditions). Subject- and condition-specific solutions were spatially normalized to MNI space and averaged across subjects, and then displayed on an MNI template (using SPM8). Figure 3.2 (column 4) shows the result of cluster-based statistical tests (essentially the same tests as used for the electrode-level data before) that yielded voxel clusters for covariation of source power

with the alpha and theta effect, respectively. This was mainly done for illustration purposes, and unlike the tests for channel–time–frequency clusters in sensor space, no strict cluster-level thresholding was applied. We plotted t values on a standard MR template, and MNI coordinates mentioned in the figure caption refer to brain structures that showed local maxima of activation.
In order to visualise the specificity of the neural networks for either alpha or theta frequency range oscillations, we calculated an index using the t values of the wordness t_α- and the ambiguity t_θ-effect and divided their difference by their sum:

$$i_{\alpha\theta} = \frac{t_\theta - t_\alpha}{|t_\theta| + |t_\alpha|} \tag{3.1}$$

The index has been calculated only for those grid points which exceeded the critical value of $t_{19} = 1.7291$ in the source space solution. As such, only areas are highlighted which either show an alpha (blue) or theta (red) effect. This resulted in a descriptive source map as shown in Figure 3.3. Values around zero indicate non-dominance for either network (i.e. green in the figure).

3.3 Results

3.3.1 Highly accurate performance

The performance of the lexical decision task after each trial revealed high accuracy overall ($>95\%$ in each condition, see Fig. 3.1B). Nevertheless, an ANOVA with the three-level factor wordness was significant ($F_{2,38} = 28.54, p < 0.001$) with lowest accuracy for ambiguous pseudowords (ambiguous vs. real: $t_{19} = -4.16, p < 0.001$; ambiguous vs. pseudo: $t_{19} = -8.01, p < 0.001$). Highest accuracy was found for proper pseudowords (vs. real: $t_{19} = 2.18, p < 0.05$, indicating some confusion of ambiguous pseudowords with real words. Since the response was prompted with delay, effects on reaction time were neither expected nor found ($F_{2,38} = 1.582, p = 0.221$, see Fig. 3.1C).

3.3.2 Sequential effects of word-pseudoword discrimination in ERPs

Overall, the ERPs over midline electrodes show the typical pattern of an N1–P2 complex followed by a later N400-like deflection in all conditions (see Fig. 3.1D).
Binning the ERP in 50 ms time windows with 25 ms overlap and testing for condition differences (repeated measures ANOVA, threefold factor wordness) showed no differences in amplitude before 500 ms post stimulus onset: There were no differences in the N1 or P2 ($F < 1$). The repeated measures ANOVA yielded significantly different amplitudes from 0.5 to 1.2 s (mean $F_{2,38} = 13.19, p < 0.01$ after FDR correction). Furthermore, post-hoc t tests on the ERP amplitudes confirmed a regrouping of conditions over time: pseudowords differed from real words over the whole time course (pseudo > real from 0.5 to 1.125 s,

mean $t_{19} = -4.62$, $p_{mean} < 0.01$); ambiguous stimuli initially differed from real words (ambiguous > real from 0.525 to 0.825 s, mean $t_{19} = -4.27, p_{mean} < 0.01$), but regrouped with real words later, differing from proper pseudowords (pseudo>ambiguous from 0.85 to 1.2 s, mean $t_{19} = 3.1, p_{mean}$ <0.01; Fig. 3.1D, gray-shaded inlay).

3.3.3 Differential signatures of wordness in time–frequency data

As seen in the grand average TFRs in Figure 3.2 top row, frequencies of the theta range (3–7 Hz) were enhanced, first phase-locked to stimulus onset around 200 ms, and, with markedly decreased phase-locking, from 400 to 1000 ms after stimulus onset. In contrast, alpha power (8–12 Hz) was suppressed during the whole time course of a trial with the lowest power around 800 ms.
For assessing relative power changes, a multi-level statistics approach was chosen as described in the methods section. A linear contrast was set on the first level for testing the wordness effect [real > ambiguous > pseudo]. On the second-level, the cluster permutation test, testing the first-level betas against zero, revealed one positive cluster ($T_{sum} = 8,319.8, p < 0.05$) covering mainly lower- and mid-alpha frequencies (peak at 9.3 Hz and 0.88 s; Fig. 3.2 bottom row). In general broadly distributed, the cluster showed the largest statistical differences over the left frontal and right and left central electrodes (Fig. 3.2 bottom row fourth column). Extracted power values from the cluster (8–12 Hz, 0.88±0.06 s) confirmed significant differences between all three conditions (post-hoc paired t tests: real vs. ambiguous: $t_{19} = 2.32, p < 0.05$; real vs. pseudo: $t_{19} = 4.66, p < 0.01$; ambiguous vs. pseudo: $t_{19} = -2.09, p < 0.05$). When using absolute power, the positive cluster ($T_{sum} = 39,928; p < 0.001$) showed a similar distribution over frequency and time with peak effects at 10.7 Hz and 0.9 s over left anterior electrodes.
Interestingly, testing the ambiguity effect [ambiguous > (pseudo, real)] using the same statistical approach revealed one positive cluster ($T_{sum} = 8134.6; p < 0.05$) in the theta frequency range (peak at 5.2 Hz, 0.94 s; Fig. 3.2 middle row). Scalp topographies suggested two foci, one at the left-central anterior electrodes and the other at the parietal electrodes. Further, post-hoc paired t tests on power values extracted from the cluster (3–7 Hz, 0.88–1.1 s) confirmed that pseudowords and real words did not differ from each other ($t_{19} = 1.72, p < 0.1$) in the theta frequency range. Testing the absolute theta power, a comparable positive cluster was identified ($T_{sum} = 17,919; p < 0.01$) with the highest effect size at 5.5 Hz and 0.92 s but with a slightly shifted topography that overlaps at the left anterior electrodes but additionally emphasizes the right temporal areas.

3.3.4 Source localization of alpha and theta power changes

With respect to scalp topography (Fig. 3.2 bottom row), alpha oscillations appeared to be distributed broadly over the scalp with a central focus and exhibited less power

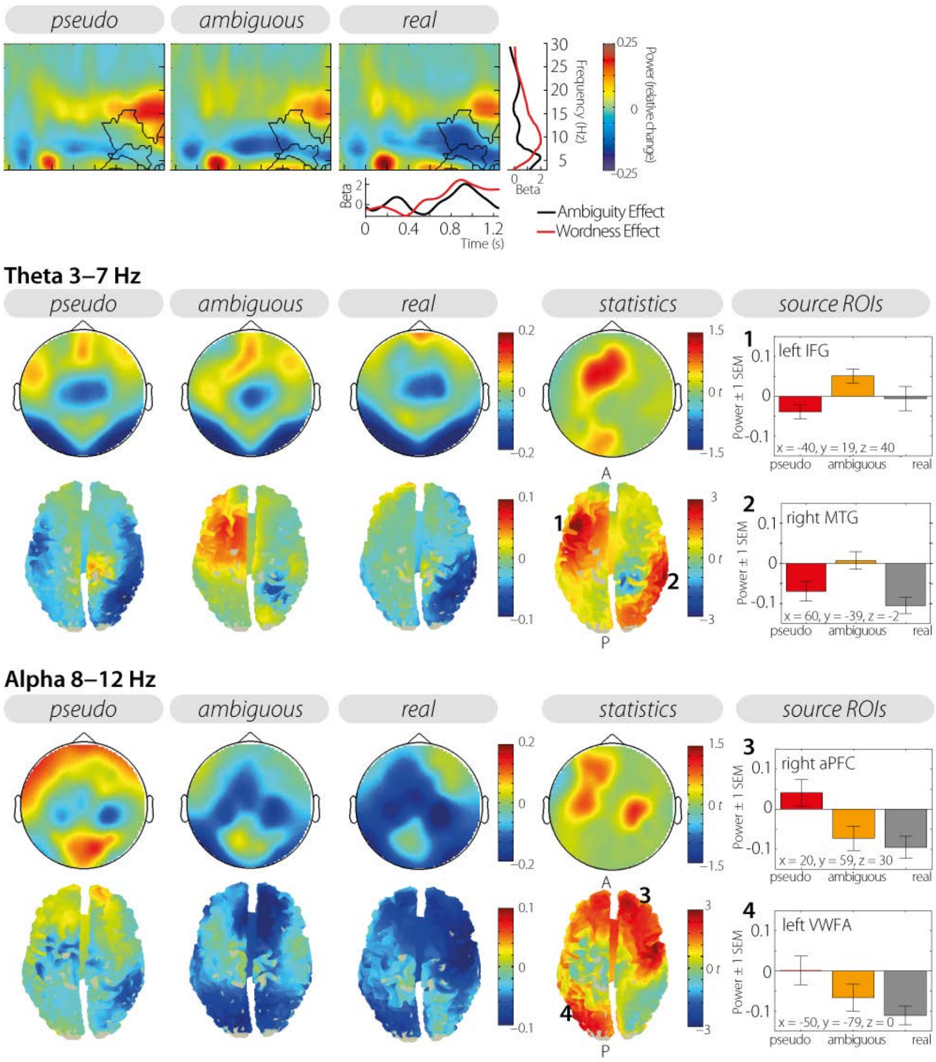

Figure 3.2: TFRs in sensor and source space. Top row shows the grand average TF power changes relative to a 500 ms pre-stimulus baseline over all electrodes for the three conditions separately: from left to right for clear pseudowords, ambiguous pseudowords, and real words. Black contours mark cluster boundaries. Middle row shows scalp topographies for relative power changes in the theta band (4.5 ±2.5 Hz, 500–1200 ms, corresponding to the time and frequency window used for the source localisation) and below the source projection. Bottom row shows the same for relative alpha power changes (10 ±2 Hz, 1000 ±350 ms). Fourth column depicts statistical differences. Fifth column are bargraphs extracted from source peaks in left IFG and right MTG for theta, and left VWFA and right aPFC for alpha, respectively.

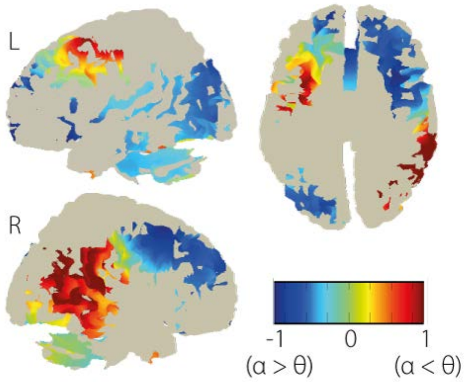

Figure 3.3: Alpha–theta index. The index compares the theta effect (Fig. 3.2 middle row) and the alpha effect (Fig. 3.2 bottom row) per source space grid point. The index has been calculated for grid points only which exceeded the critical value of $t_{19} = 1.7291$ such that only areas are highlighted which either show an alpha (blue) or theta (red) effect. Areas with index values around zero (green) show equal sensitivity to both effects, e.g., left frontal regions.

with increasing wordness. Following from the single conditions' source projections, source estimation of the alpha-driven wordness effect revealed peak activation in BA 9, right dorsolateral prefrontal cortex ($t_{19} = 3.04$; MNI = [10, 57, 40]). The cluster ($T_{sum} = 1,152.4; p < 0.05$) extended into the right primary somatosensory areas (BA 3), premotor cortex (BA 6), and motor cortex (BA 4), but also into the bilateral ventral and dorsal anterior cingulate cortex (BA 24/32), and the right inferior prefrontal gyrus (BA 47), including pars triangularis (BA 45). A second peak was found in the left occipital temporal cortex ($t_{19} = 2.88$; MNI = [–50, –79, 0]) and extending into BA 37 (fusiform gyrus) and BA 20/21 (inferior and middle temporal gyrus).

For theta power changes, the spreading of power change on the scalp (Fig. 3.2 middle row) suggested at least two generators: one with left frontal and one with right parietal origin, which had the highest relative power increase for ambiguous stimuli. Accordingly, two peak activations were found in one trend-level cluster ($T_{sum} = 341.9; p = 0.067$) for the ambiguity effect in the theta range. The first peak activation was found left anteriorly in BA 44 (pars opercularis; $t_{19} = 3.18$; MNI = [–40, 19, 40]). It extends to BA 9/46, left dorsolateral prefrontal cortex, and BA 6, premotor cortex. The second local peak was found right posteriorly in the middle temporal gyrus ($t_{19} = 3.01$; MNI = [60, –39, –2]), extending into inferior temporal gyrus (BA 20), fusiform gyrus (BA 37), supramarginal gyrus (BA 40), and posterior STG (BA 22).

3.3.5 Two separate networks disclosed by an alpha–theta index.

Calculating the alpha–theta index as shown in Figure 3.3 reveals that three of the four identified source peaks are selective for either the alpha-indexed lexical integration or the theta-indexed ambiguity resolution. Notably, the left IFG shows equally strong effects of alpha and theta activities as indicated by index values around zero.

3.4 Discussion

In order to functionally dissociate slow neural oscillations contributing to speech processing, we set up an auditory EEG study using a well-established lexical decision paradigm. Simultaneously, the data speak to theoretical controversies concerning spoken word recognition models (e.g., McClelland and Elman, 1986) by applying time–frequency analysis and revealing parallel processes of lexical integration and ambiguity resolution. Notably, alpha suppression, scaling with wordness and hence more akin to the N400, can be considered as a marker of ease in lexical integration, while theta enhancement marks the re-evaluation of the available sensory evidence. Generators of the alpha suppression effect were part of a left temporo-occipital and right frontal network. Oppositely, generators of the theta effect were localized in the left frontal and right middle temporal regions.
As we discuss below in further detail, the analysis of different oscillatory frequency bands disclosed the parallel maintenance of lexical and prelexical word versus pseudoword features in different brain regions and frequency ranges. To this end, time–frequency analysis is an important tool to inform discussions on sequential versus parallel processes in word recognition (e.g., Marslen-Wilson, 1987; for discussion see Norris et al., 2000).

3.4.1 Wordness effect in the alpha band

In line with previous findings (Obleser et al., 2012), alpha power showed the greatest suppression for real words compared to the lowest suppression (or even enhancement) for clear pseudowords. Interestingly, ambiguity leads to sub-optimal lexical integration (Friedrich et al., 2006; Proverbio and Adorni, 2008) and seems to be expressed in a state of intermediate alpha power.
Two (related) theoretical framings are relevant for this effect of wordness observed in the alpha frequency range. On the one hand, it has been emphasized that parieto-occipital alpha power reflects an inhibitory mechanism, with particular relevance for working memory and selective attention tasks (Klimesch et al., 2007; Foxe et al., 1998). On the other hand, recent findings provide more direct evidence for an influence of alpha oscillations on the timing of neural processing: Haegens et al. (2011) could show that better discrimination performance can be traced back to neuronal spiking in sensorimotor regions, which depends on the alpha rhythm not only in terms of power (firing is highest during alpha suppression) but also in terms of phase (firing is highest at the trough of a cycle; see also Spaak et al., 2012). Supporting the view put forward by Hanslmayr et al. (2012; high alpha oscillatory power mirroring reduced Shannon entropy and flow of information), Haegens et al. (2011) also found low spike-firing rates during periods of strong alpha coherence, for example, during the baseline, as opposed to the stimulus period. Both frameworks converge on predicting that low alpha power can serve as a marker of successful lexical integration.

An open issue is the potential contribution of the visual “what”-pathway to the alpha effect observed here. Particularly the left temporo-occipital source localisation peak suggests involvement of visual fields. This might be due to the fact that we used concrete nouns, which are by definition easily imaginable in comparison to the less imaginable pseudowords (for review see Binder et al., 2009). Note, however, that in a previous fMRI study using highly similar manipulations (Raettig and Kotz, 2008), no such effects even in the contrast of concrete versus abstract nouns were found. Nevertheless, the visual word form area has been found in auditory lexical decision tasks before and has been attributed to the literacy of participants (Dehaene et al., 2010; Dehaene and Cohen, 2011). Binder et al. (2006) gathered evidence that this area is especially sensitive to sublexical bigram frequency—a pivotal element of our study design. The argument of suppressed alpha power allowing lexical integration laid out above would also hold for such a traditionally more reading-related brain area.

3.4.2 Ambiguity effect in the theta band

Contrary to a previous study by Obleser et al. (2012), theta power did not scale linearly with difficulty of word processing (if defined as difficulty of lexical access). In particular, Obleser et al. (2012) found higher theta power for higher intelligibility, whereas in our case theta power was highest for the ambiguous (i.e. the most difficult) case. The data provided by Obleser et al. (2012) suggest that sufficient spectral information is needed to enable linguistic processes or lexical evaluation, which is reflected in increasing theta power. Our data extend this view by adding ambiguity on a lexical level which requires additional lexical re-evaluation. Future research needs to clarify whether these two factors, spectral detail and lexical ambiguity, might interact. Nevertheless, both results together support our interpretation of theta oscillations subserving a language-related but task-dependent mechanism and are in line with previous studies associating theta enhancement with lexico-semantic processing (Hagoort et al., 2004; Hald et al., 2006; Bastiaansen et al., 2008; Peña and Melloni, 2012).

Interestingly, a recent opinion paper by Roux and Uhlhaas (2014) suggests that theta oscillations may be involved in the phonological loop (Baddeley, 2003). The link to the phonological loop as a concept of linguistic short-term memory speaks in favor of our interpretation where lexical re-evaluation is achieved by replay of sensory evidence (Fuentemilla et al., 2010).

Furthermore, increased prefrontal theta power has been found in response to other types of ambiguous stimuli as well, and therefore might not be tied to the language domain. Specifically, increased mid-frontal theta activity has been reported in studies investigating the ambiguity induced response conflict (Hanslmayr et al., 2008; Cavanagh et al., 2009; Cohen and van Gaal, 2013) and episodic memory retrieval (Staudigl et al., 2010; Ferreira et al., 2014). Although these studies differ markedly with regard to several aspects, they all share the need for processing an ambiguous stimulus. It thus appears possible that enhanced theta oscillations during ambiguous word processing reflects enhanced conflict monitoring due to the co-activated real word ('Banene' co-activates 'Banane').

We localised the enhanced theta activity in a bilateral fronto-temporal network with peak activity in the left inferior frontal gyrus (IFG, BA 44) and the right middle temporal gyrus (MTG). Their contributions, though, to the proposed interpretation of replay need to remain speculative. Instructively, a right hemispheric advantage in tracking spectral information has been shown (Zatorre and Belin, 2001; Obleser et al., 2012; Scott et al., 2009; for review see Price, 2012) which converges with the fact that vowel differences (our crucial manipulation) are primarily spectral differences. More specifically, Carreiras and Price (2008) found in accordance with our results increased activation of right hemispheric areas when manipulating vowels. Combining both ideas, Zaehle et al. (2008) could show that the analysis of prelexical segments with respect to their spectral characteristics involved bilateral MTG activation.

The left IFG, however, has been associated with a variety of linguistic processes (see Binder et al., 2009 for a meta-analysis). The unfortunate vagueness of EEG source localisation limits functional dissociations which have been assigned to different subregions of the left IFG. Still, left IFG as a whole plays a role when monitoring auditory input (e.g., Zatorre et al., 1996; Giraud et al., 2004; Obleser et al., 2012). Other terms such as "auditory search", "auditory attention", or "auditory short-term memory" have been used to describe this function. This speaks in favor of our interpretation of auditory re-evaluation.

One might argue that our task was too easy to require top-down or re-evaluative processes. This relates to the ongoing psycholinguistic discussion whether replay or any feedback loop is really necessary in word recognition (Norris et al., 2000; McClelland et al., 2006). Since our stimuli were not phonetically ambiguous (see section 3.2.2), no perceptual confusion occurred which would have required replay (Ganong, 1980; Frauenfelder et al., 1990; Wurm and Samuel, 1997; Newman et al., 1997; Norris, 2006). However, stimuli were lexically ambiguous which led to decisional conflicts and required ambiguity resolution processes. Recall that we introduced manipulations not before the second syllable. The third (and final) syllable, however, either continued the wordness violation (clear pseudoword) or created a lexically ambiguous case by resuming to the initially pre-activated cohort (ambiguity). Mattys (1997) summarizes evidence that retrograde information, i.e. provided after the deviation point, can influence the decision on the identity of a stimulus. This may increase reaction times (Goodman and Huttenlocher, 1988; Taft and Hambly, 1986), implying some re-evaluative processes. We therefore suggest that prelexical information were maintained and replayed in order to resolve decisional ambiguity.

In sum, we argue for a theta-tuned network which is co-activating the left IFG and the right MTG in order to replay lexico-semantic information for task-relevant ambiguity resolution.

3.4.3 Relationship of evoked potentials and induced oscillations

So far, studies analysing the ERP have related the N400 to effortful processing, for example when mapping the phonological form and meaning of pseudowords, compared to real words, onto a stored representation in the mental lexicon (Friedrich et al., 2006). Recent accounts more rooted in the predictive coding framework of cortical functional organization (e.g., Summerfield and Egner, 2009) may describe the N400 as a marker of the mismatch between what is predicted and what is perceived (Lau et al., 2009, 2013). While we cannot distinguish between these explanations in a context-free setup using single word stimuli, our data more importantly show parallels in the pattern of the N400 changes over midline electrodes and the pattern of alpha oscillatory changes. Contrary to the effort- or predictive coding-hypothesis, the inhibition theory for alpha oscillations would then imply that lexical processing takes place for real words, and must be inhibited for pseudowords. Notably, only analysing the ERP would have led to the view that lexico-semantic integration in ambiguous pseudowords can be accomplished in the same way as their real word analogs. The regrouping of N400 deflections over time would have suggested a sequential change in processing strategy: First, ambiguous stimuli were analysed in the same way as proper pseudowords, but from 850 ms onwards no difference between ambiguous and real word stimuli was discernible. Thus, the conclusion derived from ERPs only would have been a sequential process of lexical access. Such time–frequency decompositions of the ERP as demonstrated here may help in the future to resolve inconsistencies in the N400 literature and its generating brain structures (Halgren et al., 2002; Khateb et al., 2010). By looking additionally at oscillatory activity, which arguably constitutes the ERP activity to large extents (Makeig et al., 2004; Mazaheri and Jensen, 2008; Min et al., 2007; Hanslmayr et al., 2007; Klimesch et al., 2007), parallel neural processes become discernible. The data suggest a combination of lexical integration and ambiguity resolution processes: wordness violations are detected (N400) and maintained (alpha power), but also re-evaluated retrieving stimulus-specific information (i.e., enhanced power of theta oscillations for ambiguous stimuli).

3.4.4 Conclusion

Time–frequency decomposition functionally separates parallel contributions of theta and alpha oscillations to speech processing, thereby fruitfully extending current frameworks based on evoked potentials. The data presented here provide evidence that lexical as well as prelexical information are maintained in spoken word recognition. The observed specificity, with theta bearing relevance to stimulus-specific, lexico-semantic processes and alpha reflecting more general inhibitory processes (thereby gating lexical integration), is a promising starting point for future studies on speech comprehension in more demanding circumstances such as peripheral hearing loss and/or noisy environments. The data furthermore shed light onto the neural bases of the lexical decision task that has been in use for decades. In sum, this approach allows for a refinement of neural models describing the complex nature of spoken word recognition.

4 Alpha oscillations as a tool for auditory selective inhibition[3]

4.1 Introduction

In ecological listening situations, auditory signals are rarely perceived in quiet due to the presence of different auditory maskers such as distracting background speech or environmental noise. Thus, sounds from different sources greatly overlap spectro-temporally at the level of the listener's ear. What are the neural correlates that facilitate selective listening to relevant target signals despite irrelevant auditory input (i.e., the "cocktail party problem"; Cherry, 1953)? At the central neural level, two complementary mechanisms of top-down control (i.e., regulation of subsidiary cognitive processes) should be considered: First, top-down selective attention to relevant information (Fritz et al., 2007) could facilitate target processing by enhancing the neural response to the attended stream (i.e., gain control; Lee et al., 2014). Second, top-down selective inhibition of maskers (Melara et al., 2002) could help to direct limited processing capacities away from irrelevant information (Desimone and Duncan, 1995), thereby avoiding full processing of distractors (Foxe and Snyder, 2011).

In this regard, interference of auditory maskers might be the result of both insufficient attention to the target and poor inhibition of noise and distractors. In this perspective article we focus on the latter, that is, neural mechanisms of auditory selective inhibition. We propose that cortical alpha (~10 Hz) oscillations are an important tool for top-down control as they regulate the inhibition of masker information during speech processing in challenging listening situations.

4.2 A framework to test auditory alpha inhibition

A common observation is a prominent increase in alpha power when participants listen to auditory materials presented against background noise (e.g., Wilsch et al., 2014). Figure 4.1A, for example, shows the grand average time–frequency representations of 11 participants during a lexical decision task on isolated words presented in quiet (data reported in Chapter 3) and in white noise. For words in quiet, alpha power at around 10 Hz did

[3] This chapter is adapted from parts of the published article by Strauß, Wöstmann, and Obleser (2014). *Front Hum Neurosci 8*, 350.

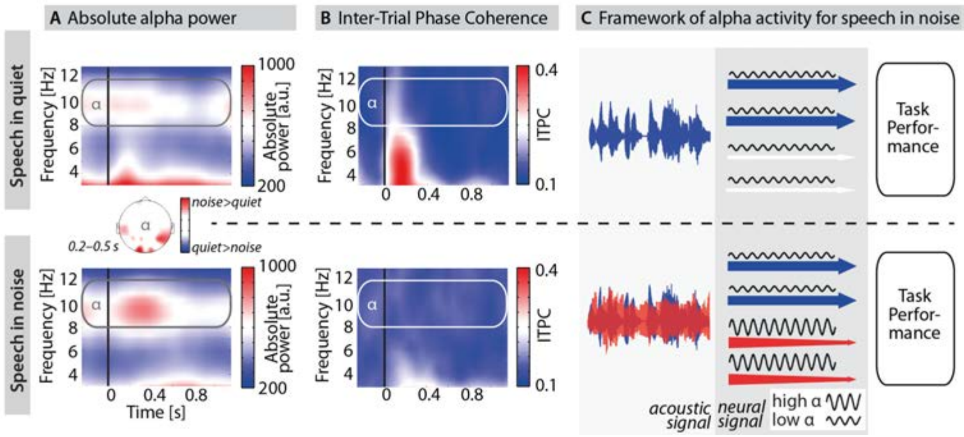

Figure 4.1: The proposed role of alpha activity for speech processing in noise. A. Average time-frequency representations (TFR) of 11 participants performing a lexical decision task on words in quiet (top) and in white noise (bottom). SNRs were titrated individually using a two-down-one-up staircase adaptive tracking procedure. Average SNR was -10.22 dB ±1.95 (*SD*) such that participants performed about 71 % correct. Speech onset is indicated by the black vertical line at 0 s; average word length = 750 ms; EEG recorded from 61 scalp electrodes; time-frequency analysis using Morlet wavelets. Plots show measures of absolute power averaged over all scalp electrodes. Topography depicts the alpha power difference for speech in noise – quiet. Data were SCD (source current density)-transformed before TFR estimation to improve spatial resolution. **B.** Inter-trial phase coherence (ITPC) as a measure of phase-locking of oscillations over trials. ITPC is bound between 0 and 1; higher ITPC values indicate stronger phase alignment across trials. **C.** A simple framework of alpha oscillations for speech processing in noise. Acoustic signals overlap energetically as they enter the ear. At the brain level, features of speech and noise are processed as far as possible in distinct processing channels (depicted here with arrows; for details see text). High alpha power inhibits channels processing noise features to allow for an optimal task performance with minimised noise interference.

not considerably increase after word onset. However, when words were presented in noise, alpha power was increased during the first 500 milliseconds after word onset corresponding to the first two thirds of the average word duration. This effect was strongest over temporal and occipital sites (topography in Fig. 4.1A) suggesting the inhibition of the task irrelevant visual modality but also compensatory mechanisms within speech-related areas. Critically, alpha power difference did not depend on ITPC (inter-trial phase coherence) differences, as indicated by the absence of a stronger ITPC in noise compared to quiet (Fig. 4.1B). We therefore presume that induced (i.e., not strictly stimulus-locked; Freunberger et al., 2009) alpha power is crucial for speech processing in challenging listening conditions as it suppresses irrelevant information.

Figure 4.1C illustrates a tentative framework for how alpha oscillations could support auditory selective inhibition. Sounds arriving at the listener's ear must be further processed in the brain to extract task-relevant information. One way to think about the proposed mechanism is in terms of auditory object selection which requires object formation in the first place (Shinn-Cunningham, 2008). An auditory object might be formed on the basis of common spectro-temporal features, harmonicity, simultaneous onsets, or spatial grouping (Griffiths and Warren, 2004; Bizley and Cohen, 2013). We refer to all these different features used to form auditory objects as "channels" of auditory information represented by the arrows in Fig. 4.1C. The concept of channels has a long tradition (Broadbent, 1958) and is inspired by the most clear distinction of target and distractor used in many dichotic listening paradigms where left and right ear channel need to be separated. Nevertheless, channels in our framework should be conceived as functional auditory processing units rather than anatomical pathways. As soon as these channels are defined, attention or inhibition can be selectively applied, given attentionally flexible fields in the auditory cortices (Petkov et al., 2004). Note that even though in the visual modality claims about alpha oscillations in feature-based (Romei et al., 2012) and object-based (Kinsey et al., 2011) attention have been made, we do not make any assumption about this distinction in our framework and use the term "channels" for both features and objects, or early and late selection.

If speech is presented in quiet (Fig. 4.1C, top panel), alpha power is low in channels processing features of the speech signal to support processing of task-relevant information. Accordingly, the net resulting alpha power in the M/EEG would continue on baseline level (Fig. 4.1A) and decrease during word integration (>400 ms). If, however, speech is presented in the presence of maskers (e.g., environmental noise, distracting talkers; Fig. 4.1C, bottom panel), alpha power needs to be up-regulated first in those channels processing noise features before it is going to be suppressed during word integration (Fig. 4.1A). Enhanced alpha activity inhibits processing of noise and thereby "protects" (Klimesch, 1999; Roux and Uhlhaas, 2014) the task- or performance-relevant information in the speech signal from noise interference.

Importantly, the up-regulation of alpha power in channels that process noise is not an automatic ("bottom-up") process but critically depends on "top-down" attentional control. For instance, in a multi-talker situation, target and distracting talker switch roles permanently, as the listener decides to change the conversational partner. In such a situation, M/EEG alpha power would be constantly at a high level; however, the deployment of alpha power onto the different processing channels would be changing continuously.

4.3 A short review of auditory alpha inhibition

What is the functional role of high alpha activity for word processing in noise? To answer this question, it is essential to distinguish between interpretations in which alpha activity is related to target processing from these related to noise processing. It is possible that the reduced intelligibility of words in noise leads to sub-optimal word processing and thus to less alpha suppression in brain areas relevant for speech processing (Strauß et al., 2014). The inverse mechanism, as we put forward in the current framework, is equally likely by which alpha power is enhanced for temporarily irrelevant information and thereby compensates for perceived cognitive effort (increased when listening to speech in noise: Larsby et al., 2005; Helfer et al., 2010; Zekveld et al., 2011). In this regard, alpha would "protect" the lexical processes from noise interference. The challenge will be to experimentally dissect these (not mutually exclusive) mechanisms. We now review initial evidence for alpha's inhibitory role in audition.

Currently, there are only few studies that show alpha power modulations when participants simultaneously listen to two auditory streams, that is, one signal and one masker. In one study by Kerlin and colleagues (2010), participants were simultaneously listening to two spatially separated speech streams. On each trial, an initial visual cue indicated whether they were supposed to attend the left or right stream. During speech presentation, EEG alpha power was enhanced over the cerebral hemisphere contralateral to the masker, while alpha power was reduced contralateral to the to-be-attended stream. The authors concluded that this alpha lateralization indexes the direction of auditory attention to speech in space. Importantly, this finding corroborates our view that enhanced alpha power in brain areas engaged in distractor processing decreases further processing of the distractor and hence, facilitates processing of the target signal. However, two questions arise from this study: First, as the direction of auditory attention was cued visually in this study, it might be that the alpha lateralization indicates the allocation of supramodal rather than auditory selective attention (Farah et al., 1989). Second, spatial attention may play a special role not least because of auditory processing models suggesting separate what- and where-pathways (Rauschecker and Scott, 2009).

In three other recent studies, alpha power modulations were consistently found during the anticipation of auditory target signals from the left or right (Müller and Weisz, 2012; Banerjee et al., 2011; Ahveninen et al., 2013). In these studies, participants were cued to attend either the auditory event on the left or right, and to ignore the distractor on the other side. Alpha power was enhanced during the anticipation of auditory stimulation contralateral to the distractor. These results demonstrate alpha lateralization effects already during the preparation for an auditory selective listening task. This is in line with studies reporting high pre-stimulus alpha power when participants are about to miss a (visual) target (van Dijk et al., 2008; Busch et al., 2009; Romei et al., 2010). In terms of our framework (Fig. 4.1C), anticipatory high alpha power successfully blocks in-depth processing of sensory information that might lead to missing the target.

However, interpretations of these studies are limited for our model, since alpha power modulations were found only during the anticipation but not during the actual processing of competing auditory streams. More data are clearly needed on the peri-stimulus alpha dynamics. As the spatial resolution of M/EEG is limited, prospective experiments could induce alpha oscillations over specific brain areas using transcranial alternating current stimulation (tACS) to assess the influence of alpha modulations on listening success under adverse acoustic conditions. Moreover, future studies could record the electrocorticogram (ECoG) directly from the cortical surface to track alpha sources and reveal the interplay between frequency bands. Such higher spatial resolution would allow to differentiate between alpha activity in brain regions associated with processing the masker or the signal. As of now, we are left to speculate how spatially specific alpha oscillations might operate, for example along a cochleotopic gradient in primary auditory cortex. The best data to infer from stems from visual cortex, where for example Buffalo and colleagues recorded with two electrode tips in attended vs non-attended receptive fields less than a millimeter apart and report attention-dependent opposing, and deep-layer-specific alpha changes (expressed as alpha spike–field coherence; Buffalo et al., 2011). Comparable data are, to our knowledge, still missing for auditory areas.

4.4 Conclusion

We have presented a framework for studying alpha oscillations as a tool for auditory selective inhibition in challenging listening situations. The data provide initial evidence qualifying alpha oscillations as a pivotal mechanism affecting listening in multi-talker situations. Future studies could expand these findings and study the role of alpha oscillations during speech perception in ecologically valid listening situations.

5 Alpha phase determines successful lexical decision in noise[4]

5.1 Introduction

Human psychophysical performance for detection and discrimination of low-level stimuli has been found to depend on slow pre-stimulus oscillatory brain states across domains (visual: Varela et al., 1981; Hanslmayr et al., 2007; van Dijk et al., 2008; Busch et al., 2009; Schubert et al., 2009; Cravo et al., 2013; Spaak et al., 2014; auditory: Lakatos et al., 2005; Henry and Obleser, 2012; audiovisual: Keil et al., 2014). These findings relate neural phase to neural excitability fluctuations, such that performance is best for targets coinciding with the excitable phase of a neural oscillation, and worst for targets coinciding with the inhibitory phase. Going beyond low-level perception, we ask here whether higher cognitive functions such as speech processing would also depend on neural phase. Although recently proposed models would predict a dependence of speech processing on neural oscillatory phase (Ghitza, 2011; Gagnepain et al., 2012; Giraud and Poeppel, 2012), no experimental evidence has been gathered so far.

One elegant task that can bridge psychophysical aspects of performance (detection or discrimination) with speech processing is the auditory lexical decision task (Marslen-Wilson, 1980): Listeners are presented with words as well as word-like stimuli (i.e., pseudowords), and have to judge whether they heard a meaningful word or not. Parallel to low-level discrimination studies, we made the lexical decision task "near-threshold" by embedding speech in individually titrated levels of white noise, which increased the difficulty of the task and, purposefully, the amount of errors. We simultaneously recorded the electroencephalogram and hypothesized that a dependence of lexical-decision accuracy on low-frequency neural oscillatory phase should be observed.

Here, we were interested in the role of alpha (8–12 Hz) and theta (3–7 Hz) neural phase for lexical decision performance. Instantaneous alpha phase has previously been linked to low-level detection and discrimination performance not only in the visual (Mathewson et al., 2009; Busch and VanRullen, 2010; Romei et al., 2010), but also in the auditory domain (Rice and Hagstrom, 1989; Neuling et al., 2012). Critically, alpha phase has been

[4]This chapter is adapted from a manuscript by Strauß, Henry, Scharinger, and Obleser (sion). *The Journal of Neuroscience.*

found to modulate neuronal firing and to determine the neural phase associated with best discrimination performance (Haegens et al., 2011). Discrimination performance in lexical decision may also depend on syllabic processing and thus potentially be indexed by oscillatory activity in the theta range ($\sim$ 4 Hz) with oscillation periods corresponding to the average syllable duration of around 250 ms (Ng et al., 2012; Peelle and Davis, 2012; Gross et al., 2013; Doelling et al., 2014; note that also Busch et al., 2009, reported a pre-stimulus phase bifurcation effect in the 7-Hz range). Similar to alpha, theta oscillations have been linked to neuronal firing (e.g., Kayser et al., 2012) and can impact auditory detection performance (Ng et al., 2013).
Our data show that the accuracy of auditory lexical decision depends on the instantaneous phase of alpha oscillations: Stimuli that were later judged incorrectly fell into an alpha phase opposite to that for stimuli that were judged correctly in a pre-stimulus time window as well as in a second, peri-stimulus time window.

5.2 Methods

5.2.1 Participants

Eleven participants (7 females; 25.1 $\pm$ 1.6 years, M $\pm$ SD) gave informed consent to take part in the experiment. All were native speakers of German, right-handed, with self-reported normal hearing abilities, and no history of neurological or language-related problems. They received financial compensation for their participation. All procedures had ethical approval from the Ethics Committee of the University of Leipzig.

5.2.2 Stimuli

Stimuli were real words and their pseudoword counterparts (Raettig and Kotz, 2008; Strauß et al., 2014). Pseudowords were created as follows: From a list of 60 tri-syllabic concrete German nouns (‘real’ words, e.g., /banane/, [engl. banana]) two types of pseudowords were derived, ‘ambiguous’ pseudowords, by exchanging only the nucleus vowel of the second syllable (e.g., /banene/), and ‘opaque’ pseudowords by scrambling the syllables across words while keeping the position-in-word fixed (e.g., /bapossner/). Furthermore, 60 ‘abstract’ real words (e.g., /botanik/, [engl. botany]) served as fillers to ensure a balanced word–pseudoword ratio. In sum, the experimental corpus consisted of 240 lexical stimuli with a mean length of 754.2 $\pm$ 83.5 ms (M $\pm$ SD). In the following, opaque pseudowords and abstract real words were not analyzed because we focused on the noise-induced vowel confusion between real words and ambiguous pseudowords that lead to “Yes” or “No textquotedblright decisions about whether an item was a word or not.
All words and pseudowords were spoken by a trained female speaker and digitized at 44.1 kHz. Post-processing included down-sampling to 22.05 kHz, cutting at zero crossings closest to articulation on- and offsets, and root mean square (RMS) normalization.

5.2.3 Experimental procedure

Prior to each experimental EEG session, individual signal-to-noise ratios (SNR) were determined by means of an adaptive tracking procedure. During adaptive tracking, participants were presented with the second syllables extracted from the real words and their ambiguous-pseudoword counterparts. On each trial, the participant heard two successive syllables embedded in white noise and indicated whether the vowels in each pair were “same” or “different”. Intensity of the white noise was adjusted according to a two-down-one-up staircase procedure that estimated the signal-to-noise ratio (SNR) targeting 70.7% accuracy (Levitt, 1971). Resulting average SNR was -10.22 ± 1.95 dB (M $\pm$ SD).
Next, a short familiarization for the trial timing was provided during which participants made lexical decisions in noise about 10 additional items from Raettig and Kotz (2008) that were not used in the present experiment. During the EEG experiment, participants heard words and pseudowords embedded in white noise and indicated via button press whether they heard a real word or not (“Yes”/“No”). Button order (left/right for “Yes”/“No” responses) was counterbalanced across participants. On each trial, the white noise started 1 sec before (pseudo)word onset, coincident with the appearance of a fixation cross, and lasted for 2.2 sec in total (see Fig. 5.1A). After 2.2 sec, the fixation cross changed to a question mark that prompted the lexical decision response. Trial timing was chosen based on a previous study in our lab using the same paradigm without noise (Strauß et al., 2014) and allowed artifact-free estimations of time–frequency representations (see Data analysis).
Each participant listened to 240 stimuli (120 words, 120 pseudowords) in an individually pseudo-randomized sequence. That is, each participant heard both the ‘real’ and the ‘ambiguous’ versions of each word. The order of occurrence for a given real word and its pseudoword counterpart was counterbalanced across participants in order to control for potential interfering effects of previous exposure to the respective complementary item. For the same reason, the distance between a word and its pseudoword counterpart was maximized (i.e., on average 120 other items in-between). Listeners paused after each block of 60 trials. Overall duration of the experimental procedure was about 30 minutes.

5.2.4 Data acquisition and preprocessing

The electroencephalogram (EEG) was recorded from 64 Ag–AgCl electrodes positioned according to the extended 10–20 standard system on an elastic cap with a ground electrode mounted on the sternum. Bipolar horizontal and vertical electrooculograms (EOG) were recorded for ocular artifact-rejection purposes. All impedances were kept below 5 kΩ. Signals were referenced online against the left mastoid, and digitized with a sampling rate of 500 Hz and a passband of DC to 140 Hz. Individual electrode positions were determined after EEG recording with the Polhemus FASTRAK electromagnetic motion tracker.

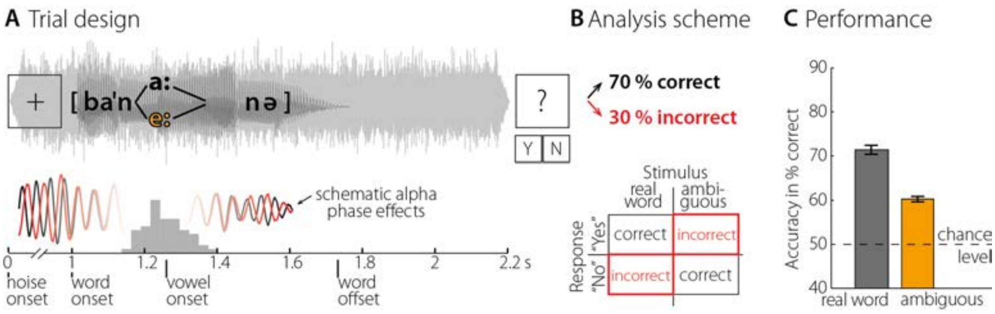

Figure 5.1: Trial design and behavioural measures. A. Trial design. Lexical stimuli were presented against a white-noise background. The distribution of critical vowel onsets is shown schematically in relation to the timing of the two alpha phase effects reported here. Average word length was 0.74 $\pm$ 0.08 s (M $\pm$ SD). Delayed lexical decision was prompted by a question mark. **B.** Analysis scheme. 70% correct was targeted with individual signal-to-noise ratios. For the analysis, correct trials comprised trials on which participants responded "Yes" to a real word or "No" to the ambiguous counterpart as illustrated by the cross-tabulation. **C.** Behavioural results. Participants performed better for real words than for ambiguous pseudowords. However, performance for both stimulus types was significantly above chance.

EEG preprocessing was done offline using the open-source Fieldtrip toolbox (Oostenveld et al., 2011) for Matlab (Mathworks). To avoid edge effects at low frequencies, broad epochs were defined ranging between −700 ms (excluding ERPs due to noise onset) and 2100 ms relative to (pseudo)word onset. Data were band-pass filtered from 0.1 Hz to 100 Hz using a two-pass Butterworth filter and, for ERP analysis only, re-referenced to combined mastoids (time–frequency analyses, see below, used re-referencing to average reference). To reject systematic artifacts, independent component analysis (ICA) was applied and components comprising eye movement, heartbeat, and muscle artifacts were rejected according to definitions provided by Debener et al. (2010). After ICA, an automatic artifact-rejection routine removed single trials for which within-channel peak-to-peak range exceeded 120 μV. On average, 2.7 $\pm$ 3.0 (M $\pm$ SD) trials were rejected per participant. The resulting clean data were used for subsequent data analyses.

5.2.5 Data analyses

Phase analysis. Time– frequency representations (TFRs) were estimated from single-trial data so that we could assess the effects of phase and power on lexical decisions. Epoched, filtered, artifact-rejected time-domain data were re-referenced to average reference (Strauß et al., 2014). Subsequently, Morlet wavelets were applied on single-trial TFRs in 20-ms steps with a frequency-specific window width to account for the trade-off between higher frequency resolution for lower frequencies and higher time resolution for higher frequencies. Therefore, TFRs for logarithmically spaced frequencies from 3 to 30 Hz were convolved with linearly increasing window widths ranging from 2 to 12 cycles. Phase and power values were then estimated at each channel × frequency × time point from the complex output of the wavelet convolution.

For the analysis of phase data, we calculated a phase bifurcation index (BI), ϕ, suggested by Busch et al. (2009). First, trials were split based on accuracy (i.e., correct versus incorrect responses) for each participant. Then, we calculated inter-trial phase coherence ($0 \leq \text{ITPC} \geq 1$) separately for correct trials, for incorrect trials, and for all trials taken together. Lastly, to compute the phase bifurcation index ϕ, the ITPC for correct, incorrect, and all trials were combined according to the following formula:

$$\phi = (ITPC_{correct} - ITPC_{all}) \times (ITPC_{incorrect} - ITPC_{all})$$

BIs were calculated separately for each channel × frequency × time bin. A positive BI indicates that both conditions have high inter-trial phase coherence but that the mean phase angles for the two conditions are anti-phase. A negative BI, by contrast, indicates that one condition is more phase locked than the other, i.e. angles of one condition are randomly distributed while angles of the other condition concentrate towards a certain direction. Further details on the BI can be found in Busch et al. (2009).

As expected, the number of trials was not balanced between correct (number of trials per subject = 75.36 ± 8.87) and incorrect trials (number of trials per subject = 39 ± 8.28; see analysis scheme in Fig. 5.1B). To account for this inequality, which can bias estimates of ITPC (Lachaux et al., 1999; Ding and Simon, 2013), we performed a randomization test analogous to the Monte Carlo method described in Maris and Oostenveld (2007) to obtain a robust measure of the BI in each participant. For each participant, the number of trials to be selected was equal to 75% of the amount of incorrect trials (the category with the smallest number of trials) resulting in 29.45 ± 6.25 trials per condition. On each of 1000 iterations, trials were randomly selected without replacement from the set of correct and incorrect trials. From ITPC estimates for correct, incorrect, and all selected trials, a single BI was calculated. The mean bifurcation index over these 1000 repetitions was used per participant for further statistical analyses. par On the group level, we tested BIs against zero separately for the alpha (8–12 Hz) and the theta (3–5 Hz) frequency bands for each time point in the range between –0.35 and 1.1 s with respect to (pseudo)word onset using the Monte Carlo randomization method (1000 repetitions) with cluster correction as implemented in FieldTrip. The time window was chosen such that edge effects of TFR estimation for lowest frequencies were avoided and stimulus offset responses at 1.2 s post-(pseudo)word onset (i.e., the end of the masking noise) were excluded.

Further analyses. In order to further characterize the phase effects found via the test of the BI and to test for potential confounds, we also evaluated alpha and theta ITPC, absolute alpha and theta power, and event-related potentials (ERPs). For the ITPC analysis, the differences between $\text{ITPC}_{\text{correct}}$ and ITPCincorrect trials (8–12 Hz and 3–5

Hz; from -0.35 to 1.1 s) as estimated for the BI calculation (i.e. 1000 iterations) were averaged per participant and submitted to a two-tailed single-sample t test against zero with cluster correction using the Monte Carlo randomization method (1000 repetitions).
For power estimates, we squared the magnitude (complex modulus) of single-trial Fourier data. Analogous to the phase analysis, the same amount of trials for correct and incorrect trials were selected as described above. Subsequently, their power difference was calculated, and the mean over 1000 of such differences was taken per subject. The group-level analysis on power was the same as described above for BI and ITPC analyses.
For analysis of ERPs, the epoched, filtered, and artifact-rejected time-domain data were filtered with a 6th-order Butterworth low-pass filter at 15 Hz. For baseline correction, a time window from -200 to 0 ms pre-(pseudo)word onset (i.e., during the masking noise) was chosen. Amplitudes were then averaged in selected time windows over selected channels. Time-window and channel selection was based on the peri-stimulus BI cluster. A pairwise t test compared ERP amplitudes of correct versus incorrect trials across subjects.

Effect sizes. For simple t statistics (dependent and independent samples t tests), we estimated the effect size measure $r_{equivalent}$, here denoted r, which is bound between 0 and 1 (Rosenthal, 1994; Rosenthal and Rubin, 2003). Effect sizes for multiple t tests (e.g., for all channel $\times$ frequency $\times$ time bins belonging to a significant cluster) were estimated by averaging r values across individual tests constituting the cluster (denoted R).

5.3 Results

5.3.1 Accuracy of lexical decisions.

As shown in Figure 5.1C, participants achieved an average accuracy near the one targeted by the adaptive tracking procedure for real words (71.4 $\pm$ 1.02%). Although slightly worse, accuracy for the ambiguous pseudowords was still better than chance (60.3 $\pm$ 0.61%; t test against 50%: $p = 0.0005$, $t_{(10)} = 5.1, r = 0.85$).

5.3.2 Neural phase in the alpha band predicts lexical-decision accuracy.

Non-parametric permutation tests of the BI against zero revealed two positive clusters (i.e., high phase concentration but opposite mean phases for correct vs. incorrect trials) in the alpha frequency range from 8–12 Hz. The first positive cluster was found in a time window ranging from –120 to 40 ms pre-(pseudo)word onset, and had a right anterior scalp distribution ($p = 0.036; T_{sum} = 124.93, R = 0.71$; Fig. 5.2A). The second positive cluster was found in a time window ranging from 420 to 580 ms post-(pseudo)word onset, and had a central-left anterior distribution ($p = 0.011; T_{sum} = 168.95, R = 0.61$; Fig. 5.2B).
In order to illustrate the nature of the phase effects underlying the significant BI results, we extracted the single-participant phase angles for both positive clusters, and plotted the

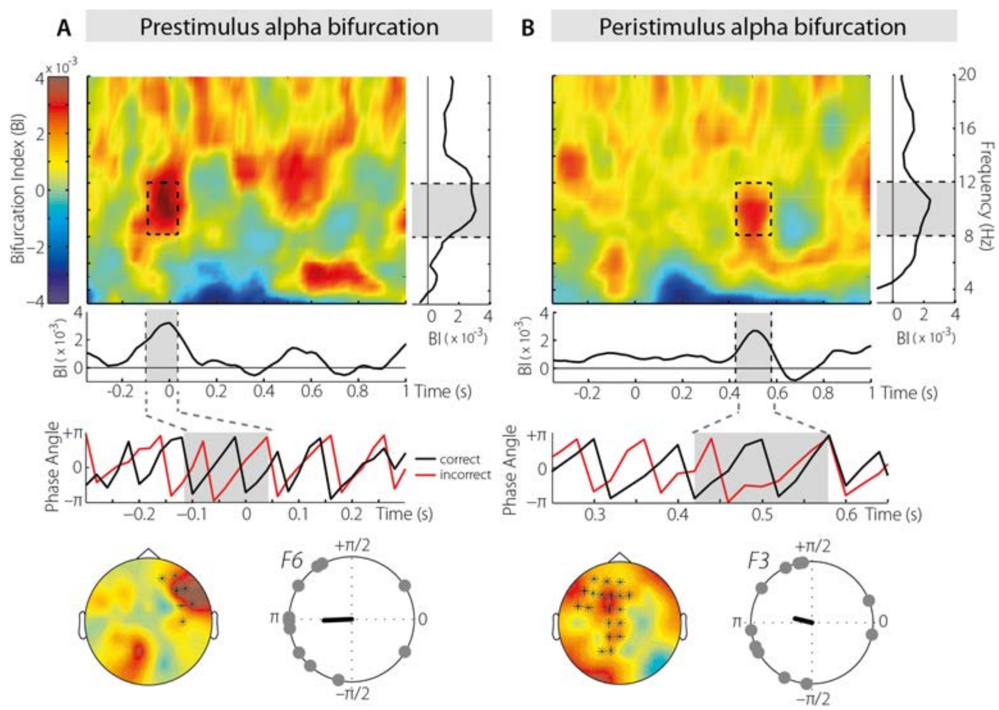

Figure 5.2: Results from BI analysis. A. Pre-stimulus alpha phase bifurcation (time–frequency plots). One cluster was found in the alpha band (8–12 Hz) with right anterior scalp distribution. Correct and incorrect trials were anti-phase between –120 and 40 ms (0 ms is (pseudo)word onset). Below, alpha phase extracted from and averaged over the cluster is shown in radians, for correct (black) and incorrect (red) trials separately as a function of time. Phase differences (per subject) are plotted for electrode F6 along with resultant vector. **B.** Peri-stimulus alpha bifurcation. Second cluster was found in the alpha band (8–12 Hz) with left anterior central scalp distribution. Conditions were anti-phase from 420 to 580 ms (0 ms is (pseudo)word onset). Alpha phase extracted from and averaged over the cluster is shown in radians as a function of time, for correct (black) and incorrect (red) trials separately. Phase differences (per subject) are plotted for electrode F3. Electrodes belonging to significant clusters are highlighted in topographies as asterisks.

circular distance between mean phase angles for correct versus incorrect trials (bottom panels in Fig. 5.2A and B). For example, at electrode F6, nine of 11 participants have a mean pre-stimulus phase distance greater than 90° ($\pi/2rad$) leading to a consistently positive BI.

5.3.3 Lexical-decision accuracy was not predictable from phase coherence, power, or ERP amplitude.

Accuracy of lexical decision could not be predicted based on any of the other neural parameters (Fig. 5.3). First, no ITPC differences were observed in a non-parametric permutation test using the same time and frequency parameters as for the BI analysis (see

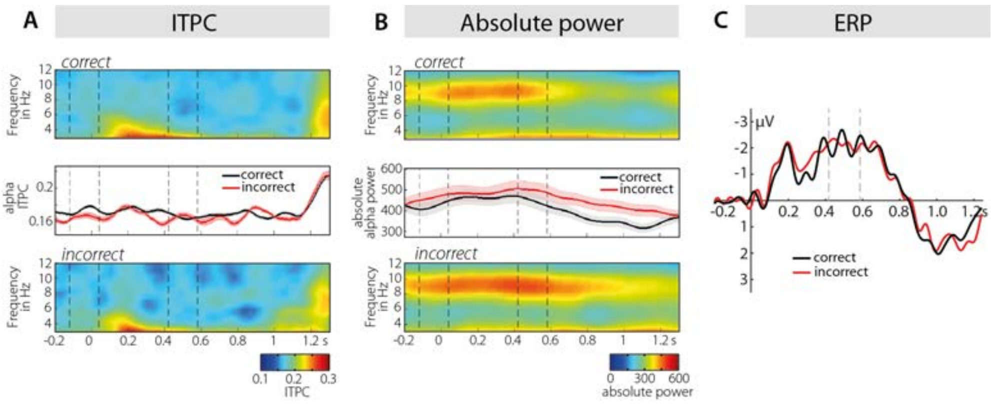

Figure 5.3: Time–frequency and time-domain analyses. A. Inter-trial phase coherence (ITPC) shown separately for correct (top) and incorrect (bottom) trials. ITPC specifically for the alpha (8–12 Hz) band is shown in the middle panel, separately for correct (black) and incorrect (red) trials. No differences were observed in the alpha or theta bands. Vertically dotted lines mark the time window of the pre- and peri-stimulus alpha phase bifurcations. Shades in time-series plot (middle) depict ± SEM. **B.** Absolute power shown for correct (top) and incorrect (bottom) trials. Conditions diverged only in a late time window after both alpha phase effects (see middle panel, which shows alpha-band power for correct trials in black and incorrect trials in red). **C.** Event-related potentials (ERPs). No effect of correct (black) versus incorrect (red) trials was found.

Fig. 5.3A; cluster closest to statistical significance with $p = 0.81; T_{sum} = 73.96, R = 0.68$). Second, one cluster was observed in which absolute alpha power was higher for incorrect than for correct trials ($p = 0.037; T_{sum} = -3619.8, R = 0.60$). However, the cluster comprised only lower alpha frequencies (peak at 8.3 Hz) in a later post-stimulus time window (peak at 0.98 s post-(pseudo)word onset) and over more posterior electrodes (peak at CP1, see Fig. 5.3B). Third, evoked potentials did not show any difference for the accuracy contrast during the same time interval and over the same electrodes as the peri-stimulus alpha phase effect ($p = 0.13; t_{(10)} = -1.65, r = 0.46$; see Fig. 5.3C). In sum, these results support the notion that neural phase in the alpha frequency range was the best predictor for lexical decisions in noise.

5.3.4 Phase effects in the theta band.

Non-parametric permutation tests of the BI against zero also revealed a negative cluster in the theta frequency range from 3–5 Hz. The negative cluster ranged between 120 and 580 ms post-(pseudo)word onset, and was broadly distributed over electrodes ($p < 0.001, T_{sum} = -1987.8, R = 0.69$; see Fig. 5.4A).

Generally, a negative BI emerges in cases where neural oscillations in one condition are more phase locked than in another condition (Busch et al., 2009). Therefore, a negative BI should be followed-up by a comparison of ITPC. In our case, it was surprising at first glance that a whole-brain cluster-based permutation test did not reveal any ITPC differences for

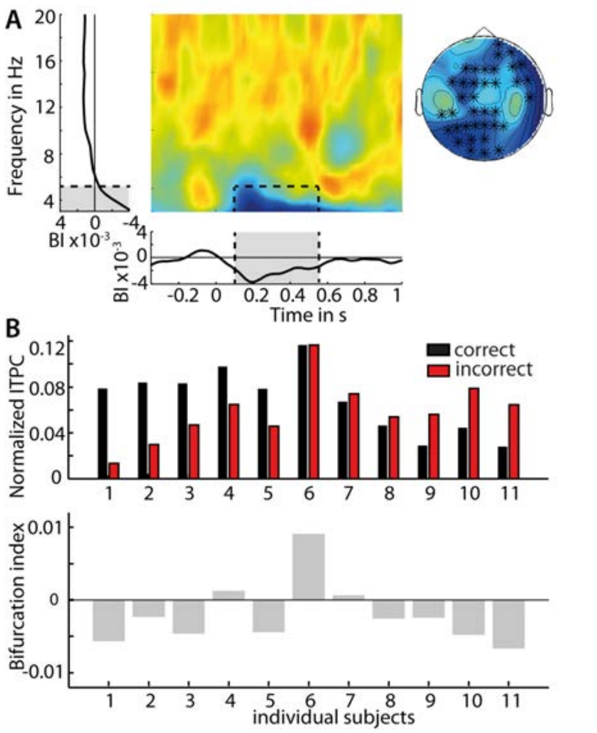

Figure 5.4: Theta-band phase effects are not consistent across participants. **A.** Negative BI cluster in the theta band extended from 120 to 580 ms with scattered scalp distribution. **B.** Normalized ITPC (middle) for correct (black) and incorrect (red) trials as required by the BI formula (see Section 5.2.5). Normalizing means subtracting the $\mathrm{ITPC}_{\mathrm{all}}$ from $\mathrm{ITPC}_{\mathrm{correct}}$ and from $\mathrm{ITPC}_{\mathrm{incorrect}}$. Conditions differ inconsistently in both directions across participants, leading to consistent but misleading negative BI (bottom).

the accuracy contrast. On closer inspection, however, the theta effect resulted from some participants showing stronger phase locking for correct than for incorrect trials and other participants showing the opposite pattern (Fig. 5.4B). This, somewhat misleadingly, led to a negative BI that survives statistical testing across participants.

5.4 Discussion

The current experiment examined the impact of slow neural oscillatory phase on word recognition. Going beyond previous work on neural phase effects in low-level perceptual tasks, we show that alpha (8–12 Hz) phase determines the accuracy of lexical decisions in perceptually uncertain situations (i.e., when stimuli are embedded in noise). The alpha phase bifurcation emerged first in a pre-stimulus (–75 ms pre-(pseudo)word onset) time window, but attained significance also in a peri-stimulus (500 ms post-(pseudo)word onset) time window.

5.4.1 Alpha phase reflects fluctuations in the probability of attentional selection

For near-threshold stimulation, pre-stimulus alpha phase has been found to determine psychophysical detection performance (e.g., Mathewson et al., 2009; Neuling et al., 2012). Consistent with and extending these results, we found a pre-stimulus alpha phase effect for a lexical decision task in noise, whereby stimuli that were judged correctly versus incorrectly coincided with opposite pre-stimulus phases of the ongoing alpha oscillation, respectively. On incorrect trials, the initial phonemes of the stimulus would thus coincide with suboptimal "windows" for sensory input (Dugué et al., 2011), that is, the inhibitory phase of an ongoing alpha oscillation. Recognition of a word-initial syllable is crucial for lexical access (Greenberg, 1999) and therefore has been emphasized in models of auditory word recognition (Taft and Forster, 1976; Marslen-Wilson, 1987). Missing the first phonemes limits the ability to enter a lexical path and recruit top-down information from the mental lexicon which is helpful in order to perform the lexical decision task in noise accurately. Thus, we suggest that coincidence of word-initial phonemes with a suboptimal phase of the ongoing alpha oscillation led ultimately to relatively poor lexical-decision performance.

We observed the pre-stimulus alpha bifurcation effect over right anterior electrodes. Although the nature of this index as a first-level statistic prevents an informed interpretation of underlying neural sources (Busch et al., 2009), this location is nevertheless consistent with previous studies that have observed the recruitment of a right frontal network (for review see Corbetta et al., 2008. Most notably, right middle frontal gyrus, frontal eye fields (Lee et al., 2014), and the right anterior insula (Eckert et al., 2009; Erb et al., 2013; Wilsch et al., 2014) have been found to be activated in particular during challenging auditory tasks. Potentially, involvement of these structures, also associated with selective attention, would have been necessary here to isolate speech from the noise background. Importantly, alpha activity has been argued to be a neural means of selecting the relevant sensory object (for more detailed discussion see Mathewson et al., 2011; Strauß et al., 2014). Moreover, the current alpha phase results are in line with the idea of Schroeder and Lakatos (2009) that low-frequency neural oscillatory phase correlates with the fluctuations of the probability that a stimulus is "selected" by attention. On this view, stimuli arriving in the optimal (excitatory) phase of the alpha oscillation are selected by attention and are thus more likely to be thoroughly processed and correctly judged than stimuli arriving in the suboptimal phase. Thus, the optimal alpha phase could have effectively increased the instantaneous signal-to-noise ratio by allowing for a more robust neural processing of the initial phonemes of the (pseudo)word.

5.4.2 Alpha phase reflects decision weighting

In the current study, we also observed an additional peri-stimulus alpha phase bifurcation over left fronto-temporal regions. Interestingly, this peri-stimulus alpha phase effect occurred directly after the crucial vowel manipulation, but was not phase-locked to the onset of the vowel. (Note that a repetition of the bifurcation-index analysis time-locked to vowel onsets did not reveal any significant clusters.) This favors a decision-related interpretation of the observed phase effect over a more stimulus-related interpretation.

The dissociation of perceptual from decisional stages and their dependence on slow neural phase is difficult in low-level detection paradigms, but has been demonstrated recently by Wyart et al. (2012) in a visual discrimination task that involved integrating visual information over approximately two seconds in order to discriminate the mean orientation of a series of Gabor patches. They found that the accumulation of perceptual evidence is not linear (as assumed previously by a number of prominent models of decision making; for review see Ratcliff and McKoon, 2008; Mulder et al., 2014) but proceeds rhythmically. Moreover, integration and weighting of decisional information were also found to be coupled to low-frequency neural phase, but were critically dissociated from (earlier) accumulation of perceptual evidence. The left anterior distribution of the later, peri-stimulus alpha bifurcation is compatible with the common finding of left inferior frontal gyrus (IFG) involvement in visual and auditory lexical decision tasks (e.g., (Fiebach et al., 2002; Xiao et al., 2005)). Especially BA 45 (i.e., pars triangularis) has been suggested to receive information from inferior temporal gyrus (Heim et al., 2009) via the ventral stream (Hickok and Poeppel, 2007) presumably to support lexical selection when lexical access is difficult (Fiebach et al., 2002). Importantly, our data suggest that this selection process in left anterior cortical structures might in part be mediated via alpha-band oscillatory activity. One remaining question concerns the relationship between pre- and peri-stimulus alpha phase in our data. We suggest that our pre- and peri-stimulus alpha phase effects reflect dissociable perceptual and decisional processes, respectively. Pre- and peri-stimulus bifurcation indices were not directly correlated (Spearman's $\rho = 0.2; p > 0.5$), suggesting at least partially independent mechanisms. Their independence is also supported by the observed difference in topographical distribution and would be in line with the interpretation of dissociable earlier perceptual and later decisional weighting (Wyart et al., 2012). In particular, our data are consistent with the necessity of achieving an optimal neural state not only during anticipation of a stimulus (for optimizing accumulation of perceptual evidence) but also during preparation of lexical decisions during and after the stimulus (for optimizing decisional weighting and integration).

5.4.3 Accuracy is not predicted by other neurophysiological measures

Strikingly, lexical decision accuracy was not predictable from other measures of neural activity such as the amplitude of the event-related potential (ERP), absolute alpha power, or inter-trial phase coherence (ITPC) in the alpha band in our data (see Fig. 5.3). That is, differences in instantaneous alpha phase seem to exhibit an independent effect on lexical decision processes and might index mechanisms that have so far not been subject to closer electrophysiological examination.

5.4.4 Theta vs alpha phase effects on lexical decision

Lastly, even though recent models of speech processing have provided good arguments to assign a crucial role to theta band oscillations (Ghitza, 2011; Gagnepain et al., 2012; Giraud and Poeppel, 2012), theta phase here was not predictive of accuracy. As a more technical aside, the multiplicative nature of the phase bifurcation index makes it insensitive to which condition is causing the negative sign of the bifurcation index. In our case, this feature could have led to the unwarranted conclusion of consistent theta-phase effects based on bifurcation statistics only. Our analysis shows that ITPC analyses are important in order to control for false positives when employing the phase bifurcation index, specifically when the observed bifurcation index is negative.

Speculatively, the current finding (i.e., consistent predictability of response accuracy by alpha, but not by theta phase) might be due in part to the type of manipulation (short-lived vowel manipulations in isolated words) or to embedding of (pseudo)words in noise, prompting an alpha- rather than theta-driven neural processing strategy (for the functional dissociation of alpha and theta activity during word recognition see Strauß et al., 2014). In sum, the available evidence from this study renders alpha but not theta phase at two separate points in time and in space a good predictor of accurate lexical decisions in noise.

5.4.5 Conclusion

This study constitutes a first step towards characterizing neural phase signatures of higher cognitive processes, such as the ones that enable spoken word recognition in noise. Our data demonstrate that alpha phase (both before and during the presentation of word or word-like stimuli) predicts the accuracy of lexical decisions in noise. The data suggest that alpha phase acts not only to select stimuli for perceptual processing, but might also underlie rhythmic fluctuations in decisional weighting. We suggest that dependence on rhythmic fluctuations in neural excitability is encouraged in particular when perceptual evidence is limited (due for example to the presence of background noise) as is often the case in naturalistic listening conditions. Therefore, both sensory processing as well as decision-making proceed coupled to ongoing internal alpha rhythms that in turn modulate performance.

5.5 Supplement: Influence of formant distances and stress patterns on behaviour

5.5.1 Introduction

In the previous chapter, we have provided evidence that accuracy of lexical decisions in noise depend on neural alpha phase. Here, correct and incorrect performance is explored in greater detail. Specifically, it is of interest how participants used phonetic, i.e. formant related, and prosodic, i.e. stress related, cues in order to support their performance in the lexical decision task in noise. These analyses were executed in order to reveal factors that should be considered in future studies of spoken word recognition in adverse listening conditions, especially to further determine the role of neural alpha phase for accurate performance.

5.5.2 Methods

Participants and Stimuli. As described in section 5.2, participants ($N = 11$) performed a lexical decision task in noise with 60 trisyllabic real words and 60 ambiguous counterparts where the nucleus vowel of the second syllable had been exchanged. The signal-to-noise ratio was individually determined by using an adaptive tracking procedure. The Two-down-one-up staircase procedure (Levitt, 1971) estimated a threshold where participants performed about 70.7% correct.
Note that in trisyllabic German words the second syllable is most frequently stressed (for discussion see (Domahs et al., 2014), which is also reflected in the frequency distribution of the current set of stimuli: $N_{1st} = 9; N_{2nd} = 35; N_{3rd} = 16$ items. At the same time, stimuli were controlled for word frequency and generally did not show much variance, ergo did not differ across stress conditions: word frequency$_{\mathrm{1st}}$ = 14.4 ± 0.9; word frequency$_{\mathrm{2nd}}$ = 13.7 ± 1.7; word frequency$_{\mathrm{3rd}}$ = 13.9 ± 1.5 (assessed by http://wortschatz.uni-leipzig.de/). Therefore, if differences in behaviour due to phonetics (i.e., formants) or prosody (i.e., stress) are found, word frequency can be ruled out as a confounding factor.

Behavioural analysis. Behavioural performance was analyzed in the signal detection theory framework (Macmillan and Creelman, 2005). This framework allows to separate perceptual sensitivity d' to discriminate words and pseudowords (zero indicates chance performance) from a response bias c to respond either "Yes, it is a word" (positive c) or "No, it is a pseudoword" (negative c). We expected a significant response bias in the sense that participants would show a preference to interpret ambiguous stimuli as meaningful words (Ganong, 1980). Both d' and c were calculated based on proportions of hits and false alarms, which were defined based on "No" responses because of our interest in detection of the critical vowel manipulation. Note that within the lexical decision task, it is not a priori clear which response category should be labeled as a 'hit'. In keeping with listeners' challenge to detect a vowel violation in a stimulus embedded in noise, we here defined hits as correct "No" responses to pseudowords.

Formant analysis. Formant analyses were based on annotated stimuli in the phonetic sound application PRAAT. Vowel portions of the stimuli underwent a linear predictive coding analysis, i.e. effectively a smoothed Fourier analysis using a 25-ms Hanning-window that slid over the vowel portion in 5-ms steps. Formants were determined as peaks of spectral power between 1 and 5000 Hz. For formant distances, only the first three formant frequencies were used. Formant frequencies were mean values of three measurements within the vowel portion (beginning, middle, end) to best capture the steady-state of the vowel. Formant distance was then calculated as Euclidean distance according to the formula introduced in Section 2.1 such that the formant distance between the first three formant values of the vowel in the real word and the vowel in its ambiguous counterpart was assessed.

5.5.3 Results

Overall accuracy of lexical decisions. Following-up the behavioural results from section 5.3.1, participants showed a mean performance of 71.4 $\pm$ 1.02 % for words and a slightly worse performance for ambiguous pseudowords 60.3 $\pm$ 0.61 % (Fig. 5.5A) which means in terms of perceptual sensitivity d' that differentiating between the two stimulus types was difficult, but better than chance ($d' = 0.86 \pm 0.12$ SEM; t test against zero: $t_{(10)} = 8.10, p < 0.001, r = 0.93$; see Fig. 5.5B). Furthermore, the difference in percentage correct for real words and ambiguous pseudowords can be explained by a slight but significant bias to respond "word" (i.e., responding "yes"; $c = 0.16 \pm 0.05$ SEM; t test against zero: $t_{(10)} = 4.47; p < 0.01, r = 0.81$; see Fig. 5.5B).

Close formant distances increase response bias. A 2×2 repeated measures ANOVA with factors wordness (ambiguous, real) and formant distance (close, far) revealed a significant main effect of wordness ($F_{(1,10)} = 11.31, p = 0.007$) confirming higher accuracy for discriminating words (correct "Yes" responses) than discriminating pseudo-words (correct "No" responses; Fig. 5.5C). The main effect of formant distance was not significant ($F_{(1,10)} = 0.54, p = 0.48$) but the interaction of both factors was ($F_{(1,10)} = 32.245, p = 0.000$). Post-hoc t tests showed that if formant distances were far apart, accuracy for both conditions was statistically indistinguishable ($t_{(10)} = 0.19, p = 0.85$). If formant distances were close, accuracy dropped for ambiguous stimuli to 53% (statistically not different from chance = 50%: $t_{(10)} = 1.31, p = 0.22$). Interestingly, this was not due to a lack of perceptual sensitivity (d' of close versus far formant distances: paired t test $t_{(10)} = -0.25, p = 0.81$) but due to a response bias to answer "Yes, it is a word" as shown in Figure 5.5D (c for close formant distances: t test against zero $t_{(10)} = 5.70, p = 0.0002$).

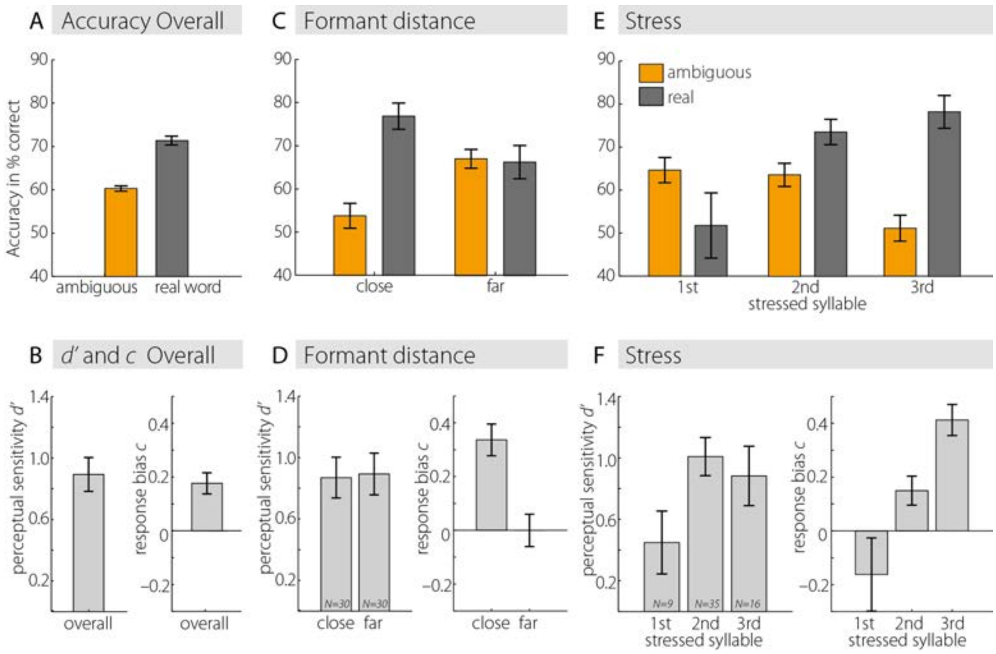

Figure 5.5: Accuracy, perceptual sensitivity d', and response bias c for the lexical decision task in noise. Bars plot the mean ± 1 SEM. **A.** Overall accuracy for ambiguous pseudowords and real words. **B.** Overall d' and c. **C.** Accuracy split by formant distance (median split). If vowel distances in formant space are far apart between real words and their ambiguous counterparts, accuracy is the same for both conditions. If they are close, performance drops for ambiguous stimuli to chance level. **D.** d' and c split formant distance. A decline in d' but not c is probably underlying the decline in accuracy when formant distances are close (**C**). **E.** Accuracy split by stressed syllable. Notably, performance declines for ambiguous stimuli if stressed on the third syllable whereas for word stimuli if stressed on the first syllable. **F.** d' and c split by stress. Chance level performance to discriminate word stimuli stressed on the first syllable (**E**) are due to a decline of d'. Chance level performance to discriminate ambiguous stimuli when stressed on the third syllable (**E**) is instead reflected in a bias to respond "Yes, it is a word".

First syllable stress reduces perceptual sensitivity. As illustrated in Figure 5.5E, a 2×3 repeated measures ANOVA with factors of wordness (ambiguous, real) and stress (1st, 2nd, or 3rd syllable) disclosed a main effect for wordness only on trend level ($F_{(1,10)} = 3.67, p = 0.085$), but a significant main effect for stress ($F_{(2,20)} = 6.03, p = 0.017$) and a significant interaction between the two factors ($F_{(2,20)} = 13.799, p = 0.001$). Post-hoc t test determined on the one hand a significant drop in accuracy if ambiguous pseudowords were stressed on the third syllable (1st vs 2nd: $t_{(10)} = 0.27, p = 0.79$; 1st vs 3rd: $t_{(10)} = 3.04, p = 0.012$; 2nd vs 3rd: $t_{(10)} = 3.81, p = 0.0034$). This is reflected in a strong bias to respond "word" for third-syllable-stressed stimuli (t test of c against zero: $t_{(10)} = 7.12, p = 0.000032$; Fig. 5.5F). On the other hand, stress on the first syllable was detrimental for word discrimination (1st vs 2nd: $t_{(10)} = -3.54, p = 0.0053$; 1st vs 3rd:

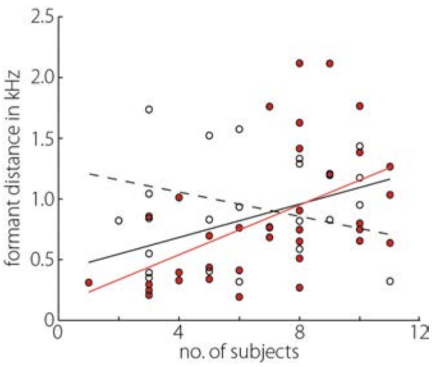

Figure 5.6: Correlation of formant distance and number of participants that correctly responded "No" to ambiguous stimuli. Black circles depict all formant distances (N=60) and the black line their positive correlation with no. of subjects. Reddish filled circles mark the subset of 2$^{\text{nd}}$-syllable-stressed stimuli (N=35) and the red line their positive correlation. Dashed line depicts absent correlation of correct "word" responses and formant distances.

$t_{(10)} = -3.83, p = 0.0033$; 2nd vs 3rd: $t_{(10)} = -1.57, p = 0.15$). This is reflected in a d' for first-syllable-stressed stimuli that is only different from zero on a trend level (t test of d' against zero: $t_{(10)} = 2.19, p = 0.053$; Fig. 5.5F).

Correlation of performance with formant distance and stress. Formant distances between conditions were correlated with the number of participants that correctly responded to the ambiguous stimulus with "No, not a word" (Spearman's $\rho = 0.39, p = 0.0018$) marked by the solid black line in Figure 5.6. Thus, the further apart real words and their ambiguous counterparts were, the more participants correctly judged ambiguous stimuli as nonwords. Interestingly, correct word discrimination did not depend statistically significantly on formant distance ($\rho = -0.20, p = 0.12$) as depicted by the dashed line. The red line shows the correlation of formant distance with a subset of ambiguous stimuli, namely the ones which had been stressed on the second syllable ($\rho = 0.56, p = 0.0005$). The correlation was not significant for the subsets stressed on the first or third syllable. Caution is advised because these non-significant correlations might be due to smaller subset sizes (N=9 and N=16 compared to N=35).

5.5.4 Discussion

Supplementary analyses of the behavioural data revealed influences of phonetic and prosodic cues on the accuracy of lexical decision performance in noise. In particular, phonetic characteristics as measured by formant distances between vowels shifted the response bias c towards "word" responses if formants were close between conditions. Formant distance was especially relevant for the judgement of pseudowords stressed on the second syllable as revealed by the correlation shown in Figure 5.6. Prosody as measured by stress pattern, instead, had a differential influence on word vs. pseudoword judgements. Stress on the first syllable decreased perceptual sensitivity d' reflected in an accuracy decrease in judging word stimuli as words. Third syllable stress increased the bias c to respond "word". This is also reflected in an accuracy decrease in judging pseudowords as not a word.

Discussing the neural phase effects in the previous section, we argued that the two successive alpha phase effects distinguish between attentional gain for sensory discrimination first and for decisional weighting later. Interestingly, sensory discrimination as assessed by perceptual sensitivity on the behavioural level is only reduced for words stressed on the first syllable. This is somewhat counterintuitive, since the stress on the first syllable should in particular support lexical access in noise (Mattys, 2004). It might be a peculiarity of our study that did not use bi- but tri-syllabic stimuli which are by default stressed on the second (but not first) syllable. It has been shown that first-syllable stress in tri-syllabic words is only preferred by subjects with higher working memory capacity (Domahs et al., 2014).

One explanation, therefore, might be that the prominent first syllable triggered an attentional-blink- or forward-masking-like phenomenon that masked the second (and crucial) syllable (Horváth and Burgyán, 2011). Such a forward masking phenomenon would have been particularly detrimental in our paradigm, which manipulated the second-syllable vowel exclusively.

Alternatively coming from the brain data results, pre-stimulus alpha phase might had additionally a detrimental effect on the recognition of the first syllable. Ongoing slow oscillatory phase has been shown to modulate sensory information processing (Lakatos et al., 2008; Mathewson et al., 2009; notably, there seems to be a relationship between alpha oscillations and the attentional blink: Hanslmayr et al., 2011). Given that stressed syllables provides the most salient cue for word recognition (Altman and Carter, 1989; Gow and Gordon, 1995), the modulation by pre-stimulus alpha phase might had a detrimental effect on word recognition especially in the case of first syllable stressed items. Thus, during the nonoptimal alpha phase perceptual sensitivity to the most important cue was diminished (compare $d'_{1st} = 0.45, d'_{2nd} = 1.01, d'_{3rd} = 0.88$) thereby reducing word ($51.8 \pm 7.6\%$) but not pseudoword recognition ($64.6 \pm 2.9\%$). A feasible hypothesis for future studies therefore would be that a nonoptimal pre-stimulus alpha phase is especially detrimental for perceptual sensitivity and recognition of first-syllable stressed words but not for other stress patterns irrespective of the vowel manipulation done in the current study.

For the later alpha phase effect around 500 ms, we argued in the previous chapter that perceptual evidence is integrated and the decision is updated accordingly (Wyart et al., 2012). This interpretation is in line with the observation that perceptual sensitivity is not different for stimuli stressed later than the first syllable, namely the second or third syllable. Errors at these later stages during the lexical stimulus are exclusively due to response bias. Responses to pseudowords stressed on the second syllable are biased towards "word" responses depending on their formant distance to their real word neighbour. Responses to pseudowords stressed on the third syllable seem to be biased towards "word" judgements simply because of their third-syllable stress. In this case, the task-relevant information had already passed which might index backward masking.

These behavioural results underline the importance of syllable stress for word recognition in noise (Mattys, 2004). Notably, failures could occur due to a decrease in perceptual discrimination or to a change in response bias. Future studies need to clarify whether alpha phase effects account for decreases in perceptual sensitivity or for biased decisions, or for both.

5.6 Supplement: The bifurcation index and its dependencies

5.6.1 Introduction

In section 5.3, we reported results when analysing the phase bifurcation index as proposed by (Busch et al., 2009). In the case of a negative bifurcation index (observed for frequencies in the theta range, see section 5.3.4), we found some inconsistencies on the single subject level when following-up consistent group statistics. To further test the limitations of the phase bifurcation index (BI), simulations have been run in Matlab^{TM} by drawing random variates from the Von Mises circular distribution (Fisher, 1993) and by systematically changing parameters of interest.

5.6.2 Methods and Results

For the first three tests, one condition was set perfectly phase locked ($\alpha = 0°$, $\kappa = 4.5$, i.e. ITPC = 0.9) while the second condition parametrically varied in different dimensions.

Number of trials. We tested the stability of the BI in terms of minimum number of trials that are needed to get a reliable estimation of phase differences. Therefore, the second condition was chosen to be perfectly phase locked to the opposite phase angle ($\alpha = 180°$; the other parameters being identical: $\kappa = 4.5$, i.e. ITPC = 0.9). The number of trials for which these random variates were drawn increased from 5 to 15 to 30 and finally to 60 for both conditions.
Results are shown in Figure 5.7A. BI estimation across 5 trials appear very unstable, whereas with 15 trials drop-outs were already tremendously reduced. At 30 trials, best performance is reached. In our case, 29.45 $\pm$ 6.25 trials were used per subject (see section 5.2.5) which is sufficient for reliable BI estimation according to the current simulation. The fact that BI calculation was repeated 1000 times, which were then averaged for the final BI submitted to statistical analyses, additionally ensured BI's reliablility. For the following simulations, trial number is fixed at 1000 to exclude any possibility of variability in the BI estimation.

Inter-trial phase coherence. We tested the influence of inter-trial phase coherence (ITPC) on BI. While the first condition remained identical in terms of mean angle and ITPC, the second condition now successively increased in phase locking to the opposite angle ($\alpha = 180°$, κ increased in 20 steps from 0 to 4.5).
Results depicted in Figure 5.7B revealed that the slope approximates linearity which means that the BI increases (almost) linearly with ITPC increase. This is an important feature for interpreting the BI such that positive BI values not only indicate an anti-phasic relationship but also indicate that if BI values are more positive, conditions had been also more phase-locked to a particular mean angle. This feature also qualifies the BI as a nonparametric

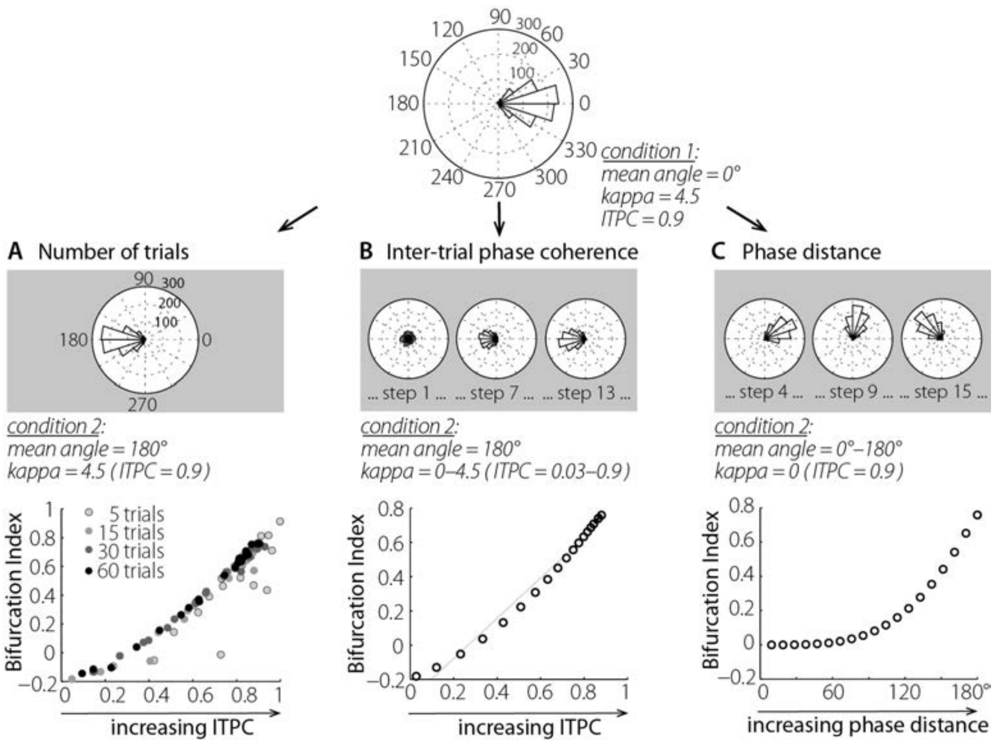

Figure 5.7: Features of the phase bifurcation index. The first condition shown above panels A, B, and C, is fixed at the parameters written on the right side. The second condition is shown on grey backgrounds. Parameter settings are written below the grey bar. Scatter plots on the bottom show the results when calculating the BI for condition one and two. **A.** Dependence of BI estimation on number of trials. Grey nuance scales with increasing number of trials that went into BI calculation. **B.** Dependence of BI on ITPC. 1000 trials were used. Scatter plots shows that the BI increases (almost) linearly with ITPC increase. **C.** Dependence of BI on phase distance. 1000 trials were used. Scatter plots shows that conditions need to be at least 90°apart in order to be detected by a positive BI.

test measure for phase analysis of time-domain brain data. More caution is advised when using negative BI values. Interestingly, BI never undershoots a value of –0.2, whereas the positive range is exhausted up to 1 under ideal conditions (these high BI values might actually never occur in empirical data), making the BI asymmetric. One has to keep in mind that this is reducing variability in the negative range with possible consequences for calculating t statistics.

Phase distance. We asked for the minimum phase distance which would be detected by the BI. Thus, the second condition varied highly phase locked ($\kappa = 4.5$, i.e. ITPC = 0.9) to a circumnavigating mean angle in 20 steps from 0°to 180°. Results in Figure 5.7C suggest that conditions need to be at least 90°apart in order to be detected by a positive BI.

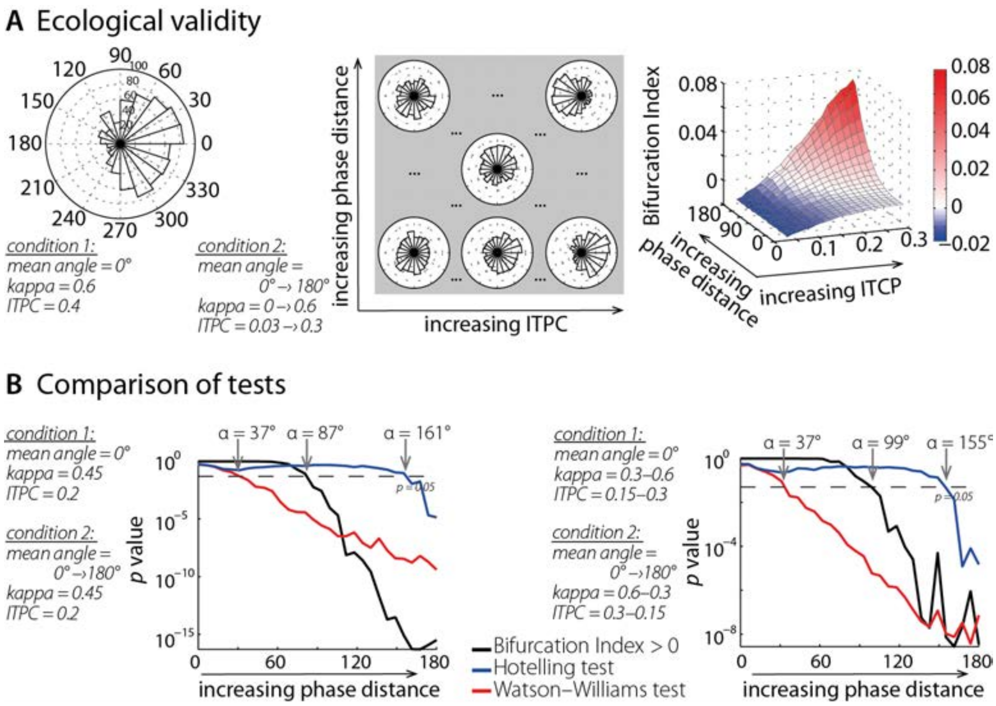

Figure 5.8: A. Ecological validity. The rose plot on the left side shows the distribution of phase angles for condition one which corresponds to high phase-locking in empirical data. Condition two varies along two dimensions, ITPC and phase distance to condition one, as depicted in the middle panel. On the right side BI is plotted depend on increasing ITPC and increasing phase distance. **B. Comparison of tests.** p Values for BI (black line), Hotelling test (blue line), and Watson-Williams test (red line) are compared. Arrows indicate the minimum of phase distance in degree needed to be able to detect a significant difference when using one of these three approaches.

These characteristics raise the question whether ITPC and phase distance interact. It could be that bigger phase distances are needed when angle dispersion is more natural. Furthermore, if in an ideal case like Figure 5.7C the minimum phase distance is already a quarter of a cycle, maybe other (parametric) tests like the Watson-Williams test or the Hotelling test would detect more subtle phase differences.

Ecological validity. Therefore, we checked next BI behaviour in more ecologically valid terms which primarily concerns much higher variance, i.e. lower κ or lower ITPC, observed in electrophysiological data. Hence, condition one was again phase locked at $\alpha = 0°$, but with $\kappa = 0.6$, which translates to an ITPC of 0.4. Both factors, angle α and variance κ, varied parametrically in condition two. Results are summarized in Figure 5.8A. The greater the angle of the two conditions is the more sensitive BI will be, which means the faster BI increases positively. If angles are less than 90°different from each other BI rather becomes zero. If condition one is phase locked with an ITPC of 0.4, condition two has to have an ITPC greater than 0.15 in order to be considered as meaningfully phase locked. In sum, at medium ITPC which is very common in natural systems the chance of BI being zero increases, thereby diminishing the probability to detect more subtle ITPC or angle changes.

Comparison to other tests. The Hotelling test as well as the Watson-Williams test are parametric tests for angle differences applicable if dispersions of the to-be-compared samples are the same. In our experiment, BI was implemented as a first level, i.e. within subject, contrast, which was tested in a paired t-test against zero on the second or group level. In contrast, the Hotelling and the Watson-Williams test provide F statistics on the group level only. The advantage of the Hotelling test is that it also considers the ITPC (like BI) whereas the Watson-Williams test only determines differences of mean angles. In order to have a fair and ecologically valid comparison between the two measures, we simulated p values on the basis of subject (N=11) and trial number (N=60) on the basis of the results reported in Section 5.3. Final p values displayed in Figure 5.8B are the mean over 100 iterations of this simulation.
Interestingly, the Watson-Williams test is the most liberal test such that only 37°are needed to detect significant differences. The Hotelling test appeared to be the most conservative measure so that only phase distances greater than 161°will be detected given ecologically valid phase-locking values. BI requires angle distances >87°to become significant and nicely integrates ITPC (please compare Fig. 5.8B left and right panels).

5.6.3 Discussion

In sum, BI was for our purposes the best way to test phase differences nonparametrically. First, testing increasing number of trials amongst which BI is calculated showed that at least 30 trials per condition should be used in order to get stable BI estimates. Second, ITPC has a quasi-linear influence on BI. Third, BI is sensitive to detect phase distances bigger than about 90°. Fourth, this is also true in more ecologically valid conditions as revealed by testing BI within lower ITPC ranges. Finally, in comparison to other phase difference tests, BI proves to be especially suitable for M/EEG data. Limitations of the BI remain mainly when considering the negative BI. Here, the simulations suggested to preferably use ITPC measures.

6 Narrowed expectancies in degraded speech[5]

6.1 Introduction

When hearing speech, listeners can use at least two streams of information: perceptual information provided by the speech signal itself, sometimes referred to as the "bottom-up" stream; and predictions or expectancies, denoted as the "top-down" stream. This top-down stream is, of course, commonly dependent on global discourse knowledge, but in this study it is used in a more specific sense of accumulated semantic context as the signal unfolds.

It is unclear how these two streams interact, particularly at the neural processing level. An intuitive assumption would be one of top-down expectancies becoming dominant whenever bottom-up perceptual evidence is ambiguous. Without doubt, top-down phenomena, where patchy or ambiguous perceptual evidence is filled in, are a powerful mechanism (in the visual domain: Tallon-Baudry and Bertrand, 1999; in the auditory domain: e.g., Sivonen et al., 2006; Riecke et al., 2009). However, recent psycholinguistic and psychoacoustic research has emphasized that the opposite may also be true: Acoustic challenges have been shown to prompt listeners to focus on the bottom-up perceptual evidence, as opposed to mainly relying on top-down (contextual, i.e., semantic) cues (Mattys et al., 2009).

Of the few studies on semantic cues in degraded speech that exists, most have operated with a unitary, simplified concept of "context". This, necessarily, confounded various linguistic aspects that might be differentially affected by speech degradation – only an experimental separation into various "levels" of context would allow investigation of whether the expectancy forming neural mechanisms are differentially susceptible to degradation.

The current study attempts to study the specific interactions of degradation and expectancy formation at the neural level, using a simple and well-established marker, the N400 component of the event-related potential (ERP; Kutas and Hillyard, 1980; see below).

[5] This chapter is adapted from the published article by Strauß, Kotz, and Obleser (2013). *J Cogn Neurosci 25(8)*, 1383–1395.

6.1.1 Semantic context

As outlined above, expectancies can derive from various linguistic factors. Early experiments provided evidence that the recognition of a word occurs faster in a sentence context, compared to isolated or listed word presentations (Miller et al., 1951). Besides the syntactic structure that sentences provide (Miller and Isard, 1963), the semantic context of congruent or predictable sentences facilitates processing (Kalikow et al., 1977; Kutas and Hillyard, 1980; Stanovich and West, 1983).
How can the benefits from semantic context be measured? In their seminal study, Kutas and Hillyard (1980) introduced the "cloze test" (originally developed by Taylor, 1953) to find a quantitative evaluation of sentence ending probability. In this test, participants have to complete sentences with the most likely word that comes to their mind, capturing implicit knowledge about contextual suitability. A number of studies have consistently replicated the benefit of high over low sentence ending probability (e.g., Connolly et al., 1990; van den Brink et al., 2006; Van Petten and Luka, 2006; Friedrich and Kotz, 2007; Obleser and Kotz, 2010). Jurafsky (2003) argued that one reason for this could be that people make overly crude distinctions between congruence and incongruence or high and low predictability. In fact, neither of these concepts is categorical but, rather, they operate on a continuum in natural languages. In the past, there was a lack of a priori criteria and measures of congruency and predictability to allow for the parametric variation of such concepts. As a result, many studies confined themselves to investigating effects of single word frequency on spoken word recognition (Howes, 1957; Luce and Cluff, 1998; Benki, 2003; Cleland et al., 2006), or—more complexly—to looking at the effects of bigram frequency (e.g., Ferrand et al., 2011).
During the last decade, the collection of huge text corpora and the establishment of computational tagging algorithms have made it possible to calculate several frequency-based interdependencies of words. This puts us in the position of being able to generate a continuum of context-based typicality, the probability of a word given some previous context, which not only respects the single lexical frequency, but also bigram probabilities and lexical class probabilities (Geyken, 2009; a psycholinguistic term for this being collocation). The sensitivity gained by quantifying the contextual relation within a sentence will be utilized in the present study.

6.1.2 Neural signatures of context in language comprehension

A prominent component of the ERP in response to words, as measured by electroencephalography (EEG), is the N400. This negativity, peaking at around 400 ms after word onset, is used as a neural indicator of context-based expectations and actual word input. Kutas and Hillyard (1980) reported the first observation of increased amplitude in response to an incongruent sentence ending word (for reviews see: (Kutas and Federmeier, 2000;

Lau et al., 2008; Van Petten and Luka, 2012)). Van Petten and Kutas (1990) found a general positive shift of ERP amplitudes, i.e., a reduction of the N400, the later a word appeared in an unfolding sentence. Halgren et al. (2002) showed that when a open-class content word appears in earlier sentence positions, the brain activation in the N400 time range is less wide spread in left temporal cortices. Both studies interpreted their results as reflecting the insufficient amount of predictive context up to this point in the sentence. This could also explain the high sensitivity to (semantic) violation at sentence endings.

Besides effects of repetition, word frequency, and sentential context on the amplitude of the N400, Federmeier and Kutas (1999) also found an influence of categorical typicality. This allows differentiation of sentential semantic context from expected semantic features. An example of a sentence context in their study was: “They wanted to make the hotel look more like a tropical resort, so along the driveway they planted rows of...”. “Palms” would be the highest cloze probability completion because, first, the context constrains the sentence ending to a tropical plant and, second, palms are prototypical representatives of tropical plants. Moreover, the authors found not only a reduced N400 in response to “palms”, but a moderately reduced N400 in response to “pines”, and the most pronounced N400 in response to “tulips”, suggesting that palms and pines share more semantic features than palms and tulips. Both, palms and pines, belong to the same category “tree”, but in the context of the tropics, palms are more typical than pines. However, the term categorical typicality does not describe the relationship between sentence constituents, but, rather, relies on prototype theory and feature semantics, which is why it is hard to extend this to other word classes, such as verbs or adjectives.

Therefore, we looked for a measure of typicality based on collocation statistics that would capture the distribution of a word and its contextual co-occurrence probabilities. This would relate our findings back to sentential semantic constraints and not to the hierarchical organization of prototypes in the mental lexicon (even though this hierarchy is, to some extent, context dependent, as D’Arcy et al., 2004, and Federmeier and Kutas, 1999, have shown). In short, the current study focuses not on the categorical typicality but on the sentence context-based typicality.

6.1.3 Semantic benefits in adverse listening

While a whole tradition of behavioural studies have laid the ground work for understanding cognitive processes in adverse listening conditions (Miller et al., 1951; Kalikow et al., 1977; Stickney and Assmann, 2001; Pichora-Fuller, 2003; Davis and Johnsrude, 2003; Mattys et al., 2009), only a few neuroimaging studies (e.g., Obleser et al., 2007; Obleser and Kotz, 2010; Davis et al., 2011; McGettigan et al., 2012) and EEG studies (e.g., Connolly et al., 1992; Aydelott et al., 2006; Boulenger et al., 2011; Obleser and Kotz, 2011; Romei et al., 2011) have taken on the issue of semantic or expectancy benefits in degraded speech.

For example, Aydelott et al. (2006) contrasted natural with low-pass filtered speech signals and showed a reduced N400 effect in response to incongruent sentence-final words under acoustic degradation. Likewise, Obleser and Kotz (2011) used very simple German sentences, varying in cloze probability under three degradation levels, and found the cloze-driven N400 amplitude decreased linearly with more signal degradation. In an fMRI version of this paradigm, the same authors showed that the cortical extent of activation in the superior temporal cortex not only varied with degradation (better signals yielding stronger and more extended activations along the entire superior temporal gyrus and sulcus; STG/STS), but that this degradation effect was modulated by contextual predictability: For high-cloze sentences, the degradation effects were confined to areas within and surrounding primary auditory areas, in contrast to the wide-spread bilateral anterolateral STG/STS activation for low-cloze sentences. This hints to a narrowing or pruning of brain activity, dependent on good predictions and moderate signal quality (Obleser and Kotz, 2010).

The present study aimed to specify how expectancies from context are formed and adjusted over the time course of a sentence under various degrees of acoustic degradation. We aimed to study this phenomenon by using an established, time-sensitive, and comparably simple-to-acquire neural parameter (the N400 component of the ERP).

The design crossed a three-fold factor degradation with a three-fold factor semantic expectancy, which combined "context" and "typicality" manipulations (Fig. 6.1). "Context" of a sentence-final keyword was manipulated, as in a large number of previous studies, via the preceding sentence context: highly constraining verbs often co-occur with fewer specific nouns than low constraining ones (± con). In the strong context, however, we additionally varied what we refer to as the "typicality" of the sentence-final object: We distinguished between high and low frequency co-occurrences of the $verb - object_{[AKK]}-$relation (+con ± typ). These choices were validated using corpus analysis (collocations) and empirical cloze tests (see section 6.2).

We first hypothesized that the presence of any expectancy effect (i.e., context or typicality) in the N400 window would depend on signal quality. We, thus, expected strongest N400 effects under clear-speech conditions. Second, we expected the context manipulation to be more salient than the comparably subtle typicality manipulation. Our third question, however, was the pivotal one: Would broad effects of "context" and more subtle effects of "typicality" behave the same under degraded and clear speech conditions? If acoustic degradation elicits a sharper or more specific adjustment of linguistic predictions as a sentence unfolds in time, then only the most typical word in a given context should match a formed prediction and effectively reduce the N400 effect.

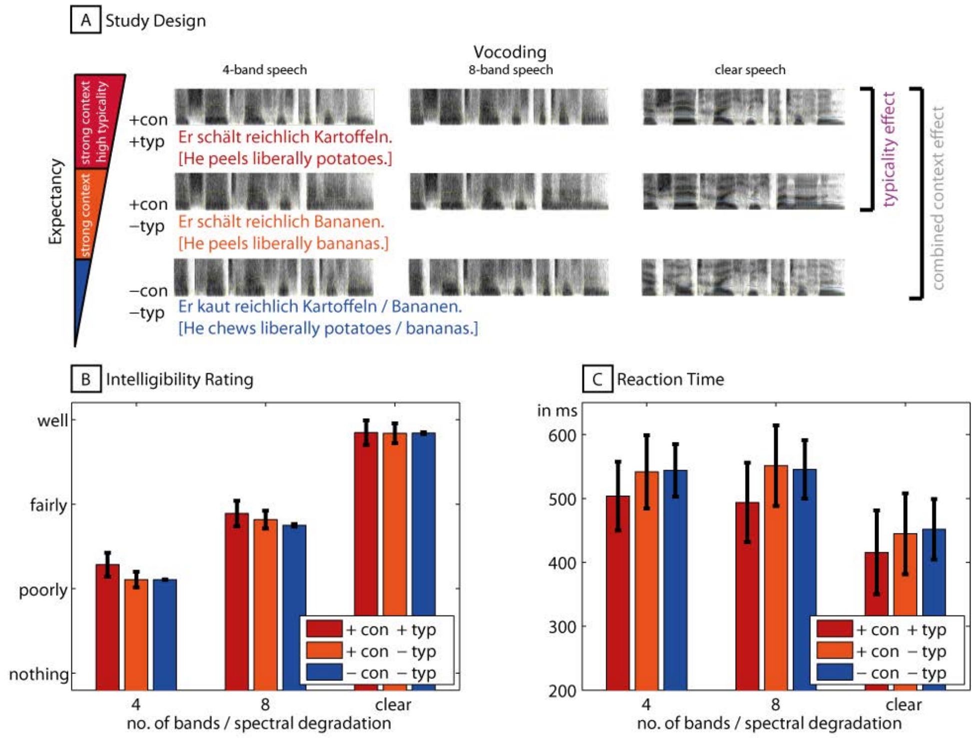

Figure 6.1: Study design and behavioural data. A. Study design with factors Degradation and Expectancy and the two differences of interest. The effect of combined context is defined by condition -con -typ minus condition +con +typ, and the effect of typicality only by +con -typ minus +con +typ. **B** and **C.** Behavioural results of the EEG experiment (mean ± 1 SEM).

6.2 Methods

6.2.1 Participants

Twenty participants (13 female, 7 male; mean age = 25.7 years, S.D. ± 2.64) took part in the auditory electroencephalography (EEG) experiment. All of them were native speakers of German and right-handed, with self-reported normal hearing abilities, no history of neurological or language-related problems, and no prior experience with vocoded speech. They gave their informed consent and received financial compensation for their participation. Thirty different participants were recruited for a behavioural pilot study. All procedures were approved by the ethics committee of the University of Leipzig.

6.2.2 Stimuli and design

The study design was based upon three kinds of German sentences, varying in semantic context (± con) and context-based typicality (± typ), which will be outlined below in more detail. These were presented at three levels of speech signal degradation (severely degraded 4-band speech, moderately degree 8-band speech, and clear speech).

All sentences consisted of pronoun (“er” masc. vs. “sie” fem.), verb (in the present tense), adverb, and object. The neutral bi- or tri-syllabic adverb (e.g., “häufig” [often]), was inserted to temporally separate the two parts of interest (verb and object). Part of the material had already been used in previous studies on cloze probability (Gunter et al., 2000; Obleser and Kotz, 2010, 2011). For the present study, the material was revised based on collocation statistics in the DWDS-Corpus (Digitales Wörterbuch der deutschen Sprache: www.dwds.de, edited by Berlin-Brandenburgische Akademie der Wissenschaften).

The corpus provides a measure of salience (Lin, 1998), on the basis of mutual information (MI; i.e., whether a combination of words co-occurs more often than chance). Different from the MI, however, the relative frequencies are not calculated over the whole text corpus but with respect to the syntactic relation (Geyken, 2009). This is especially relevant for the German language because the simple KWIC (Key Word In Context; also concordance), often used in English corpora, is inappropriate due to the less constrained word order and case syncretism in German (Geyken, 2009). In order to determine a meaningful measure of salience, word combinations have to co-occur at least four times in a specific syntagmatic relation in the DWDS corpus.

The semantic context was evoked by verbs with either strong or weak collocations: An ideal strongly determining context would have few co-occurring accusative objects and only one very frequent accusative object (e.g., “schält – Kartoffeln” [peels – potatoes]), whereas a weakly determining context would have a lot of equally (low) frequent alternative accusative objects (e.g., “kaut – Brot / Kaugummi / Fingernägel / Kartoffeln / etc.” [chews – bread / chewing gum / fingernails / potatoes / etc.]). The context-based typicality was manipulated within the same semantic frame that each verb required. In the case of a strong context, this would be the contrast between the one very high-frequency candidate and the one very low-frequency candidate (but nevertheless co-occurring). In the case of a weak context, both candidates were selected to be equally probable. Therefore, we defined “high-typical” in the present study as a frequency tagging of the $verb - object_{[AKK]}$−relation greater than four (e.g., “schält – Kartoffeln”, [peels – potatoes]) whereas “low-typical”, that is, non-salient combinations would be tagged fewer than four times in the DWDS corpus (e.g., “schält – Bananen”, [peels – bananas]).

The less typical object of the semantic frame (e.g., “Bananen”) always differed from the more typical target from the first phoneme on; where possible, the syllabic structure and stress pattern of the high-typical and low-typical objects were matched. In sum, 160 different sentences (40 themes × all 4 possible verb-object-combinations) were created.

We also checked for single-word frequency: In order to estimate spoken word frequencies for all verbs and objects, we used the corpus of German movie subtitles, SUBTLEX, which has been shown to correlate better with lexical decision times than CELEX measures (Brysbaert et al., 2011). Word frequency = log10(item count + 2). High and low constraining verbs did not differ in their single word frequency (high: 1.49 ± 0.77, low: 1.69 ± 0.87), but typical objects were more frequent than untypical objects (typical: 2.41 ± 0.64, untypical: 1.94 ± 0.83). Note that, by manipulating the verb, we varied the occurrence probabilities of the objects so that single word frequencies would not play a role. More specific verbs are likely to be used in fewer contexts, which is why they are semantically constraining. Single word frequency of an object suitable for a highly constrained context can be high, which in our case condenses in a higher collocation frequency of the verb-object-combination, i.e., our main manipulation. This is because "Kartoffeln" [potatoes] and "schälen" [to peel] predict each other equally well, irrespective of what occurs first in a sentence.

All sentences were spoken by a phonetically trained female speaker and recordings were digitized at 44.1 kHz. Post-editing included down sampling to 22.05 kHz, cutting at zero crossing, and RMS normalization. Additionally, each of the clear speech sentences was spectrally degraded using a Matlab^{TM}-based noise-band vocoding algorithm (70–9000 Hz, all vocoding-band envelopes smoothed with a 256-Hz zero-phase Butterworth low-pass filter). Levels of spectral degradation, that is, numbers of bands, were chosen according to a behavioural pilot study (see below).

6.2.3 Pilot study

In order to select appropriate vocoding levels, we used a procedure for pre-testing degraded speech stimuli (previously conducted by e.g., (Obleser et al., 2007; Eisner et al., 2010; Obleser et al., 2012): Participants (N = 30; 15 females), who were not part of the EEG study described here, listened to all sentences at 5 different degradation levels (2-, 4-, 8-, 16- and 32-band speech) and were instructed to type what they just heard. The first trial always picked a stimulus of the least degraded signal quality and was followed by a pseudo-randomized order of sentences and degradation level. Feedback about correctness of response was provided only for the first ten trials. Breaks were possible at participants' own discretion, resulting in an experimental duration of around 45 minutes. Accuracy was measured by taking the mean number of verb and object matches between each played sentence and the typed input.

For the EEG experiment, we identified 8-band speech as the critical condition, flanked by clear speech (no degradation) and 4-band speech (hardly intelligible). These were selected because, first, it was at 4 and 8 bands that the expectancy manipulation influenced comprehension. This was not, or only weakly, the case for 2-, 16- and 32-band speech because of ceiling effects. 4-band speech: +con+typ, e.g., "schälen – Kartoffeln" [peels – potatoes] = 63.8%; +con–typ, e.g., "schälen – Bananen" [peels – bananas] = 38.6%; –con+typ, e.g., "kauen – Kartoffeln" [chews – potatoes] = 41.3%; –con–typ, e.g., "kauen – Bananen" [chews – bananas] = 40.2%. 8-band speech: +con+typ, e.g., "schälen – Kartoffeln" [peels – potatoes] = 93%; +con–typ, e.g., "schälen – Bananen" [peels – bananas] = 88.6%; –con+typ, e.g., "kauen – Kartoffeln" [chews – potatoes] = 91.8%; –con–typ, e.g., "schälen – Bananen" [chews – bananas] = 73.6%. Second, 8-band speech yielded levels of comprehension that were approximately intermediate between 4-band and highly intelligible 32-band speech.

6.2.4 Electroencephalogram acquisition

The electroencephalogram (EEG) was recorded from 64 Ag-AgCl electrodes, positioned according to the extended 10-20 standard system, on an elastic cap with a ground electrode mounted on the sternum. Electrooculogram (EOG) was acquired bipolar at a horizontal (left and right eye corner) and a vertical (above and below left eye) line. All impedances were kept below 5kΩ. Signals were referenced against the left mastoid and digitized online with a sampling rate of 500 Hz.

In an electrically shielded and sound-proof EEG cabin, participants were instructed to listen carefully to sentences and rate them according to their intelligibility on a scale from 1 to 4, where "1" meant "not at all comprehensible" and "4" meant "perfectly understandable" (see Davis and Johnsrude, 2003; Obleser et al., 2012; Obleser and Kotz, 2011 for previous use of this rating task and close correspondence to actual comprehension). Responses were given via button press and the button order was counterbalanced across participants. Seated comfortably in front of a computer screen, each participant listened to all 160 sentences at 3 degradation levels (in total 480 trials).

After each sentence, a question mark appeared on the screen prompting participants to give a rating. Subsequently, an eye-symbol marked the time period for a "blink break". Duration of the blink break and onset of the next sentence were jittered to avoid a contingent negative variation. Before the actual experimental trials, a short familiarization session was provided consisting of 10 trials (excluded from the analysis). Overall duration of the experimental procedure was about 1 hour.

Sentences were presented in a pseudo-randomized order so that no more than 2 stimuli of the same signal quality were presented in succession and a clear-speech sentence was heard in at least every fifth trial; also, the different expectancy manipulations belonging to one theme (i.e., one set of 4 sentences) were also presented at least 20 trials apart. The order of the clear speech and the degraded speech versions of one theme were changed across subjects to counteract facilitation through repetition. Nevertheless, contexts and objects were heard twice for each theme and, additionally, at three degradation levels, which is, in total, 6 times across the whole experiment. Despite carefully randomization, it remains possible that repetition lead to a reduction of the present effects. However, splitting trials would have prohibitively lowered the signal-to-noise ratio.

Individual electrode positions were determined after EEG recording with the Polhemus FASTRAK electromagnetic motion tracker (Polhemus, Colchester, VT, USA).

6.2.5 Data analysis

Offline pre-processing of data included: re-referencing to linked mastoids, a finite impulse response (FIR) high pass filter at 0.03 Hz for drift removal, and automatic artifact rejection when EOG channels exceeded $\pm$ 60 μV. Two different trigger points were used to average the EEG Signal: For early event related potentials (ERP) extraction, epochs of 3.2 s (200 ms prestimulus baseline) were averaged, centered around the onset of sentences. For N400 analyses, the mean of the 2.2 s epochs (200 ms pre-stimulus baseline), centered on the beginning of the sentences' final keywords, were considered.
Early ERP responses were analyzed at the Cz. For the N100, two time windows of interest were defined: 50–100 ms, and 100–150 ms, splitting the N100 into an early and a late time window in order to derive conclusions about the latency differences of the acoustic manipulation. For the P200, one time window from 150–300 ms was identified. A repeated measures analysis of variance (ANOVA) with the three-fold factor of degradation (4-, 8-band, and clear speech) was calculated for each time window separately.
For the later ERPs associated with semantic processing, we merged both weak-context versions (e.g., "Er kaut reichlich Kartoffeln" [He chews liberally potatoes], "Er kaut reichlich Bananen" [He chews liberally bananas]) into one "weak-context, low-typicality" condition, because it is not possible to have a more or less typical completion in a low constrained context. In order to match the number of trials of this resultant –con –typ control condition to the other two conditions, a random selection of trials was chosen. Thus, the final three conditions for data analysis each contained an equal number of trials. These final three conditions tested semantic context and typicality not in an orthogonal way, but rather, as a continuum of semantic expectation.
Generally, the N400 effect is defined as the difference between an expected, easy-tointegrate standard condition and some less expected, harder-to-integrate deviant condition. Therefore, we treated the +con +typ condition as standard, and calculated the difference waves for the "typicality-only effect" [(+con –typ) – (+con +typ)] and the "combined context + typicality effect" ([(–con –typ) – (+con +typ)], for reasons of readability, this is referred to as the "combined context effect"; see Fig. 6.1A on page 71 for contrasts of interest). As the –con –typ-condition was a merged condition and consisted of $\pm$ typ-nouns, the latter contrast combined a context effect with some portion of typicality effects. This is important to note and will be addressed when interpreting the results of the data.
In line with the N400 literature, we defined the scalp midline as the region of interest (ROI) and averaged the signal condition-wise across the electrodes Fz, FCz, Cz, CPz, Pz, POz. Confined to the midline ROI, we applied a time series analysis on the difference waves of the combined context effect and the typicality-only effect in all three degradation levels separately. By taking the mean over 50 ms time windows, we calculated 10 t tests against zero from 200–700 ms after object onset (Fig. 6.2C) and corrected for multiple comparisons

using the false discovery rate (FDR). To describe the differential N400 effects in response to clear versus medium degraded speech, 2 × 2 repeated measures ANOVAs with factors degradation (8-band, clear speech) and expectancy effect (typicality-only effect, combined context + typicality effect) were applied, first, on the N400 peak latencies which were extracted between 300–600 ms after object onset, and second, on the amplitude over a time window of 450–500 ms after object onset.

For the behavioural measures (reaction times and accuracy), we performed a 3 × 3 repeated measures ANOVA with factors degradation (4-, 8- band, and clear speech) and expectancy (+con +typ, +con –typ, and –con –typ). p Values were always acquired with Greenhouse-Geisser corrected degrees of freedom. Nevertheless, degrees of freedoms are reported uncorrected throughout the manuscript for readability purposes. Where indicated by a significant interaction, further post hoc ANOVAs and t tests were calculated. Post-hoc tests were corrected for multiple comparisons using FDR.

6.3 Results

6.3.1 Intelligibility rating and reaction time

Results of the intelligibility rating analysis show the two main effects of degradation ($F_{2,38} = 51.58, p < 0.001$) and expectancy ($F_{2,38} = 30.06, p < 0.001$), and a significant interaction of both factors ($F_{4,76} = 12.31, p < 0.001$; Fig. 6.1B). For each vocoding level, sentence types differed significantly (clear speech: $F_{2,38} = 3.44, n.s.$ after FDR correction; 8-band speech: $F_{2,38} = 17.09, p < 0.001$; 4-band speech: $F_{2,38} = 21.19, p < 0.001$).

At 8-band speech, intelligibility ratings linearly increased with semantic expectancy, i.e., there was not only a benefit of context (+con –typ vs. –con –typ: $t_{19} = 3.55, p < 0.01$; +con +typ vs. –con –typ: $t_{19} = 4.85, p < 0.001$), but also of typicality (+con +typ vs. +con –typ: $t_{19} = 3.15, p < 0.01$). At 4-band speech, only the strong-context, high-typicality condition differed from the other sentences and received markedly higher intelligibility ratings (+con +typ vs. +con –typ, $t_{19} = 4.59, p < 0.001$, and +con +typ vs. –con –typ, $t_{19} = 5.54, p < 0.001$).

As Figure 6.1C suggests, reaction times showed main effects of degradation ($F_{2,38} = 18.45, p < 0.001$) and of expectancy ($F_{2,38} = 18.19, p < 0.001$), but no interaction ($F < 1$). These main effects are founded in faster responses under clear speech conditions (clear vs. 8-band speech: $t_{19} = -5.04, p < 0.001$, clear vs. 4-band speech: $t_{19} = -5.25, p < 0.001$), and faster button presses for strong-context, high-typicality sentences compared to the other sentences (+con +typ vs. +con –typ: $t_{19} = -5.35, p < 0.001$, +con +typ vs. –con –typ: $t_{19} = -5.02, p < 0.001$, +con –typ vs. –con –typ: $t_{19} = -0.14, n.s.$).

6.3.2 Event related potentials to sentence onset: N100–P200

To assess the effects of degradation, disregarding the sentence-level expectancy manipulation, we first analyzed the evoked potential in response to sound (i.e., sentence) onset. Results are shown in Figure 6.2A. The N100 has a steeper slope and greater negative amplitude for clear speech than for degraded speech (at Cz 50–100 ms: main effect of degradation $F_{2,38} = 4.97, p < 0.05$; clear vs. 8-band: $t_{19} = -2.72, p < 0.05$; clear vs. 4-band: $t_{19} = -3.23, p < 0.01$; 8-band vs. 4-band: $t_{19} = 0.17, n.s.$; at Cz 100–150 ms: main effect of degradation $F_{2,38} = 3.67, p < 0.05$; clear vs. 8-band: $t_{19} = -2.38, p < 0.05$; clear vs. 4-band: $t_{19} = -1.67, n.s.$; 8-band vs. 4-band: $t_{19} = 1.13, n.s.$).
The P200 shows a stepwise amplitude reduction depending on degradation level, with the highest amplitude in response to clear speech and the lowest in response to 4-band speech (at Cz 150–300 ms: main effect of degradation $F_{2,38} = 56.67, p < 0.001$; clear vs. 8-band: $t_{19} = 7.39, p < 0.001$; clear vs. 4-band: $t_{19} = 9.91, p < 0.001$; 8-band vs. 4-band: $t_{19} = 2.64, p < 0.05$).

6.3.3 Event related potentials to sentence-final word: N400

Our main hypotheses were focused on the N400 component of the evoked potential elicited at the sentence-final object. The typicality and combined context effects in the N400 were calculated as difference potentials to the +con +typ (standard) condition (see Methods). We began the N400 analysis by a series of 10 t tests over the midline ROI with a window length of 50 ms from 200–700 ms (testing the N400 difference waves against zero; Fig. 6.2C). Only p values that survived FDR correction are shown.
Confirming our hypothesis, this analysis revealed that a weak N400-like effect in response to 4-band speech (Fig. 6.2C) was not significantly and consistently different from zero (see Fig. 6.2C for details). For the typicality-only effect in 4-band speech, there was no p value below 0.05, not even in one time window (the closest was found from 600–650 ms, $t_{19} = -2.06, p = 0.053$), and for the combined context effect, three time windows differed from zero (the highest t value was from 550–600 ms, $t_{19} = -2.92, p < 0.01$), but they did not survive the FDR correction. Hence, all further analyses reported here focuses on 8-band and clear speech only. Second, there was no consistent N400 typicality-only effect in response to clear speech, but the N400 typicality-only effect was strong and long-lasting in response to moderate degradation (8-band speech; Fig. 6.2C).
Generally, the N400 time window appeared to be delayed in response to 8-band speech. This was corroborated by a significant difference in N400 peak latency: For each subject and condition, the peak latencies of the N400 difference waves between 300 and 600 ms post word onset were extracted. A 2 × 2 repeated measures ANOVA with factors degradation level (8-band and clear speech) and expectancy difference (typicality and combined context + typicality) showed the N400 to peak around 78 ms earlier in response to clear speech

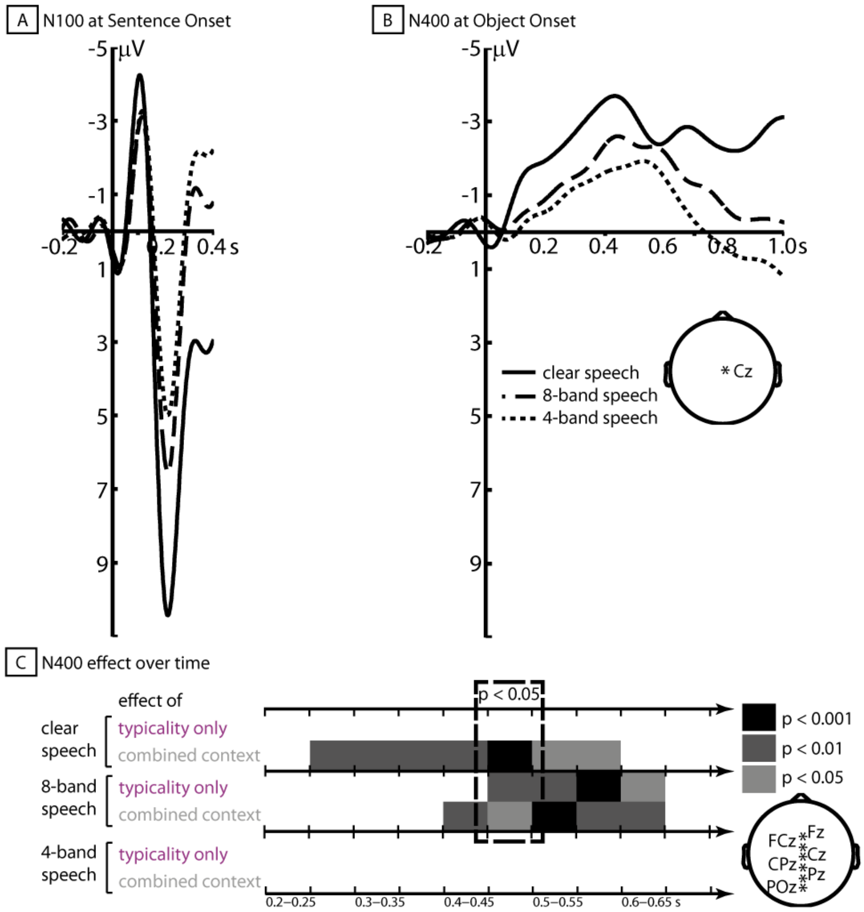

Figure 6.2: Grand-averaged ERP responses. A. N1–P2 complex at sentence onset, by degradation level, shown at electrode Cz. **B.** N400 at object onset, by degradation level, shown at electrode Cz. **C.** Statistical analysis in 50 msec time windows after object onset. t Test of N400 difference waves against zero. Note that in 4-band speech no robust N400 effect is detectable independent of condition. Also note that, in clear speech, no typicality only effect occurs whereas the context manipulation is robust and long-lasting. p Values shown survive FDR correction.

(average peaks around 458 ms and 460 ms, respectively) than in response to 8-band speech (average peaks around 553 ms and 522 ms, respectively; $F_{1,19} = 8.28, p < 0.01$; Fig. 6.3B). A time window from 450–500 ms, which covers the N400 amplitudes regardless of the delay (as indicated by the dashed box in Fig. 6.2C), was chosen and a 2 × 2 repeated measures ANOVA with factors of degradation level (8-band and clear speech) and expectancy difference (typicality and combined context + typicality) was calculated (Fig. 6.3). We found a significant interaction ($F_{1,19} = 5.91, p < 0.05$). Post hoc t-tests confirmed that the N400 effects of typicality and context differed significantly in response to clear speech

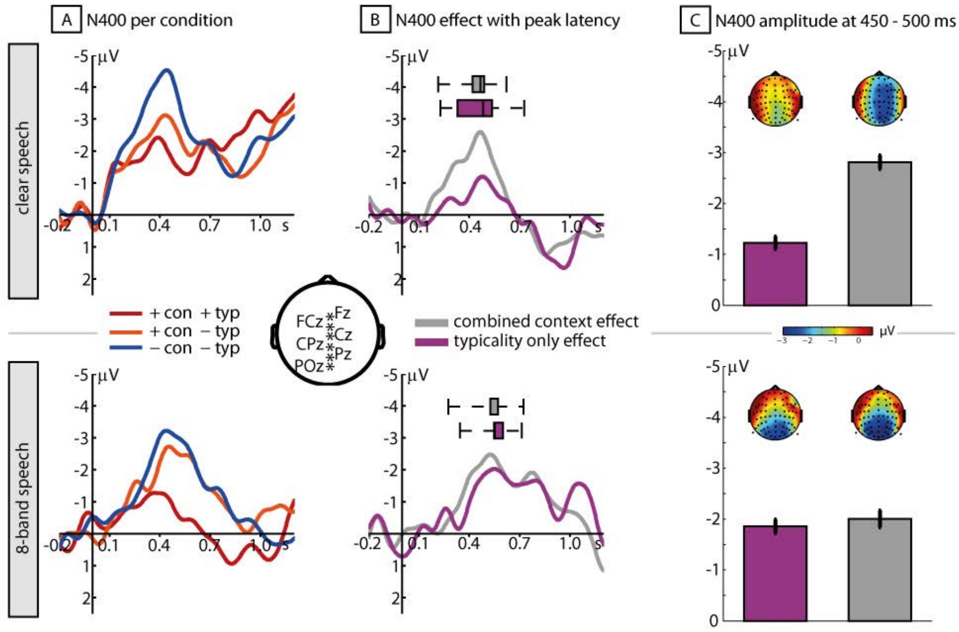

Figure 6.3: Differential N400 effects for clear and degraded speech. A. Grand-averaged ERP responses at midline electrodes after object onset. **B.** Difference waves of the ERPs and their N400 peak latency distributions. Note the significant delay of N400 peaks in 8-band speech. **C.** Bar graph of the N400 effects for combined context and typicality only at 450–500 msec after object onset.

(i.e., only the low context condition evoked an N400, while the typicality manipulation did not; $t_{19} = 2.79, p < 0.05$), whereas in response to 8-band speech, there was no significant difference in the strength of the effects ($t_{19} < 1, n.s.$; Fig. 6.3C).
To summarize, speech degradation reduced the amplitude of the early, as well as the later, ERP responses, which is consistent with the stepwise decrease in the behavioural intelligibility ratings. Furthermore, signal degradation not only delayed the N400 component, but also interacted with the expectancy manipulation.

6.4 Discussion

The goal of the present study was to specify how expectancies may be formed from context and adjusted as a sentence unfolds over time, under various degrees of acoustic degradation. The central research question concerned how broad effects of context and more subtle effects of typicality may influence a neural marker of effortful integration, the N400, under degraded speech conditions. In contrast to a general broadening of semantic predictions under degradation, we hypothesized that acoustic degradation would, instead, elicit a sharpening and more narrow adjusting of linguistic predictions: Only the most typical word in a given context should match a formed expectancy and effectively reduce the N400 effect.

First, the occurrence of an N400 effect depended on the extent of signal degradation. There was no semantic modulation of the N400 in the severely degraded (4-band) speech condition, suggesting that fast linguistic processes were effectively hindered. In moderately degraded (8-band) speech, the N400 amplitude was attenuated and the peak was significantly delayed for $\sim$ 78 ms; this is in line with previous studies (Connolly et al., 1992; D'Arcy et al., 2005; Holcomb, 1993; Obleser and Kotz, 2011).

Second, the N400 reflected fine semantic differentiations of context strength of a sentence and the typicality of a particular word in this context (Kutas and Hillyard, 1984; Connolly et al., 1992; Connolly and Phillips, 1994; Federmeier and Kutas, 1999). The combined context effect was generally more pronounced than the typicality-only effect and showed the known posteriorcentral scalp topography.

In addition, however, we found an expectancy differentiation in the N400 that was dependent on signal degradation: In the clear speech condition, a strong-context, low-typicality object appeared to be compatible with the predictions formed by the context, and the neural effort of integration (as reflected by the N400) was low in amplitude (Fig. 6.3C; see also schematic display in Fig. 6.4A). Note that unlike previous studies with similar manipulations (Connolly and Phillips, 1994; Desroches et al., 2009; Newman and Connolly, 2009), we were unable to show N200- or PMN-like effects. This might be due to our task, which guided participants to focus on semantic rather than segmental information. In the moderate degradation condition (8-band speech), however, the same strong-context, low-typical word triggered a pronounced N400 response that was statistically indistinguishable from the response to an unpredictable (i.e., weak-context, low-typical) word.

The topography of the N400 combined context effect at 450–500 ms was more frontally distributed in response to clear speech than degraded speech. A tentative explanation could be that, for clear speech, N400 sources might be more anterior and widespread than for acoustically degraded speech. In relation to spatially more precise functional MRI work on expectancies under degraded speech conditions, Obleser and Kotz (2010) found that the anterior superior temporal sulcus/gyrus (STS/STG) showed a linear increase of activation with a more intelligible signal. Interestingly, the same study generally reported more spatially constrained intelligibility activations in response to highly predictable sentences. A more widespread N400 in response to degraded speech was also reported recently (Romei et al., 2011), suggesting that additional attention or working memory processes are needed in adverse listening conditions. In contrast to our manipulation, however, this study used isolated words without sentential context and investigated the N400 not in response to the last word, but the intermediate word in a list of three.

Note that our manipulation (i.e., restricting the relevant semantic context to a single word: the preceding verb) also allows interpretations in terms of lexical priming. The

adjectives inserted between the verb and target object were included to minimize this possibility. Nevertheless, associative priming may be observed even when an intervening item is presented (Joordens and Besner, 1992). Thus, the current results may not differentiate between sentential and lexical semantics but, instead, hint to the interesting fact that, even though participants focused on semantic information because of the task, they were utilizing it differently under degraded speech conditions. Somewhat counter-intuitively, (Mattys et al., 2009) showed that listeners in adverse hearing situations tend to rely less upon lexical-semantic cues and more on acoustic-phonetic detail. These results suggest that perceptual load might have narrowed the expectancy to an acoustic-phonetic focus, that is, unexpected segmental information could not be compensated by top-down knowledge (see section on “Prediction capacities and other cognitive resources” below).
The fact that less typical objects also had lower word frequency (see Methods and Materials) allows an alternative interpretation: comprehension and integration of lower-frequency words, in general, might benefit from strong contextual constraints in clear speech, but not in degraded speech. Note, however, that this interpretation would be also consistent with a basic conjecture of this study—namely, that the effects of contextual constraint on target processing differ depending upon the intelligibility of the acoustic signal and the probability of the target.
Overall, the present findings confirm that a degraded context is, in absolute terms, less effective at activating compatible semantic features than clear speech (cf. Aydelott et al., 2012), such that contextual facilitation of unexpected but semantically consistent words is reduced. However, the results indicate that constraint-based expectancies that favor high-probability completions are relatively maintained under moderate perceptual degradation; consistent with previous findings that listeners' use of semantic cues in strongly biasing sentence contexts is relatively robust in adverse conditions (e.g., (Kalikow et al., 1977; Bilger et al., 1984)).

6.4.1 N400 and behavioural responses: fast vs. delayed processes

A remaining open question is how to reconcile behavioural effects in the intelligibility ratings and brain effects in the N400 time range. The N400 component has been understood as a marker of context integration effort, but not context integration failure.
Thus, the absence of a semantic modulation of the N400, as the current data show in 4-band speech, might be due to the lack of fast mapping capacities under poor acoustics. However, an intelligibility gain of strong-context, high-typical sentences in the ensuing behavioural response was still present. This suggests that the time it took for participants to generate a behavioural response allowed for retrospective semantic analysis of the degraded signal, and affected intelligibility ratings for these sentences.
At intermediate signal degradation (8-band speech), one could argue that the N400 is equally sensitive to different expectancy manipulations, whereas the behavioural data show

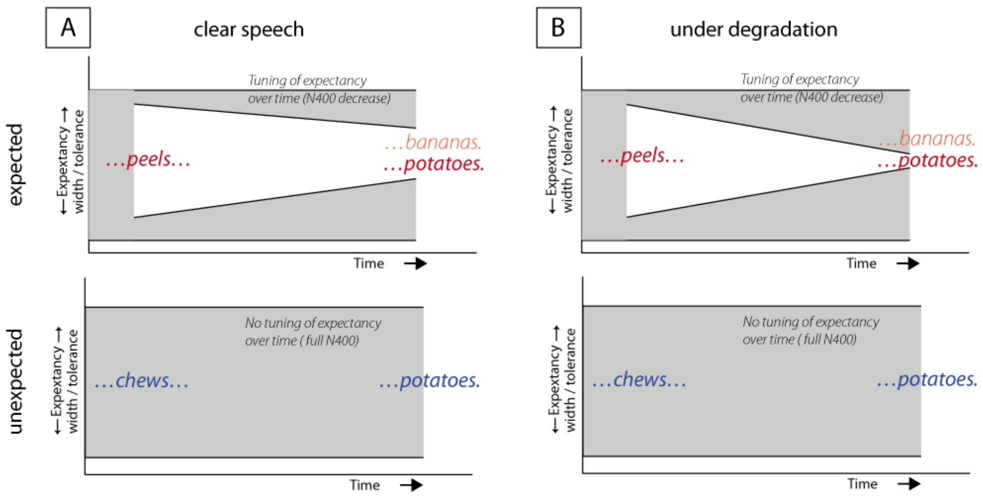

Figure 6.4: Model of expectancy searchlight. A. Fast processing and context abstraction in clear speech, which can be characterized as a more tolerant "searchlight" process. **B.** Fast recognition but decelerated context abstraction in moderate degradation (i.e., a narrowing of the expectancy searchlight). (bottom) No expectancy searchlight is formed when low context is provided independent of signal quality. See Discussion for details.

a stepwise increase of intelligibility with growing expectation. Also, the N400 response is generally delayed under degraded speech compared to clear speech conditions. The benefit in performance for strong context, low typical sentences against unrelated sentences again suggests some later recovery, if at least some expectations could be formed.

This indicates that, under intermediate degradation, fast recognition and integration processes are possible (sensitivity to semantic expectations, i.e., reduced N400 in response to strong-context, high-typical words), but that they are still delayed when recognition and integration have to be based on less typical words (enhanced N400 in response to strongcontext, low-typicality words). Put differently, an "expectancy searchlight" can be formed, based on sufficient perceptual evidence, but it will be narrowed because of limited cognitive resources (Fig. 6.4B).

Finally, under clear speech conditions, we found an N400 combined context effect that was absent in the behavioural data. The N400, therefore, seems to reflect successful, albeit effortful, comprehension. Figure 6.4A displays it as a liberal expectancy searchlight where less thoroug sentence processing in clear speech results from fast cue integration and context abstraction.

As Figure 6.4 summarizes, we suggest a tentative interpretation of our main findings in terms of an "expectancy searchlight". If listening conditions are ideal, expectancies are more liberal and the "searchlight" in a strong context is focused, but tolerant. The clear-speech N400 effects were reduced in the strong context, irrespective of low or high typicality, compared to weakcontext sentences. When dealing with acoustic limitations, however, this searchlight is narrowed down, and only the most typical sentence ending is facilitated in this case (Fig. 6.4B).

6.4.2 Prediction capacities and other cognitive resources

The present data deliver important evidence for a trade-off between acoustics and semantics: First, the results of the N100-P200-complex at sentence onset suggest familiarization and categorization difficulties with degraded signals. We found the highest N100 amplitude in response to clear speech, and, in line with Obleser and Kotz (2011), no significant difference between 4-band and 8-band speech. Moreover, Obleser and Kotz (2011) found the N100 response to be most pronounced in the 1-band speech condition, a highly unintelligible signal. Further testing indicated that the N100 amplitude has a u-shaped relation to speech intelligibility (Obleser et al., in preparation). The N100 is thought to index an initial allocation of resources and formation of a sensory memory trace (e.g., (Schröger et al., 2004)), but it is unclear whether higher familiarity and easier categorization, as in clear speech, should lead to an increased or reduced N100. The current data, together with previous observations, suggest that the measured N100 amplitude is under the joint influence of low-level acoustic factors, such as perceived loudness and spectral resolution, and cognitive factors, such as familiarity. Furthermore, we observed the strongest P200 amplitude in response to clear speech and the weakest in response to 4-band speech. Paulmann et al. (2011) related their differential P200 responses to salient acoustic features (e.g., pitch, voice quality, and loudness) of a stimulus. Less spectral information, and, thus, reduced saliency of important acoustic features, may lead to greater variance in the neural processing due to wide spread resource allocation for processing, which condenses in a reduced time-locked ERP response.

Second, all N400-like processes in response to 8-band speech were delayed in time, as shown by a significant difference in N400 peak latency that was driven by degradation. Also, reaction times were longer in response to degraded sentences (Fig. 6.1C). This is compatible with results by D'Arcy et al. (2005) who reported a reduced and delayed ($\sim$ 51 ms) N400 response to incongruent sentence ending words when working memory load was increased. More directly, evidence on the detrimental effects of speech degradation on working memory processes has accumulated (e.g., Obleser et al., 2012; Piquado et al., 2010; Rabbitt, 1968).

To accomplish rapid speech comprehension in everyday communication, a languagefamiliarized listener constantly predicts forthcoming linguistic input (Gagnepain et al., 2012). The adjustment of these predictions may be partly explained by psycholinguistic models that describe auditory language comprehension as a trade-off between perceptual evidence and other cognitive resources.

Norris and McQueen (2008)'s Shortlist B model (2008), for example, takes into account perceptual ambiguities and their interaction with word frequency. Recall that the conditional probability of the target word in a given context varied ("typicality"; [peel ... potatoes] vs. [peel ... bananas]). Consequently, an explanation arising from the Shortlist B model would be the following: Under clear speech, probability or typicality differences would be assumed to play only a negligible role, because all words would be correctly identified, and performance would approach ceiling. However, when the perceptual information is sparse (i.e., under degraded speech), the listener will have to resort to established word probabilities, and these probabilities would also affect the neural processes reflected by the N400. Thus, from such a cognitive psychology angle, it would be expected that acoustic degradation would narrow the range of lexical items that an automatic neural integration mechanism will pre-activate (and that will, hence, elicit only a small N400 amplitude; Fig. 6.4).

Another interpretation of the observed adjusting of the range of expected words would be that perceptual load (i.e., the resources used for effortful processing of the signal itself) limits the resources a listener has available for forming predictions as the sentence unfolds. In this case, word probabilities would always be used (as they would under clear speech conditions) but are less accessible in adverse listening conditions because of shared resources (auditory and lexical analysis). Therefore, only the most probable ones would be pre-activated.

Thus, in a Shortlist B framework, probabilities are used as an active compensation, whereas in a capacity-limitation framework, shared cognitive capacities inevitably lead to a limited evaluation of context suitability. Both concepts leave open the question of whether the typicality judgment in a strong context, but adverse listening, condition should be understood as a poorly generated or a more specific prediction. With our model, we suggest that these perspectives are two sides of the same coin.

To summarize, processing efforts in response to degraded auditory sentences capture resources that would normally be available for predictive processes in the mental lexicon in response to non-degraded sentences. Thus, we propose that adverse listening conditions limit the ability to form abstract expectancies from context, which leads to stronger reliance on acoustic–phonetic rather than lexical cues. This is in line with Mattys et al. (2009), who also demonstrated that, listeners, confronted with energetically masked speech, rely more on segmental (rather than lexical) information. Furthermore, studies on time-compressed speech (another form of speech degradation) have shown that listeners can recover intelligibility (i.e., access their mental lexicon) in severely time-compressed speech, as long as silent breaks are inserted at clause boundaries, such that listeners gain processing time intermittently (Wingfield et al., 1999). Bringing together this limited-resources account and the current results, an experimental prediction can be formed: By allowing the listener to free-up resources by allowing more time for processing the degraded sentence, the typicality-only effect in intermittent-delay degraded speech should be reduced.

To conclude, we propose a simplified, yet testable expectancy searchlight model (Fig. 6.4), which aims to bring together the different aspects discussed here. While inevitably leaving open many questions (e.g., no assumption is formulated on how pre-lexical, acoustic–phonetic processing information enters the post-lexical stage), such a searchlight model is able to capture how expectancy is modulated by semantics and acoustics. It, thereby, combines top-down and bottom-up approaches.

6.4.3 Conclusion

The present study investigated the relative importance of different sources of information in speech comprehension under adverse listening conditions. Do we rely more on top-down context or on bottom-up perceptual input? The data show that semantic context plays a crucial role, but deficient perceptual evidence in a degraded signal leads to more conservative, more narrowly adjusted expectancies on the forthcoming acoustic–phonetic information. Only common sentence endings are facilitated in the processing of moderately degraded speech. These results, thus, provide a starting point to better understand and aid speech comprehension in hearing-impaired and ageing listeners.

6.5 Supplement: Theta power and phase coherence dissociate two kinds of semantic integration during sentence processing

6.5.1 Introduction

In the previous chapter we have shown that the event-related potentials, namely the N400, is sensitive to both acoustic degradation of the speech signal and the semantic predictability of the sentence final word. As argued in Section 2.3, ERPs represent only a limited part of the EEG signal. Especially the differences in N400 latency as well as N400 amplitude suggest that effects might be more appropriately described by looking at time–frequency power and phase separately. In accordance with previous findings when manipulating sentence semantics (Hagoort et al., 2004; Bastiaansen et al., 2005; Hald et al., 2006), we hypothesized that varying the cloze probability of sentence final words should be reflected in power and phase-coherence differences of theta oscillations ($\sim$4 Hz).

6.5.2 Methods

Data from Chapter 6 were re-analyzed for the current research question. In order to obtain time–frequency representations (TFRs), clean data (as described in Section 6.2) were re-referenced to average reference. For single-trial power estimates, Morlet wavelets were applied in 20-ms steps from -1 to 2 ms for time-locked responses to object onsets with a frequency-specific window width. This accounts for the trade-off between higher frequency resolution for lower frequencies and higher time resolution for higher frequencies. Therefore, TFRs for logarithmically spaced frequencies from 3 to 30 Hz were convolved with linearly increasing window widths from 2 to 12 cycles.

Subsequently, absolute power estimates of single trial TFRs were submitted to a multi-level or "random effects" statistics approach (for application to time–frequency data see e.g., Obleser and Weisz, 2012; Henry and Obleser, 2012): On the first or individual level, massed independent samples regression coefficient t tests with condition as dependent variable and contrast weights as independent variable (zero-centred values linearly increasing with cloze probability) were calculated. Uncorrected regression t values and *beta*s were obtained for all time–frequency bins. On the second level, *beta*s were tested against zero in a one-tailed dependent-samples t test for low frequencies (3–30 Hz) in the time range from –0.5 to 1.5 s. Monte-Carlo non-parametrical permutation method (1000 randomisations) implemented in the Fieldtrip toolbox (Oostenveld et al., 2011) estimated type I-error controlled cluster probabilities ($\alpha < 0.05$).

6.5.3 Results

For assessing absolute power time-locked to object onset, a multi-level statistics approach was chosen. The main effect of cloze probability was set on the first level by contrasting

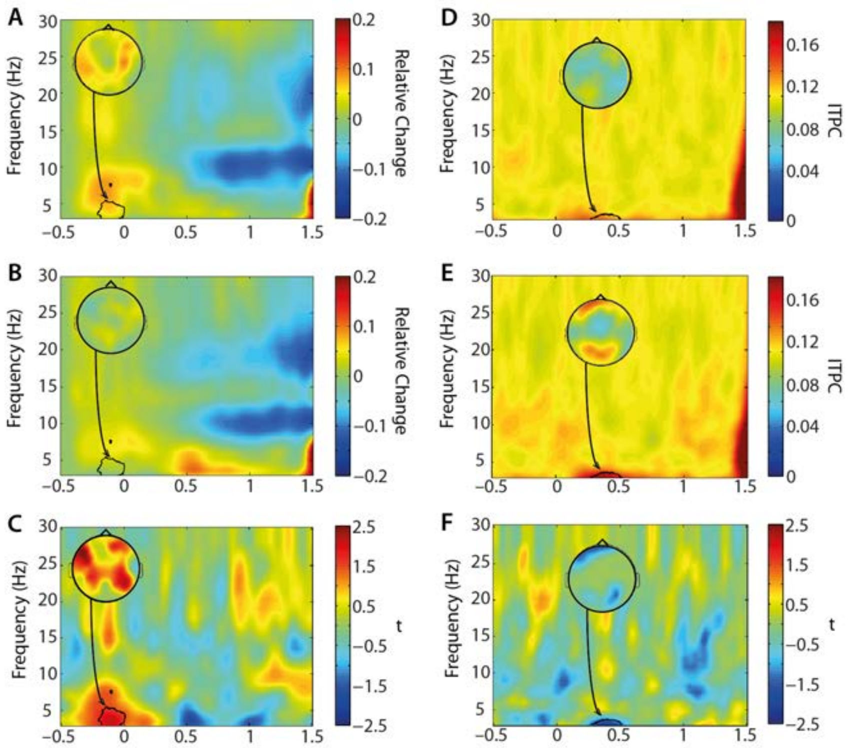

Figure 6.5: Effect of cloze probability on theta activity. Black contours frame found clusters; insets show respective cluster topographies. **A.** TFR of high cloze sentences. For illustration, TFRs were baseline corrected (baseline from –500 to –400 ms). **B.** TFR of low cloze sentences. **C.** Statistical contrast of **A** vs. **B** reveal an increase in theta power before object onset (0s) over bilateral temporal areas. **D.** ITPC of high cloze sentences. **E.** ITPC of low cloze sentences. **F.** Statistical contrast of **D** vs. **E**. In line with N400 effects in Section 6.3.3, ITPC is increased for low cloze sentences. Note a slight, statistically nonsignificant theta power increase at the same time in **B**.

high and low cloze probability sentences ($\pm con$). On the second-level, first-level betas were tested against zero. The cluster permutation test revealed one positive cluster ($p = 0.012; T_{sum} = 12,062; R = 0.47$) in the lower theta frequencies (3–5 Hz) which shows a bilateral fronto-temporal scalp distribution and covers a time window of –200 ms and object onset (Fig. 6.5C).

To control whether the increase in power might be accompanied by an increase in phase-locking, the cluster permutation test was repeated for inter-trial phase coherence measures. A negative cluster was found ($p = 0.015; T_{sum} = -2,388; R = 0.54$) also in the lower theta frequencies (3–5 Hz) but during a later time window from 250 to 500 ms and with a fronto-parietal scalp distribution (Fig. 6.5D). These results go hand in hand with the N400 results reported in the previous section which also showed more negative N400 amplitudes for low cloze probability sentences.

6.5.4 Discussion

Theta power was found to be modulated during sentence processing depending on the predictability of the sentence-final word. Most importantly, theta power was enhanced just before the onset of the sentence-final word if the preceding sentence context was semantically constraining. We interpret this finding in terms of increased memory demands in case the context provides beneficial semantic information. In line with results in Section 6.3.3, theta oscillations might tentatively reflect a neural means to implement the phonological loop (Baddeley, 2003; Roux and Uhlhaas, 2014), such that phonemically coded information is periodically re-(or pre-) activated (Fuentemilla et al., 2010). If the semantic context is predictive, sentence final words can be pre-activated as indexed by theta enhancement. Subsequent lexical access is facilitated as indicated by decreased inter-trial phase coherence in a later time window corresponding to the N400 effect reported in the previous section.

There is one other recent study by Meyer et al. (2013) which observed increased slow oscillatory power increase just before the onset of the sentence-final word. However, this was in the alpha frequency band. The authors manipulated the distance (short vs. long) between the verb and its object. Results were interpreted such that prior to memory retrieval, premature object release needs to be inhibited via enhanced alpha power. In contrast to this study, no syntactic but semantic contraints were manipulated in the current experiment. Interestingly, differences were thus observed in the theta but not in the alpha frequency range right before the onset of the sentence-final word. We therefore argue that prior to the actual memory retrieval lexico-semantic information does not need to be inhibited as it would have been indexed by increased alpha activity but is actually semantically pre-activated. This interpretation is in line with the topography of the anticipatory theta enhancement which extends over bilateral temporal regions suggesting the activation of middle temporal gyrus associated with word processing (Kotz et al., 2002; Minicucci et al., 2013).

If lexico-semantic retrieval cannot be initiated in advance as in low cloze probability sentences, later lexico-semantic integration is more effortful as indexed by the peri-stimulus increase of theta inter-trial phase coherence. This finding corresponds to the more negative N400 amplitude for low compared to high cloze probability sentences reported in Section 6.3.3. Furthermore, the topographical distribution of the peri-stimulus increase in inter-trial phase coherence over parietal areas is in line with the centro-parietal scalp topography of the N400 effects (see Fig. 6.3). Interestingly, it has been shown that the sentence-N400 effect reflects parallel processes of lexical selection and semantic context integration (van den Brink et al., 2006). Our data extend this view by adding that stimulus-driven synchronization in the theta frequency range is required to accomplish both processes simultaneously.

In sum, we suggest that theta power enhancement over bilateral temporal regions reflects the pre-activation of lexico-semantic information. Increased peri-stimulus phase-locking in the theta band, instead, reflects stimulus-driven synchronization to accumulate phonetic information for lexical selection and to integrate these information on-line into sentence context.

7 General Discussion

The current thesis was concerned with the neural oscillatory dynamics of spoken word recognition. In particular, slow neural temporal dynamics of lexico-semantic processing were investigated when the speech signal was ambiguous or degraded. To this end, EEG data was acquired using two established experimental manipulations to study lexical access, namely the lexicality of isolated items and the cloze probability of words in sentence context. Accuracy was examined when speech was intact, embedded in white noise, or spectrally degraded. After summarizing the experimental findings in section 7.1, the main results of the current thesis will be highlighted and discussed. In section 7.2, implications of the current evidence are discussed for models of spoken word recognition. In particular, evidence of parallel and nonlinear oscillatory patterns is considered. In section 7.4, the relationship between the N400 and slow neural oscillations are surveyed. In particular, it is suggested that processes so far subsumed to be reflected in one ERP component, namely the N400, can be dissected into alpha- and theta-related processes. Finally, an outlook for future research is provided. First in section 7.5, the role of alpha activity is discussed as a mechanism of selective noise inhibition and selective signal enhancement along the auditory pathway. Second in section 7.6, the question how to further investigate the role of theta oscillations in speech processing is elaborated.

7.1 Summary of experimental findings

Chapter 3 asked about the functional dissociation of oscillatory power modulations in the alpha and theta frequency bands in spoken word recognition. Interestingly, time–frequency power analysis revealed parallel processes of lexical integration in the alpha band and of ambiguity resolution in the theta band. Post-lexical alpha power suppression scaled with wordness such that real words showed the lowest, ambiguous pseudowords intermediate and opaque pseudowords the highest alpha power. Thus, alpha was interpreted to index the gradually increasing difficulty of lexical integration (Obleser and Weisz, 2012). Source localisation of the alpha power revealed left occipito-temporal cortex and right anterior prefrontal cortex. These results were supported by the gradual increase of the N400 magnitude showing the most negative amplitude for opaque pseudowords. Usually, the N400 component is interpreted as indexing the neural effort of lexical search (Kutas and Van Petten, 1994). Here, reduced N400 magnitude and alpha power suppression for real words suggested to be an index of lexical integration whereas enhanced N400 magnitude and alpha power are rather indicating the inhibition of lexical integration. Furthermore, theta power was found to be selectively enhanced for ambiguous pseudowords that differed only in one vowel from their real word neighbours. Source localisation revealed left inferior frontal gyrus and right middle temporal gyrus. These results were interpreted to reflect ambiguity resolution of the response conflict induced by the proximity to real words.

In Chapter 4, data from Chapter 3 were compared with data acquired from the same participants doing the lexical decision task in noise. The comparison showed higher induced alpha power in noise than in quiet. At the same time and in line with reduced ERP amplitudes of the N1-P2-complex for degraded speech (see for example also Chapter 6), alpha inter-trial phase coherence showed the opposite pattern and was higher for speech in quiet than in noise. Results suggested the inhibition of task-irrelevant noise by means of induced alpha power increase (Jensen and Mazaheri, 2010). A framework was developed to further systematically study the functional role of alpha activity during speech processing in noise.

Chapter 5 explored mechanisms of alpha phase for spoken word recognition in noise by re-analyzing the data of the lexical decision task in noise acquired in Chapter 4. The accuracy of lexical decisions in noise critically depended on pre- and peri-stimulus alpha phase. These investigations provided a significant link between research on low-level psychophysical performance that is modulated by neural phase (i.e., neural excitability; Henry and Obleser, 2012) and higher-level word recognition. In particular, pre-stimulus alpha phase was anti-phase for correct and incorrect trials over right frontal areas. We interpreted this finding to reflect selective inhibition in the sense that stimuli coinciding with the excitatory phase were more likely to be thoroughly processed than when coinciding with the inhibitory phase and were thus ultimately judged correctly (Schroeder and Lakatos, 2009). Peri-stimulus alpha phase bifurcation for correct and incorrect trials over left fronto-temporal regions was interpreted to reflect decisional weighting during lexical selection if lexical access is difficult as in adverse listening conditions (Wyart et al., 2012). No phase effects were found in the theta band, although theta oscillations have been shown to modulate neural firing as well (Kayser et al., 2012). Supplementary behavioural results in section 5.5 yielded differential influences of lexical stress pattern and formant information on perceptual sensitivity and response bias. If ambiguous pseudowords were acoustically closer to their real word neighbour (measured by formant distances of the manipulated vowels), lexical decisions were biased towards word-judgements. This relationship was most pronounced when stimuli were stressed on the second syllable which was the crucial syllable in the current experimental design. Additionally in section 5.6, features of the phase bifurcation index (Busch et al., 2009) used for the current analysis of neural phase were elaborated by simulations to support the validity of the current methodological approach.

Chapter 6, finally, asked about the temporal dynamics of spoken word recognition in sentence context. Here, the signal was compromised by noise-vocoding (i.e., spectrally degrading) the speech signal in three levels of severity. The cloze probability of sentence final words was manipulated to be high or low. The magnitude of N400 was reduced in intact as well as degraded speech, if sentence-final words were expected versus unexpected from semantic context (Kutas and Hillyard, 1980). This is in line with the well-established semantic benefit for intelligibility in adverse listing situations (Kalikow et al., 1977). Beyond this replication, the N400 magnitude was also reduced for less typical sentence-final words in clear speech indicating facilitated lexical access for an extended semantic field. In degraded speech, however, highly predictable contexts were not benefical for less typical sentence-final words as indexed by higher N400 magnitude thus disclosing narrowed semantic expectations. In Section 6.5, these results were elaborated by analyzing slow oscillatory dynamics. Importantly, semantic context modulated theta but not alpha oscillations. In particular, theta power was enhanced before the onset of the sentence-final word if the context was predictable. In turn, if the context was not predictable peri-stimulus theta inter-trial phase coherence was increased. Results were interpreted in terms of lexico-semantic pre-activation in highly predictable contexts indexed by anticipatory theta power increase whereas without predictable context lexical access and context integration were accomplished simultaneously via increased theta inter-trial phase coherence.

7.2 The dissociation of alpha and theta activity in spoken word recognition

The current thesis aimed at determining neural oscillatory signatures of spoken word recognition. Alpha oscillations (8–12 Hz) as the predominant rhythm in human EEG have been observed in diverse cognitive functions (amongst others auditory processing, e.g. Hartmann et al., 2012 or attention Klimesch, 2012). Alpha activity is presumably a neural means to implement the general cognitive function of gating information flow (Jensen and Mazaheri, 2010; Hanslmayr et al., 2012). The current thesis found alpha oscillations to play a role during spoken word recognition in three possible ways:
First in Chapter 3, induced alpha power was found to be post-lexically suppressed when words were recognized, thus when lexico-semantic information was processed (Hanslmayr et al., 2012). Second in Chapter 4, induced alpha power was found to be enhanced at the beginning of words embedded in noise suggesting that the noise was identified as the task-irrelevant auditory object and thus selectively inhibited for further processing by alpha power enhancement (Klimesch et al., 2007). Third in Chapter 5, pre-stimulus alpha phase was found to modulate lexical decision accuracy in noise. We interpreted this finding to reflect selective attention insofar as stimuli coinciding with the excitatory phase were more likely to be thoroughly processed and ultimately judged correctly (Schroeder and Lakatos, 2009; Mathewson et al., 2011).

Another frequency band of interest had been theta oscillations because of their association with long-term (thus semantic) memory (Fell and Axmacher, 2011) and their presumablely chunking role in speech processing due to their correspondance to the syllabic rate (Ghitza, 2013). Theta activity might control the timing of periodic re-activation of memory content (Fuentemilla et al., 2010) and has been found during lexico-semantic memory retrieval (Bastiaansen et al., 2008). At this point, the data speaks in favor of theta oscillations playing a role for lexico-semantic access although no direct evidence could be found to support ideas about chunking linguistic content for further processing:
First in Chapter 3, induced theta power was found to be post-lexically enhanced for ambiguous pseudowords. Because of their proximity to real words (only one vowel exchanged), ambiguous pseudowords induced response conflicts when judging their lexicality. In line with Fuentemilla et al. (2010), we suggest that phonemic information needed to be "replayed" in order to re-compare it with long-term memory representations and thus resolve ambiguity. Second in Chapter 6, in high cloze probability sentences theta power was found to be enhanced just before the onset of the sentence-final word, thus indicating the anticipatory activation of long-term memory content, i.e. lexico-semantics. Third in Chapter 5, theta phase was not found to modulate lexical decision accuracy so that its chunking role for speech processing remains elusive. To follow-up this null-finding, a research programm will be suggested in Section 7.6.

7.3 Spoken word recognition as a nonlinear process

Our findings extend current knowledge on spoken word recognition gained previously by N400 analysis. We provide two arguments that challenge the linearity of spoken word recognition (as for example modelled in Cohort; Marslen-Wilson and Tyler, 1980). First, we showed the simultaneous occurrence of alpha and theta power modulation, indexing lexical integration and ambiguity resolution respectively. By looking at the N400 response only, lexical access would have appeared as a sequential process which is effortlessly accomplished for real words (reduced N400 magnitude) and is at first not successful for both types of pseudowords. For ambiguous pseudowords, however, word recognition might occur delayed (indexed by an intermediate N400 magnitude) whereas lexical search for opaque pseudowords continues (indexed by a permanently increased N400 magnitude; see Fig. 3.1 in Chapter 3). Intermediate alpha power suppression for ambiguous pseudowords in turn suggested suboptimal lexical integration whereas high alpha power for opaque pseudowords indexed inhibited lexical integration. By looking at slow neural oscillations, alpha power scaled with wordness comparable to the N400. At the same time, theta power was found selectively enhanced for ambiguous pseudowords. This is compatible with models that assume a dual route of word recognition where lexical and segmental information are both held in memory at the same time (Norris et al., 2000).

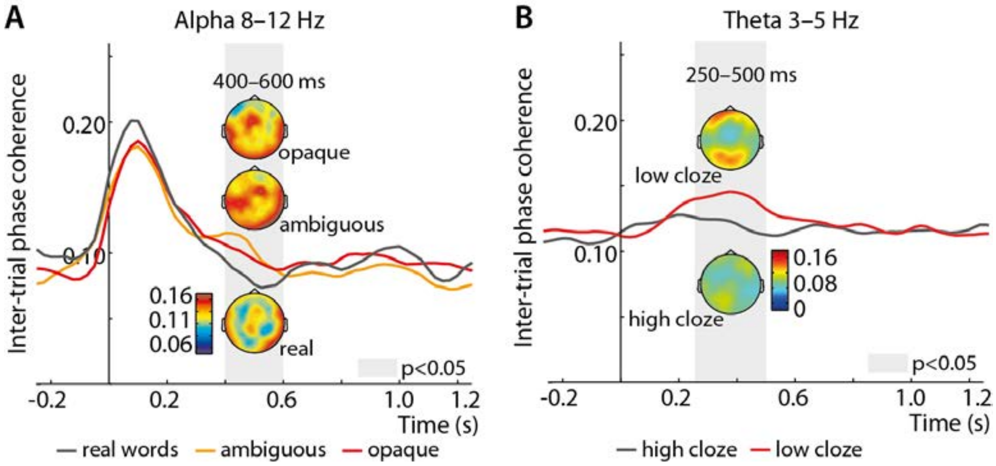

Figure 7.1: Effects of inter-trial phase coherence (ITPC) during the time window of the N400. Grey background highlight significant cluster. **A.** Time-line of alpha ITPC dependent on lexicality (data from Chapter 3). **B.** Time-line of theta ITPC dependent on the cloze probability of a sentence (data from Chapter 6)

The second piece of evidence that questions the linearity of spoken word recognition is provided by the alpha phase bifurcation (Busch et al., 2009). Previously, a number of prominent models of decision making assumed the linear accumulation of perceptual evidence (for review see Ratcliff and McKoon, 2008; Mulder et al., 2014). In Chapter 5, we observed instead that pre-stimulus alpha was anti-phase for correct and incorrect trials over right anterior electrodes. This rather indicates in line with Schroeder and Lakatos (2009) that a stimulus is rhythmically "selected" by attention via aligning with the high excitable alpha phase. Furthermore, peri-stimulus alpha phase bifurcated over left fronto-temporal regions. Because the phase effect was not locked to the critical vowel manipulation, we interpreted the effect in line with Wyart et al. (2012) to reflect rhythmical integration and weighting of decisional information. In sum, this data provide first evidence that perceptual evidence in lexical decision tasks is not accumulating linearly but proceeds rhythmically. However, future research needs to further determine the relationship between slow neural oscillations and speech processing.

7.4 Are N400 effects better explained by alpha and theta inter-trial phase coherence?

Both experimental paradigms used in the current thesis, namely the lexicality as well as the cloze probability manipulation, replicated common N400 effects. First, the magnitude of the N400 was increased in case of processing an isolated pseudoword compared to a real word (see Chapter 3). Second, the magnitude of the N400 was increased in case of processing a word in a low compared to a word in a high cloze probability sentence (see Chapter 6; for review see Kutas and Federmeier, 2011; Van Petten and Luka, 2012). Although attempts to map semantics onto pseudowords and efforts to semantically integrate sentence-final words into preceding contexts obviously impose differential psycholinguistic challenges, both processes are consistently reflected in the N400 component which has led to the unsatisfactory conclusion about common neurolinguistic processes. Thus, an enhanced N400 magnitude is usually interpreted as a general increased effort of lexico-semantic processing.

The results of the current thesis suggest that there might be different oscillatory activities underlying the two N400 effects. First, when comparing words and pseudowords, the N400 effect is accompanied by a decrease in the inter-trial phase coherence (ITPC) for real words compared to pseudowords in the alpha frequency range as summarized in Figure 7.1A. Hence, Chapter 3 more appropriately framed the lexicality-N400 effect in terms of lexico-semantic integration in line with interpreting alpha desynchronization as an index of successful information flow (Hanslmayr et al., 2012). The increase in N400 magnitude together with the increased alpha power for pseudowords, accordingly, was interpreted as indicating the inhibition of lexico-semantic processing (Jensen and Mazaheri, 2010).

Second, when comparing sentence-final words in low versus high cloze probability sentences, ITPC was enhanced for words in low cloze contexts in the theta frequency range (Fig. 7.1B). Accordingly, the sentence-N400 effect might be re-considered as reflecting simultaneous lexico-semantic retrieval and semantic integration. To coordinate both processes on-line at the same time, synchronization via theta oscillations might be necessary. If, instead, lexico-semantic information can be pre-activated via pre-stimulus theta power enhancement, peri-stimulus synchronization might be reduced because semantic integration is facilitated as shown in Section 6.5.

The proposed functional distinction between alpha and theta inter-trial phase coherence underlying the N400 effects is also in line with the idea of different informational time windows that are considered for speech analysis (Ghitza, 2011). Thus, in pseudowords phonemic information is primarily analyzed so that lexicality effects would be reflected in the faster alpha frequency band. For the semantic integration of several words in a sentence, however, longer analysis time windows as provided by slower theta oscillations might be considered.

In order to test whether alpha and theta ITPC indeed reflect distinct functional mechanisms and are able to dissect N400 effects, future research might combine lexicality (word, pseudoword) and cloze probability (high, low) manipulations in a 2×2 design. Low cloze probability sentences that are completed by word-like pseudowords should show both effects simultaneously: higher alpha ITPC compared to low cloze sentences with real word endings, and higher theta ITPC compared to high cloze sentences with pseudoword completions. N400 effects should be analyzed accordingly.

7.5 Alpha activity along the auditory pathway.[6]

In Chapter 4, we have shown that alpha oscillations are an attractive neural candidate mechanism of selective auditory inhibition. There are different aspects which need to be systematically investigated in order to determine the role of alpha: Which neural circuits "deploy" or trigger high-alpha states? And in terms of the current framework: What kind of channels can be attenuated by enhanced alpha power?

Currently, there are few studies mapping the sources of alpha power during masked auditory processing. Some evidence has accumulated showing noise-invariant representations of the signal in auditory cortices (Chang et al., 2010; Ding and Simon, 2012a) with the degree of invariance increasing from peripheral to cortical processing stages (Rabinowitz et al., 2013). If we assume that alpha is an important central mechanism to inhibit various types of maskers, these studies suggest that masking release via alpha enhancement might occur as early as in primary auditory cortex. A first direct hint to this idea might be the case of an illusory sound percept like tinnitus, which can be centrally suppressed by means of increasing alpha power in primary auditory cortex (Leske et al., 2013; Weisz et al., 2014). This is in line with research showing that attention modulates activity in sensory cortices corresponding to the modality of the stimulus (e.g., Heinrich et al., 2011; Wild et al., 2012). Thus, alpha activity in primary auditory cortex might be crucially contributing to inhibiting the formation of auditory objects.

In future studies investigating underlying alpha sources, a distinction between energetic and informational masking might be crucial (Brungart et al., 2001; Mattys et al., 2009; Scott and McGettigan, 2013; for a more comprehensive overview of adverse listening conditions see Mattys et al., 2012). Energetic masking describes the competition of auditory target and masker in the auditory periphery due to spectro-temporal overlay of the two signals, causing an overlap of excitation patterns in the cochlea and auditory nerve (Durlach et al., 2003). One type of background signal often assumed to cause primarily energetic masking is white noise (e.g., Arbogast et al., 2005) which is quasi-stationary and has high energy in a broad frequency range (for discussion see Stone et al., 2012). Although informational masking is sometimes defined only negatively as all masking effects not accounted for by energetic masking (cf. Gutschalk et al., 2008), a more refined definition is required, especially when it comes to speech processing. When target speech is masked by a competing talker, it is not just the energetic overlap of two signals that causes masker interference. Rather, the speech masker initiates phonetic and semantic processing that interferes with the linguistic processing of the target (Schneider et al., 2007). Thus, informational masking describes the interference of target and masker at a more central, cognitive level, whereas energetic masking refers to energetic overlap in the auditory periphery.

[6]This chapter is adapted from parts of the published article by Strauß, Wöstmann, and Obleser (2014). *Front Hum Neurosci 8*, 350.

According to the framework developed in Chapter 4, alpha oscillations might be important for inhibition of both types of maskers, however in different brain areas. We presume that energetic maskers are inhibited by enhanced alpha activity in auditory cortex (Müller and Weisz, 2012). In contrast, processing of informational maskers like competing speech should rather be inhibited by alpha activity in higher auditory areas such as posterior superior temporal gyrus (pSTG) and beyond, relevant for linguistic processing (Scott et al., 2004, 2009).

7.6 The relationship of theta oscillations and speech processing

In Chapter 5, results on alpha phase were reported showing its modulatory influence on lexical decision accuracy. Optimal alpha phase led to more thoroughly processing of the input signal thus constituting a mechanism of sensory selection (Schroeder and Lakatos, 2009). Unfortunately, the role of theta phase could not be determined although recent models of neural speech processing have provided good arguments to assign a special role to it (Ghitza, 2011; Gagnepain et al., 2012; Giraud and Poeppel, 2012). The following section elaborates a framework to further investigate theta oscillations in speech processing and discusses along the way why theta phase coherence was not predictable for the accuracy of lexical decisions in our data.

Neural oscillations, especially in the theta (and delta) range, have been demonstrated to track rhythmic stimulation, through a process referred to as cortical *entrainment* (for review see Ding and Simon, 2014). For example, amplitude modulations—for example expressed in the speech envelope—have been shown to drive slow neural oscillations (Ahissar et al., 2001; Luo and Poeppel, 2007; Aiken and Picton, 2008; Nourski et al., 2009; Henry and Obleser, 2013). The underlying cognitive functions remain unclear and several theoretical accounts have been suggested ranging from auditory encoding (Howard and Poeppel, 2010; Ding and Simon, 2012b) to informational chunking (Lakatos et al., 2008; Giraud and Poeppel, 2012; Ghitza, 2013). On the one hand, interpretations of auditory encoding are supported by the fact that entrainment is observed by non-speech sounds as well (e.g., Henry and Obleser, 2012; Steinschneider et al., 2013). On the other hand, higher informational processing accounts are corroborated by evidence showing that entrainment can be enhanced by non-acoustic factors such as speech intelligibility (Peelle et al., 2013) and attention (Kerlin et al., 2010; Zion Golumbic et al., 2013).

The framework of informational chunking is based on the fact that oscillations in cortical layer IV are mainly driven by external stimulation frequencies (Oberlaender et al., 2012). Hence, if layer IV of the auditory cortex is faced with speech, theta oscillations are induced because the average syllabic rate is about 4 Hz (i.e., the average syllable length is about 250 ms). However, the framework furthermore suggests that the auditory cortex is not only entraining to the syllabic rate (maybe because it is its prefered resonance frequency anyways) but at the same time fulfills the higher cognitive function of chunking the incoming speech information into segments which can be further processed by higher language related areas (Ghitza, 2011; Giraud and Poeppel, 2012). This idea relates to fluctuations in neural excitability which have been found to be reflected by theta oscillations (Kayser et al., 2012; Ng et al., 2013). Thus, linguistic information arriving in the non-excitable, i.e. inhibitory, phase are thought to be less thoroughly processed, leading to discretized informational chunks.

Although the sketched informational chunking function is very inspiring, important limitations have been raised from a neuroscientific as well as a phonetic perspective. First from the neuroscientific perspective, the framework relies on cortical entrainment to speech envelope. Importantly, the cortex does not only entrain to amplitude fluctuations but also to other rhythmic cues such as frequency modulations (for discussion see Obleser et al., 2012; Ghitza et al., 2013). Additionally and associated with this line of argumentation, speech envelope indeed contributes to speech intelligibility yet spectral content is at least as decisive for comprehension (Xu et al., 2005) and oscillations entrain better to speech with full spectral content even when envelopes are identical (Ding et al., 2013). That means the set of speech-specific factors (another most likely candidate being F0 contour, see for example Cummins, 2009; Spinelli et al., 2010) that facilitates entrainment needs yet to be determined and are underspecified in this framework. Second from the phonetic perspective, in this framework cortical theta oscillations are in fact linked to the syllabic rate extracted from speech envelope. However, this view oversimplifies speech rhythmicity as syllable boundaries are not easily deducible from speech amplitude fluctuations (Cummins, 2012). The speech signal itself might not be actually rhythmic enough to support entrainment and justify neural entrainment as a useful mechanism for speech comprehension (Cummins, 2012).

Following this fundamental critique especially made by phoneticians, it is an important future endeavour to specifically test the relationship of theta oscillations and syllables and their importance for speech comprehension. If theta activity actually relates to the syllable rate and its function is about packaging speech input into meaningful units, theta phase should be important for segmenting a speech stream. One reason, therefore, of why there was no dependence of lexical decision accuracy on theta phase in Chaper 5 could be that participants did not have to segment lexical stimuli and use syllabification as a strategy to accomplish the task. Note that the missing syllabifying strategy in our data might also be instrinsic to the German language: For example, native speakers of French have been shown to base speech segmentation on syllables (Cutler et al., 1986). The authors argued that languages which prefer stress patterns for segmentation (like English and German) show greater ambiguity in their syllable boundaries, therefore preventing native speakers of such languages to apply any syllabifying segmentation strategies. However, in order to test whether theta actually chunks linguistic information, a speech stream with ambiguous segmentation boundaries could be used (Dilley and Pitt, 2010; Baese-Berk et al., 2014) to see whether the segmentation decision depends on theta phase. For example, Spinelli and colleagues (2010) used French sentences which were phonemically identical but their meaning depended on the time point of segmentation: "C'est l' | affiche" versus "C'est la | fiche" (the vertical line marks the earlier or later time point of segmentation leading to sentence meaning "This is a poster" or "a sheet"). The research question would be whether theta phase aligns with segmentation boundaries such that the inhibitory theta phase peaks earlier in sentences which were segmented as "l'affiche" than "la fiche".

Besides the clarification of the relationship between theta oscillations and syllabic rate, the influence of the articulatory motor representation on speech segmentation and lexical access processes needs to be tested. While the acoustic speech signal is actually very complex and aperiodic (although some authors would argue that the cochlea filter "generates" rhythmic fluctuations within a single pass band, see Ghitza et al., 2013), the articulatory motor system and the jaw opening in particular oscillates quasi-periodically in the theta frequency range (Dohen et al., 2004; unfortunately, the relationship between jaw opening, syllable onsets, and speech envelope is not straightforward, see Benus and Pouplier, 2011; Cummins, 2012). It has been suggested that the periodicity of the articulatory motor system might even be the evolutionary reason for the importance to follow slow oscillatory activity for auditory comprehension (Morillon et al., 2010; Giraud and Poeppel, 2012; Schwartz et al., 2012).

Observing a talker, the jaw opening provides some information on the articulated consonant to the listener which might be quite specific for bilabials (e.g., /p/) or underspecified for back consonants (e.g., /k/). In everyday life, audio and visual information are usually congruent and support each other to facilitate comprehension (van Wassenhove et al., 2005), especially relevant in suboptimal hearing situations (Ten Oever et al., 2014). If audio and visual information are incongruent, a merged percept is induced, the so-called McGurk effect (e.g., hearing /ka/ while seeing /pa/ leads to comprehending /ta/; McGurk and MacDonald, 1976). Interestingly, some evidence is suggestive for a possible involvement of theta oscillations during audiovisual integration. For example, it has been shown that there is a time-window of about 250 ms tolerance ($\pm 180ms$) to integrate audio and visual information (Munhall et al., 1996). Furthermore, if audio and visual information are congruent, the phase-locking of theta oscillations has been found to be increased (Arnal et al., 2011).

In sum, bistable speech streams could be used (e.g., a sequence of /tapatapata/ which can be segmented into /tapa-tapa-ta/ or /ta-pata-pata/; Sato et al., 2006, 2007) to test whether entrainment to visually presented jaw cycles might provide regular onset cues such that segmentation is supported and biased depending on neural theta phase.

References

Ahissar, E., S. Nagarajan, M. Ahissar, A. Protopapas, H. Mahncke, and M. M. Merzenich (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America 98*(23), 13367–13372.

Ahveninen, J., S. Huang, J. W. Belliveau, W.-T. Chang, and M. Hämäläinen (2013). Dynamic oscillatory processes governing cued orienting and allocation of auditory attention. *Journal of Cognitive Neuroscience 25*(11), 1926–1943.

Aiken, S. J. and T. W. Picton (2008). Human cortical responses to the speech envelope. *Ear and Hearing 29*(2), 139–157.

Akam, T. E. and D. M. Kullmann (2012). Efficient "communication through coherence" requires oscillations structured to minimize interference between signals. *PLoS Computational Biology 8*(11), e1002760.

Altman, G. and D. Carter (1989). Lexical stress and lexical discriminability: Stressed syllables are more informative, but why? *Computer Speech & Language 3*(3), 265–275.

Arbogast, T. L., C. R. Mason, and G. J. Kidd (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America 117*(4 Pt 1), 2169–2180.

Arnal, L. H., V. Wyart, and A.-L. Giraud (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience 14*(6), 797–801.

Awh, E., E. K. Vogel, and S.-H. Oh (2006). Interactions between attention and working memory. *Neuroscience 139*(1), 201–208.

Aydelott, J., D. Baer-Henney, M. Trzaskowski, R. Leech, and F. Dick (2012). Sentence comprehension in competing speech: Dichotic sentence-word priming reveals hemispheric differences in auditory semantic processing. *Language and Cognitive Processes 27*, 1108–1144.

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

Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communication Disorders 36*(3), 189–208.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology 63*, 1–29.

Baddeley, A. D. and G. Hitch (1974). Working Memory. In Gordon H. Bower (Ed.), *Psychology of Learning and Motivation*, Volume 8, pp. 47–89. Academic Press.

Baese-Berk, M. M., C. C. Heffner, L. C. Dilley, M. A. Pitt, T. H. Morrill, and J. D. McAuley (2014). Long-Term Temporal Tracking of Speech Rate Affects Spoken-Word Recognition. *Psychological Science 25*(8), 1546–1553.

Balota, D. A. and J. I. Chumbley (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology. Human Perception and Performance 10*(3), 340–357.

Banerjee, S., A. C. Snyder, S. Molholm, and J. J. Foxe (2011). Oscillatory Alpha-Band Mechanisms and the Deployment of Spatial Attention to Anticipated Auditory and Visual Target Locations: Supramodal or Sensory-Specific Control Mechanisms? *The Journal of Neuroscience 31*(27), 9923–9932.

Bastiaansen, M. C. M., R. Oostenveld, O. Jensen, and P. Hagoort (2008). I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. *Brain and Language 106*, 15–28.

Bastiaansen, M. C. M., M. van der Linden, M. Ter Keurs, T. Dijkstra, and P. Hagoort (2005). Theta responses are involved in lexical-semantic retrieval during language processing. *Journal of Cognitive Neuroscience 17*(3), 530–541.

Benki, J. R. (2003). Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition. *The Journal of the Acoustical Society of America 113*, 1689–1705.

Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain and Language 31*(2), 308–327.

Benus, S. and M. Pouplier (2011). Jaw Movement in Vowels and Liquids Forming the Syllable Nucleus. In P. Cosi, R. De Mori, G. Di Fabbrizio, and R. Pieraccini (Eds.), *Proceedings of Interspeech 2011. 12th Annual Conference of the International Speech Communication Association*, pp. 396–399. Florence, Italy.

Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software 31*(10), 1—21.

Berger, H. (1931). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten 94*(1), 16–60.

Bilger, R. C., J. M. Nuetzel, W. M. Rabinowitz, and C. Rzeczkowski (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research 27*, 32–48.

Binder, J. R., R. H. Desai, W. W. Graves, and L. L. Conant (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex 19*, 2767–2796.

Binder, J. R., D. A. Medler, C. F. Westbury, E. Liebenthal, and L. Buchanan (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *NeuroImage 33*, 739–748.

Bizley, J. K. and Y. E. Cohen (2013). The what, where and how of auditory-object perception. *Nature Reviews. Neuroscience 14*(10), 693–707.

Bloom, P. A. and I. Fischler (1980). Completion norms for 329 sentence contexts. *Memory & Cognition 8*(6), 631–642.

Blume, W. T., M. Kaibara, and G. B. Young (2002). *Atlas of adult electroencephalography.* Philadelphia [u.a.]: Lippincott Williams & Wilkins.

Boulenger, V., M. Hoen, C. Jacquier, and F. Meunier (2011). Interplay between acoustic/phonetic and semantic processes during spoken sentence comprehension: an ERP study. *Brain and Language 116*, 51–63.

Broadbent, D. E. (1958). *Perception and communication.* Oxford, UK: Pergamon Press.

Brungart, D. S., B. D. Simpson, M. A. Ericson, and K. R. Scott (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America 110*(5), 2527–2538.

Brysbaert, M., M. Buchmeier, M. Conrad, A. M. Jacobs, J. Bölte, and A. Böhl (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology 58*, 412–424.

Buchsbaum, B. R. and M. D'Esposito (2008). The search for the phonological store: from loop to convolution. *Journal of Cognitive Neuroscience 20*(5), 762–778.

Buffalo, E. A., P. Fries, R. Landman, T. J. Buschman, and R. Desimone (2011). Laminar differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America 108*(27), 11262–11267.

Busch, N. A., J. Dubois, and R. VanRullen (2009). The phase of ongoing EEG oscillations predicts visual perception. *The Journal of Neuroscience 29*(24), 7869–7876.

Busch, N. A. and R. VanRullen (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences of the United States of America 107*(37), 16048–16053.

Buzsáki, G. and A. Draguhn (2004). Neuronal oscillations in cortical networks. *Science 304*(5679), 1926–1929.

Cantero, J. L., M. Atienza, R. Stickgold, M. J. Kahana, J. R. Madsen, and B. Kocsis (2003). Sleep-Dependent θ Oscillations in the Human Hippocampus and Neocortex. *The Journal of Neuroscience 23*(34), 10897–10903.

Carreiras, M. and C. J. Price (2008). Brain activation for consonants and vowels. *Cerebral Cortex 18*, 1727–1735.

Cavanagh, J. F., M. X. Cohen, and J. J. B. Allen (2009). Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring. *The Journal of Neuroscience 29*(1), 98–105.

Chang, E. F., J. W. Rieger, K. Johnson, M. S. Berger, N. M. Barbaro, and R. T. Knight (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience 13*(11), 1428–1432.

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Accoustical Society of America 25*(5), 975–979.

Chwilla, D. J., C. M. Brown, and P. Hagoort (1995). The N400 as a function of the level of processing. *Psychophysiology 32*, 274–285.

Cleland, A. A., M. G. Gaskell, P. T. Quinlan, and J. Tamminen (2006). Frequency effects in spoken and visual word recognition: evidence from dual-task methodologies. *Journal* *of Experimental Psychology: Human Perception and Performance 32*, 104–119.

Cluff, M. S. and P. A. Luce (1990). Similarity neighborhoods of spoken two-syllable words: retroactive effects on multiple activation. *Journal of Experimental Psychology. Human Perception and Performance 16*(3), 551–563.

Cohen, M. X., C. E. Elger, and J. Fell (2009). Oscillatory activity and phase-amplitude coupling in the human medial frontal cortex during decision making. *Journal of Cognitive Neuroscience 21*, 390–402. 2.

Cohen, M. X. and S. van Gaal (2013). Dynamic interactions between large-scale brain networks predict behavioral adaptation after perceptual errors. *Cerebral Cortex 23*(5), 1061–1072.

Cole, R. A., J. Jakimik, and W. E. Cooper (1978). Perceptibility of phonetic features in fluent speech. *The Journal of the Acoustical Society of America 64*(1), 44–56.

Connine, C. M. and C. Clifton (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology. Human Perception and Performance 13*(2), 291–299.

Connolly, J. F. and N. A. Phillips (1994). Event-Related Potential Components Reflect Phonological and Semantic Processing of the Terminal Word of Spoken Sentences. *Journal of Cognitive Neuroscience 6*, 256–266.

Connolly, J. F., N. A. Phillips, S. H. Stewart, and W. G. Brake (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language 43*, 1–18.

Connolly, J. F., S. H. Stewart, and N. A. Phillips (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language 39*, 302–318.

Corbetta, M., G. Patel, and G. L. Shulman (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron 58*(3), 306–324.

Cravo, A. M., G. Rohenkohl, V. Wyart, and A. C. Nobre (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *The Journal of Neuroscience 33*(9), 4002–4010.

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics 37*(1), 16–28.

Cummins, F. (2012). Oscillators and syllables: a cautionary note. *Frontiers in Psychology 3*, 364.

Cutler, A., J. Mehler, D. Norris, and J. Segui (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language 25*(4), 385–400.

D'Arcy, R. C. N., J. F. Connolly, E. Service, C. S. Hawco, and M. E. Houlihan (2004). Separating phonological and semantic processing in auditory sentence processing: a high-resolution event-related brain potential study. *Human Brain Mapping 22*, 40–51.

D'Arcy, R. C. N., E. Service, J. F. Connolly, and C. S. Hawco (2005). The influence of increased working memory load on semantic neural systems: a high-resolution event-related brain potential study. *Cognitive Brain Research 22*, 177–191.

Davis, M. H., M. A. Ford, F. Kherif, and I. S. Johnsrude (2011). Does semantic context benefit speech understanding through "top-down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience 23*, 3914–3932.

Davis, M. H. and I. S. Johnsrude (2003). Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience 23*(8), 3423–3431.

Debener, S., J. Thorne, T. R. Schneider, and F. C. Viola (2010). *Using ICA for the analysis of multi-channel EEG data*, pp. 121–134. Oxford University Press.

Dehaene, S. and L. Cohen (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences 15*, 254–262.

Dehaene, S., F. Pegado, L. W. Braga, P. Ventura, G. Nunes Filho, A. Jobert, G. Dehaene-Lambertz, R. Kolinsky, J. Morais, and L. Cohen (2010). How learning to read changes the cortical networks for vision and language. *Science 330*, 1359–1364.

Desimone, R. and J. Duncan (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience 18*, 193–222.

Desroches, A. S., R. L. Newman, and M. F. Joanisse (2009). Investigating the time course of spoken word recognition: electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience 21*, 1893–1906.

Diependaele, K., M. Brysbaert, and P. Neri (2012). How Noisy is Lexical Decision? *Frontiers in Psychology 3*, 348.

Dilley, L. C. and M. A. Pitt (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science 21*(11), 1664–1670.

Ding, N., M. Chatterjee, and J. Z. Simon (2013). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage 88C*, 41–46.

Ding, N. and J. Z. Simon (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences 109*(29), 11854–11859.

Ding, N. and J. Z. Simon (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology 107*(1), 78–89.

Ding, N. and J. Z. Simon (2013). Power and phase properties of oscillatory neural responses in the presence of background activity. *Journal of Computational Neuroscience 34*(2), 337–343.

Ding, N. and J. Z. Simon (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Frontiers in Human Neuroscience 8*, 311.

Doelling, K. B., L. H. Arnal, O. Ghitza, and D. Poeppel (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage 85 Pt 2*, 761–768.

Dohen, M., H. Lœvenbruck, M.-A. Cathiard, and J.-L. Schwartz (2004). Visual perception of contrastive focus in reiterant French speech. *Speech Communication 44*(1–4), 155–172.

Domahs, F., M. Grande, W. Huber, and U. Domahs (2014). The direction of word stress processing in German: evidence from a working memory paradigm. *Frontiers in Psy-*

chology 5, 574.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology 92 Part 1*, 53–78.

Dufau, S., J. Grainger, and J. C. Ziegler (2012). How to say "no" to a nonword: a leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology. Learning, Memory, and Cognition 38*(4), 1117–1128.

Dufour, S., A. Brunellière, and U. H. Frauenfelder (2013). Tracking the time course of word-frequency effects in auditory word recognition with event-related potentials. *Cognitive Science 37*(3), 489–507.

Dugué, L., P. Marque, and R. VanRullen (2011). The phase of ongoing oscillations mediates the causal relation between brain excitation and visual perception. *The Journal of Neuroscience 31*(33), 11889–11893.

Durlach, N. I., C. R. Mason, G. K. Jr, T. L. Arbogast, H. S. Colburn, and B. G. Shinn-Cunningham (2003). Note on informational masking (L). *The Journal of the Acoustical Society of America 113*(6), 2984–2987.

Düzel, E., W. D. Penny, and N. Burgess (2010). Brain oscillations and memory. *Current Opinion in Neurobiology 20*(2), 143–149.

Eckert, M. A., V. Menon, A. Walczak, J. Ahlstrom, S. Denslow, A. Horwitz, and J. R. Dubno (2009). At the heart of the ventral attention system: the right anterior insula. *Human Brain Mapping 30*(8), 2530–2541.

Eisner, F., C. McGettigan, A. Faulkner, S. Rosen, and S. K. Scott (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *The Journal of Neuroscience 30*, 7179–7186.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences 8*(7), 301–306.

Erb, J., M. J. Henry, F. Eisner, and J. Obleser (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *The Journal of Neuroscience 33*(26), 10688–10697.

Farah, M. J., A. B. Wong, M. A. Monheit, and L. A. Morrow (1989). Parietal lobe mechanisms of spatial attention: Modality-specific or supramodal? *Neuropsychologia 27*(4), 461–470.

Federmeier, K. D. and M. Kutas (1999). Right words and left words: electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research 8*, 373–392.

Fell, J. and N. Axmacher (2011). The role of phase synchronization in memory processes. *Nature Reviews Neuroscience 12*(2), 105–118.

Felty, R. (2007). *Context Effects in Spoken Word Recognition of English and German by Native and Non-native Listeners (Unpublished doctoral dissertation)*. Michigan: The University of Michigan.

Ferrand, L., M. Brysbaert, E. Keuleers, B. New, P. Bonin, A. Méot, M. Augustinova, and C. Pallier (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: evidence from chronolex. *Frontiers in Psychology 2*, 306–306.

Ferreira, C. S., A. Marful, T. Staudigl, T. Bajo, and S. Hanslmayr (2014). Medial Prefrontal Theta Oscillations Track the Time Course of Interference during Selective Memory Retrieval. *Journal of Cognitive Neuroscience 26*(4), 777–791.

Fiebach, C. J., A. D. Friederici, K. Müller, and D. Y. von Cramon (2002). fMRI evidence for dual routes to the mental lexicon in visual word recognition. *Journal of Cognitive Neuroscience 14*(1), 11–23.

Fisher, N. I. (1993). *Statistical Analysis of Spherical Data*. Cambridge University Press.

Foxe, J. J., G. V. Simpson, and S. P. Ahlfors (1998). Parieto-occipital approximately 10 Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport 9*(17), 3929–3933.

Foxe, J. J. and A. C. Snyder (2011). The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. *Frontiers in Psychology 2*, 154.

Frauenfelder, U. H., J. Segui, and T. Dijkstra (1990). Lexical effects in phonemic processing: facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance 16*(1), 77–91.

Freunberger, R., R. Fellinger, P. Sauseng, W. Gruber, and W. Klimesch (2009). Dissociation between phase-locked and nonphase-locked alpha oscillations in a working memory task. *Human Brain Mapping 30*(10), 3417–3425.

Friederici, A. D. (1997). Neurophysiological aspects of language processing. *Clinical Neuroscience 4*, 64–72.

Friedrich, C. K., C. Eulitz, and A. Lahiri (2006). Not every pseudoword disrupts word recognition: an ERP study. *Behavioral and Brain Functions 2*, 36.

Friedrich, C. K. and S. A. Kotz (2007). Event-related potential evidence of form and meaning coding during online speech recognition. *Journal of Cognitive Neuroscience 19*, 594–604.

Friedrich, C. K., U. Schild, and B. Röder (2009). Electrophysiological indices of word fragment priming allow characterizing neural stages of speech recognition. *Biological Psychology 80*, 105–113.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences 9*(10), 474–480.

Fritz, J. B., M. Elhilali, S. V. David, and S. A. Shamma (2007). Auditory attention-focusing the searchlight on sound. *Current Opinion in Neurobiology 17*(4), 437–455.

Fuentemilla, L., W. D. Penny, N. Cashdollar, N. Brunzeck, and E. Düzel (2010). Theta-coupled periodic replay in working memory. *Current Biology 20*(7), 606–612.

Gagnepain, P., R. N. Henson, and M. H. Davis (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology 22*(7), 615–621.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance 6*(1), 110–125.

Geyken, A. (2009). Statistische Wortprofile zur schnellen Analyse der Syntagmatik in Textkorpora. *Berlin-Brandenburgische Akademie der Wissenschaften*, 115–137.

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology 2*, 130.

Ghitza, O. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Frontiers in Psychology 4*, 138.

Ghitza, O., A.-L. Giraud, and D. Poeppel (2013). Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Frontiers in Human Neuroscience 6*, 340.

Ghitza, O. and S. Greenberg (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica 66*(1-2), 113–126.

Giraud, A.-L., C. Kell, C. Thierfelder, P. Sterzer, M. O. Russ, C. Preibisch, and A. Kleinschmidt (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex 14*, 247–255.

Giraud, A.-L. and D. Poeppel (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience 15*(4), 511–517.

Goldinger, S. (1996). Auditory lexical decision. *Language and Cognitive Processes 11*(6), 559–567.

Goldinger, S. D., P. A. Luce, and D. B. Pisoni (1989). Priming Lexical Neighbors of Spoken Words: Effects of Competition and Inhibition. *Journal of Memory and Language 28*(5), 501–518.

Goodman, J. C. and J. Huttenlocher (1988). Do we know how people identify spoken words? *Journal of Memory and Language 27*(6), 684–698.

Gow, D. W. and P. C. Gordon (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance 21*(2), 344–359.

Green, J. D. and A. A. Arduini (1954). Hippocampal Electrical Activity in Arousal. *Journal of Neurophysiology 17*(6), 533–557.

Greenberg, S. (1999). Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation. *Speech Communication 29*(2–4), 159–176.

Griffiths, T. D. and J. D. Warren (2004). What is an auditory object? *Nature Reviews. Neuroscience 5*(11), 887–892.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics 28*(4), 267–283.

Gross, J., N. Hoogenboom, G. Thut, P. Schyns, S. Panzeri, P. Belin, and S. Garrod (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology 11*(12), e1001752.

Gross, J., J. Kujala, M. Hamalainen, L. Timmermann, A. Schnitzler, and R. Salmelin (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences of the United States of America 98*, 694–699.

Grossberg, S. and S. Kazerounian (2011). Laminar cortical dynamics of conscious speech perception: Neural model of phonemic restoration using subsequent context in noise. *The Journal of the Acoustical Society of America 130*(1), 440–460.

Gunter, T. C., A. D. Friederici, and H. Schriefers (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience 12*, 556–568.

Gutschalk, A., C. Micheyl, and A. J. Oxenham (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biology 6*(6), e138.

Haegens, S., L. Luther, and O. Jensen (2012). Somatosensory anticipatory alpha activity increases to suppress distracting input. *Journal of Cognitive Neuroscience 24*(3), 677–685.

Haegens, S., V. Nácher, R. Luna, R. Romo, and O. Jensen (2011). α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences of the United States of America 108*(48), 19377–19382.

Haegens, S., D. Osipova, R. Oostenveld, and O. Jensen (2010). Somatosensory working memory performance in humans depends on both engagement and disengagement of regions in a distributed network. *Human Brain Mapping 31*(1), 26–35.

Hagoort, P., L. Hald, M. Bastiaansen, and K. M. Petersson (2004). Integration of word meaning and world knowledge in language comprehension. *Science 304*, 438–441.

Hald, L. A., M. C. M. Bastiaansen, and P. Hagoort (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language 96*(1), 90–105.

Halgren, E., R. P. Dhond, N. Christensen, C. Van Petten, K. Marinkovic, J. D. Lewine, and A. M. Dale (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage 17*, 1101–1116.

Hanslmayr, S., A. Aslan, T. Staudigl, W. Klimesch, C. S. Herrmann, and K.-H. Bäuml (2007). Prestimulus oscillations predict visual perception performance between and within subjects. *NeuroImage 37*(4), 1465–1473.

Hanslmayr, S., J. Gross, W. Klimesch, and K. L. Shapiro (2011). The role of α oscillations in temporal attention. *Brain Research Reviews 67*(1-2), 331–343.

Hanslmayr, S., W. Klimesch, P. Sauseng, W. Gruber, M. Doppelmayr, R. Freunberger, T. Pecherstorfer, and N. Birbaumer (2007). Alpha phase reset contributes to the generation of ERPs. *Cerebral Cortex 17*, 1–8.

Hanslmayr, S., B. Pastötter, K.-H. Bäuml, S. Gruber, M. Wimber, and W. Klimesch (2008). The electrophysiological dynamics of interference during the Stroop task. *Journal of Cognitive Neuroscience 20*(2), 215–225.

Hanslmayr, S., B. Spitzer, and K.-H. Bäuml (2009). Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cerebral Cortex 19*, 1631–1640.

Hanslmayr, S., T. Staudigl, and M.-C. Fellner (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Frontiers in Human Neuroscience 6*.

Hartmann, T., W. Schlee, and N. Weisz (2012). It's only in your head: expectancy of aversive auditory stimulation modulates stimulus-induced auditory cortical alpha desynchronization. *NeuroImage 60*(1), 170–178.

Heim, S., S. B. Eickhoff, A. K. Ischebeck, A. D. Friederici, K. E. Stephan, and K. Amunts (2009). Effective connectivity of the left BA 44, BA 45, and inferior temporal gyrus during lexical and phonological decisions identified with DCM. *Human Brain Mapping 30*(2), 392–402.

Heinrich, A., R. P. Carlyon, M. H. Davis, and I. S. Johnsrude (2011). The continuity illusion does not depend on attentional state: FMRI evidence from illusory vowels. *Journal of Cognitive Neuroscience 23*(10), 2675–2689.

Helfer, K. S., J. Chevalier, and R. L. Freyman (2010). Aging, spatial cues, and single- versus dual-task performance in competing speech perception. *The Journal of the Acoustical Society of America 128*(6), 3625–3633.

Helmholtz, H. (1853). Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche. *Annalen der Physik 165*(6), 211–233.

Henry, M. J. and J. Obleser (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences of the United States of America 109*(49), 20095–20100.

Henry, M. J. and J. Obleser (2013). Dissociable neural response signatures for slow amplitude and frequency modulation in human auditory cortex. *PloS One 8*(10), e78758.

Herrmann, C., M. Grigutsch, and N. A. Busch (2005). EEG oscillations and wavelet analysis. In T. C. Handy (Ed.), *Event-Related Potentials*, pp. 229–259. Cambridge, Mass. [u.a.]: MIT Press.

Hickok, G. and D. Poeppel (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience 8*(5), 393–402.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology 30*(1), 47–61.

Horváth, J. and A. Burgyán (2011). Distraction and the auditory attentional blink. *Attention, Perception & Psychophysics 73*(3), 695–701.

Howard, M. F. and D. Poeppel (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology 104*(5), 2500–2511.

Howes, D. (1954). On the interpretation of word frequency as a variable affecting speed of recognition. *Journal of Experimental Psychology 48*(2), 106–112.

Howes, D. (1957). On the relation between the probability of a word as an association and in general linguistic usage. *Journal of Abnormal Psychology 54*, 75–85.

Jarvis, M. R. and P. P. Mitra (2001). Sampling properties of the spectrum and coherency of sequences of action potentials. *Neural Computation 13*(4), 717–749.

Jensen, O., M. Bonnefond, and R. VanRullen (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences 16*(4), 200–206.

Jensen, O., J. Gelfand, J. Kounios, and J. E. Lisman (2002). Oscillations in the Alpha Band (9–12 Hz) Increase with Memory Load during Retention in a Short-term Memory Task. *Cerebral Cortex 12*(8), 877–882.

Jensen, O. and A. Mazaheri (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Frontiers in Human Neuroscience 4*, 186.

Johnson, E. L. and R. T. Knight (2015). Intracranial recordings and human memory. *Current Opinion in Neurobiology 31*, 18–25.

Joordens, S. and D. Besner (1992). Priming effects that span an intervening unrelated word: implications for models of memory representation and retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition 18*, 483–491.

Jung, R. and A. E. Kornmüller (1938). Eine Methodik der Ableitung Iokalisierter Potentialschwankungen aus subcorticalen Hirngebieten. *Archiv für Psychiatrie und Nervenkrankheiten 109*(1), 1–30.

Jung, T. P., S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology 37*(2), 163–178.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, and S. Jannedy (Eds.), *Probabilistic linguistics*, pp. 39–95. Cambridge, Mass. u.a.: MIT Pr.

Jusezyk, P. W. and P. A. Luce (2002). Speech perception and spoken word recognition: past and present. *Ear and Hearing 23*(1), 2–40.

Kaiser, J., T. Heidegger, M. Wibral, C. F. Altmann, and W. Lutzenberger (2007). Alpha synchronization during auditory spatial short-term memory. *Neuroreport 18*(11), 1129–1132.

Kalikow, D. N., K. N. Stevens, and L. L. Elliott (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America 61*, 1337–1351.

Kayser, C., R. A. A. Ince, and S. Panzeri (2012). Analysis of Slow (Theta) Oscillations as a Potential Temporal Reference Frame for Information Coding in Sensory Cortices. *PLoS Computational Biology 8*(10), e1002717.

Keil, J., N. Müller, T. Hartmann, and N. Weisz (2014). Prestimulus beta power and phase synchrony influence the sound-induced flash illusion. *Cerebral Cortex 24*(5), 1278–1288.

Kerlin, J. R., A. J. Shahin, and L. M. Miller (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *The Journal of Neuroscience 30*(2), 620–628.

Keuleers, E., P. Lacey, K. Rastle, and M. Brysbaert (2012). The British Lexicon Project: lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods 44*(1), 287–304.

Khateb, A., A. J. Pegna, T. Landis, M. S. Mouthon, and J.-M. Annoni (2010). On the origin of the N400 effects: an ERP waveform and source localization analysis in three matching tasks. *Brain Topography 23*, 311–320.

Kinsey, K., S. J. Anderson, A. Hadjipapas, and I. E. Holliday (2011). The role of oscillatory brain activity in object processing and figure-ground segmentation in human vision. *International Journal of Psychophysiology 79*(3), 392–400.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews 29*(2–3), 169–195.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences 16*(12), 606–617.

Klimesch, W., P. Sauseng, and S. Hanslmayr (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Research Reviews 53*(1), 63–88.

Klimesch, W., P. Sauseng, S. Hanslmayr, W. Gruber, and R. Freunberger (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neuroscience and Biobehavioral Reviews 31*, 1003–1016.

Kotz, S. A., S. F. Cappa, D. Y. von Cramon, and A. D. Friederici (2002). Modulation of the lexical-semantic network by auditory semantic priming: an event-related functional MRI study. *NeuroImage 17*(4), 1761–1772.

Kutas, M. and K. D. Federmeier (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences 4*(12), 463–470.

Kutas, M. and K. D. Federmeier (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology 62*, 621–647.

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

Kutas, M. and S. A. Hillyard (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature 307*, 161–163.

Kutas, M. and C. Van Petten (1994). Psycholinguistics electrified. Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics*, pp. 83–143. San Diego, CA: Academic Press.

Lachaux, J.-P., A. Lutz, D. Rudrauf, D. Cosmelli, M. Le Van Quyen, J. Martinerie, and F. Varela (2002). Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence. *Neurophysiologie Clinique/Clinical Neurophysiology 32*(3), 157–174.

Lachaux, J. P., E. Rodriguez, J. Martinerie, and F. J. Varela (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping 8*(4), 194–208.

Ladefoged, P. (2005). *Vowels and consonants: an introduction to the sounds of languages.* Malden, Mass. [u.a.]: Blackwell.

Lakatos, P., G. Karmos, A. D. Mehta, I. Ulbert, and C. E. Schroeder (2008). Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection. *Science 320*(5872), 110–113.

Lakatos, P., A. S. Shah, K. H. Knuth, I. Ulbert, G. Karmos, and C. E. Schroeder (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology 94*(3), 1904–1911.

Larsby, B., M. Hällgren, B. Lyxell, and S. Arlinger (2005). Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology 44*(3), 131–143.

Laszlo, S., M. Stites, and K. D. Federmeier (2012). Won't get fooled again: An event-related potential study of task and repetition effects on the semantic processing of items without semantics. *Language and Cognitive Processes 27*, 257–274.

Lau, E., D. Almeida, P. C. Hines, and D. Poeppel (2009). A lexical basis for N400 context effects: evidence from MEG. *Brain and Language 111*(3), 161–172.

Lau, E. F., P. J. Holcomb, and G. R. Kuperberg (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience 25*(3), 484–502.

Lau, E. F., C. Phillips, and D. Poeppel (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews. Neuroscience 9*, 920–933.

Lee, A. K. C., E. Larson, R. K. Maddox, and B. G. Shinn-Cunningham (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hearing Research 307*, 111–120.

Lehtelä, L., R. Salmelin, and R. Hari (1997). Evidence for reactive magnetic 10-Hz rhythm in the human auditory cortex. *Neuroscience Letters 222*, 111–114.

Leiberg, S., W. Lutzenberger, and J. Kaiser (2006). Effects of memory load on cortical oscillatory activity during auditory pattern working memory. *Brain Research 1120*(1), 131–140.

Leske, S., A. Tse, N. N. Oosterhof, T. Hartmann, N. Müller, J. Keil, and N. Weisz (2013). The strength of alpha and beta oscillations parametrically scale with the strength of an illusory auditory percept. *NeuroImage 88C*, 69–78.

Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America 49*(2B), 467–477.

Lin, D. (1998). Automatic retrieval and clustering of similar words. In *Proceedings of the 17th International Conference on Computational Linguistics*, Volume 2 of *COLING '98*, Stroudsburg, PA, USA, pp. 768–774. Association for Computational Linguistics.

Ling, S. and M. Carrasco (2006). Sustained and transient covert attention enhance the signal via different contrast response functions. *Vision Research 46*(8-9), 1210–1220.

Lisman, J. E. and O. Jensen (2013). The θ–γ neural code. *Neuron 77*(6), 1002–1016.

Luce, P. A. and M. S. Cluff (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception and Psychophysics 60*(3), 484–490.

Luce, P. A. and D. B. Pisoni (1998). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing 19*, 1–36.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, Mass: MIT Press.

Luo, H., F. T. Husain, B. Horwitz, and D. Poeppel (2005). Discrimination and categorization of speech and non-speech sounds in an MEG delayed-match-to-sample study. *NeuroImage 28*(1), 59–71.

Luo, H. and D. Poeppel (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron 54*(6), 1001–1010.

Lutzenberger, W., T. Elbert, and B. Rockstroh (1985). *Das EEG: Psychophysiologie und Methodik von Spontan-EEG und ereigniskorrelierten Potentialen.* Berlin [u.a.]: Springer.

Ma, W. J., M. Husain, and P. M. Bays (2014). Changing concepts of working memory. *Nature Neuroscience 17*(3), 347–356.

Macmillan, N. A. and C. D. Creelman (2005). *Detection theory: a user's guide.* Mahwah, NJ [u.a.]: Erlbaum.

Makeig, S., S. Debener, J. Onton, and A. Delorme (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences 8*(5), 204–210.

Maris, E. and R. Oostenveld (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods 164*(1), 177–190.

Marslen-Wilson, W. and L. K. Tyler (1980). The temporal structure of spoken language understanding. *Cognition 8*, 1–71.

Marslen-Wilson, W. and P. Zwitserlood (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance 15*, 576–585.

Marslen-Wilson, W. D. (1980). Speech Understanding as a Psychological Process. In J. C. Simon (Ed.), *Spoken Language Generation and Understanding*, Number 59 in NATO Advanced Study Institutes Series, pp. 39–67. Springer Netherlands.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition 25*(1-2), 71–102.

Mathewson, K. E., G. Gratton, M. Fabiani, D. M. Beck, and T. Ro (2009). To see or not to see: prestimulus alpha phase predicts visual awareness. *The Journal of Neuroscience 29*(9), 2725–2732.

Mathewson, K. E., A. Lleras, D. M. Beck, M. Fabiani, T. Ro, and G. Gratton (2011). Pulsed out of awareness: EEG alpha oscillations represent a pulsed-inhibition of ongoing cortical processing. *Frontiers in Psychology 2*, 99.

Mattys, S. (1997). The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin and Review 4*(3), 310–329.

Mattys, S. L. (2004). Stress versus coarticulation: toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology. Human Perception and Performance 30*(2), 397–408.

Mattys, S. L., J. Brooks, and M. Cooke (2009). Recognizing speech under a processing load: dissociating energetic from informational factors. *Cognitive Psychology 59*(3), 203–243.

Mattys, S. L., M. H. Davis, A. R. Bradlow, and S. K. Scott (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes 27*(7-8), 953–978.

Mazaheri, A. and O. Jensen (2006). Posterior alpha activity is not phase-reset by visual stimuli. *Proceedings of the National Academy of Sciences of the United States of America 103*(8), 2948–2952.

Mazaheri, A. and O. Jensen (2008). Asymmetric amplitude modulations of brain oscillations generate slow evoked responses. *The Journal of Neuroscience 28*, 7781–7787.

McClelland, J. L. and J. L. Elman (1986). The TRACE model of speech perception. *Cognitive Psychology 18*(1), 1–86.

McClelland, J. L., D. Mirman, and L. L. Holt (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences 10*(8), 363–369.

McGettigan, C., A. Faulkner, I. Altarelli, J. Obleser, H. Baverstock, and S. K. Scott (2012). Speech comprehension aided by multiple modalities: behavioural and neural interactions. *Neuropsychologia 50*(5), 762–776.

McGurk, H. and J. MacDonald (1976). Hearing lips and seeing voices. *Nature 264*(5588), 746–748.

Medendorp, W. P., G. F. I. Kramer, O. Jensen, R. Oostenveld, J.-M. Schoffelen, and P. Fries (2007). Oscillatory activity in human parietal and occipital cortex shows hemispheric lateralization and memory effects in a delayed double-step saccade task. *Cerebral Cortex 17*, 2364–2374.

Melara, R. D., A. Rao, and Y. Tong (2002). The duality of selection: Excitatory and inhibitory processes in auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance 28*(2), 279–306.

Meyer, L., J. Obleser, and A. D. Friederici (2013). Left parietal alpha enhancement during working memory-intensive sentence processing. *Cortex 49*(3), 711–721.

Miller, G. A., G. A. Heise, and W. Lichten (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology 41*, 329–335.

Miller, G. A. and S. Isard (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior 2*, 217–228.

Miller, G. A. and P. E. Nicely (1955). An Analysis of Perceptual Confusions Among Some English Consonants. *The Journal of the Acoustical Society of America 27*(2), 338–352.

Min, B.-K., N. A. Busch, S. Debener, C. Kranczioch, S. Hanslmayr, A. K. Engel, and C. S. Herrmann (2007). The best of both worlds: phase-reset of human EEG alpha activity and additive power contribute to ERP generation. *International Journal of Psychophysiology 65*, 58–68.

Minicucci, D., S. Guediche, and S. E. Blumstein (2013). An fMRI examination of the effects of acoustic-phonetic and lexical competition on access to the lexical-semantic network. *Neuropsychologia 51*(10), 1980–1988.

Mitra, P. P. and B. Pesaran (1999). Analysis of dynamic brain imaging data. *Biophysical Journal 76*, 691–708. 2.

Morillon, B., K. Lehongre, R. S. J. Frackowiak, A. Ducorps, A. Kleinschmidt, D. Poeppel, and A.-L. Giraud (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences of the United States of America 107*(43), 18688–18693.

Mulder, M. J., L. van Maanen, and B. U. Forstmann (2014). Perceptual decision neurosciences – A model-based review. *Neuroscience*.

Müller, N. and N. Weisz (2012). Lateralized Auditory Cortical Alpha Band Activity and Interregional Connectivity Pattern Reflect Anticipation of Target Sounds. *Cerebral Cortex 22*(7), 1604–1613.

Munhall, K. G., P. Gribble, L. Sacco, and M. Ward (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics 58*(3), 351–362.

Neuling, T., S. Rach, S. Wagner, C. H. Wolters, and C. S. Herrmann (2012). Good vibrations: oscillatory phase shapes perception. *NeuroImage 63*(2), 771–778.

Newman, R. L. and J. F. Connolly (2009). Electrophysiological markers of pre-lexical speech processing: evidence for bottom-up and top-down effects on spoken word processing. *Biological Psychology 80*, 114–121.

Newman, R. S., J. R. Sawusch, and P. A. Luce (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology. Human Perception and Performance 23*(3), 873–889.

Ng, B. S. W., N. K. Logothetis, and C. Kayser (2013). EEG phase patterns reflect the selectivity of neural firing. *Cerebral Cortex 23*(2), 389–398.

Ng, B. S. W., T. Schroeder, and C. Kayser (2012). A Precluding But Not Ensuring Role of Entrained Low-Frequency Oscillations for Auditory Perception. *The Journal of Neuroscience 32*(35), 12268–12276.

Norris, D. (2006). The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychological Review 113*(2), 327–357.

Norris, D. (2009). Putting it all together: a unified account of word recognition and reaction-time distributions. *Psychological Review 116*(1), 207–219.

Norris, D. and J. M. McQueen (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review 115*(2), 357–395.

Norris, D., J. M. McQueen, and A. Cutler (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences 23*(3), 299–325; discussion 325–370.

Nourski, K. V., R. A. Reale, H. Oya, H. Kawasaki, C. K. Kovach, H. Chen, M. A. Howard, and J. F. Brugge (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *The Journal of Neuroscience 29*(49), 15564–15574.

Oberlaender, M., C. P. J. d. Kock, R. M. Bruno, A. Ramirez, H. S. Meyer, V. J. Dercksen, M. Helmstaedter, and B. Sakmann (2012). Cell Type–Specific Three-Dimensional Structure of Thalamocortical Circuits in a Column of Rat Vibrissal Cortex. *Cerebral Cortex 22*(10), 2375–2391.

Obleser, J., F. Eisner, and S. A. Kotz (2012). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *The Journal of Neuroscience 28*(32), 8116–8123.

Obleser, J., T. Elbert, A. Lahiri, and C. Eulitz (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Brain Research. Cognitive Brain Research 15*(3), 207–213.

Obleser, J., B. Herrmann, and M. J. Henry (2012). Neural Oscillations in Speech: Don't be Enslaved by the Envelope. *Frontiers in Human Neuroscience 6*, 250.

Obleser, J. and S. A. Kotz (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex 20*, 633–640.

Obleser, J. and S. A. Kotz (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage 55*(2), 713–723.

Obleser, J. and N. Weisz (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebral Cortex 22*(11), 2466–2477.

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

Obleser, J., M. Wöstmann, N. Hellbernd, A. Wilsch, and B. Maess (2012). Adverse Listening Conditions and Memory Load Drive a Common Alpha Oscillatory Network. *The Journal of Neuroscience 32*(36), 12376–12383.

Oostenveld, R., P. Fries, E. Maris, and J.-M. Schoffelen (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience 2011*, 156869.

Oostenveld, R. and P. Praamstra (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology 112*(4), 713–719.

Oostenveld, R., D. F. Stegeman, P. Praamstra, and A. van Oosterom (2003). Brain symmetry and topographic analysis of lateralized event-related potentials. *Clinical Neurophysiology 114*(7), 1194–1202.

Paulesu, E., C. D. Frith, and R. S. J. Frackowiak (1993). The neural correlates of the verbal component of working memory. *Nature 362*(6418), 342–345.

Paulmann, S., D. V. M. Ott, and S. A. Kotz (2011). Emotional speech perception unfolding in time: the role of the basal ganglia. *PloS One 6*, e17694–e17694.

Peelle, J. E. and M. H. Davis (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology 3*, 320.

Peelle, J. E., J. Gross, and M. H. Davis (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex 23*(6), 1378–1387.

Peña, M. and L. Melloni (2012). Brain oscillations during spoken sentence processing. *Journal of Cognitive Neuroscience 24*(5), 1149–1164.

Petkov, C. I., X. Kang, K. Alho, O. Bertrand, E. W. Yund, and D. L. Woods (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience 7*(6), 658–663.

Pfurtscheller, G., A. Stancák Jr., and C. Neuper (1996). Event-related synchronization (ERS) in the alpha band — an electrophysiological correlate of cortical idling: A review. *International Journal of Psychophysiology 24*(1–2), 39–46.

Phatak, S. A. and J. B. Allen (2007). Consonant and vowel confusions in speech-weighted noise. *The Journal of the Acoustical Society of America 121*(4), 2312–2326.

Phatak, S. A., A. Lovitt, and J. B. Allen (2008). Consonant confusions in white noise. *The Journal of the Acoustical Society of America 124*(2), 1220–1233.

Pichora–Fuller, M. K. and G. Singh (2006). Effects of Age on Auditory and Cognitive Processing: Implications for Hearing Aid Fitting and Audiologic Rehabilitation. *Trends in Amplification 10*(1), 29–59.

Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *International Journal of Audiology 42 Suppl 2*, 2S26–32.

Picton, T., S. Bentin, P. Berg, E. Donchin, S. Hillyard, R. Johnson, G. Miller, W. Ritter, D. Ruchkin, M. Rugg, and M. Taylor (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology 37*(2), 127–152.

Piquado, T., K. A. Q. Cousins, A. Wingfield, and P. Miller (2010). Effects of degraded sensory input on memory for speech: behavioral data and a test of biologically constrained computational models. *Brain Research 1365*, 48–65.

Pizzagalli, D. A. (2007). Electroencephalography and High-Density Electrophysiological Source Localization. In J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (Eds.), *Handbook of Psychophysiology*, pp. 56–84. Cambridge [u.a.]: Cambridge Univ. Press.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication 41*(1), 245–255.

Price, C. J. (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.

Proverbio, A. M. and R. Adorni (2008). Orthographic familiarity, phonological legality and number of orthographic neighbours affect the onset of ERP lexical effects. *Behavioral and Brain Functions 4*, 27.

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

Rabinowitz, N. C., B. D. B. Willmore, A. J. King, and J. W. H. Schnupp (2013). Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biology 11*(11), e1001710.

Raettig, T. and S. A. Kotz (2008). Auditory processing of different types of pseudo-words: an event-related fMRI study. *NeuroImage 39*, 1420–1428.

Ralph, M. A. L. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical Transactions of the Royal Society B: Biological Sciences 369*(1634), 20120392.

Ratcliff, R., P. Gomez, and G. McKoon (2004). A diffusion model account of the lexical decision task. *Psychological Review 111*(1), 159–182.

Ratcliff, R. and G. McKoon (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation 20*(4), 873–922.

Rauschecker, J. P. and S. K. Scott (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience 12*(6), 718–724.

Rice, D. M. and E. C. Hagstrom (1989). Some evidence in support of a relationship between human auditory signal-detection performance and the phase of the alpha cycle.

Perceptual and Motor Skills 69(2), 451–457.

Riecke, L., F. Esposito, M. Bonte, and E. Formisano (2009). Hearing illusory sounds in noise: the timing of sensory-perceptual transformations in auditory cortex. *Neuron 64*, 550–561.

Romei, L., I. J. A. Wambacq, J. Besing, J. Koehnke, and J. Jerger (2011). Neural indices of spoken word processing in background multi-talker babble. *International Journal of Audiology 50*, 321–333.

Romei, V., J. Gross, and G. Thut (2010). On the role of prestimulus alpha rhythms over occipito-parietal areas in visual input regulation: correlation or causation? *The Journal of Neuroscience 30*(25), 8692–8697.

Romei, V., G. Thut, R. M. Mok, P. G. Schyns, and J. Driver (2012). Causal implication by rhythmic transcranial magnetic stimulation of alpha frequency in feature-based local vs. global attention. *The European Journal of Neuroscience 35*(6), 968–974.

Rönnberg, J., T. Lunner, A. Zekveld, P. Sörqvist, H. Danielsson, B. Lyxell, O. Dahlström, C. Signoret, S. Stenfelt, M. K. Pichora-Fuller, and M. Rudner (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience 7*, 31.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper and L. V. Hedges (Eds.), *The Handbook of Research Synthesis*, pp. 231–244. Russell Sage Foundation.

Rosenthal, R. and D. B. Rubin (2003). r equivalent: A Simple Effect Size Indicator. *Psychological Methods 8*(4), 492–496.

Roux, F. and P. J. Uhlhaas (2014). Working memory and neural oscillations: alpha–gamma versus theta–gamma codes for distinct WM information? *Trends in Cognitive Sciences 18*(1), 16–25.

Roux, F., M. Wibral, W. Singer, J. Aru, and P. J. Uhlhaas (2013). The Phase of Thalamic Alpha Activity Modulates Cortical Gamma-Band Activity: Evidence from Resting-State MEG Recordings. *The Journal of Neuroscience 33*(45), 17827–17835.

Rudner, M., T. Lunner, T. Behrens, E. S. Thorén, and J. Rönnberg (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology 23*(8), 577–589.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory & Cognition 18*, 367–379.

Samuel, A. G. and W. H. Ressler (1986). Attention within auditory word perception: insights from the phonemic restoration illusion. *Journal of Experimental Psychology. Human Perception and Performance 12*(1), 70–79.

Sato, M., J.-L. Schwartz, C. Abry, M.-A. Cathiard, and H. Loevenbruck (2006). Multistable syllables as enacted percepts: a source of an asymmetric bias in the verbal transformation effect. *Perception & Psychophysics 68*(3), 458–474.

Sato, M., N. Vallée, J.-L. Schwartz, and I. Rousset (2007). A perceptual correlate of the labial-coronal effect. *Journal of Speech, Language, and Hearing Research 50*(6), 1466–1480.

Scharinger, M. and V. Felder (2011). ERP signatures of cross-modal semantic fragment priming: early context effects in speech perception. *International Journal of Psychophysiology 80*, 19–27.

Schneider, B. A., L. Li, and M. Daneman (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology 18*(7), 559–572.

Schroeder, C. E. and P. Lakatos (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences 32*(1), 9–18.

Schröger, E., M. Tervaniemi, and M. Huotilainen (2004). Bottom-up and top-down flows of information within auditory memory: electrophysiological evidence. In C. Kaernbach, E. Schröger, and H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition*, pp. 389–407. Mahwah, NJ: Erlbaum.

Schubert, R., S. Haufe, F. Blankenburg, A. Villringer, and G. Curio (2009). Now you'll feel it, now you won't: EEG rhythms predict the effectiveness of perceptual masking. *Journal of Cognitive Neuroscience 21*(12), 2407–2419.

Schwartz, J.-L., A. Basirat, L. Ménard, and M. Sato (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics 25*(5), 336–354.

Scott, S. K. and C. McGettigan (2013). The neural processing of masked speech. *Hearing research 303*, 58–66.

Scott, S. K., S. Rosen, C. P. Beaman, J. P. Davis, and R. J. S. Wise (2009). The neural processing of masked speech: evidence for different mechanisms in the left and right temporal lobes. *The Journal of the Acoustical Society of America 125*(3), 1737–1743.

Scott, S. K., S. Rosen, L. Wickham, and R. J. S. Wise (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *The Journal of the Acoustical Society of America 115*(2), 813–821.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences 12*(5), 182–186.

Sivonen, P., B. Maess, and A. D. Friederici (2006). Semantic retrieval of spoken words with an obliterated initial phoneme in a sentence context. *Neuroscience Letters 408*(3), 220–225.

Slepian, D. (1978). Prolate spheroidal wave functions, fourier analysis, and uncertainty – V: the discrete case. *The Bell System Technical Journal 57*(5), 1371–1430.

Spaak, E., M. Bonnefond, A. Maier, D. A. Leopold, and O. Jensen (2012). Layer-specific entrainment of gamma-band neural activity by the alpha rhythm in monkey visual cortex. *Current Biology 22*(24), 2313–2318.

Spaak, E., F. P. de Lange, and O. Jensen (2014). Local entrainment of alpha oscillations by visual stimuli causes cyclic modulation of perception. *The Journal of Neuroscience 34*(10), 3536–3544.

Speckmann, E.-J. and C. E. Elger (2005). Introduction to the Neurophysiological Basis of the EEG and DC Potentials. In E. Niedermeyer and F. H. Lopes da Silva (Eds.), *Electroencephalography: Basic principles, clinical applications, and related fields*, pp. 17–29. Philadelphia: Lippincott Williams & Wilkins.

Spinelli, E., N. Grimault, F. Meunier, and P. Welby (2010). An intonational cue to word segmentation in phonemically identical sequences. *Attention, Perception & Psychophysics 72*(3), 775–787.

Stanovich, K. E. and R. F. West (1983). On priming by a sentence context. *Journal of Experimental Psychology: General 112*, 1–36.

Staudigl, T., S. Hanslmayr, and K.-H. T. Bäuml (2010). Theta oscillations reflect the dynamics of interference in episodic memory retrieval. *The Journal of Neuroscience 30*(34), 11356–11362.

Steinschneider, M., K. V. Nourski, and Y. I. Fishman (2013). Representation of speech in human auditory cortex: is it special? *Hearing Research 305*, 57–73.

Stickney, G. S. and P. F. Assmann (2001). Acoustic and linguistic factors in the perception of bandpass-filtered speech. *The Journal of the Acoustical Society of America 109*, 1157–1165.

Stone, M. A., C. Füllgrabe, and B. C. J. Moore (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America 132*(1), 317–326.

Strauß, A., M. J. Henry, M. Scharinger, and J. Obleser (In revision). Alpha phase determines successful lexical decision in noise.

Strauß, A., S. A. Kotz, and J. Obleser (2013). Narrowed expectancies under degraded speech: revisiting the N400. *Journal of Cognitive Neuroscience 25*(8), 1383–1395.

Strauß, A., S. A. Kotz, M. Scharinger, and J. Obleser (2014). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *NeuroImage 97*, 387–395.

Strauß, A., M. Wöstmann, and J. Obleser (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience 8*, 350.

Strauss, D. J., F. I. Corona-Strauss, C. Trenado, C. Bernarding, W. Reith, M. Latzel, and M. Froehlich (2010). Electrophysiological correlates of listening effort: neurodynamical modeling and measurement. *Cognitive Neurodynamics 4*(2), 119–131.

Summerfield, C. and T. Egner (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences 13*, 403–409.

Taft, M. and K. I. Forster (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior 15*, 607–620.

Taft, M. and G. Hambly (1986). Exploring the Cohort Model of spoken word recognition. *Cognition 22*(3), 259–282.

Tallon-Baudry, C. and O. Bertrand (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences 3*(4), 151–162.

Tallon-Baudry, C., O. Bertrand, C. Delpuech, and J. Permier (1997). Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in humans. *The Journal of Neuroscience 17*(2), 722–734.

Tavabi, K., D. Embick, and T. P. L. Roberts (2011). Word repetition priming-induced oscillations in auditory cortex: a magnetoencephalography study. *Neuroreport 22*, 887–

891.

Taylor, W. L. (1953). "Cloze procedure": a new tool for measuring readability. *Journalism Quarterly 30*, 415–433.

Ten Oever, S., C. E. Schroeder, D. Poeppel, N. van Atteveldt, and E. Zion-Golumbic (2014). Rhythmicity and cross-modal temporal cues facilitate detection. *Neuropsychologia 63C*, 43–50.

Tiesinga, P. H. and T. J. Sejnowski (2010). Mechanisms for Phase Shifting in Cortical Networks and their Role in Communication through Coherence. *Frontiers in Human Neuroscience 4*, 196.

van den Brink, D., C. M. Brown, and P. Hagoort (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition 32*, 364–372.

van Dijk, H., J.-M. Schoffelen, R. Oostenveld, and O. Jensen (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *The Journal of Neuroscience 28*(8), 1816–1823.

Van Petten, C. and M. Kutas (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition 18*, 380–393.

Van Petten, C. and B. J. Luka (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language 97*, 279–293.

Van Petten, C. and B. J. Luka (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology 83*(2), 176–190.

Van Veen, B., W. Van Drongelen, M. Yuchtman, and A. Suzuki (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering 44*(9), 867–880.

van Wassenhove, V., K. W. Grant, and D. Poeppel (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America 102*(4), 1181–1186.

Varela, F. J., A. Toro, E. R. John, and E. L. Schwartz (1981). Perceptual framing and cortical alpha rhythm. *Neuropsychologia 19*(5), 675–686.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science 167*(3917), 392–393.

Weisz, N., T. Hartmann, N. Müller, I. Lorenz, and J. Obleser (2011). Alpha rhythms in audition: cognitive and clinical perspectives. *Frontiers in Psychology 2*.

Weisz, N., C. Lüchinger, G. Thut, and N. Müller (2014). Effects of individual alpha rTMS applied to the auditory cortex and its implications for the treatment of chronic tinnitus. *Human Brain Mapping 35*(1), 14–29.

Welch, P. D. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics 15*(2), 70–73.

Wild, C. J., A. Yusuf, D. E. Wilson, J. E. Peelle, M. H. Davis, and I. S. Johnsrude (2012).

Effortful listening: the processing of degraded speech depends critically on attention. *The Journal of Neuroscience 32*(40), 14010–14021.

Wilsch, A., M. J. Henry, B. Herrmann, B. Maess, and J. Obleser (2014). Alpha Oscillatory Dynamics Index Temporal Expectation Benefits in Working Memory. *Cerebral Cortex*.

Wingfield, A., P. A. Tun, C. K. Koh, and M. J. Rosen (1999). Regaining lost time: adult aging and the effect of time restoration on recall of time-compressed speech. *Psychology and Aging 14*, 380–389.

Wurm, L. H. and A. G. Samuel (1997). Lexical Inhibition and Attentional Allocation during Speech Perception: Evidence from Phoneme Monitoring. *Journal of Memory and Language 36*, 165–187.

Wyart, V., V. de Gardelle, J. Scholl, and C. Summerfield (2012). Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron 76*(4), 847–858.

Xiao, Z., J. X. Zhang, X. Wang, R. Wu, X. Hu, X. Weng, and L. H. Tan (2005). Differential activity in left inferior frontal gyrus for pseudowords and real words: an event-related fMRI study on auditory lexical decision. *Human Brain Mapping 25*(2), 212–221.

Xu, L., C. S. Thompson, and B. E. Pfingst (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *The Journal of the Acoustical Society of America 117*(5), 3255–3267.

Zaehle, T., E. Geiser, K. Alter, L. Jancke, and M. Meyer (2008). Segmental processing in the human auditory dorsal stream. *Brain Research 1220*, 179–190.

Zatorre, R. J. and P. Belin (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex 11*, 946–953.

Zatorre, R. J., E. Meyer, A. Gjedde, and A. C. Evans (1996). PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cerebral Cortex 6*, 21–30.

Zekveld, A. A., S. E. Kramer, and J. M. Festen (2011). Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing 32*(4), 498–510.

Zion Golumbic, E. M., N. Ding, S. Bickel, P. Lakatos, C. A. Schevon, G. M. McKhann, R. R. Goodman, R. Emerson, A. D. Mehta, J. Z. Simon, D. Poeppel, and C. E. Schroeder (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron 77*(5), 980–991.

List of Figures

List of words and pseudowords

Real word	Syllabification	Stress	**Ambiguous**	Syllabification	Stress
Adjutant	Ad · ju · tant	3	Adjatant	Ad · ja · tant	3
Advokat	Ad · vo · kat	3	Advekat	Ad · ve · kat	3
Ameise	A · mei · se	1	Amerse	A · mer · se	1
Ananas	A · na · nas	1	Aninas	A · ni · nas	1
Anorak	A · no · rak	1	Anirak	A · ni · rak	1
Antenne	An · ten · ne	2	Antanne	An · tan · ne	2
Apostel	A · pos · tel	2	Apastel	A · pas · tel	2
Attrappe	At · trap · pe	2	Atroppe	At · trop · pe	2
Banane	Ba · na · ne	2	Banene	Ba · ne · ne	2
Baracke	Ba · ra · cke	2	Baricke	Ba · ri · cke	2
Bariton	Ba · ri · ton	1	Baruton	Ba · ru · ton	1
Batterie	Bat · te · rie	3	Battorie	Bat · to · rie	3
Elefant	E · le · fant	3	Elufant	E · lu · fant	3
Etikett	E · ti · kett	3	Etukett	E · tu · kett	3
Experte	Ex · per · te	2	Expirte	Ex · pir · te	2
Forelle	Fo · rel · le	2	Foralle	Fo · ral · le	2
Fregatte	Fre · gat · te	2	Fregutte	Fre · gut · te	2
Genosse	Ge · nos · se	2	Genasse	Ge · nas · se	2
Geselle	Ge · sel · le	2	Gesille	Ge · sil · le	2
Getreide	Ge · trei · de	2	Getraude	Ge · trau · de	2
Granate	Gra · na · te	2	Granete	Gra · ne · te	2
Hebamme	Heb · am · me	2	Hebomme	Heb · om · me	2
Herberge	Her · ber · ge	1	Herbarge	Her · bar · ge	1
Hospital	Hos · pi · tal	3	Hospatal	Hos · pa · tal	3
Kabeljau	Ka · bel · jau	1	Kaboljau	Ka · bol · jau	1
Kalender	Ka · len · der	2	Kalunder	Ka · lun · der	2
Kamerad	Ka · me · rad	3	Kamirad	Ka · mi · rad	3
Kardinal	Kar · di · nal	3	Kardunal	Kar · du · nal	3
Kartoffel	Kar · tof · fel	2	Karteffel	Kar · tef · fel	2
Kassette	Kas · set · te	2	Kassutte	Kas · sut · te	2
Kavalier	Ka · va · lier	3	Kavolier	Ka · vo · lier	3
Kollege	Kol · le · ge	2	Kolloge	Kol · lo · ge	2
Kommode	Kom · mo · de	2	Komide	Kom · mi · de	2
Komplize	Kom · pli · ze	2	Komploze	Kom · plo · ze	2
Kontrahent	Kon · tra · hent	3	Konträhent	Kon · trü · hent	3

Krokodil	Kro · ko · dil	3	Kroküdil	Kro · kü · dil	3
Laterne	La · ter · ne	2	Lateune	La · teu · ne	2
Magister	Ma · gis · ter	2	Magester	Ma · ges · ter	2
Märtyrer	Mär · ty · rer	1	Märtorer	Mär · to · rer	1
Matratze	Ma · trat · ze	2	Matretze	Ma · tret · ze	2
Matrose	Ma · tro · se	2	Matruse	Ma · tru · se	2
Melone	Me · lo · ne	2	Meline	Me · li · ne	2
Offizier	Of · fi · zier	3	Offazier	Of · fa · zier	3
Orange	O · ran · ge	2	Orenge	Oren · ge	2
Palette	Pa · let · te	2	Palütte	Pa · lüt · te	2
Paprika	Pa · pri · ka	1	Papraka	Pa · pra · ka	1
Papagei	Pa · pa · gei	3	Papugei	Pa · pu · gei	3
Patrone	Pa · tro · ne	2	Patrene	Pa · tre · ne	2
Pelikan	Pe · li · kan	1	Pelekan	Pe · le · kan	1
Perücke	Pe · rü · cke	2	Peracke	Pe · ra · cke	2
Plakette	Pla · ket · te	2	Plakatte	Pla · kat · te	2
Posaune	Po · sau · ne	2	Poseine	Po · sei · ne	2
Rabbiner	Rab · bi · ner	2	Rabboner	Rab · bo · ner	2
Rakete	Ra · ke · te	2	Rakote	Ra · ko · te	2
Salami	Sa · la · mi	2	Salomi	Sa · lo · mi	2
Samurai	Sa · mu · rai	3	Samerai	Sa · me · rai	3
Schatulle	Scha · tul · le	2	Schatelle	Scha · tel · le	2
Sekretär	Se · kre · tär	3	Sekratär	Se · kra · tär	3
Veteran	Ve · te · ran	3	Vetiran	Ve · ti · ran	3
Walküre	Wal · kü · re	2	Walkere	Wal · ke · re	2

Opaque	Syllabification	Stress	**Filler**	Syllabification	Stress
Ajekle	A · jek · le	2	Adjektiv	Ad · jek · tiv	1
Anlarie	An · la · rie	3	Akustik	A · kus · tik	2
Antulpher	An · tul · pher	2	Allegro	Al · le · gro	2
Aspiom	As · pi · om	3	Allergie	Al · ler · gie	3
Atimi	A · ti · mi	1	Alphabet	Al · pha · bet	3
Atzeran	At · ze · ran	3	Amnestie	Am · nes · tie	3
Axopol	Ak · so · pol	3	Anarchie	A · nar · chie	3
Bamagro	Ba · ma · gro	2	Antike	An · ti · ke	2
Baposner	Ba · pos · ner	2	Apathie	A · pa · thie	3
Blamäthie	Bl · mä · thie	3	Appetit	Ap · pe · tit	3
Blamitrik	Bla · mi · trik	2	Askese	As · ke · se	2
Bomete	Bo · me · te	2	Attribut	At · tri · but	3
Charoleg	Cha · ro · leg	3	Blamage	Bla · ma · ge	2
Delotte	De · lo · tte	2	Blasphemie	Blas · phe · mie	3
Denarze	De · nar · ze	2	Botanik	Bo · ta · nik	2
Enobut	E · no · but	3	Charisma	Cha · ris · ma	1
Fodine	Fo · di · ne	2	Debakel	De · ba · kel	2
Foltrappai	Fol · tra · ppei	3	Dezibel	De · zi · bel	1
Frewoda	Fre · wo · da	2	Didaktik	Di · dak · tik	2
Gearkum	Ge · ar · kum	1	Diskrepanz	Dis · kre · panz	3
Gerügei	Ge · rü · gei	2	Domäne	Do · mä · ne	2
Getarak	Ge · ta · rak	3	Elektrik	E · lek · trik	2
Grabade	Gra · ba · de	2	Epilog	E · pi · log	3
Henoket	He · no · ket	3	Exotik	E · xo · tik	2
Herlite	Her · li · te	2	Folklore	Fol · klo · re	2
Hoditik	Ho · di · tik	2	Hierarchie	Hie · rar · chie	3
Hysteltit	Hys · tel · tit	3	Hysterie	Hys · te · rie	3
Ikustal	I · kus · tal	3	Idiom	I · di · om	3
Inlekan	In · le · kan	3	Intellekt	In · tel · lekt	3
Inpetät	In · pe · tät	3	Intrige	In · tri · ge	2
Kagiste	Ka · gis · te	2	Kalauer	Ka · lau · er	1
Kargane	Kar · ga · ne	2	Kollektiv	Kol · lek · tiv	3
Karnirad	Kar · ni · rad	3	Konjunktiv	Kon · junk · tiv	1
Konkrinym	Kon · kri · nym	3	Litanai	Li · ta · nei	3
Kotikaat	Ko · ti · kaat	3	Maxime	Ma · xi · me	2
Lafijau	La · fi · jau	3	Metapher	Me · ta · pher	2
Likebat	Li · ke · bat	3	Minimum	Mi · ni · mum	1
Märantik	Mä · ran · tik	2	Monarchie	Mo · nar · chie	3
Metyrne	Me · tyr · ne	2	Monopol	Mo · no · pol	3
Mimaffel	Mi · ma · ffel	2	Nostalgie	Nos · tal · gie	3
Moplibel	Mo · pli · bel	3	Parabel	Pa · ra · bel	2
Motritik	Mo · tri · tik	2	Parodie	Pa · ro · die	3
Otebel	O · te · bel	2	Parole	Pa · ro · le	2
Palige	Pa · li · ge	2	Prädikat	Prä · di · kat	3
Pamange	Pa · man · ge	2	Privileg	Pri · vi · leg	3

Pelede	Pe · le · de	2	Prozedur	Pro · ze · dur	3
Plabagie	Pla · ba · gie	3	Pseudonym	Pseu · do · nym	3
Poterse	Po · ter · se	2	Quantität	Quan · ti · tät	3
Prisiske	Pri · sis · ke	2	Resistenz	Re · sis · tenz	3
Prokane	Pro · ka · ne	2	Sakrileg	Sa · kri · leg	3
Ranatel	Ra · na · tel	2	Satire	Sa · ti · re	2
Ratalge	Ra · tal · ge	2	Schlamassel	Schla · mas · sel	2
Saklope	Sa · klo · pe	2	Semantik	Se · man · tik	2
Schararer	Scha · ra · rer	2	Stakato	Sta · ka · to	2
Schlalecke	Schla · le · cke	2	Symmetrie	Sym · me · trie	3
Sereka	Se · re · ka	1	Synonym	Sy · no · nym	3
Setalle	Se · ta · le	2	Terminus	Ter · mi · nus	1
Statrodur	Sta · tro · dur	3	Trilogie	Tri · lo · gie	3
Vetroge	Ve · tro · ge	2	Unikum	U · ni · kum	1
Zöberhent	Zö · ber · hent	3	Zölibat	Zö · li · bat	3

List of cloze probability sentences

		Pronoun	Verb	Adverb	Object
1.	Hh	Er	kapert	hinterrücks	Schiffe
	Hl	Er	kapert	hinterrücks	Boote
	Lh	Er	löchert	hinterrücks	Schiffe
	Ll	Er	löchert	hinterrücks	Boote
2.	Hh	Er	bereist	freudig	Länder
	Hl	Er	bereist	freudig	Staaten
	Lh	Er	befährt	freudig	Länder
	Ll	Er	befährt	freudig	Staaten
3.	Hh	Er	schlachtet	sorgfältig	Schweine
	Hl	Er	schlachtet	sorgfältig	Bullen
	Lh	Er	bewacht	sorgfältig	Schweine
	Ll	Er	bewacht	sorgfältig	Bullen
4.	Hh	Sie	näht	sonntags	Kleider
	Hl	Sie	näht	sonntags	Stücke
	Lh	Sie	klaut	sonntags	Kleider
	Ll	Sie	klaut	sonntags	Stücke
5.	Hh	Sie	mäht	ständig	Rasen
	Hl	Sie	mäht	ständig	Flächen
	Lh	Sie	zertritt	ständig	Rasen
	Ll	Sie	zertritt	ständig	Flächen
6.	Hh	Sie	baut	clever	Häuser
	Hl	Sie	baut	clever	Villen
	Lh	Sie	beantragt	clever	Häuser
	Ll	Sie	beantragt	clever	Villen
7.	Hh	Er	bohrt	gezielt	Löcher
	Hl	Er	bohrt	gezielt	Bretter
	Lh	Er	kittet	gezielt	Löcher
	Ll	Er	kittet	gezielt	Bretter
8.	Hh	Sie	spült	umsichtig	Geschirr
	Hl	Sie	spült	umsichtig	Porzellan
	Lh	Sie	empfiehlt	umsichtig	Geschirr
	Ll	Sie	empfiehlt	umsichtig	Porzellan

9.	Hh	Er	singt	täglich	Lieder
	Hl	Er	singt	täglich	Schlager
	Lh	Er	produziert	täglich	Lieder
	Ll	Er	produziert	täglich	Schlager
10.	Hh	Er	reitet	gekonnt	Pferde
	Hl	Er	reitet	gekonnt	Zebras
	Lh	Er	pflegt	gekonnt	Pferde
	Ll	Er	pflegt	gekonnt	Zebras
11.	Hh	Sie	dirigiert	kurzfristig	Konzerte
	Hl	Sie	dirigiert	kurzfristig	Musiker
	Lh	Sie	leitet	kurzfristig	Konzerte
	Ll	Sie	leitet	kurzfristig	Musiker
12.	Hh	Er	zapft	fröhlich	Bier
	Hl	Er	zapft	fröhlich	Wein
	Lh	Er	leert	fröhlich	Bier
	Ll	Er	leert	fröhlich	Wein
13.	Hh	Sie	malt	fachmännisch	Bilder
	Hl	Sie	malt	fachmännisch	Arbeiten
	Lh	Sie	fertigt	fachmännisch	Bilder
	Ll	Sie	fertigt	fachmännisch	Arbeiten
14.	Hh	Sie	schickt	häufig	Pakete
	Hl	Sie	schickt	häufig	Sendungen
	Lh	Sie	erhält	häufig	Pakete
	Ll	Sie	erhält	häufig	Sendungen
15.	Hh	Er	tapeziert	momentan	Wände
	Hl	Er	tapeziert	momentan	Mauern
	Lh	Er	bekritzelt	momentan	Wände
	Ll	Er	bekritzelt	momentan	Mauern
16.	Hh	Sie	jätet	dauernd	Unkraut
	Hl	Sie	jätet	dauernd	Blumen
	Lh	Sie	schnippelt	dauernd	Unkraut
	Ll	Sie	schnippelt	dauernd	Blumen
17.	Hh	Sie	liest	massenhaft	Bücher
	Hl	Sie	liest	massenhaft	Werke
	Lh	Sie	liebt	massenhaft	Bücher
	Ll	Sie	liebt	massenhaft	Werke
18.	Hh	Sie	lutscht	immerzu	Bonbons
	Hl	Sie	lutscht	immerzu	Süßes
	Lh	Sie	verschenkt	immerzu	Bonbons
	Ll	Sie	verschenkt	immerzu	Süßes
19.	Hh	Er	zerbricht	leichtsinnig	Gläser
	Hl	Er	zerbricht	leichtsinnig	Schüsseln
	Lh	Er	öffnet	leichtsinnig	Gläser

	Ll	Er	öffnet	leichtsinnig	Schüsseln
20.	Hh	Er	buchstabiert	kurzerhand	Wörter
	Hl	Er	buchstabiert	kurzerhand	Sätze
	Lh	Er	überfliegt	kurzerhand	Wörter
	Ll	Er	überfliegt	kurzerhand	Sätze
21.	Hh	Er	schält	reichlich	Kartoffeln
	Hl	Er	schält	reichlich	Bananen
	Lh	Er	kaut	reichlich	Kartoffeln
	Ll	Er	kaut	reichlich	Bananen
22.	Hh	Sie	kämmt	vorsichtig	Haare
	Hl	Sie	kämmt	vorsichtig	Frisuren
	Lh	Sie	schwärzt	vorsichtig	Haare
	Ll	Sie	schwärzt	vorsichtig	Frisuren
23.	Hh	Er	schreibt	schwungvoll	Briefe
	Hl	Er	schreibt	schwungvoll	Akten
	Lh	Er	trägt	schwungvoll	Briefe
	Ll	Er	trägt	schwungvoll	Akten
24.	Hh	Sie	tanzt	allein	Walzer
	Hl	Sie	tanzt	allein	Polkas
	Lh	Sie	spielt	allein	Walzer
	Ll	Sie	spielt	allein	Polkas
25.	Hh	Er	pflügt	behände	Äcker
	Hl	Er	pflügt	behände	Wiesen
	Lh	Er	bestellt	behände	Äcker
	Ll	Er	bestellt	behände	Wiesen
26.	Hh	Er	knallt	lautstark	Türen
	Hl	Er	knallt	lautstark	Pforten
	Lh	Er	schraubt	lautstark	Türen
	Ll	Er	schraubt	lautstark	Pforten
27.	Hh	Sie	klärt	grimmig	Fragen
	Hl	Sie	klärt	grimmig	Themen
	Lh	Sie	brummt	grimmig	Fragen
	Ll	Sie	brummt	grimmig	Themen
28.	Hh	Er	raucht	heimlich	Zigaretten
	Hl	Er	raucht	heimlich	Päckchen
	Lh	Er	beschafft	heimlich	Zigaretten
	Ll	Er	beschafft	heimlich	Päckchen
29.	Hh	Er	stiftet	begierig	Unruhe
	Hl	Er	stiftet	begierig	Randale
	Lh	Er	erträumt	begierig	Unruhe
	Ll	Er	erträumt	begierig	Randale
30.	Hh	Er	schluckt	blindlings	Pillen
	Hl	Er	schluckt	blindlings	Drogen

	Lh	Er	tauscht	blindlings	Pillen
	Ll	Er	tauscht	blindlings	Drogen
31.	Hh	Sie	schmiert	mittags	Brote
	Hl	Sie	schmiert	mittags	Schnitten
	Lh	Sie	speist	mittags	Brote
	Ll	Sie	speist	mittags	Schnitten
32.	Hh	Er	schmiedet	verlegen	Pläne
	Hl	Er	schmiedet	verlegen	Ideen
	Lh	Er	stottert	verlegen	Pläne
	Ll	Er	stottert	verlegen	Ideen
33.	Hh	Sie	tankt	stillschweigend	Benzin
	Hl	Sie	tankt	stillschweigend	Diesel
	Lh	Sie	stiehlt	stillschweigend	Benzin
	Ll	Sie	stiehlt	stillschweigend	Diesel
34.	Hh	Sie	steigt	pausenlos	Treppen
	Hl	Sie	steigt	pausenlos	Absätze
	Lh	Sie	läuft	pausenlos	Treppen
	Ll	Sie	läuft	pausenlos	Absätze
35.	Hh	Sie	summt	betörend	Melodien
	Hl	Sie	summt	betörend	Harmonien
	Lh	Sie	haucht	betörend	Melodien
	Ll	Sie	haucht	betörend	Harmonien
36.	Hh	Er	verschrottet	illegal	Autos
	Hl	Er	verschrottet	illegal	Wracks
	Lh	Er	beseitigt	illegal	Autos
	Ll	Er	beseitigt	illegal	Wracks
37.	Hh	Sie	schleckt	genüsslich	Eis
	Hl	Sie	schleckt	genüsslich	Pudding
	Lh	Sie	verdrückt	genüsslich	Eis
	Ll	Sie	verdrückt	genüsslich	Pudding
38.	Hh	Sie	häkelt	geschwind	Deckchen
	Hl	Sie	häkelt	geschwind	Aufleger
	Lh	Sie	erspäht	geschwind	Deckchen
	Ll	Sie	erspäht	geschwind	Aufleger
39.	Hh	Sie	windelt	eigentlich	Babys
	Hl	Sie	windelt	eigentlich	Kinder
	Lh	Sie	überwacht	eigentlich	Babys
	Ll	Sie	überwacht	eigentlich	Kinder
40.	Hh	Er	hobelt	geschäftig	Bretter
	Hl	Er	hobelt	geschäftig	Möhren
	Lh	Er	verpackt	geschäftig	Bretter
	Ll	Er	verpackt	geschäftig	Möhren

Summary

Introduction

Speech comprehension is often challenging. Background noise, e.g. caused by traffic, or spectral degradations, for example in cochlea implants, impose problematic hearing situations. Still, native speakers are able to compensate for the sparse perceptual evidence. The current thesis aims at determining neural temporal dynamics underlying spoken word recognition in adverse listening conditions.

We used words and word-like pseudowords in a lexical decision task (Goldinger, 1996) to compare successful and failed lexico-semantic processing, since mapping of semantic meaning onto unknown pseudowords is not possible. Also, we used sentences in which the cloze probability was manipulated so that the preceding context is more or less semantically predictive for the sentence-final word (Kutas and Hillyard, 1980). The robustness to dissociate words and pseudowords on the one hand and the robustness to predict words from context on the other hand was tested by introducing background noise and by degrading the spectral information of the speech signal.

One general hypothesis of the current thesis is that processes of spoken word recognition are reflected in slow neural oscillations. Oscillatory mechanisms indicate dynamic synchronization of brain areas in certain frequency bands, thus temporarily enabling or inhibiting information processing. Moreover, oscillations can index fluctuations in neural excitability (Haegens et al., 2011). In particular, synchronization in the alpha frequency band ($\sim$10 Hz) has been associated with selective inhibition of task-irrelevant noise (Jensen and Mazaheri, 2010) and alpha desynchronization with successful memory encoding (Hanslmayr et al., 2012). Furthermore, oscillations in the theta frequency band ($\sim$4 Hz) have been related to periodic reactivation of maintained information (Fuentemilla et al., 2010) and lexico-semantic memory retrieval (Bastiaansen et al., 2008).

Slow neural oscillations might also be important for the acoustic analysis of the incoming speech signal (Ghitza, 2011). They might chunk speech into smaller units by temporally aligning peaks of neural excitability with the most informative acoustic cues (Giraud and Poeppel, 2012). Hence, effects in the slightly faster alpha frequency range might be observed if segmental information such as vowels had been manipulated. Effects in the theta band, however, might be observed if sentence semantics had been manipulated.

In sum, the current thesis sets the stage for investigating neural oscillatory dynamics in spoken word recognition. Different methodological approaches are applied to reveal temporal signatures of lexico-semantic processing. Importantly, interpretations of commonly studied event-related potentials are reassessed. Thus, the results will have important implications for the comprehension of neuropsychological mechanisms underlying spoken word recognition.

EXPERIMENTS AND RESULTS

In the first experiment, we aimed at showing parallel processes of lexical integration and ambiguity resolution during spoken word recognition. We hypothesized that alpha and theta frequency bands would dissociate when comparing words, ambiguous pseudowords, and opaque pseudowords in a lexical decision task. Real words were 60 three-syllabic, concrete German nouns (e.g., "Banane" [banana]; adapted from Raettig and Kotz, 2008). Ambiguous pseudowords were derived by exchanging the core vowel of the second syllable (e.g., "Banene"). For opaque pseudowords, syllables were scrambled across words, while keeping their position-in-word fixed (e.g., "Bapossner").
Post-lexical alpha power suppression scaled with wordness such that real words showed the lowest, ambiguous pseudowords intermediate and opaque pseudowords the highest alpha power. Source localisation of the alpha power revealed left occipito-temporal cortex and right anterior prefrontal cortex. These results were supported by the gradual increase of the N400 magnitude showing the most negative amplitude for opaque pseudowords. Furthermore, theta power was found to be enhanced for ambiguous pseudowords in left inferior frontal gyrus and right middle temporal gyrus.

In a second step, we asked how these oscillatory patterns are changed in adverse listening conditions. Embedding stimuli in white noise increases the difficulty to discriminate vowels (Phatak and Allen, 2007) which is important to discriminate words and ambiguous pseudowords and thus perform accurately in the current lexical decision task. Signal-to-noise ratios were determined individually by means of an adaptive tracking procedure targeting 70.7 % accuracy. We hypothesized to observe neural mechanisms of selective inhibition when comparing the data from the lexical decision task in quiet with data from the same participants doing the lexical decision task in noise. The comparison between quiet and noise showed higher induced alpha power in noise than in quiet after word onset. At the same time and in line with reduced N1-P2-magnitudes in degraded compared to intact speech, alpha inter-trial phase coherence showed the opposite pattern and was higher for speech in quiet than in noise.

Next, we hypothesized to find accuracy of lexical decisions in noise to be modulated by neural phase as it has been shown for the detection of low-level auditory targets (Henry and Obleser, 2012). Only lexical decisions to words and ambiguous pseudowords were considered because of their difference in one vowel only. The accuracy was modulated by pre- and peri-stimulus alpha phase. Pre-stimulus alpha phase was anti-phase for correct and incorrect trials over right frontal areas. Peri-stimulus alpha was anti-phase over left fronto-temporal areas. No phase effects were found in the theta band, although theta oscillations have been shown to modulate neural firing as well (Kayser et al., 2012). Supplementary behavioural results yielded differential influences of lexical stress pattern and formant information on perceptual sensitivity and response bias. If ambiguous pseudowords were acoustically closer to their real word neighbour (measured by formant distances of the manipulated vowels), lexical decisions were biased towards word judgements. This relationship was most pronounced when stimuli were stressed on the second syllable which was the crucial syllable in the current experimental design. Additionally, features of the phase bifurcation index (Busch et al., 2009), used for the current analysis of neural phase, were explored by simulations to support the validity of the current methodological approach.

Finally, we asked about the temporal dynamics of spoken word recognition in sentence context when the signal is compromised, here operationalised by noise-vocoding the speech signal in three severity level. The cloze probability of semantic contexts was manipulated (high vs low) and the typicality of the sentence-final word was varied (high vs low). The magnitude of the N400 was reduced in intact as well as degraded speech in high cloze compared to low cloze probability sentences. Furthermore in clear speech, the N400 was reduced for more typical as well as for less typical sentence-final words. In degraded speech, however, the N400 was reduced for typical sentence-final words only. N400 results were accompanied by effects in the theta but not the alpha band. In particular, theta power was enhanced before the onset of the sentence-final word in high cloze probability sentences. In low cloze probability sentences, in turn, peri-stimulus theta inter-trial phase coherence was increased in line with the increased N400 magnitude.

Discussion

The current thesis aimed at determining neural oscillatory signatures of spoken word recognition. Alpha oscillations (8–12 Hz) as the predominant rhythm in human EEG are presumably a neural means to implement the general cognitive function of gating information flow (Jensen and Mazaheri, 2010; Hanslmayr et al., 2012). In line with this notion, the current thesis found alpha oscillations to play a role during spoken word recognition in three possible ways:
First, induced alpha power scaled with wordness, that is, with the difficulty to map the phonological representation onto meaning. Post-lexical alpha power was suppressed for words indicating processing of lexico-semantic information. In turn, alpha power was enhanced for opaque pseudowords indicating the inhibition of lexico-semantic processing. Second, induced alpha power was found to be enhanced at the beginning of words em-

bedded in noise compared to clear speech for which we proposed a framework to further assess the presumable role of alpha in selectively inhibiting task-irrelevant auditory objects (Klimesch et al., 2007). Third, pre-stimulus alpha phase was found to modulate lexical decision accuracy in noise. We interpreted this finding to reflect selective attention insofar as stimuli coinciding with the excitatory phase were more likely to be thoroughly processed and ultimately judged correctly (Schroeder and Lakatos, 2009; Mathewson et al., 2011).

Besides alpha activity, theta oscillations were of interest because of their association with long-term (thus semantic) memory (Fell and Axmacher, 2011) and their presumed role in speech processing due to the correspondance to the syllabic rate (Ghitza, 2011). At this point, the current data speak in favor of theta oscillations playing a role for lexico-semantic mapping but no direct evidence could be found to support ideas about chunking linguistic content for further processing:
First, induced theta power was found to be post-lexically enhanced for ambiguous pseudowords. We interpreted this finding in terms of ambiguity resolution. Because of their proximity to real words (only one vowel exchanged), ambiguous pseudowords induced response conflicts when judging their lexicality. In line with Fuentemilla et al. (2010), we suggest that phonemic information needed to be "replayed" in order to re-compare it with long-term memory representations and thus resolve ambiguity. Second, in high cloze probability sentences theta power was found to be enhanced just before the onset of the sentence-final word, thus indicating the anticipatory activation of lexico-semantics in long-term memory (Bastiaansen et al., 2008). Third, theta phase was not found to modulate lexical decision accuracy in a consistent manner so that its chunking role for speech processing remains elusive (Ghitza, 2011; Giraud and Poeppel, 2012).

The results of the current thesis suggest that there might be different oscillatory activities underlying the N400 component. First, when comparing words and pseudowords, the N400 effect was accompanied by a decrease in the inter-trial phase coherence for real words compared to pseudowords in the alpha frequency range. Hence, the increase in N400 magnitude, together with the increased alpha inter-trial phase coherence for pseudowords could accordingly be interpreted as indicating the inhibition of lexico-semantic processing. Second, when comparing sentence-final words in low- versus high-cloze probability sentences, the N400 effect was accompanied by an increase in the inter-trial phase coherence for words in low-cloze contexts in the theta frequency range. We suggest that, in contrast to the alpha-N400 effect, the theta-N400 effect might be reconsidered as reflecting simultaneous lexico-semantic retrieval and semantic integration. To coordinate both processes on-line at the same time, synchronization via theta oscillations might be necessary. If, instead, lexico-semantic information can be pre-activated via pre-stimulus theta power enhancement, peri-stimulus synchronization might be reduced because semantic integration is facilitated.

Our findings extend current knowledge on spoken word recognition gained previously by N400 analysis. We provide two arguments that challenge the linearity of spoken word recognition (as for example modelled in COHORT; Marslen-Wilson and Tyler, 1980). First, we showed the simultaneous occurrence of alpha and theta power modulation, indexing lexical integration and ambiguity resolution respectively. By looking at the lexicality-N400 effect only, word recognition would have appeared as a sequential process which is effortlessly accomplished for real words (reduced N400 magnitude) and is at first not successful for both types of pseudowords. For ambiguous pseudowords, however, word recognition might occur delayed (indexed by an intermediate N400 magnitude) whereas lexical search for opaque pseudowords continues (indexed by a permanently increased N400 magnitude).

By looking at slow neural oscillations, alpha power scaled with wordness comparable to the N400. At the same time, theta power was found selectively enhanced for ambiguous pseudowords. This is compatible with models that assume a dual route of word recognition where lexical and segmental information are both held in memory at the same time (Norris et al., 2000).

The second piece of evidence that questions the linearity of spoken word recognition is provided by the alpha phase bifurcation showing that lexical evidence is not accumulated linearly but rather rhythmically (Ghitza, 2011). In particular, lexical decision accuracy in noise was modulated by alpha phase in a pre-stimulus and a peri-stimulus time window. Both times, correct and incorrect lexical decisions yielded opposite phase patterns. This speaks in favor of rhythmic accumulation of perceptual evidence to arrive at decisions (Wyart et al., 2012). However, these data are the first evidence to show this and future research needs to further determine the relationship between slow neural oscillations and speech processing.

Taken together, the present thesis elucidates neural oscillatory dynamics underlying spoken word recognition. We showed that alpha and theta oscillations play important and complementary roles for word comprehension in quiet as in noise. In particular, we demonstrated that alpha phase modulates the accuracy of lexical decisions. Theta power, instead, is involved in processing lexico-semantic information.

Zusammenfassung

Einleitung

Das Verstehen gesprochener Sprache ist oft herausfordernd. Hintergrundlärm, wie zum Beispiel durch Verkehr verursacht, oder spektrale Einschränkungen, wie etwa bei Cochlea Implantaten, stellen problematische Hörsituationen dar. Dennoch ist es gerade Muttersprachlern möglich, den Verlust von Sprachinformation zu kompensieren. Die vorliegende Dissertation untersucht die zeitlichen Dynamiken von Hirnprozessen, die der Worterkennung gesprochener Sprache in schwierigen Hörsituationen zugrunde liegen.
Um die Verarbeitung von lexiko-semantischen Informationen zu untersuchen, werden Wörter mit wortartigen Pseudowörtern während einer auditiven lexikalischen Entscheidungsaufgabe kontrastiert (Goldinger, 1996). Da Pseudowörter im Gegensatz zu Wörtern keine Bedeutung tragen, bleibt die lexiko-semantische Verarbeitung erfolglos, sodass die Prozesse der Zuweisung von Semantik zur phonologischen Repräsentation sichtbar werden. Darüberhinaus werden Sätze benutzt, die bezüglich ihrer *cloze probability* manipuliert sind, d.h. die sich hinsichtlich ihrer Vorhersagekraft bezüglich des letzten Wortes des Satzes unterscheiden (stark oder schwach vorhersagende kontextuelle Semantik; Kutas and Hillyard, 1980). Die Robustheit, einerseits Wörter und Pseudowörter zu unterscheiden und andererseits Wörter aus ihrem Kontext vorherzusagen, wird überprüft, indem Hintergrundrauschen zum Sprachsignal hinzugefügt und indem das Sprachsignal selbst spektral eingeschränkt wird.
Der vorliegenden Arbeit liegt die allgemeine These zugrunde, dass sich die Prozesse der Worterkennung in langsamen neuralen Oszillationen widerspiegeln. Oszillatorische Mechanismen bestehen in der dynamischen Synchronisierung von Hirnregionen, womit vorübergehend der Informationsaustausch zwischen den Regionen freigegeben oder inhibiert wird. Außerdem können Oszillationen die Fluktuation der neuralen Reizbarkeit reflektieren (Haegens et al., 2011). Synchronisierungen im Alpha-Frequenzband ($\sim$10 Hz) sind mit der selektiven Inhibierung der Verarbeitung irrelevanter Informationen (Rauschen) in Verbindung gebracht worden (Jensen and Mazaheri, 2010). Alpha-Desynchronisierungen sind dagegen mit der erfolgreichen Gedächtnisenkodierung assoziiert (Hanslmayr et al., 2012). Oszillationen im Theta-Frequenzband ($\sim$4 Hz) sind im Zusammenhang mit der periodischen Reaktivierung von gespeicherten Informationen gezeigt worden (Fuentemilla et al., 2010)

und spielen eine Rolle beim Gedächtnisabruf lexiko- semantischer Information (Bastiaansen et al., 2008). Langsame neurale Oszillationen sind möglicherweise auch für die akustische Analyse des Sprachsignals wichtig (Ghitza, 2011). Sie zerlegen möglicherweise das Sprachsignal in kleinere Einheiten, indem die Phase der höchsten neuralen Reizbarkeit an dem Zeitpunkt der wichtigsten akustischen Information ausgerichtet wird (Giraud and Poeppel, 2012). Folglich würden Effekte im Alpha-Frequenzbereich beobachtet, wenn segmentale Information wie Vokale manipuliert würden, und Effekte im langsameren Theta-Frequenzband, wenn die Satzsemantik manipuliert würde.
Zusammenfassend erforscht die vorliegende Arbeit neurale oszillatorische Dynamiken bei der Erkennung gesprochener Worte. Verschiedene methodische Herangehensweisen werden genutzt, um die zeitlichen Signaturen der lexiko-semantischen Verarbeitung zu bestimmen. Nebenbei werden die Interpretationen der üblicherweise erhobenen Ereignis-korrelierten Potentiale wie der N400 neu bewertet. Dadurch werden die Ergebnisse wichtige Implikationen für das Verstehen von neuropsychologischen Mechanismen haben, die der Erkennung gesprochener Worte zugrunde liegen.

Experimente und Ergebnisse

Das erste Experiment zielte auf den Nachweis von parallelen Prozessen bei der Erkennung gesprochener Worte, und zwar den Prozessen der lexikalischen Integration und den der Ambiguitätsauflösung. Wir nahmen die Dissoziation von Alpha- und Theta-Frequenzen an, wenn Wörter, ambige Pseudowörter und opake Pseudowörter in der lexikalischen Entscheidungsaufgabe verglichen würden. Die echten Wörter bestanden aus 60 dreisilbigen, konkreten, deutschen Substantiven (z.B. "Banane"; entnommen von Raettig and Kotz, 2008). Von diesen wurden die ambigen Pseudowörter abgeleitet, indem der Nukleusvokal der zweiten Silbe ausgetauscht wurde (z.B. "Banene"). Zur Erstellung der opaken Pseudowörter wurden die Silben über Wörter hinweg ausgetauscht, wobei ihre jeweilige Position innerhalb des Wortes erhalten blieb (z.B. "Bapossner").
Die Unterdrückung der post-lexikalischen Alpha-Power skalierte mit der Wortheit der Stimuli und zwar so, dass echte Wörter die niedrigste, ambige Pseudowörter mittlere und opake Pseudowörter die höchste Alpha-Power zeigten. Die Quelllokalisation der Alpha-Power ergab den linken temporo-okzipitalen Cortex und den rechten anterioren präfrontalen Kortex. Die Ergebnisse wurden unterstützt vom graduellen Anstieg der N400 Komponente, die die negativste Amplitude für opake Pseudowörter zeigte. Darüberhinaus war die Theta-Power selektiv für ambige Pseudowörter im linken inferioren Frontalgyrus und im rechten mittleren Temporalgyrus erhöht.
In einem zweiten Schritt wurde untersucht, wie sich diese oszillatorischen Muster in schwierigen Hörsituationen ändern. Die Einbettung der Stimuli in weißes Rauschen erhöht die Schwierigkeit der Vokaldiskriminierung (Phatak and Allen, 2007), die hier wichtig

ist, um Wörter und ambige Pseudowörter zu unterscheiden und damit die Aufgabe der lexikalischen Entscheidung richtig zu lösen. Das Signal-zu-Rausch-Verhältnis wurde mit Hilfe einer adaptiven Prozedur individuell festgelegt, die die Schwelle für eine Korrektheit von 70.7 % ermittelt. Der Vergleich von der lexikalischen Entscheidungsaufgabe mit und ohne Hintergrundrauschen sollte neurale Mechanismen der selektiven Inhibition aufdecken. Die Ergebnisse zeigen, dass die induzierte Alpha-Power gleich nach dem Wortbeginn im Rauschen erhöht ist. Zeitgleich—und im Einklang mit den reduzierten N1-P2-Amplituden für degradierte im Vergleich zur intakten Sprache—zeigt die Alpha-Phasenkohärenz das entgegengesetzte Muster: Sie ist größer für Sprache ohne Rauschen.

Als nächstes wurde getestet, ob die Korrektheit von lexikalischen Entscheidungen im Rauschen von der neuralen Phase moduliert wird, wie es für die Detektierung von einfachen auditiven Stimuli gezeigt wurde (Henry and Obleser, 2012). Nur Antworten auf Wörter und ambige Pseudowörter wurden hier analysiert, weil sie von der erfolgreichen Vokaldiskriminierung abhängen. Die Korrektheit wurde durch die prä- und die peri-lexikalische Alpha-Phase moduliert. Die prä-lexikalische Alpha-Phase war für korrekte und inkorrekte Entscheidungen anti-phasisch über rechts frontalen Arealen. Die peri-lexikalische Alpha-Phase war über links fronto-temporalen Arealen anti-phasisch. Keine konsistenten Phaseneffekte wurden im Theta-Band gefunden. Zusätzliche Analysen der Verhaltensdaten ergaben, dass das Wortbetonungsmuster und die Formanteninformation unterschiedliche Auswirkungen auf die perzeptuelle Sensitivität und die Antwortpräferenz hatten. Wenn ambige Pseudowörter ihrem Echtwortnachbarn akustisch sehr ähnlich waren (gemessen in Formantendistanz der manipulierten Vokale), dann tendierten die lexikalischen Entscheidungen zu "Wort"-Antworten. Dieses Verhältnis war am stärksten ausgeprägt, wenn die Stimuli auf der zweiten, d.h. der kritischen, Silbe betont wurden. Außerdem wurden die Eigenschaften des Phasenbifurkationsindexes (Busch et al., 2009) exploriert, der für die Analyse der neuralen Phase angewendet wurde, um die Validität dieses methodischen Ansatzes anhand von Simulationen zu überprüfen.

Schließlich wurden die zeitlichen Dynamiken der Wortverarbeitung im Satzkontext untersucht. Das Signal wurde hier spektral durch Vocodieren in 3 Schwierigkeitsgraden beeinträchtigt. Die *cloze probability* des Satzes (stark vs. schwach vorhersagend) und die Typikalität der Satz finalen Wörter wurden manipuliert (typisch vs. untypisch). Die Amplitude der N400 war reduziert im intakten genauso wie im degradierten Sprachsignal, wenn eine hohe im Vergleich zu einer niedrigen Erwartbarkeit des letzten Wortes bestand. Des Weiteren war bei intakter Sprache die N400 für typische und untypische Wörter am Satzende reduziert. Bei degradierter Sprache war die N400 nur für typische Satzendungen reduziert. Neben der N400-Effekte war die Theta-Power bereits vor dem Beginn des Satz finalen Wortes erhöht, wenn das Wort aufgrund des Satzkontextes stark erwartbar war. War das letzte Wort dagegen nur schwach erwartbar, war die Theta-Phasenkohärenz —entsprechend der erhöhten N400-Amplitude—wortbegleitend erhöht.

Diskussion

Die vorliegende Dissertation zielte auf die Bestimmung der neuralen, oszillatorischen Signaturen bei der Erkennung gesprochener Worte ab. Alpha-Oszillationen (8–12 Hz) als der vorherrschende Rhythmus im menschlichen EEG sind vermutlich ein neurales Mittel, um die allgemeine kognitive Funktion des Taktens des Informationsstroms umzusetzen (Jensen and Mazaheri, 2010; Hanslmayr et al., 2012). Im Einklang mit dieser Ansicht hat diese Arbeit gezeigt, dass Alpha-Oszillationen auch eine wichtige Rolle bei der Verarbeitung gesprochener Worte spielt:
1) *Induzierte Alpha-Power skalierte mit der Wortheit der Stimuli*, d.h. mit der Schwierigkeit, eine Bedeutung auf eine phonologische Representation abzubilden. Die post-lexikalische Alpha-Power war bei Wörtern unterdrückt, was die Verarbeitung von lexiko-semantischen Information indiziert (Obleser and Weisz, 2012). Im Gegensatz dazu war die Alpha-Power bei opaken Pseudowörtern erhöht, was die Inhibierung von lexiko-semantischer Informationsverarbeitung indiziert. 2) *Induzierte Alpha-Power war am Anfang der Wörter erhöht, wenn diese nicht klar sondern im Rauschen eingebettet waren*, im Einklang mit der inhibitorischen Funktion von Alpha (Klimesch et al., 2007). Um die Rolle von Alpha bei der selektiven Inhibierung von irrelevanten Informationen systematisch zu erforschen, wurde ein theoretischer Rahmen entworfen. 3) *Die prä-lexikalische Alpha-Phase modulierte die lexikalische Entscheidungskorrektheit im Rauschen.* Dieser Effekt wurde als eine Reflektion der selektiven Aufmerksamkeit interpretiert, insofern als dass Stimuli die auf die exzitatorische Phase treffen mit größerer Wahrscheinlichkeit sorgfältiger verarbeitet werden als die, die mit der inhibitorischen Phase zusammenfallen, und damit letztendlich korrekt bewertet werden (Schroeder and Lakatos, 2009; Mathewson et al., 2011).
Neben der Alpha-Aktivität waren auch Theta-Oszillationen Untersuchungsgegenstand aufgrund ihrer Assoziation mit der lexiko-semantischen Verarbeitung (Bastiaansen et al., 2008) und ihrer hypothetisch wichtigen Rolle bei der Zerlegung von linguistischer Information (Ghitza, 2011). Die vorliegenden Daten sprechen bislang nur für die Assoziation von Theta mit der lexiko-semantischen Verarbeitung:
1) *Induzierte Theta-Power war post-lexikalisch selektiv bei ambigen Pseudowörtern erhöht.* Dieser Effekt wurde als Indikation von Ambiguitätsauflösung interpretiert. Wegen ihrer Ähnlichkeit zu echten Wörtern (nur ein Vokal verschieden) induziert die lexikalische Entscheidungsaufgabe einen Antwortkonflikt bei ambigen Pseudowörtern. Im Einklang mit Fuentemilla et al. (2010) wird vorgeschlagen, dass die phonemische Information "wieder abgespielt" wird, um sie nochmals mit der Langzeitrepräsentation abzugleichen und damit die Ambiguität aufzulösen. 2) *Bei Sätzen mit hoher semantischer Erwartbarkeit war die Theta-Power kurz vor dem Beginn des Satz finalen Wortes erhöht.* Dieser Effekt wurde als Voraktivierung der lexiko-semantischen Information interpretiert (Bastiaansen et al., 2008). 3) *Die Theta-Phase stand nicht in einem konsistenten Zusammenhang zur Mod-*

ulierung der Entscheidungskorrektheit. Dadurch bleibt die hypothetische Rolle bei der Zerlegung von linguistischer Information weiter offen (Ghitza, 2011; Giraud and Poeppel, 2012).

Die Ergebnisse erweitern das bestehende Wissen über die Wortverarbeitung gesprochener Sprache, die im Vorfeld durch die N400-Analyse gewonnen wurden. Es werden zwei Argumente geliefert, die die Linearität von Wortverarbeitung in Frage stellen (wie z.B. im COHORT-Modell; Marslen-Wilson and Tyler, 1980):

Zunächst wurde gezeigt, dass Alpha- und Theta-Power gleichzeitig moduliert wurden und damit die Prozesse der lexikalischen Integration und der Ambiguitätsauflösung dissoziieren. Bei der ausschließlichen Betrachtung der N400-Effekte wäre die Worterkennung als sequentieller Prozess erschienen, der bei Wörtern mühelos erfolgt (indiziert durch die reduzierte N400-Amplitude) und der bei beiden Arten von Pseudowörtern zunächst scheitert. Bei ambigen Pseudowörtern erfolgt die Worterkennung allerdings dennoch nur mit Verzögerung (indiziert durch die mittlere N400-Amplitude), während die lexikalische Suche bei opaken Pseudowörtern anhält (indiziert durch die dauerhaft erhöhte N400-Amplitude).

Bei der Betrachtung von langsamen neuralen Oszillationen zeigt sich dagegen, dass zwei Prozesse parallel ablaufen. Zum einen skaliert die Alpha-Power (wie die N400) mit der Wortheit der Stimuli. Zum anderen ist die Theta-Power selektiv erhöht bei ambigen Pseudowörtern. Diese Ergebnisse sind mit Modellen kompatibel, die eine zweifache Route der Worterkennung annehmen, eine, die die lexikalische und eine die die segmentelle Information verarbeiten (Norris et al., 2000).

Das zweite Argument gegen die Linearität der Worterkennung wird durch die Bifurkation der Alpha-Phase geliefert. Diese Ergebnisse zeigen, dass die Evidenz für einen lexikalischen Stimulus nicht linear akkumuliert wird, sondern vielmehr rhythmisch (Ghitza, 2011). Die Korrektheit der lexikalischen Entscheidung im Rauschen wurde durch die Alpha-Phase in einem prä-lexikalischen und in einem peri-lexikalischen Zeitfenster moduliert. In beiden Zeitfenstern zeigten korrekte und inkorrekte lexikalische Entscheidungen anti-phasische Muster, was für die rhythmische Akkumulation von perzeptueller Evidenz spricht, um zur lexikalischen Entscheidung zu gelangen (Wyart et al., 2012). Diese Daten stellen die ersten Hinweise auf diese Art von Rhythmizität bei der Worterkennung dar, weshalb zukünftige Forschung das Verhältnis zwischen langsamen neuralen Oszillationen und der Sprachverarbeitung weiter bestimmen muss.

Zusammenfassend erhellt die vorliegende Dissertation die Dynamiken neuraler Oszillationen, die der Worterkennung gesprochener Sprache zugrunde liegen. Es konnte gezeigt werden, dass Alpha- und Theta-Oszillationen wichtige und komplementäre Rollen spielen beim Verstehen von Worten in idealen und eingeschränkten Hörsituationen. Insbesondere wurde demonstriert, dass die Alpha-Phase die Korrektheit der lexikalischen Entscheidung moduliert. Theta-Power wurde hingegen mit der Verarbeitung von lexiko-semantischen Information assoziiert.

Curriculum Vitae

Antje Strauss

born in Germany, 8th of June 1985

Education

2011–2014	Doctoral Studies (Psychology) *Max Planck Institute, Leipzig*
2004–2010	Magistra Artium (German Philology, Philosophy) *Albert Ludwig University, Freiburg*
2008–2010	Undergraduate Research Assistant *Freiburg Institute for Advanced Studies, Freiburg*
2006–2007	Erasmus stipendiary (Philosophy) *Universidad Complutense, Madrid*

Publications

Strauß A, Henry MJ, Scharinger M, & Obleser J (In revision). Alpha phase determines successful lexical decision in noise. *The Journal of Neuroscience.*

Strauß A, Wöstmann M, & Obleser J (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience* 8:350. doi: 10.3389/fnhum.2014.00350.

Strauß A, Kotz SA, Scharinger M, & Obleser J (2014). Functional dissociation of alpha and theta brain oscillations in speech processing. *NeuroImage* 97, 387-395. doi: 10.1016/j.neuroimage.2014.04.005.

Bendixen A, Scharinger M, **Strauß** A, & Obleser J (2014). Prediction in the service of speech comprehension: modulated early brain responses to omitted speech segments. *Cortex* 53:9-26. doi: 10.1016/j.cortex.2014.01.001.

Strauß A, Kotz SA, & Obleser J (2013). Narrowed expectancies under degraded speech: Revisiting the N400. *Journal of Cognitive Neuroscience* 25:8, 1383-1395. doi: 10.1162/jocn_a_00389.

Selbstständigkeitserklärung
gemäß § 8(2) der Promotionsordnung

Hiermit erkläre 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 01. Oktober 2014

Antje Strauß

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

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

159 Julia Erb
The neural dynamics of perceptual adaptation to degraded speech

160 Philipp Kanske
Neural bases of emotional processing in affective disorders

161 David Moreno-Dominguez
Whole-brain cortical parcellation: A hierarchical method based on dMRI tractography

162 Maria Christine van der Steen
Temporal adaptation and anticipation mechanisms in sensorimotor synchronization