Universität Konstanz

Mathematisch - Naturwissenschaftliche Sektion
Fachbereich Psychologie

Deficient Speech Repaired:
Evidence from Event Related Brain Responses.

Dissertation zur Erlangung des akademischen Grades
des Doktors der Naturwissenschaften

vorgelegt von Dipl.-Psych. Ronny Hannemann
Oktober 2008

Tag der mündlichen Prüfung: 12.12.2008
1. Referent: Prof. Dr. Carsten Eulitz
2. Referent: Dr. Nathan Weisz

Dank

An erster Stelle möchte ich Carsten Eulitz für sein Vertrauen und die Möglichkeit eigene Ideen zu entwickeln und verwirklichen zu können danken. Ohne seinen Rat und die Unterstützung in allen Belangen wäre diese interessante Arbeit nicht möglich gewesen. Weiterhin gebührt Dank meinen Kollegen, ohne die das interdisziplinäre Gemeinschaftsprojekt einer (linguistisch geprägten) auditorischen und kognitiven Neurowissenschaft nicht funktionieren würde: Aditi Lahiri, Alex Bobrov, Claudia Massau, Katrin Preller, Marcus Meinzer, Mathias Scharinger, Nathan Weisz, Sarah dos Santos Sequeira, Sonia Cornell, Verena Felder und vielen Anderen.

Besonders möchte ich Jonas Obleser danken, der trotz räumlicher Entfernung immer ein offenes Ohr für mich hatte und mir mit kleineren und größeren Ratschlägen stets zur Seite stand. Sascha Otterbein, Winfried Schlee und anderen Freunden gebührt ebenfalls ein großes Dankeschön für die vielen guten Stunden trotz oder gerade wegen der gemeinsamen „Leidenszeit“. Von Herzen Danken möchte ich auch Martina Ziegler für die sonnige Begleitung auf dieser spannenden Etappe und meiner Familie für ihren Rückhalt und ihr Vertrauen.

Contents

Glossary (abbreviation index)

ANOVA	analysis of variance
ART	Adaptive Resonance Theory
dB	decibel
EEG	electroencephalogram or electroencephalography
EMG	electromyogram
EOG	electrooculogram
ERP	event-related potential
fMRI	functional magnetic resonance imaging
FUL	featurally underspecified lexicon
GBA	gamma band activity
Hz	Hertz
IFG	inferior frontal gyrus
MEG	magnetoencephalogram or magnetoencephalography
MMN	mismatch negativity
µV	Microvolt
ms	milliseconds
MUM	Match and Utilisation Model
PET	positron emission tomography
RMS	root mean square

Figure index

Abstract

The comprehension of speech is an extraordinary ability of the human cognitive system. Listeners are capable of recognizing speech efficiently and rapidly under an exceedingly broad range of acoustic environmental conditions. Even if segments of the speech signal are acoustically not present or heavily distorted these disturbances may not impede the comprehension of speech.

In a series of EEG experiments different repair processes in speech perception using lexical top-down information were examined. It was hypothesized that a successful recovery of deficient speech results in resonant states which are reflected in the induced gamma band activity (GBA). Instead, if the general conditions of an ambiguous speech event prevent the repair process, this failure should be reflected in an error signal, i.e. the mismatch negativity (MMN). All these processes were expected to be influenced by the overall structure of representations in the mental lexicon.

The findings of the present thesis revealed a direct correlate for the facilitating influence of top-down knowledge on speech comprehension under adverse listening conditions in the induced GBA. Twice, a modulation in the 40 Hz range over left anterior temporal electrode sites was reported. Further, the present results evinced deep insights in the general conditions that allow for a repair of deficient speech. Only if the expected phonemic information matches somewhat the characteristics of the incoming stimulus a successful repair was evident. On the contrary, an explicit difference between the sensory and the expected information prevent the recovery of deficient speech and revealed a distinct MMN over frontal electrode sites. Moreover, at least for the phonemic restoration illusion, the induced GBA as well as the MMN responses corroborated the hypothesis of sparse representations in the mental lexicon (Lahiri and Reetz, 2002).

Taken together, the present data show experimentally how comprehending speech and the bottom-up brain processes mediating it depend highly on memory-driven (i.e., top-down) expectancies. Finally, the present data are in line with recent models of speech perception relying on Bayesian statistics (Friston, 2005; Norris and McQueen, 2008) by demonstrating how the cognitive system adapts in an optimal way to a complex and constant changing environment.

Zusammenfassung

Die Fähigkeit Sprache zu verstehen ist eine außergewöhnliche Eigenschaft des menschlichen kognitiven Systems. Menschen können Sprache unter den verschiedensten Bedingungen schnell und effizient verstehen und verarbeiten. Selbst fehlende oder stark verfremdete Sprachsegmente verhindern das Verständnis der gesprochenen Sprache nicht.

Diese Arbeit untersucht mit Hilfe von EEG-Experimenten Reparaturprozesse, die trotz unzureichenden akustischen Signalen mittels Top-Down-Informationen die Wahrnehmung von gesprochener Sprache ermöglichen. Ausgangspunkt ist hierbei die Annahme, dass der sensorische Input mit Top-Down-Informationen abgeglichen wird und bei Übereinstimmung miteinander verknüpft wird. Diese Integration resultiert in einer resonanten Aktivität in Neuronenverbänden, die sich als induzierte Gammaband-Aktivität (GBA) mit dem EEG messen lassen. Wenn jedoch die Reparatur eines unzureichenden Sprachsignals aufgrund verschiedenster Rahmenbedingungen nicht möglich ist, wird erwartet, dass dieses Scheitern durch eine Art Fehlersignal in Form einer Mismatch-Negativity (MMN) beobachtbar ist. Weiterhin wird angenommen, dass sowohl die Prozesse der Integration als auch diejenigen des Scheiterns von der Struktur der Repräsentationen im mentalen Lexikon bzw. im Langzeitgedächtnis beeinflusst werden.

Die Ergebnisse dieser Arbeit zeigen ein direktes Gammaband-Korrelat im 40 Hz Bereich für die Unterstützung von Sprachwahrnehmungsprozessen durch Top-Down-Informationen in schwierigen akustischen Kontexten. In zwei Experimenten konnte über einem links fronto-temporalen Elektrodencluster ein Zuwachs an induzierter Aktivität im Gammaband festgestellt werden. Außerdem zeigen die vorliegenden Resultate eine deutliche MMN über frontalen Elektroden, wenn eine Integration von aktuellem Input und Top-Down-Informationen aufgrund von eindeutigen (akustischen) Unterschieden zwischen diesen beiden nicht möglich ist. Somit veranschaulichen die Ergebnisse dieser Arbeit, dass die Reparatur eines unzureichenden Sprachsignals nur dann erfolgreich ist, wenn das akustische Signal zumindest teilweise die gleichen (akustischen) Charakteristiken aufweist, die auch aufgrund von Top-Down-Informationen erwartetet werden. Weiterhin unterstützen die vorliegenden Resultate im induzierten Gammaband und der MMN die Annahme

eines sparsamen mentalen Lexikon (Lahiri and Reetz, 2002) – zumindest für die Restauration einzelner Phoneme.

Die vorliegende Arbeit veranschaulicht mit Hilfe experimenteller Daten, wie lexikalische Top-Down-Informationen die Wahrnehmung von gesprochener Sprache und die dafür verantwortlichen Signalverarbeitungsprozesse beeinflussen. Außerdem unterstützen die vorliegenden Ergebnisse die Annahme, dass die menschliche Sprachwahrnehmung durch Modelle beschrieben werden kann, die auf dem Bayesschen Theorem aufbauen (Friston, 2005; Norris and McQueen, 2008). Somit demonstrieren die Ergebnisse dieser Arbeit, wie sich das kognitive System auf optimale Weise an eine komplexe und sich konstant ändernden Umwelt anpassen kann.

1 General Introduction

In a complex and constant changing environment adaptive behaviour requires fast and efficient processing of sensory information. Therefore incoming information is rapidly structured and processed according to current needs, long term goals and situational demands by specialized mechanisms implemented in the human cognitive system.

Also the ability to communicate with members of the own species is important to ensure the survival. Thus, aside from speech production the ability to perceive spoken language is one of the essentials providing humans with a better chance to survive the struggle for existence (Darwin, 1874). Unfortunately, most of these speech conversations take place under conditions that are less than ideal. Extraneous sounds interfere with or occasionally obliterate individual speech sounds or even whole words. According to models of brain functioning emphasizing predictive coding, imperceptible or ambiguous properties of a (speech) stimulus, such as features that are occluded or masked, cannot be resolved by simply analyzing the sensory input. In this case, the nature of the speech stimulus can only be resolved by inferring the imperceptible or ambiguous properties of the acoustical input based on information derived from previous experience (Friston, 2005). Hence, instead of relying solely on the sensory speech information, the human cognitive system enables rapid understanding of spoken language by using additional information. Aside from knowledge about the world in general (Jackendoff, 2003), information about the speaker (Clark, 1996) and information from other sensory modalities (e.g. Tanenhaus et al., 1995) listeners rely also on stored meanings of words in the mental lexicon when interpreting the speech input. Additionally whole discourse information is taken into account when perceiving language (e.g. Otten et al., 2007). Together these kinds of information can be subsumed as top-down factors and they help to cope with shortcomings in the sensory input and to make sense of the acoustic speech signal.

The aim of the present thesis is to gain new insights in the perceptual mechanisms that allow coping with sketchy speech input and its underlying cognitive mechanisms. Top-down influences on the perception of clear and distorted speech seem to be inevitable but a reliable brain measure identifying the interaction of those influences with speech input is still lacking. Therefore, this thesis is set out to describe a reliable measure for the integration of top-down knowledge with insufficient speech input. This is accomplished by means of investigating modulations in the induced gamma band activity (GBA), which had been shown to be first of all signatures of object recognition (Tallon-Baudry and Bertrand, 1999) and mnemonic processes (e.g. Gruber et al., 2008), thus serving as ideal candidate for describing integration and unification processes (Hagoort, 2005; Bastiaansen and Hagoort, 2006) in speech perception.

However, beside fast integration of sensory input and lexical expectancies in order to facilitate speech and language comprehension, also the detection of errors in the acoustic signal is inevitable to avoid miscommunication. Thus, this thesis also aims to answer the question about the general conditions in the perception of deficient speech that allow for a repair and what conditions prevent it.

First, an experiment intended to investigate the facilitating influence of lexical top-down knowledge on the comprehension of degraded or distorted speech measured as signature of induced GBA in the electroencephalogram (EEG) is reported. Taken the experimental results of this study as a basis, a series of two EEG studies investigated handling of missing segmental information in speech perception. This so called phonemic restoration illusion is a powerful illusion in which listeners hear spoken words as intact, even though parts of them have been replaced by an extraneous sound (Warren, 1970). Examining for the first time directly the brain correlates of this phonemic repair process, first of all the induced GBA, the question of the influence of the fine structure of the mental lexicon on this process will be discussed. Further, this thesis tries to disentangle detection of errors and repair processes which handle deficient speech segments by means of differential event related brain responses.

From the perspective of linguistics, the current dissertation focuses mostly on the level of phonology rather than semantics or syntax when investigating the repair processes in speech perception. Overall, the thesis does not aim to localize or identify brain areas responsible for the announced cognitive coping mechanisms for insufficient speech input. Instead, the current array of experiments allows only very superficial conclusions about the underlying brain structures, but they will demonstrate the complementary role of time-sensitive brain analyses in discerning the functional neuroanatomy of speech.

1.1 Short introduction into models of speech perception

Human listeners are capable of recognizing speech spoken under an exceedingly broad range of acoustic environmental conditions efficiently and rapidly. Such sources of acoustic inference as background conversation, computer fan noise or telephone ringing rarely impede the ability to decode successfully the speech signal. This capability is an exceptional computational achievement of the human cognitive apparatus. How does the human cognitive system accomplish this remarkable feat? Recent developments in the field of cognitive neuroscience suggest that the cognitive brain relies on memory based predictions that approximate the relevant future to minimize the processing effort of perception in general (Bar, 2007; Kveraga et al., 2007). Does this linking between representations in memory and the information extracted from the acoustical input allow this robustness in speech perception? Modern theories of spoken word recognition typically postulate a number of hierarchically organized processing stages that mediate between acoustic analyses of the speech signal and higher level representations of the meaning of an utterance (e.g. McClelland and Elman, 1986; Norris, 1994; Gaskell and Marslen-Wilson, 1997). These models are generally in agreement about the existence of a mental lexicon that represents the words we know and that these lexical representations possess abstract and smaller units (e.g., syllables, phonemes, phonemic features) that are activated in the course of recognizing a spoken word. Admittedly, the theorists disagree about whether spoken word recognition is a fully bottom-up, autonomous process or whether a more interactive architecture allows for top-down mediated processing.

In a purely bottom-up or autonomous model (e.g. Shortlist, Norris, 1994; Merge, Norris et al., 2000) the smallest units of speech are processed first. Then the results of these computations are sent to more abstract levels of processing culminating finally in word recognition. For example, the bottom-up speech perception system would first recognize certain phonemic features (e.g. voicing, manner and place of articulation), then make out phonemes comprising these features and finally recognize the word that is spelled out by the sequence of extracted phonemes. Models of speech recognition of isolated words usually assume such kind of bottom-up analysis of the acoustic signal. They emphasize the importance of the words onset which activates a group of candidate words and the analysis of the auditory signal continues until one word matches with the signal (e.g. Cohort, Marslen-Wilson and Welsh, 1978; Shortlist, Norris, 1994). In these models the perception of spoken word emerges from competition amongst lexical hypotheses. This competition is said to be dependent on the goodness of fit between the acoustic signal and the actual candidate and the number of competitors and their goodness of fit respectively. Autonomous models account for top-down lexical effects on lower processing levels by proposing parallel, competing lexical and sub-lexical routes (e.g. Cutler and Norris, 1979) or by proposing that the lexical context biases the responses post-perceptually (Merge, Norris et al., 2000). The latter view, for example, proposes additional decision stages, which integrate phonological and lexical knowledge without changing the pre-lexical interpretation of the sensory input.

In contrast, models of speech perception which argue for top-down processing in speech perception (e.g. Trace, McClelland and Elman, 1986; Samuel and Kat, 1996) state that the computation of the units that are logically less abstract (i.e. phonemes) can be influenced by more abstract (i.e. lexical) representations. If the representation of a word becomes active, its lexical activation might increase the activation of the phonemes that make up the word, for instance. Thus, single phonemes or even words can be recognized faster and with less acoustic evidence than they would need without this top-down lexical influence. This view is supported by the fact that listeners do not rely solely on the acoustical input and stored word meanings when perceiving speech. For example, information about the speaker (Clark, 1996), knowledge about the world (Jackendoff, 2003) and the whole discourse (e.g. Otten et al., 2007) influence the understanding of a spoken utterance. Consequential the sentential context influences the recognition point of a word in making it earlier for the expected word (Zwitserlood, 1989). Even further, the language processing system is operating under unification principles in which linguistic information (e.g. phonology, syntax, semantics) as well as pragmatic information derived from the context, the speaker and states of affairs in the world are handled in parallel, with a direct mapping onto an event structure (or discourse model) that goes beyond the actual sentence to create the meaning of an utterance (Hagoort and Van Berkum, 2007).

Meanwhile, because of evidence for short-term changes in phonemic categories based on implicit perceptual learning, proponents of the autonomous approach acknowledged the need for some feedback in speech perception (Norris et al., 2003). But they stress that it need not occur mandatory online. In their study, adult listeners appear to be able to adjust their phonetic categories of ambiguous sounds as a result of combining prior lexical knowledge and limited exposure to a talker with an unusual way of speaking. Hence, Norris et al. (2003) made the concession for feedback for learning but they further refused the necessity or existence of online feedback in the speech perception process. However, another study investigates recently the top-down lexically driven mechanisms which are involved in perceptual learning of degraded speech input. According to this investigation the listener's ability to adjust perceptual processes rapidly and automatically depends clearly on the top-down influence of higher lexical learning mechanisms (Davis et al., 2005).

At present there seems to be a generally agreement in the speech processing literature that higher levels of processing can alter the processing of speech information at lower levels. The crux of this interaction debate concerns the question when the integration or the alignment of bottom-up and top-down information occurs. Is it a process occurring online with immediate influence of top-down knowledge on the processing of the sensory input or is it rather offline without a direct influence? The next section will shortly review the experimental findings regarding the perception of compromised speech signals, especially the effect of the phonemic restoration, and discuss their contribution to the top-down debate in speech perception models.

1.2 Perception of distorted or otherwise compromised speech signals

However the cognitive speech processing system works in detail, one of its tremendous features is the ability to guarantee a very robust speech perception over widely varying natural conditions of acoustic masking and distortion. In natural settings for example the ability to track an individual's speech amongst other sounds and speech is characterized as cocktail party problem (Cherry, 1953). Here the cognitive system is able to understand spoken messages despite concurrent intrusions of acoustic elements into the target speech stream. Experimental work has shown that speech perception remains robust even when challenged with extreme forms of artificial distortion. For instance, speech remains comprehensible when it is embedded in noise (Miller et al., 1951; Boothroyd and Nittrouer, 1988) and also when its formants were re-synthesized as sinusoids (e.g. Remez et al., 1981; Remez et al., 1983), a manipulation that removes most of the natural qualities of the human voice from a speech signal. Other manipulations have demonstrated that dramatic alterations to both the spectral (e.g. Shannon et al., 1995; Warren et al., 1995; Smith et al., 2002) and the temporal (Mehler et al., 1993; Saberi and Perrott, 1999) properties of the speech signal do not impair substantially the intelligibility of spoken language.

Common to these impairments is that the speech processing system receives at least residual speech information. However, in everyday listening speech signals are frequently interrupted by extraneous sounds and yet listeners are able to understand it quite effortlessly. Even if segments of the speech signal are not acoustically present these disturbances may remain unnoticed (Warren, 1970). First, this so called phonemic restoration illusion was examined behaviourally. Subjects listened to spoken sentences in which single phonemes were completely replaced by coughs. In that study subjects reported hearing the missing phoneme clearly and if asked for localizing the cough most of them failed to do so (see also Ladefoged and Broadbent, 1960). However, while replicating the main finding using noises with different characteristics, the phonemic restoration illusion did not occur if the missing phoneme was replaced with silence (Warren and Obusek, 1971). Thus, simply having an appropriate context is not sufficient for the speech recognition system to repair the deficient input on the phonemic level. Instead the effectiveness of restoration depends upon the similarity of the replacement sound to the phoneme it replaces (Layton, 1975; Samuel, 1981b). White noise for example was shown to be more effective in restoring fricatives than pure tones. Consequently the phonemic restoration illusion was said to arise in terms of the top-down flow of expectations from the lexical level which receive sufficient information from the actual signal to induce the perception of an intact word (Samuel, 1981a). With respect to the role of attentional processes it was shown that only the cuing of the identity, the position and the test word containing the disturbing sound can inhibit the phonemic restoration. As long as subjects attend to the lexical level of processing and that processing is not redirected in some way, the absence of the critical phoneme is overwritten by the expectancy of a certain phoneme. Thus, the structural properties of the word impair attention to the individual phonemes and therefore the phonemic restoration is possible (Samuel and Ressler, 1986).

Unfortunately, as long as an experimental task requires listeners to make phonemic decisions, the results of the former experiments did not allow ruling out a decision-level locus for any lexical influence on phoneme processing. With this constraint in mind the direct top-down lexical influences on phoneme perception by looking at the consequences of having heard a lexically driven phoneme were examined. A series of experiments using the selective adaptation procedure (Eimas and Corbit, 1973) demonstrated convincingly that restored phonemes sound and act like real phonemes by reporting an adaptation shift occurring at early acoustic-phonetic levels of perceptual analysis (Samuel, 1997; Samuel, 2001). Thus, these results clearly indicate that lexical activation can cause the perceptual system to synthesize a highly functional phonemic code.

Subsuming the results of above mentioned studies, the goodness of the speech signal alone can not be the only prerequisite for successful perception of deficient speech. They clearly demonstrate the inevitability of the integration of bottom-up and top-down information in successful speech perception. How does the cognitive system achieve this integration? To answer this question and to build the theoretical framework for this thesis the process of information-integration in speech perception and its possible underlying cortical mechanisms is outlined in the next two subsections.

1.3 Processes of integration and unification in speech perception

Beside the phonemic restoration illusion (e.g. Warren, 1970; Samuel, 1981a; Samuel, 1997) mentioned above, the psycholinguistic literature is full of examples for the integration and influence of lexical top-down and bottom-up information. One further example is the word superiority effect, in which phonemes in words are detected more quickly than in nonwords (Rubin et al., 1976). Also helpful visual context is quickly integrated to resolve ambiguities in sentence processing (Tanenhaus et al., 1995). Or, if an ambiguous sound is added to a context that makes one endpoint a nonword and the other a word (e.g. a sound in the middle of the /d/-/t/ continuum is attached to –ash or –ask), the interpretation of the ambiguous sound is consistent with the lexical endpoint (dash or task, Ganong, 1980). Quite recently a neuroimaging study examining this so called Ganong effect with the functional magnetic resonnace imaging (fMRI) technology adds further strong evidence that lexical information influences the perception of ambiguous phonetic tokens directly (Myers and Blumstein, 2008).

So, as questioned above: How does the cognitive speech processing system manage this task of integrating different sources of information (e.g. information about phonology, semantics and syntax) efficiently? This question can not be answered without looking at the underlying architecture of the speech processing system and considering its functional principles, at least briefly. In recent years a growing body of literature discloses processes of speech and language comprehension as widely distributed over the human cortex. Traditionally, the cortical regions associated with the comprehension of language are Wernicke's area and Broca's area. However, evidence suggests that the complex process of speech and language processing involves several cortical and subcortical areas and relies on multiple bilateral organized processing pathways (e.g. Dronkers et al., 2004; Scott and Wise, 2004; Petkov et al., in press). The fact that single cortical areas are involved in more than one function adds further complexity to the issue of describing the speech perception system. Hence, to give an answer to the question how to manage efficient speech perception, different types and levels of information processed in a spatial distributed cortical network have to be integrated.

It has been suggested that the binding of different sources of information into coherent objects may be propagated by synchronization and desynchronization of oscillatory neural activity. In this view, (de)synchronization links spatially distributed brain areas together to form transient functional networks (Singer, 1993; 1999). Generally, the idea is that the activation of a functional network is facilitated by synchronous and repetitive firing of its neurons. Consequently, neurons pertaining to one and the same functional network are identifiable as such by virtue of their synchronous firing at a given frequency. This frequency specificity allows one neuron to participate in different representations at different times. Besides linking areas that are part in one functional network, this oscillatory neuronal synchrony serves also to bind together information represented within these different parts of the network (Gray et al., 1989; Fries, 2005). Taken together, these long-range synchronizations across remote areas of the brain have been proposed to be relevant for conscious perception of a stimulus or an object (e.g. Singer, 1999; Engel et al., 2001; Dehaene et al., 2006; Melloni et al., 2007).

Referring to the underlying functional principles of (de)synchronization in distributed neuronal networks concisely reviewed above, the idea of predictive coding (Friston, 2005) and the Adaptive Resonance Theory (ART, Grossberg, 1980; Grossberg, 1999) offer intriguing computational approaches for integrating lexical top-down with bottom-up sensory information. These models generally posit that predictions based on prior experience are generated in higher-level areas and projected to lower-level areas to guide the (speech) recognition process driven by sensory information. Accounting for speech and language comprehension in particular (e.g. ARTphone, Grossberg et al., 1997; ARTword, Grossberg and Myers, 2000; for review see Grossberg, 2003) amongst others, the idea of resonant states in neuronal networks may clarify the achievement of successful speech perception in a variety of acoustical difficult situations. Generally the idea behind ART predicts that resonant states arise when bottom-up signals activated by environmental events interact with top-down expectations or prototypes that have been learned from prior experience. In this model the top-down expectations control a matching process that selects those combinations of bottom-up features which are consistent with the learned prototype and inhibits the inconsistent ones. ART states that the top-down expectations can not create something out of nothing by itself and that the different levels of speech processing interact reciprocally with the working memory. Further, the resonant processes are said to enable the efficient integration of information across several processing levels. ART suggests that all conscious states in the human brain are resonant states and that these trigger the learning of sensory and cognitive representations about a changing world in a stable fashion throughout life. According to ART, this resonant type of matching of top-down expectations with sensory input is evident in the phonemic restoration or the perception of otherwise distorted speech. Therefore it illustrates the underlying brain mechanisms for the rapid learning of speech and language representations (Grossberg, 2003) and its integration into a coherent conscious percept.

1.4 Integration or detection of change

What happens if a lexical top-down prediction does not fully match the bottom-up representation, i.e. in the phonemic restoration illusion? From an evolutionary point of view, beside fast integration of sensory input and lexical expectancies in order to facilitate speech and language comprehension, also the detection of errors in the acoustic signal is inevitable to avoid miscommunication. As memory based lexical predictions approximate the sensory input, the speech processing system should be able to evaluate the appropriateness of its top-down hypothesises and adjust its actual predictions if necessary.

According to the idea of predictive coding employing a Bayesian statistical framework (Friston, 2005), an error signal is generated which reflects any mismatch between the predicted signal and the actual stimulus-generated activity. As long as the prediction does not match the identity of the actual object, this error signal is projected to a higher cognitive level where a new prediction about the actual input is generated. Thus, in order to adapt the new prediction to the actual requirements, it is refined by the error signal. This process itself may be considered a learning process whereby an optimal representation of an input is “learned” through a process of iterative refinement that takes place via the integration of top-down and bottom-up information (Friston, 2005; see also Grossberg, 1999). In other words, the experience-based information is used to minimize the prediction error when processing sensory information. Following the idea of predictive coding (Friston, 2005), the speech processing system should detect unpredicted changes in the sensory input if it is unable to minimize the prediction error or if the effort of minimizing it is too high. On the contrary, the system should integrate the lexical top-down expectation and the ambiguous sensory input into a coherent object if it is able to resolve the prediction error slightly.

Claiming that humans adopt a near optimal strategy for perceiving speech, Norris and McQueen (Shortlist B, Norris and McQueen, 2008) offer another approach to speech perception using Bayesian statistics. In their view, word recognition is optimal in the face of ambiguous input when available perceptual evidence is combined with knowledge of the prior probabilities of words. Central to this approach is the calculation of conditional probabilities for words or phonemes given the actual sensory evidence according to the theorem of Bayes (1).

$$(1) \qquad \mathrm{P(Hypothesis_i/Evidence)} = \frac{\mathrm{P(Evidence/Hypothesis_i)} \ * \mathrm{P(Hypothesis_i)}}{\sum_{j=1}^{j=n} \mathrm{P(Evidence/Hypothesis_j)} * \mathrm{P(Hypothesis_j)}}$$

These conditional probabilities are calculated by reckoning the probability that the actual acoustical input is a realization of a certain phoneme or word (*P(Evidence/Hypothesis$_i$)*) and the a-priory probabilities or initial beliefs about the sensory input (*P(Hypothesis$_i$)*), which involve effects of word frequency, effects of the sentential or discourse context and the listener’s beliefs about the acoustical signal. It is essential to stress that the calculation of conditional probabilities work in parallel at all the different levels in the hierarchically organized processing stages that mediate between acoustic analyses of the speech signal and higher level representations.

Hence, this model also accounts for effects of co-articulation and within phoneme variability. Most important, this model offers a formal account of how the likelihood of different realizations of phonemes or words can modulate word recognition. The underlying assumption of Shortlist B is that the input signal is inherently ambiguous due to the noisy environment and that the listener’s knowledge of phoneme (or word) realizations is represented as likelihood functions instead of single absolute probabilities. For example, each phoneme might be described in terms of Gaussian distributions over several dimensions, each characterizing a phonemic feature. As in ART (e.g. Grossberg, 1999), the listeners crucially are able to learn and continuously update the likelihood functions for the phonemes in their language. Therefore this model can explain for example effects of mispronunciation and dialectal variance in speech perception. During online speech processing, the flow of information in Shortlist B is strictly forward from the lower to the higher lexical levels. However, the model admits feedback for perceptual learning which allows for instance the retuning of phonemic categories (Norris et al., 2003) and thus benefits speech processing. In brief, this Bayesian approach tries to explain how the speech processing system updates its initial predictions or beliefs in the light of new evidence to allow rapid and efficient speech perception.

Taking Shortlist B (Norris and McQueen, 2008) into account, the question about the integration of incongruous speech input and lexical predictions into a coherent percept or the detection of a deviance in the acoustics might be also resolvable. With respect to the assumption of likelihood functions representing possible instances of a phoneme for instance, the integration of lexical top-down influences and sensory information depicted as resonant states (Grossberg, 2003) should take place if the characteristics of the bottom-up signal are covered at least partly by the distribution of the expected phonemic features. On the other hand, if the signal and the prediction are too unusual or distinct, a mispronunciation or error in the signal ought to be detected. Obviously this assumption should hold not only for speech perception in general but also for the cases of communication, where the acoustics are interrupted by extraneous sounds, i.e. in the phonemic restoration illusion.

Combining the ideas behind predictive coding (Friston, 2005) and Shortlist B (Norris and McQueen, 2008), the phonemic restoration effect should occur only if the substituting or masking sound is in somewhat similar to the expected phoneme. In other words, if the characteristics of the sound did not fit in the likelihood functions of the predicted phoneme, the integration of pre-lexical and lexical information will fail, at least partly. This is exactly what has been observed behaviourally (e.g. Layton, 1975; Samuel, 1981b). Instead of causing resonant states as outcome of successful integration (Grossberg, 2003), the mismatch should result in an enhanced prediction error (Friston, 2005) which in turn should lead to a refining of the prediction to achieve an optimal representation of the flawed sensory input. In order to make the assumptions made by the above reviewed models revisable the next section will introduce event related brain responses as tool to verify them.

1.5 Brain responses accompanying speech perception

The aim of thesis is to gain new insights in the perceptual mechanisms that cope with sketchy speech input and its underlying cognitive mechanisms. To unravel the assumed integration of pre- and lexical information and the detection of incongruous signals in speech a tool is needed, which fits to the rapid temporal dynamics of the acoustic signal and which is able to depict the linking of distributed information in the cognitive speech perception network. As this thesis does not aim to localize or identify brain areas responsible for the handling of deficient speech, the measuring of electric currents from the head surface with the electroencephalogram (EEG) using up to 256 single electrodes (and also of the accompanying magnetic fields via the magnetoencephalography (MEG)) is most appropriate for this task.

Since the invention of the EEG in the late 1920s by Hans Berger, the EEG has become an invaluable tool for cognitive neuroscience research which delivers recordings of ongoing neuronal brain activity with a temporal resolution of milliseconds and a restricted (because of the blurring as a consequence of high skull resistance) but sufficient spatial resolution. With respect to the underlying neuronal dynamics of error detection and multilevel information integration in speech comprehension, two kinds of analyses of the obtained electric signals seem to be fruitful: the analysis of the event related potentials (ERP) and most important the analysis of synchronous oscillatory activity, first of all the activity in the so called gamma band (GBA). First, with respect to the history of EEG analyses in the literature, some implementations of ERPs in investigating speech perception will be introduced shortly before the focus is turned on oscillatory neuronal dynamics.

1.6 Event related potentials in speech perception

In general, the ERPs reflect the sum of simultaneous postsynaptic activity, which is time locked to sensory, motor or cognitive processes, of a large population of perpendicular aligned mostly pyramidal neurons. These small voltage fluctuations are recorded at the scalp surface in the EEG. To improve the signal-to-noise ratio in ERP research, usually single events are presented repeatedly and the belonging EEG signals are averaged. The resulting ERP is commonly divided into components that exhibit a typical polarity when recorded at the scalp surface and a peak, which are denoted as a P or N for the polarity and a number reflecting the latency of that maximum. Next some of the components typically investigated in speech perception studies related to the integration of different informational levels are mentioned. For instance, the auditory N100 is most reliable elicited by virtually any auditory speech stimulus (e.g. Naatanen and Picton, 1987; Naatanen and Winkler, 1999). This component reflecting auditory pattern recognition peaks usually around 100 ms after stimulus onset and is sensitive to attention.

A further frequently investigated ERP component is the mismatch negativity or MMN. Most relevant to the current thesis, the MMN is an event-related brain response elicited by infrequent acoustic events, the so-called deviant stimuli, occasionally occurring among frequently repeated sounds, called the standard stimuli and can be observed approximately after 140–200 ms (Naatanen and Michie, 1979). This component of the evoked auditory potential indicates automatic change detection in the absence of subjects' attention (Schroger, 1997). It is suggested, that the MMN also reflects early automatic processes of lexical access and selection, semantic information processing and syntactic analysis (for review see Pulvermuller and Shtyrov, 2006). On top of its known role as an automatic index of acoustic change detection, the MMN has been shown to reflect the existence of learned neuronal representations or memory traces for native language sounds. For instance, the MMN has been suggested to echo long-term memory for phonemes (e.g. Naatanen et al., 1997; Naatanen, 2001), syllables (e.g. Shtyrov et al., 2000) and words (e.g. Pulvermuller et al., 2001; Endrass et al., 2004; Shtyrov et al., 2008). In general, it can be said that MMN studies probe the interaction of cognitive and lexical top-down influences and basic auditory integration of speech input (Schroger et al., 2003).

Another component which is related to speech perception with respect to contextual integration processes is the N400 (Kutas and Hillyard, 1980; Kutas and Federmeier, 2000). The N400 is most sensitive to semantic relations between individual words, or between words and their sentence and discourse context (for review see Hagoort and Van Berkum, 2007). Modulations of the amplitude of the N400 are generally viewed as related to the processing costs of integrating the meaning of a word into the overall representation of meaning that is built up on the basis of the preceding language input. Overall the time course of the different ERP effects support the view that the different information types (lexical, syntactic, phonological, pragmatic) are processed in parallel and influence the speech perception process incrementally, that is as soon as the relevant pieces of information are available (for review see Hagoort, 2008).

According to the idea of predictive coding (Friston, 2005) evoked cortical responses can be understood as transient expressions of the prediction error. Particularly, the MMN might illustrate the error signal which reflects the mismatch between the predicted input and the actual stimulus-generated activity. Therefore the MMN qualifies as well-operationalizable dependent variable for investigating prediction errors in speech perception.

1.7 Oscillatory neuronal dynamics

As introduced above, the integration or binding of pre- and lexical as well as different sources of information is most likely propagated by synchronization of oscillatory neural activity (Gray et al., 1989; Fries, 2005). These fast oscillatory neuronal dynamics can be captured by the EEG (and MEG). Different from the ERP method, the EEG signals are not averaged as first processing step to improve the signal-to-noise ratio, because this averaging procedure destroys any temporal fine structure of the EEG signal that is not phase locked to the investigated experimental event (Pfurtscheller and Lopes da Silva, 1999b). However, a differentiation between early phase-locked and late non-phase locked or induced component can be made (Galambos, 1992). The first is typically referred to as evoked activity (comparable with the ERPs), which appears within the first 100 ms after the stimulus onset, is strictly phase-locked to it and reflect the encoding of some physical attributes of the stimulus. On the contrary, the so-called induced activity shows a higher phase jitter and occurs usually somewhat later after stimulus onset (Pantev, 1995). However, the induced gamma band oscillations should not be considered as a unique or stereotyped brain response derived from the same set of neural sources. On the contrary, they reflect interactions within networks organized both in space and time. Thus, fast oscillatory gamma synchronizations are proposed as a very general neural mechanism underlying sensory integration and object representation (Tallon-Baudry and Bertrand, 1999; Gruber et al., 2006) and several other cognitive functions.

Oscillatory synchrony in the brain can be studied in different ways. The first applies to the synchronization of information within a certain part of a functional network. Because of the increase in local synchronous firing of a neuronal population which is due to spatial summation of postsynaptic potentials the amplitude of the scalp-recorded EEG oscillations increases. Different methods for the analysis of these amplitude or power changes have been used in the literature, wherein the single-trial wavelet analysis (e.g. Bertrand et al., 1994; Tallon-Baudry et al., 1997; Tallon-Baudry et al., 1998) is mostly used. Other utilized methods are the event related desynchronization or the more sophisticated multitapper analysis (Pfurtscheller and Lopes da Silva, 1999a; Mitra and Pesaran, 1999). In contrast, the analysis of the phase relationship between signals recorded from different electrode sites reveals a measure of phase coherence, which depicts the degree of synchronization between different nodes of a neuronal network (e.g. Andrew and Pfurtscheller, 1996; Lachaux et al., 1999).

With respect to speech perception, there have not been very much studies investigating oscillatory brain responses. However, oscillatory neuronal responses were studied referring to the retrieval of lexical information and to unification and integration of linguistic information. They have been observed in different frequency ranges, for example in the theta (3 - 7 Hz) and the gamma band range (above 30 Hz) whereas the latter were related to unification operations in language comprehension (Bastiaansen and Hagoort, 2006). In the processing of words and nonwords, a modulation in GBA was observed for the differentiation between words and pseudowords with an augmentation in GBA for words (Lutzenberger et al., 1994; Eulitz et al., 1996; Fiebach et al., 2005; for review see Pulvermuller et al., 1997; Pulvermuller et al., 1999).

According to the previous sections of the present thesis, repair processes in speech perception are supposed to result of merging anticipated lexical and memory contents with sketch sensory input. As the processes of unification and integration in speech perception involve the maintaining and active manipulating of the sensory input in the working memory amongst others (Bastiaansen and Hagoort, 2006), they seem to be functionally related to the binding of information within and between other sensory modalities. So far, the synchronizations in the gamma band constitute as central in unveiling the processes that allow the comprehension of deficient speech.

For the sake of completeness it has to be mentioned that most of the work in the gamma band research has been done outside the domain of language comprehension, mostly in the visual domain and with respect to memory processes. In the recent years, the cognitive neuroscience literature has shown that induced gamma band activity serves as a signature of the recognition and representation of a coherent object respectively. This object representation can be either single- or multimodal and consist of perceptual, semantic, and task-related features. In general, familiar objects induced a stronger gamma band activity compared to unfamiliar objects (Tallon-Baudry and Bertrand, 1999; Busch et al., 2006a; Gruber et al., 2006). Induced GBA enhancements have also been reported during gestalt perception and multi-sensory integration (e.g. Tallon-Baudry et al., 1996; Kaiser et al., 2004; Kaiser et al., 2005; Kaiser et al., 2006a; Yuval-Greenberg and Deouell, 2007; Schneider et al., 2008). Furthermore brain oscillations play a significant role in different short- and long-term memory processes. Modulated induced activity in the gamma band is a key-player in explicit and implicit memory tasks and mirrors early encoding- and retrieval-related processes. Also it was observed as correlate for working and associative memory processes. Therefore it can be concluded, that the investigation of modulations in the gamma band can reveal deep insight in (human) memory processing and formation (e.g. Tallon-Baudry et al., 1998; Miltner et al., 1999; Basar et al., 2001; Keil et al., 2001; Gruber et al., 2002; Pesaran et al., 2002; Kaiser et al., 2003; Axmacher et al., 2006; Busch et al., 2006b; Leiberg et al., 2006; Osipova et al., 2006; Lenz et al., 2007a; Gruber et al., 2008).

The match-and-utilisation model (MUM, Herrmann et al., 2004b) assumes that a match between a sensory stimulus and perceptual memory traces enhances the early evoked gamma activity by means of enhanced synaptic connections and stronger feedback from higher into lower (visual) areas. A similar matching mechanism for the induced GBA was proposed by Gruber and colleagues. In particular, they pointed out that the induced GBA reflect the activation and integration of perceptual stimulus properties and semantic object features. Furthermore, they demonstrate that the activation of an object representation is not merely an automatic consequence of incoming sensory information, but is influenced by implicit and explicit mnemonic requirements (Gruber and Muller, 2005; 2006). The latter interpretation agrees with some aspects of MUM, in which it is suggested that induced gamma oscillations signify the utilization of a cortical network reflecting processes of readout and the utilization of earlier matches between incoming information and memory. With respect to the differentiation between neuronal synchrony within and between cortical areas it was proposed that changes in phase locking may reflect rather bottom-up processing, whereas power changes in oscillatory activity may reflect top-down processing (Busch et al., 2006b; Schneider et al., 2008). As the current thesis aims to examine the top-down guided processes of the reparation of deficient speech, the hypotheses will be therefore formulated with the focus on modulations of power changes in oscillatory activity.

Besides being a correlate for memory processes or object representation, GBA effects of modulations in selective attention were also observed (e.g. Tiitinen et al., 1993; Fries et al., 2001; Herrmann and Mecklinger, 2001; Debener et al., 2003; Tallon-Baudry et al., 2005). For example, enhanced GBA was found in response to target but not to novel stimuli. In early 1980 it was claimed, that the neural mechanisms that support voluntary shifts of attention are also involved in perceptual binding (Treisman and Gelade, 1980). Following this idea, influences of voluntary deployment of attention in perceptual binding depicted in the induced gamma band were reported (Landau et al., 2007).

As indicated shortly above the studies examining cognitive mechanism by means of modulations in the gamma band in the auditory domain, especially in speech perception, are sparse compared to the large amount of studies in the visual domain. However, there is evidence which suggest that modulations in the auditory GBA are highly functional related to the visual findings. With respect to selective attention and task demands comparable activity in the gamma band can be observed for example (Tiitinen et al., 1993; Yordanova et al., 1997; Debener et al., 2003). Amongst others, in response to novel auditory stimuli early gamma oscillations were also reported (Haenschel et al., 2000). Further, intracranial recordings during auditory discrimination tasks showed enhancements of GBA in favour of phonemes compared with simple tones (Crone et al., 2001). Signatures of GBA were also associated with the perception of coherent auditory objects (Knief et al., 2000). With regard to memory processes dealing with auditory input an enhanced GBA is associated with the memorization of syllables, the reckoning of pattern mismatches in passive oddball tasks and during auditory spatial short-term memory (e.g. Kaiser et al., 2002; Kaiser et al., 2003; Kaiser and Lutzenberger, 2004; Kaiser et al., 2008b). Furthermore, high-frequency oscillations were observed as a correlate of dynamics in cortical networks serving auditory decision making which were influenced by the task performance (Kaiser et al., 2008a; Kaiser et al., 2006b). With respect to the perception and the processing of the fine structure in music, higher degrees of synchrony in the gamma band over distributed cortical areas in musicians compared to non-musicians were reported (Bhattacharya et al., 2001). In addition, a recent study demonstrated an enhancement in induced GBA that reflects matches between sounds and their representations in long term memory in an auditory recognition task (Lenz et al., 2007b). Recently, the integration of auditory expectation based on visual symbolic information and auditory sensory information was examined (Widmann et al., 2007). In that study induced GBA was only elicited when the visual and auditory information was congruent. The authors concluded that these induced GBA reflect the integration of the anticipated auditory information based on the available visual symbolic information into a unitary event representation of a correct sound. The claim that induced GBA reflects expectancy based representations was supported by Snyder and Large (2005). They demonstrated that induced oscillations in the gamma band range persist, when expected tones in rhythmic tone sequences were omitted. Oscillatory neuronal dynamics have also been claimed to underlie clinically phantom perceptions in general (Llinas et al., 1999). For example, it has been shown that the laterality of the perception of a tinnitus is determined by enhanced activity in the gamma band range (Weisz et al., 2007).

Subsuming the above reviewed findings, synchronous neuronal activity in the gamma band range first of all depicts successful representations of coherent single- and multimodal objects as well as memory processes. As claimed by the Adaptive Resonance Theory (ART, Grossberg, 1980; Grossberg, 1999) this synchronous neuronal activity is the result of matching bottom-up sensory information with top-down expectations and learned representations, e.g., lexical memory traces. The neurobiological requirements for this unification of speech signals are those of working memory, which include that lexical building blocks are kept activated for some time while the unification operations take place (Hagoort, 2005). Assuming this kind of neuronal activity as a basis for information integration and feature binding into object representations, one could expect modulations in GBA when comprehension of degraded speech succeeds due to lexically guided predictions and previously strengthened lexical memory traces.

1.8 Outline

The current thesis examines repair processes in speech perception. Concluding on the literature presented previously in the short review, especially the synchronizations in the gamma band range as well as the MMN seem to be central in unveiling the processes that allow the comprehension of deficient speech. The ability to understand speech in a variety of situations might depend on the comparison and, when appropriate, the merging of anticipated lexical memory contents with sketch sensory input. Thus, depending on whether this comparison results in a match or a mismatch, different sets of cognitive processes are expected to be elicited, with the MMN should be more sensitive to processes following a mismatch and the induced GBA should be indicative of processes following a match of expectation and sensory stimulation. While the MMN should reflect the prediction error (Friston, 2005), the induced GBA is expected to depict the binding or unification of anticipated information with deficient speech input to build up a coherent speech object representation (Tallon-Baudry and Bertrand, 1999).

First, an experiment (study A) is reported, which was aimed to identify the induced GBA as a correlate of merging expected lexical information with sketchy speech input. So far, most of the work done in the field of oscillatory neural dynamics dealt not with speech stimuli. Using the EEG to uncover the interaction of auditory input and lexical memory traces this study investigates for the first time the modulation in the gamma range in relation to the intelligibility of degraded speech signals.

Second, two experiments (studies B and C) were conceived to elucidate the cognitive processes underlying the phonemic restoration illusion. To discern processes of unification (Grossberg, 2003; Hagoort, 2005) from the refinement of a-priory hypothesis through prediction error processes (Friston, 2005) the acoustical similarity between the expected and the actual sensory stimulus was manipulated while the overall experimental setup was held constant. In the second experiment of this thesis (study B) words with noise-replaced segments were contrasted with words containing a noise-overlaid segment. This results in a small acoustical difference between the expectation and the sensory information, which was expected to result in modulations in the induced GBA depicting the repairing of the deficient speech segments. This expected modulation should be comparable to the results of study A, if it refers to a similar underlying cognitive repair mechanism. On the contrary, in the third experiment (study C) the acoustical difference between the actual and the expected segments was larger. Here words with a noise-replaced segment were contrasted with normal spoken words containing no manipulation. Thus, following the assumption of likelihood functions representing possible instances of a phoneme (Norris and McQueen, 2008) this larger mismatch should prevent the successful integration of lexical top-down influences and sensory information at least partly and thus result in a MMN depicting the prediction error instead. Further, two small behavioural studies were conducted to gain knowledge about the attentive processing of the noise-replaced items. By examining response times and identification rates, the notion of multi-level mechanisms involved in the repair of deficient speech was examined (Results of these small experiments are reported together with the main results of Study B).

While study A was set out to identify the induced GBA as measure for binding and unification processes in speech perception without specific assumptions about the fine structure of the mental lexicon, the studies B and C assume an abstract and sparse representation of language sounds in the mental lexicon particularly (fully underspecified lexicon, FUL, Lahiri and Reetz, 2002). To account for variations in the sounds of natural languages FUL claims that only distinctive phonological information is stored in the mental lexicon. Thus, for example, the phonological feature of the coronal place of articulation has been proposed to be not explicitly represented in the long-term memory and to rather function as an articulatory and perceptual default in German and other languages (cf. Paradis and Prunet, 1991). This assumption of an underspecified mental lexicon was confirmed for instance by findings of Eulitz and Lahiri (2004) which showed that the suspending of a repeated presentation of fully specified (i.e. a phoneme with a dorsal place of articulation) with underspecified items lead to larger phonological conflicts mapped in differential MMNs than vice versa. Consequently, the second and the third study of this thesis were also conducted to test if the repair process inherent in the phonemic restoration depends on the fine structure of the mental lexicon. Therefore the specificity of the anticipated phoneme in the phonemic restoration in each of these two experiments was manipulated to test for that influence. Thus studies B and C were also intended to discern influences of the mental lexicon on the lexical top-down factors, which in turn are integrated or mismatch with the actual stimulus-generated activity to allow an efficient handling of deficient speech input.

In summary, the present thesis aims to examine the following experimental questions:

1.) Is the successful recovery of deficient speech input accompanied by modulations in the induced gamma band as measure of the facilitating influence of top-down knowledge on speech comprehension under adverse listening conditions? Does the expected facilitating top-down influence in speech comprehension generalize to the phonemic restoration?
2.) What are the constraints of repairing deficient speech input? Does an explicit difference between sensory and expected information prevent the integration of these into one coherent speech percept and does this mismatch result in differential brain responses depicting this prediction error?
3.) Is the repairing of missing segments in the phonemic restoration or the failure of it influenced by the fine structure of the mental lexicon and is this influence depicted in the particular brain responses also?

2 General Methods

This chapter will shortly introduce some methodological aspects valid for all studies. Details about stimulus preparation, experimental design or epoch length are subject to the sections dealing with the specific experiments.

2.1 Subjects

All subjects who took part in the experiments of the current thesis were healthy students or research staff without any neurological, psychiatric or otological disease. It was assured that only monolingual speakers of German and right-handers were included in the data analysis, as ascertained by the 10-item Edinburgh Handedness Questionnaire (Oldfield, 1971). All participants were debriefed about the nature of the experiment and the EEG measurements by the principal investigator or by a research assistant. They signed an informed consent form and were paid dependent on the experiment's duration. All studies were conducted in compliance with the Declaration of Helsinki and the standards established by the local ethics committee. During the EEG studies all subjects were asked to sit quietly and avoid excessive eye and other movements.

2.2 Electrophysiological recordings

EEG recordings were made using a 64-channel system (TMS international, Type Porti S/64) in an electric shielded and sound attenuated chamber. The system consists of 62 Ag-AgCl electrodes mounted in an elastic cap (Electro Cap International, Inc.) according to the 10-10 system and two additional electrodes placed below the eyes for monitoring eye movements. The EEG was recorded continuously with a sampling rate of 250 (study A) or 512 Hz (studies B and C). All electrodes were referenced against the average reference (except study A, were the Cz electrode was chosen as reference, the data were then re-referenced offline to average reference) and the impedances were kept below 5 kΩ. All channels were pre-processed online by means of 0.1 Hz high- and low-pass filtering, with low-pass cut-off frequency being set to values well below the Nyquist frequency.

The experimental data were corrected for eye artefacts using BESA 5 (Berg and Scherg, 1994) and artefact-flawed epochs were rejected by visual inspection and if epochs exceeded a maximum of 60 µV in amplitude or a gradient of >75 µV. Overall, if the artefact-rejection procedures rejected more than 20% of the trials belonging to one experimental condition of one subject, this subject was excluded from all further analyses. To reduce computational time and computer-memory demands for the analyses of spectral changes, the data of study B and C were down sampled to 250 Hz prior to the next analyses.

2.3 Analyses of the spectral data

In the current thesis, complex Morlet wavelets (Bertrand et al., 1994) were used for the analysis of the spectral changes in the artefact-free epochs. This procedure has the advantage that the time resolution for high frequencies is better compared to low frequency ranges, where frequency resolution is better, but time resolution is coarse. Thus, by forming a good compromise between frequency and time resolution (Sinkkonen et al., 1995), this technique is especially suited for detecting induced high-frequency oscillations that may occur during brief periods of time. It provides a time-varying magnitude of the signal in each frequency band which leads to time frequency representations of the signal and is described in detail elsewhere (e.g. Tallon-Baudry et al., 1997; Tallon-Baudry et al., 1998). In short, the complex Morlet wavelets g can be computed in the time domain for different analysis frequencies f_0 according to

$$(2) \qquad g(t, f_0) = A' e^{\frac{-t^2}{2\sigma_t^2}} e^{2i\pi f_0 t}$$

with A' depending on the parameter σ_f, specifying the width of the wavelet in the frequency domain, the analysis frequency f_0 and the user-selected ratio m:

$$(3) \qquad A' = \sigma_f \sqrt{2\pi^3} \sqrt{\frac{m}{f_0 \sqrt{\pi}}}$$

with

$$(4) \qquad m = \frac{f_0}{\sigma_f}$$

Hence, given a constant ratio m, the width of the wavelets in the frequency domain, σ_f, changes as a function of the analysis frequency f_0. In the present experiment a ratio of m = 7 was used, because this ratio provides adequate time resolution and frequency accuracy in a frequency range between 10 and 100 Hz, while introducing no distortion due to differences in the integral size of wavelets at different frequencies in the frequency domain. Both the signal and the wavelets, which were generated using equation (2), were transformed to the frequency domain and multiplied individually for different analysis frequencies. After retransformation, the energy of the signal at different frequencies was obtained as the complex modulus (magnitude) of the real and imaginary part of the result matrix. Then, the time by frequency energy was averaged across single trials, allowing one to analyze non-phase-locked frequency components.

Following the wavelet analysis in each study, the mean spectral power was averaged across six electrode arrays (three in each hemisphere) each consisting of 6 electrodes (Fig. 1). Next the raw wavelet-data was normalized by computing the relative power change for every time by frequency bin compared to the according median of the baseline (-200 to -100 ms prior to stimulus onset; see the Methods sections of the single studies for further details).

Because of having no a-priori knowledge about exact latencies and frequencies at which lexical expectations matches with sensory input a permutation test (Blair and Karniski, 1993) was applied to compare the experimental conditions in study A and B

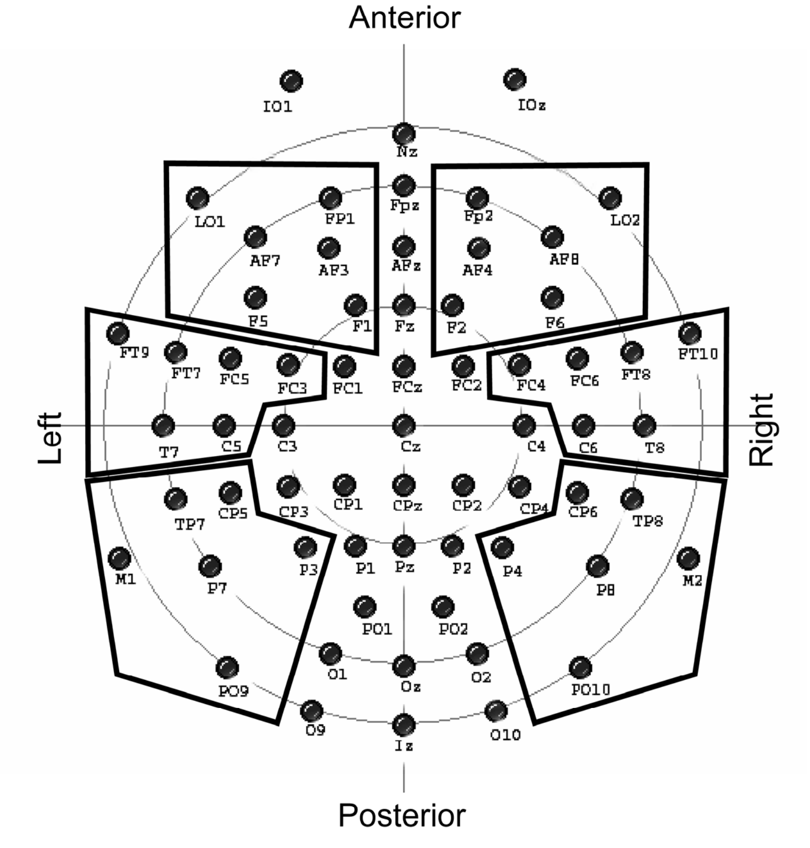


Figure 1. *Electrode montage and groupings used for statistical analyses are shown.*

respectively. These tests were applied to each frequency bin in the gamma range. With respect to the experimental hypothesises of study A and B correspondingly, only contiguous bins for an adequate duration per frequency band which showed a p-value $p < 0.01$ (uncorrected) were taken into account for further consideration to make sure that no time-frequency bins passed the test by chance (see the Methods sections of the studies A and B for details in latencies etc.). Finally, the gamma band results were subjected to repeated-measures ANOVAs to substantiate the findings and further analyses were restricted to the contiguous time by frequency bins that survived the initially permutation test. This procedure induced greater statistical power to all following analyses, because it ideally avoids a blowing up of the factorial design in a classical analysis of variance (ANOVA).

Because study C represents a perfect clone of study B with slightly variations in the stimulus material, the analyses of the GBA in study C were restricted to that group of electrodes to allow for more general conclusions about the functional meaning of the obtained gamma band data in the current thesis.

2.4 ERP analyses

Epochs for the analysis of ERPs in studies B and C were derived by averaging single epochs of the continuous EEG recording after applying a 30 Hz low-pass filter belonging to a certain experimental condition. These epochs lasted from -200 to 1000 ms post stimulus onset. The interval from -100 ms to stimulus onset served as baseline respectively. Before subjecting the ERP to further testing they were re-referenced to linked left and right mastoids. The latencies for examining the assumed prediction error are chosen with respect to the identified latencies of the spectral analyses to allow for a direct comparison of repair and error detection processes in speech perception.

2.5 Statistical analyses

All statistical analyses reported in this thesis are performed with SPSS 15 (SPSS Inc.) and MatLab 7.1 (MathWorks, Inc.). The approach used in the current series of experiments to assess the spectral or ERP data is the analysis of variance (ANOVA), which is a subtype of the general linear model $Y = \beta_0 + \beta_1 X + \varepsilon$. The ANOVA returns an F-distributed ratio of explained to total variance as a test statistic. All observations are required to be independent of one another and to be normally distributed. However, in the cognitive neuroscience most factorial variables are conceptualized as repeated or within-subject measures: different levels of the independent variable are not measured in independent samples, i.e. the restoration of an expected specified phoneme is not measured in other subjects than the restoration of not fully specified ones. This represents an important alteration and requires additional assumptions for the F test to be valid. The most important and most frequently violated assumption of the repeated measures ANOVA (besides normal distribution and independence of subjects) is the homogeneity of the covariance matrix. This so-called sphericity means that the means of differences between the repeated measures levels are equally distributed. With increasing deviance from sphericity, the F test becomes too liberal (Keselman, 1998). In psycho-physiological studies, researchers commonly account for this problem by applying Greenhouse and Geisser's sphericity estimate ε in order to reduce the degrees of freedom. Admittedly, in recent years multivariate ANOVA approaches have become a powerful alternative to univariate testing because they do not require sphericity among repeated measures levels and have been shown to be more sensitive to within subject effects for the typical demands in cognitive neuroscience research (Keselman, 1998). Therefore, following testing rationale was adopted in all three studies: First, the data matrix was tested on violations of sphericity using Mauchly's criterion. If the test indicated no violation, the degrees of freedom were not corrected and the uncorrected results of the univariate ANOVA were reported. Alternatively, if the sphericity assumption was invalid, the multivariate test using a Wilks λ-approximated F ratio was considered and reported.

3 Experiments

3.1. Study A: Top-down knowledge supports the retrieval of lexical information from degraded speech.[1]

How is it that the human brain is capable of making sense from speech under many acoustically compromised conditions? The support through top-down knowledge is inevitable, but can we identify brain measures of this matching process between degraded auditory input and possible meaning? To answer these questions, study A investigated the modulation of the induced GBA in the auditory domain in response to degraded speech.

During an EEG experiment subjects first listened to digitally degraded unintelligible speech signals (derived from German nouns). In an exposure sequence, half of the nouns were presented in a non-degraded intelligible format and memorized, while in the crucial test sequence subjects listened to all degraded speech signals again and were asked to identify the words. The induced GBA (40-Hz range) showed an increase at left temporal electrode sites around 350 ms only for words correctly identified in the test sequence. No differences in induced GBA were evident in the baseline sequence; neither did the evoked brain potentials yield any comparable effect.

It is concluded that the observed enhancement in induced gamma band activity reflects a matching process of top-down lexical memory traces with degraded sensory input to form a comprehendible speech percept. The findings are highly corroborant to analogous studies in the visual system. They lend further plausibility to a left-lateralized fronto-temporal network enabling lexically guided speech perception, and they demonstrate the complementary role of time-sensitive brain analyses in discerning the functional neuroanatomy of speech.

---

[1] This section is published in modified form in *Brain Research (2007), 1153, 134-143* (co-authored by Jonas Obleser and Carsten Eulitz). The collection of the EEG data was part of the diploma thesis of Ronny Hannemann: “Induzierte Hirnaktivität bei auditorischer Sprachwahrnehmung – Eine EEG Studie”. All analyses and the evaluation of the data presented here were not part of that diploma thesis. The author wishes to thank B. Awiszus and C. Wolf for their help in EEG data acquisition and A. Keil for helpful comments.

3.1.1 Background

Successful speech perception is an extraordinary human ability, and the human brain is capable of creating a meaningful speech percept under acoustically adverse conditions. However, this successful comprehension cannot be achieved on the basis of sensory information alone (see General Introduction for experimental evidence). Study A was designed to explore the contribution of lexical memory traces to the intelligibility of distorted speech. Listening to a spoken word (clear or acoustically degraded) might by defined operationally as an acoustic experience producing a two-dimensional image with a time and a frequency dimension (Griffiths and Warren, 2004). It can be assumed that its acoustic features such as temporal structure and contained frequencies as well as its linguistic features are processed across multiple cortical areas.

It has been shown that successful representation of objects in the human brain is accompanied by synchronous neuronal activity in the gamma band range between different cortical areas (Tallon-Baudry and Bertrand, 1999) which was consecutively interpreted as a correlate of sensory awareness for these perceived objects (e.g. Engel and Singer, 2001). According to the Adaptive Resonance Theory (ART, Grossberg, 1980; Grossberg, 1999) this synchronous neuronal activity is the result of matching bottom-up sensory information with top-down expectations and learned representations, e.g., lexical memory traces. Assuming this kind of activity as a basis for information integration and feature binding into object representations, one could expect synchronous neuronal activity in gamma range when comprehension of degraded speech succeeds due to previously strengthened lexical memory traces.

As outlined in the General Introduction, gamma band correlates of different cognitive processes and tasks were identified, most of them in visual domain. Enhanced GBA in 40 Hz range has been reported for perceiving a coherent object (e.g. Tallon-Baudry et al., 1996; Gruber et al., 2002; Goffaux et al., 2004; Busch et al., 2006a), while other studies emphasize its role in memory processes, particularly in matching stimuli to memory templates (e.g. Tallon-Baudry et al., 1998; Herrmann et al., 2004a; Gruber et al., 2004; Osipova et al., 2006) as well as in language processing (for review, see Pulvermuller et al., 1997; Pulvermuller, 1999). Studies examining selective (e.g. Herrmann and Mecklinger, 2001) and visuo-spatial attention (e.g. Muller and Gruber, 2001) as well as differentiation of words and pseudo words (Lutzenberger et al., 1994; Pantev, 1995) also observed modulation in GBA. Most relevant to study A, an enhanced induced GBA power was observed in a perceptual learning task in the visual domain (Gruber et al., 2002). In this study the identification of a fragmented picture was associated with stronger induced gamma band responses after rapid perceptual learning has taken place upon the prior experience with an unfragmented version of the same picture, while mere repetition of (other) fragmented pictures did not affect the gamma band power.

To date, as outlined above, the number of studies examining gamma band activity related to auditory tasks, especially speech is sparse compared to the visual domain. It is known, though, that selective attention to tone pips enhances GBA (Tiitinen et al., 1993), and that during perception of normal speech an enhancement in GBA is evident if subjects have to detect target words (Eulitz et al., 1996). Comparing attention to target sounds with (unattended) new environmental sounds, enhanced GBA was observed only for target sounds (Debener et al., 2003). In simple and choice reaction tasks early auditory GBA was associated with focused attention while later GBA was stated as meaningful parameter varying with different processing demands (Yordanova et al., 1997). In response to novel auditory stimuli Haenschel and colleagues (Haenschel et al., 2000) reported early gamma and beta oscillations preceding changes in broad-band event-related potentials. Intracranial recordings during auditory discrimination tasks showed enhancements of GBA in favor of phonemes compared with simple tones (Crone et al., 2001). Comparable to the visual domain, signatures of GBA was also associated with the perception of coherent auditory objects (Knief et al., 2000). In auditory pattern memory tasks, the memorization of syllables seems also to be accompanied by enhanced GBA (Kaiser et al., 2003). In addition, these patterns of gamma band activity were observed in passive as well as active oddball tasks for auditory pattern mismatch detection (Kaiser et al., 2002; Kaiser and Lutzenberger, 2004) and as a correlate of dynamics in cortical networks serving auditory decision making (Kaiser et al., 2006b). In a recent study, Lenz and colleagues were able to demonstrate how the enhancement in induced GBA reflects matches between sounds and their representations in long term memory in an auditory recognition task (Lenz et al., 2007b).

To date, however, there is no comparable study to the perceptual learning task of Gruber and colleagues (Gruber et al., 2002) as well as speech perception in the auditory domain to explore the functional significance of induced GBA in the perception of degraded auditory input. This study has been designed to fill this gap.

Applying a highly analogous task using degraded speech items, study A investigates the modulation of induced GBA during an auditory learning task. The absence of intelligibility for the degraded items (based on bisyllabic German nouns) was established by means of an independent pre-test. It is assumed that a short auditory rehearsal phase, that is, prior experience with non-manipulated words will strengthen their lexical memory traces. When listening to the degraded counterparts of these words after an exposure phase, residual features in degraded speech should enable a top-down mediated match of bottom-up degraded speech information with learned representations. This match should allow the identification of degraded speech. No such pattern is expected for non-identifiable degraded speech. Specifically, one would expect an enhancement in the induced (i.e., not phase-locked) GBA for successful matches, partly because evoked responses might fail to capture the slightly varying individual and trial-dependent latencies of such successful matching processes.

To test this hypothesis, study A was designed as an EEG experiment consisting of three experimental sequences. As a baseline measure the subjects listened passively to all stimuli in an acoustically degraded format (baseline sequence, see also Methods and Fig. 2). After being exposed to half of the stimuli in their unaltered, intelligible format in the second experimental sequence (exposure sequence), the subjects listened to all degraded stimuli again in the third sequence (test sequence). There they had to indicate by button press if they could identify a given stimulus or not (Condition = correctly identified or not identified). If the assumption about enhanced induced GBA as a correlate for a successful match holds, one would expect a significant Condition x Sequence interaction.

3.1.2 Materials and Methods

Subjects

Twenty right-handed university students (ten female; mean age: 24.6 years, standard deviation [SD] = 2.8 years) without audiological or neurological pathology participated in this study. They received class credits or a small financial bonus. Another 15 (seven female; mean age: 25.6 years, SD = 2.2 years) subjects took part in a 40-minute pre-test and received the same reimbursement. Informed consent was obtained from all participants. All 35 subjects were monolingual Germans.

Stimuli

The degraded speech signals were derived from natural recordings of 42 German concrete nouns. Original sound recordings of a male speaker were digitized with 44.1 kHz at 16 bit and edited using the Cooledit 2000 audio editing software package. Nouns were bisyllabic, with each syllable following a consonant-vowel pattern (i.e., CVCV). The consonants of the first syllable were all plosives (see Eulitz et al., 1996). To degrade the sounds, the resolution of the original files was reduced by replacing n sampling points at a time by their average value using MatLab 6.5 (Fig. 2). This procedure resembles an insufficient scanning of the digital signal without further down sampling. In a pre-test all words were presented in 14 different degrees of degradation and an item-wise threshold of identification was determined. Eight items were excluded for the subsequent EEG study for showing a value larger than 2 in mean-median difference or a standard deviation larger than 5 for identification threshold level. Degradation level for the remaining 34 items for the EEG study was chosen for each item separately to ensure that all words were not intelligible when presented for the first time. To this end we used the item-wise intelligibility threshold plus two times the standard deviation as the actual degradation level for each item in the EEG study, respectively. All stimuli were normalized for 99% of peak amplitude and presented in comfortable loudness (approx. 50 dB SPL) via headphones (Sony MDR-CD570).

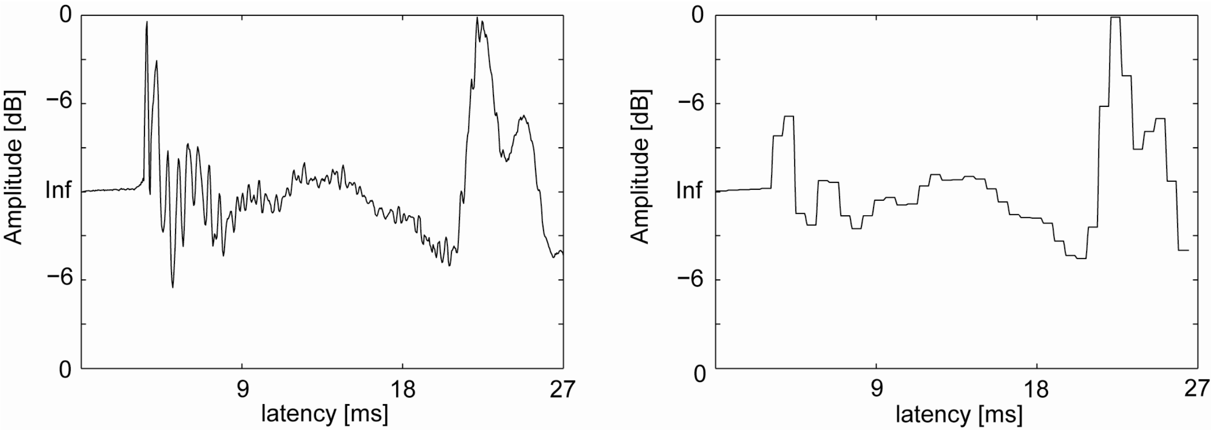


Figure 2. *Illustration of the speech-degradation-algorithm used. First 27 ms of one experimental item are depicted in original (left) and degraded form (right).*

Experimental Design

Study A comprised three sequences. In the first sequence (baseline sequence) all 34 items were presented in an unintelligible degraded format. Each item was presented ten times in randomized order (i.e., 340 trials). The stimulus onset asynchrony (SOA) varied between 2200 and 2600 ms. Subjects here were only asked to "listen attentively to the auditory items" without any further task. In the second sequence (exposure sequence) one half of the items were presented in an intelligible format, six times each (i.e., 102 trials, also randomized, same SOA). Subjects were asked to silently "memorize the randomized words for a later recall test". In the exposure sequence no association to the corresponding degraded items was established neither by instruction nor task. To control for effects related to auditory properties of the items per se, one half of the subjects were presented intelligible versions of items 1-17 while the other half of subjects listened to intelligible versions of items 18-34. No EEG data are available for the exposure sequence. The third sequence (test sequence) was similar in structure to the baseline sequence (i.e., 340 trials) for the only exception that after every item presentation subjects had to indicate whether the degraded item could represent intelligible speech by pressing a mouse button. Due to considerations concerning the overall length of the experiment, it was decided to use the task to press a button whether recognizing a word or not. After the experiment, however, subjects indeed had to name all recognized items once. To control for laterality of the motor responses use the assignment of left and right thumb to "intelligible" and "not intelligible" was balanced across subjects respectively. Between the button press and the next stimulus presentation there was a variable delay of 1000 to 1200 ms.

Subjects were instructed to avoid eye movements and blinks during the EEG recordings. After finishing the test sequence, subjects were presented once more with those items they had just indicated as intelligible and were ask to name them. Two experimental conditions were defined on the basis of their performance ("correctly identified" – heard in intelligible format in the exposure sequence and correctly named after test sequence; "not identified" – never heard in intelligible format and indicated as not intelligible in the test sequence). Misnamed items were excluded from further analyses.

Data acquisition and analysis

Continuous EEG was recorded with an elastic cap (EASY Cap) with 62 scalp electrodes at international 10-10 system locations and 2 additional electrodes for controlling eye movements below both eyes (see Fig. 1, General Methods, for a schematic representation of the recording array). The EEG data were sampled with 250 Hz and band-filtered from 0.1 to 100 Hz. All electrodes were online referenced to Cz and impedances were kept below 5 kΩ. After re-referencing to original average reference continuous EEG data were segmented to obtain epochs 500 ms prior and 1000 ms following stimulus onset. Experimental data were corrected for eye artifacts using BESA 5 (Berg and Scherg, 1994) and artifact-flawed epochs were rejected by visual inspection or if epochs exceeded a maximum of 120 µV in amplitude or a gradient of > 75 µV.

To analyze the induced spectral changes in GBA, a wavelet analysis using Morlet wavelets was performed on the artifact-free epochs of baseline and test sequence (for a description of the Morlet wavelets see General Methods). In the present study the wavelets were computed for a range from 9.76 to 87.84 Hz in 1.95-Hz steps in order to achieve a good time and frequency resolution. Next we normalized the raw wavelet-data by computing the relative power change for every time by frequency bin compared to the according median of the baseline. After wavelet analysis, mean spectral power in the baseline and test sequence for correctly identified and not identified items was averaged across six electrode arrays (three in each hemisphere) each consisting of 6 electrodes (Fig. 1). The averages for the baseline sequence were kept equal in signal to noise ratio compared to the test sequence and were generated on the basis of ratings given in the test sequence. As there is no a-priori knowledge about exact latencies and frequencies at which lexical identification is achieved by brain processes depicted in the gamma range we applied a permutation test (Blair and Karniski, 1993). To compare the difference of correctly identified minus not identified items in the test sequence with the corresponding difference in the baseline sequence (according to our hypothesis a difference should be observable in the former but not the latter). These tests were applied to each frequency bin between 17 and 48 Hz (beta and gamma band range) for latencies ranging from 200 to 900 ms after stimulus onset. For reducing computational demands we performed this on a temporally smoothed data set (obtained by applying a moving average of 16 ms with a step size of 8 ms). As we expected an enhancement in gamma band power for the correctly identified items in the test sequence, we considered only time-frequency bins showing an enhancement of at least one standard deviation over the baseline power change into account for further analyses. Further statistical analyses were performed only on time-frequency bins for which the permutation test shows at least a p-value of $p < 0.01$ (uncorrected). Finally, to exclude any time by frequency bins in the original data that passed our criteria by chance, we finally excluded all time-frequency bins in the dataset that didn't form a cluster of at least 4 bins in the temporally smoothed data set before we calculated any ANOVA.

To substantiate the findings of study A, a four-way repeated-measures ANOVA Condition (correctly identified vs. not identified) x Sequence (baseline vs. test) x Hemisphere (left vs. right) x Position (anterior, medial, posterior) was performed on the original (unsmoothed) data for all time-frequency regions clusters surviving the initial selection process. For all analyses involving the factor Position, we checked for violations of the sphericity assumption (using Mauchly's criterion), and in case of violations report multivariate testing (using Wilks Lamba) instead (Keselman, 1998; Obleser et al., 2003).

To rule out confounds by electromyographic (EMG) artefacts (Pulvermuller et al., 1997), we analyzed the frequency range (76-85 Hz) which is best reflecting EMG power for facial and head muscles (Cacioppo J.T. et al., 1990). Albeit effects in similar frequency ranges might reflect neuronal activity as well (e.g. Eckhorn et al., 1993), an absence of effects in these higher frequency bands would indicate that effects at lower frequencies are unaffected by EMG artefacts.

Further, to ensure that the induced GBA effects are independent of the differences in the evoked brain activity, mean amplitudes in the same latency windows as those for the induced GBA across the same electrode positions were calculated and statistically analyzed using the same factorial design. These calculations were implemented for evoked GBA as well as evoked potentials.

To analyze the behavioral difference in intelligibility perceived by the subjects between items whose non-degraded form had been memorized in the exposure sequence with those who were not been, the number of trials correctly identified was compared with the number of those trials, which were incidental intelligible but had not been memorized in the exposure sequence, by means of paired t-tests. This was done separately for subjects who had been memorized items 1-17 and 18-34 in exposure sequence.

3.1.3 Results

Behavioral responses

Overall correct identification of degraded speech signals in the test sequence increased significantly for items which had been presented in their unaltered intelligible format in the exposure sequence ($t(19) = 6.64$, $p < 0.01$). Expectedly, the selection of items for this exposure sequence had no influence ($t(9) = 5.55$; $p < 0.01$ for selection of items 1–17; ($t(9) = 7.88$; $p < .01$ for selection of items 18–34) in the exposure sequence. Reaction times in the test sequence differed significantly ($F(1,18) = 64.9$, $p < 0.01$) between correctly identified (1051 ms, SD = 298.9 ms) and not identified items (1542.9 ms, SD = 407.6 ms). It turned out that items not identified were slightly more degraded ($n = 15.5$) compared to correctly identified items ($n = 13.9$) (Wilcoxon rank sum = 106, $p < 0.01$). Despite these differences, however, the intra-individual variance of degradation level for correctly and not identified items showed no difference ($F(1,19) = 0.84$; $p < 0.37$).

Induced brain responses

In absence of a precise presumption about the topography of speech-induced GBA enhancement, six electrode clusters (frontal, anterior temporal and posterior temporal sites; Fig. 1) were defined to capture a wide range of possible cortical sources of brain activity in our analysis. Permutation tests (Blair and Karniski, 1993) identified one cluster of time-frequency (TF) bins over anterior temporal electrode sites in the left hemisphere that passed all criteria (see Methods) for further statistical analyses. The time-frequency plots in Figure 3 depict the differences of correctly identified over not identified items in the test sequence compared to the baseline sequence averaged into six electrode groupings (see Fig. 1). The time-frequency bins which passed the permutation test criteria are shown in color. Notably, one extensive cluster at middle to anterior left hemispheric electrode sites fulfills all criteria, and it exhibits a prominent increase in the 38–41 Hz range from 330–360 ms.

Figure 4a shows the scalp topography and time course of the induced changes in the 38–41 Hz range for the left anterior electrode group during the test sequence (activity from corresponding trials in the baseline sequence has been subtracted). The correctly identified items in the test sequence (solid black) show a substantial enhancement compared to the not identified items in the test sequence (solid gray) in the latency range of 330–360 ms as identified by the permutation test.

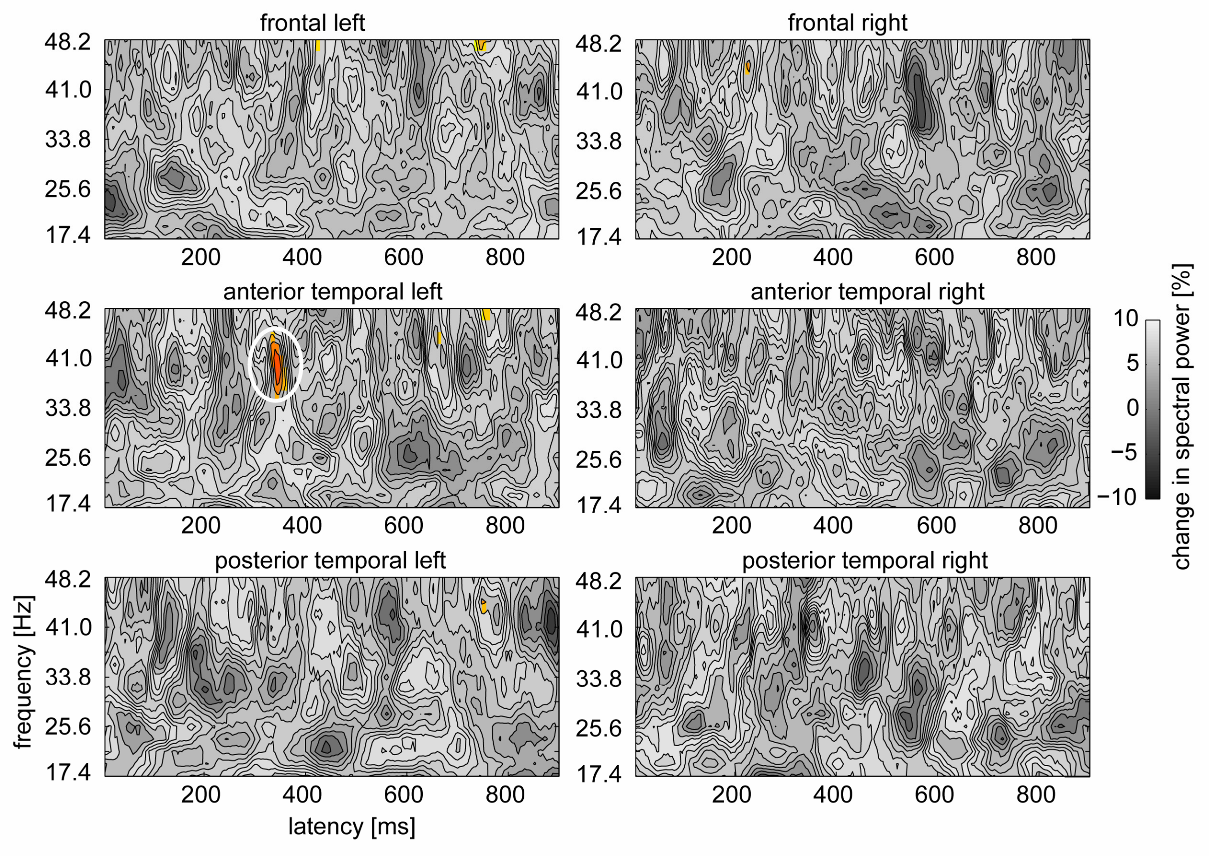


Figure 3. *Grand mean TF plots of percent change in induced brain activity over six brain regions (see Fig. 1 for electrode specifications). Depicted is the difference of correctly identified minus not identified items in the test sequence relative to the corresponding difference of the baseline sequence. Colored areas reflect TF bins identified as different by the permutation test with a p-value of at least p < 0.01. Only the big colored spot between 330 and 360 ms fulfills all criteria for further statistical analyses.*

A four-way repeated-measures ANOVA for this time and frequency range revealed a significant Condition x Sequence x Hemisphere x Position Interaction, (Wilks Lambda F(1,18) = 9.94, $p < 0.01$) for the change in induced spectral power. This confirms the permutation test results (Fig. 3) in that the prominent increase is focused on the gamma band and that it is left lateralized and most prominent at anterior temporal electrode sites. The changes in spectral power across sequences and conditions for the tested time-frequency range are summarized in Table 1 and support this assumption.

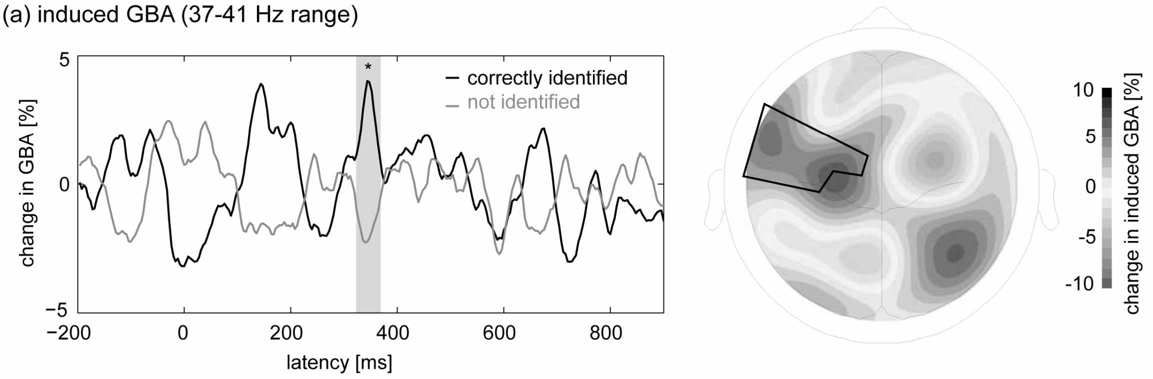
Further post-hoc comparisons were performed for this left anterior temporal spot. Here, a two way repeated-measures ANOVA revealed a significant Condition x Sequence interaction for the change in spectral power, F(1,19) = 17.72, $p < 0.01$ for left anterior temporal electrode sites. For these electrode sites the mean spectral

			Correctly identified		**Not identified**	
Position	**Hemisphere**	**Sequence**	%GBA change	(± SEM)	%GBA change	(± SEM)
Frontal	Left	Baseline	1.72	(1.13)	2.23	(1.17)
		Test	1.34	(0.96)	1.13	(0.91)
	Right	Baseline	-0.09	(0.61)	-1.14	(0.84)
		Test	0.36	(0.77)	0.16	(0.68)
Anterior temporal	Left	Baseline	-0.36	(0.80)	1.24	(1.05)
		Test	**2.81**	**(0.94)**	**-0.67**	**(0.73)**
	Right	Baseline	-0.36	(0.57)	-0.48	(0.83)
		Test	-0.05	(0.89)	1.20	(0.79)
Posterior temporal	Left	Baseline	0.09	(1.08)	0.05	(1.01)
		Test	0.52	(0.75)	-0.28	(0.73)
	Right	Baseline	-1.12	(1.03)	2.11	(1.10)
		Test	0.58	(1.03)	0.87	(0.78)

Table 1. *Mean spectral power for the 37–41 Hz / 330–360 ms range, averaged across six electrode sites and standard error of mean (SEM) in % change. Figures in bold type indicate the main result of this study.*

power is depicted in Figure 5. Further post-hoc comparisons showed a significant main effect for Sequence for correctly identified items, F(1,19) = 9.25; p < 0.01, and the mean values show clearly an enhancement for the test sequence for correctly identified items. In addition post-hoc comparisons revealed a significant main effect for Condition in the test sequence, F(1,19) = 7.44; p < 0.05 as well as barely significance in the baseline sequence, F(1,19) = 4.44, p < 0.49. The mean values for the change spectral in power show a clear enhancement for correctly identified items in the test sequence whereas in the baseline sequence this pattern is reversed. Importantly, the positive power change in the test sequence in favor of the correctly identified items exceeds the enhancement shown by the later not identified items in the baseline sequence. It is also important to note that the change in spectral power between baseline and test sequence for not identified items doesn't show a significant main effect.

(a) induced GBA (37-41 Hz range)



(b) evoked GBA (37-41 Hz range)

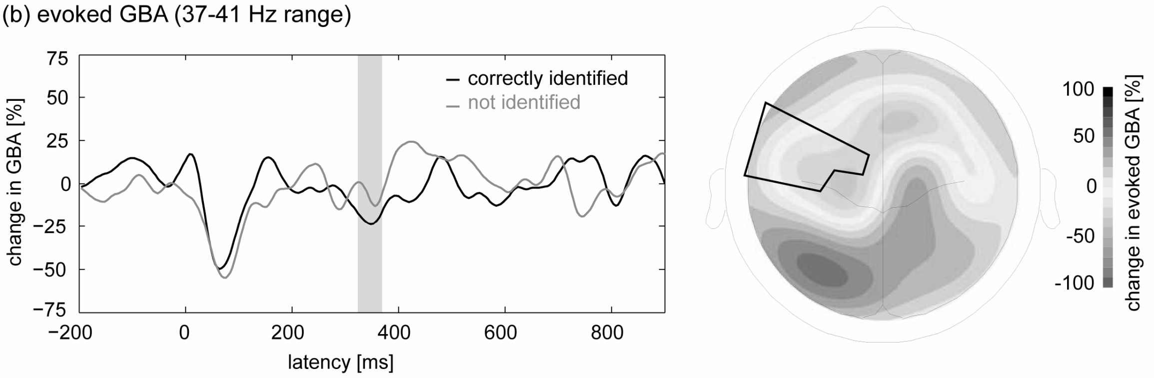


Figure 4. *Gamma band activity in the 37-41 Hz range. Left: Comparison of the time courses of induced (upper panel) and evoked GBA (lower panel) over left anterior temporal electrode sites. Black lines represent the correctly identified items (test minus baseline sequence) and gray lines show the not identified items (test minus baseline sequence). The grey boxes depict the time course for applied statistical analyses. Right: Shown is the difference topography of the test sequence (correctly identified - not identified) over the baseline sequence (correctly identified - not identified) in the 330-360 ms latency range for the induced (upper panel) and the evoked GBA (lower panel). The area of left anterior temporal electrodes sites (cf. Fig. 1) is outlined in black. Depicted latency range is the same as indicated by the grey boxes on the left (identified by the permutation tests).*
Note: TF time courses for the baseline sequence were generated on the basis of individual responses given in the test sequence.

Although our predictions were only specific with respect to processes of lexical memory (which are expected not to appear before onset of the second syllable would allow identification of a word) the time course of the induced 38-41 Hz changes in Figure 4a also points to an earlier enhancement in the latency range between 160-200 ms. Repeated-measures ANOVA confirmed that this effect is also elicited over left anterior temporal sites (Condition x Sequence x Hemisphere x Condition F(2,38) = 5.77, p < 0.01) and shows a response pattern that closely resembles the ensuing robust GBA effect (Sequence x Condition interaction F(1,19) = 10.68, p<0.01). However, this effect seems to be rather fragile, as suggested by the outcome of the permutation tests. The robustness criterion of an enhancement of at least one standard deviation compared to the baseline activity for the correctly identified items was clearly violated for this early activation.

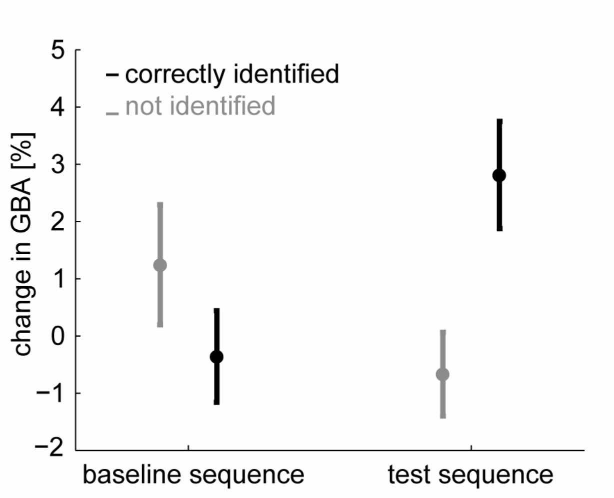


Figure 5 *Mean values of the change in induced GBA for the 37-41 Hz range in the 330-360 ms latency range.*

Post-hoc analysis of evoked brain responses and control for possible EMG confounds

To ascertain that our results are indeed changes in induced brain activity, we also analyzed the changes in the evoked GBA in the same latency and frequency ranges. Figure 4 contrasts the time course of induced and evoked changes in GBA for the left anterior temporal electrode sites. As illustrated, only the induced brain responses show an enhancement for correctly identified items in the test sequence. On the contrary, Figure 4b might suggest that the evoked responses for the correctly identified items are reduced in that time range, but statistical test analogous to the analysis reported above did not reveal a 4-way interaction ($p < 0.1$). Figure 4b also suggests early differences between correctly and not identified items for the test sequence after subtracting the corresponding trials from the baseline sequence. Four consecutive 50 ms latency windows between 20 and 220 ms were analyzed. None of these latency windows exhibited a significant Condition x Sequence interaction or main effect of Condition, which could interpreted as showing differences in acoustic features between correctly and not identified items. Though, a significant main effect of Sequence was observed for the latency window of 70-120 ms ($F(1,19) = 6.12$; $p < 0.05$).

Further tests were applied to the evoked potentials averaged over the same electrode sites as the ones used in frequency domain analyses. This was done to ensure that the enhancements in gamma band range are not solely due to an old-new effect of the stimulus items (for review, see Rugg, 1995). Notably, the Condition x Sequence interaction constituting the main result in the induced GBA power changes was not evident ($p < 0.2$) in the evoked potential. A four-way ANOVA testing spectral power in the 76–86 Hz range (in which the peak of the spectral density function of muscular contamination would be expected; Cacioppo J.T. et al., 1990) yielded no results as obtained for the 38–41 Hz frequency range (specifically no Condition x Sequence x Hemisphere x Position interaction $F<1$).

3.1.4 Discussion

Study A was set out to investigate the facilitating influence of top-down knowledge on speech comprehension under adverse listening conditions. We specifically hypothesized that one prominent correlate of such successful recovery of degraded auditory input would be an enhancement in the induced gamma band activity (GBA) as measured with EEG. One marked increase in induced gamma power at left anterior temporal electrode sites was found when subjects listened to correctly identified degraded speech signals compared to those not identified. Between about 330 ms and 360 ms we observed an enhancement in the 38–41 Hz frequency band in the test sequence, which was not evident at any other electrode location (Fig. 3). This enhancement in induced GBA in response to correctly identified items was also observable when comparing the test sequence against the baseline sequence. Neither evoked gamma band responses nor the evoked potential yielded a similar modulation through experimental conditions (Fig. 4) thus supporting the notion that induced GBA delivers additional insights about information processing in the brain (Eulitz et al., 2000). Behavioral results verified the effectiveness of the study design insofar as hearing half of the items in non-degraded form in the exposure sequence rendered them significantly more likely to be perceived as intelligible in the consecutive test sequence.

In short, the gamma band enhancement is thought to reflect the matching process between (degraded) auditory input and lexically driven expectancies (built up during the exposure sequence). This process results in an enhanced induced GBA as a correlate of resonant states in cortical networks (Grossberg, 1999). Elaborating on this, one would argue that the lexical memory traces representing items memorized for an expected recall test in the exposure sequence were strengthened and refreshed (i.e., learning took place; Gruber et al., 2002). In the consecutive test sequence, residual features in the degraded speech items were sufficient to activate these strengthened traces retained in working memory. This match between sensory information and memory-driven expectancies serves as a basis for auditory object recognition: It is only through residual traces of lexical memory that the correct word can be identified despite a degraded and highly ambiguous input signal. Furthermore, it is fruitful to think of the induced GBA changes as a correlate of binding the auditory input and available lexical information into a coherent perception, that is, a recognized word. Fundamental here is the assumption that GBA is a neural signature of cortical object representation (Tallon-Baudry and Bertrand, 1999). This notion is supported for the visual domain (Busch et al., 2006a) as well as for the auditory domain by recent results of Lenz et al. (2007b) which show enhanced induced GBA for a match of sounds and auditory long term memory traces (i.e., meaning). However, the study of Lenz et al. (2007b) did not incorporate the exposure (learning) aspect inherent in the present study and did not deal with speech stimuli, which can be conceived of as a special class amongst auditory objects.

As to the observed topographical distribution, the results of recent MEG studies gain relevance to the present study. GBA changes over fronto-temporal regions could be related to the maintenance of auditory information in short-term memory in an auditory pattern memory task (Kaiser et al., 2003) and decision making relevant to auditory pattern discrimination (Kaiser et al., 2006b). In study A, such working memory processes may be especially relevant to support the identification of the degraded speech items. Subjects had to retain the auditory information in working memory until a match with lexical memory traces and therefore comprehension was achieved, i.e. unification took place (Hagoort, 2005). Therefore it is possible that the actual results are not restricted to speech but could be a general signature of successful matching auditory information with long-term memory traces.

Another interesting support mechanism in speech perception (especially under difficult conditions) is attention, and it has been argued that some working memory processes seen in prefrontal cortex might be also explained as effects of selective attention (Lebedev et al., 2004). While GBA changes have been observed in studies of selective attention (Tiitinen et al., 1993; Debener et al., 2003), the differential pattern we observed for correctly identified versus non-identified items rules out a simple up-regulation due to attention and calls for more comprehensive explanation, such as the lexical selection and matching processes discussed above. (Note that both correctly identified and unidentifiable signals received more attention in the test sequence compared to the baseline sequence due to the imposed recognition task). The same arguments rule out task specific explanations (Yordanova et al., 1997) for the observed enhancement in induced GBA in the present study.

According to a whole body of neuroimaging studies using degraded speech (e.g. Scott et al., 2000; Giraud et al., 2004; Obleser et al., 2007), the intelligibility of degraded speech stimuli is accompanied by brain activity in inferior frontal as well as anterior and medial temporal areas in passive listening tasks. Giraud and colleagues (2004) also reported an increase in cortical activity in the anterior cingulate as well as superior to middle temporal regions during the perception of degraded but intelligible speech. A recent PET study observed increasing left inferior frontal brain responses when perceiving speech becomes more difficult due to masking with noise (Scott et al., 2004). The authors speculated that these responses might demonstrate top-down efforts to support speech comprehension with semantic information, and a recent study by Obleser et al. (2007) did demonstrate fronto-temporal and prefrontal activations (amongst others, all strongly left lateralized) when speech comprehension of degraded speech was likely to succeed solely due to semantic context cues being available. Using EEG's superior temporal resolution, the current study adds temporal resolution to these top-down modulated processes seen in neuroimaging studies. To exploit the full potential of this technique, however, further examinations of the cortical sources underlying the GBA are required, and tying fMRI and EEG measures in speech perception more closely is a stringent next step. Those experiments might also help to further investigate the functional role of the induced brain response in the latency window from 150-200 ms which turned out to be too fragile in the present study.

Design and hypotheses about gamma band responses of the present study, albeit using degraded speech items studying the auditory domain, strongly resemble a perceptual learning task in the visual domain (Gruber et al., 2002). Our study also identifies one significant effect in the gamma band range with slight differences in latency. This is most likely due to the different time course in perceiving auditory and visual information. While visual information (e.g., pictures) appears at once, auditory information (e.g., speech) becomes available in serial order and it takes more time to collect sufficient information to perceive an auditory object.

Generally speaking, it is highly unlikely that the reported effects are an artifact of specific stimuli (e.g., wordness; Lutzenberger et al., 1994; Pantev, 1995) per se because (i) the signals chosen for the exposure sequence were counterbalanced across subjects and (ii) no difference in induced GBA between later correctly identified and not identified signals could be observed in the baseline sequence (Fig. 4). In contrast, the increase in GBA power was confined to the test sequence and to correctly identified items. (iii) Although there were differences in the level of degradation between correctly identified and not identified items, the most sensitive parameter to differences in stimulus properties, the evoked GBA data, did not indicate any differences between correctly identified and not identified items. Thus, it is unlikely that the small differences in degradation level were causing the induced gamma band effects at later latencies.

In sum, the observed spectral power changes can be tied closely to the successful lexical access for stimuli that the listeners had previously been exposed to in a non-degraded format.

Summary

The present results show that the human ability to understand speech even under acoustically much compromised conditions relies on the interaction of auditory input and lexical traces (i.e., top-down influences). Consistent with current models of auditory word recognition (Scott and Johnsrude, 2003; Giraud et al., 2004) and lexical memory processes (e.g. Friederici and Kotz, 2003) left lateralized gamma band activity over anterior temporal electrode sites discriminates degraded items heard before in a non-degraded format and hence correctly identified from those never heard in an intelligible format. These findings of study A are highly corroborant to an analogous study in the visual system (Gruber et al., 2002) and extend previous auditory work into higher-order speech perception processes. Finally these data show experimentally how comprehending speech and the bottom-up brain processes mediating it depend highly on memory-driven (i.e., top-down) expectancies.

3.2 Study B: On the matching of top-down knowledge with sensory input in the perception of ambiguous speech.[2]

How does the brain repair obliterated speech and cope with acoustically ambivalent situations? A widely discussed possibility is to use top-down information for solving the ambiguity problem. In the case of speech, this may lead to a match of bottom-up sensory input with lexical expectations resulting in resonant states which are reflected in the induced gamma band activity (GBA).

In the present EEG study, the specificity of lexical top-down information was modulated (specified vs. underspecified phonological information) and compared the subject’s pre-attentive GBA responses to obliterated speech segments in German nouns. The induced GBA was larger when the expected lexical information was fully specified compared to the underspecified condition. Thus, the present results evince a direct correlate of merging lexical top-down expectancies dependent on the fine structure of information in the mental lexicon with obliterated speech input. Results of two behavioural control experiments support the notion of multi-level mechanisms involved in the repair of deficient speech. Further, the delineated alignment of pre-existing knowledge with sensory input is in accordance with recent ideas about the role of internal forward models in speech perception.

3.2.1 Background

At the level of speech, most conversations are considerably unclear. How does the brain cope with partly obliterated speech information and how does pre-existing knowledge support these coping-processes? It has been suggested that lexical information can be restored by using top-down lexical knowledge. Here the phonemic restoration illusion is used, where listeners hear spoken words as intact even though parts of them have been replaced by an extraneous sound (Warren, 1970), to study this repair processes in detail.

Given the top-down lexical influences on phonemic processing (Samuel, 1997; Samuel, 2001) the phonemic restoration illusion can be described as a match of bottom-up sensory input with lexical expectations resulting in resonant neural

---

[2] This section has been submitted in slightly modified form to a peer-reviewed journal (August 2008, co-authored by C. Eulitz). The author acknowledges gratefully the help of A. Bobrov, C. Massau and K. Preller in collecting and pre-processing the data.

dynamics (Grossberg, 1999; Grossberg, 2003). Similar, resonant states were first described in studies of feature binding in animals (Gray et al., 1989). In humans, such correlates can be measured as an enhancement in the gamma band which is discussed among others as a signature of object recognition (e.g. Tallon-Baudry and Bertrand, 1999; Gruber et al., 2006) and in relation with several mnemonic processes (Leiberg et al., 2006; Lenz et al., 2007a; Gruber et al., 2008). In language processing, a modulation in GBA was observed for the differentiation between words and pseudowords (Lutzenberger et al., 1994; Eulitz et al., 1996; see Pulvermuller et al., 1997 for review) as well as a correlate of merging expected lexical information with degraded speech input (see study A; Hannemann et al., 2007).

The present experiment was designed to depict directly the pre-attentive mechanisms of “filling-in” of phonemic information in the course of phonemic restoration without relying on sentential context by examining the induced GBA. Thus, the induced GBA was used as a measure for the pre-attentive formation of an auditory object (Griffiths and Warren, 2004). Presumably at the level of working memory anticipated phonemic information is unified with deficient sensory input to build a phoneme. To guide the phonemic restoration into one or another direction the experience-setting context was established by one of two words being a minimal pair which preceded a target word containing a noise replaced segment. The words setting the context differed in just one segment in no more than one phonological feature characterizing the place of articulation. According to actually discussed models, which assume abstract and sparse representations of language in the mental lexicon (Lahiri and Reetz, 2002) all nondistinctive and predictable information is not stored in the declarative memory. Thus, in study B, the place of articulation information of the contrastive segment in the minimal pair can be either retrieved from stored knowledge in the mental lexicon (i.e. the [dorsal] place which is called specified) or derived from redundant information (i.e. the underspecified [coronal] place; Lahiri and Reetz, 2002).

To test whether the specification of phonological details modulate the restoration of phonemes, the induced GBA to the noise-replaced stimuli was investigated. The induced GBA was expected to be larger in case of lexically specified compared to underspecified information in the precursor. This enhanced GBA might be a correlate of “filling in” the expected and lexically specified information to form a perceivable auditory object.

Alternatively, if the process of phonemic restoration does not differ according to the specification of the phoneme to be restored or if there is no (online) top-down influence on phonemic restoration (e.g. Norris et al., 2000; Norris et al., 2003), no differential modulation of induced GBA should be observable. To substantiate the induced brain activity as a correlate of merging lexical top-down expectancies with obliterated speech input it should be also dissociable from evoked brain activity.

3.2.2 Materials and Methods

Subjects

Nineteen healthy right-handed monolingual German-speaking volunteers without otological or neurological diseases participated in this study. Due to bad signal-to-noise ratio three subjects had to be excluded and all further analyses were performed for 16 subjects (eight female; mean age = 23.8 years, standard deviation [SD] = 3.1 years). All participants gave their written consent and received class credits or a small financial bonus.

Stimuli

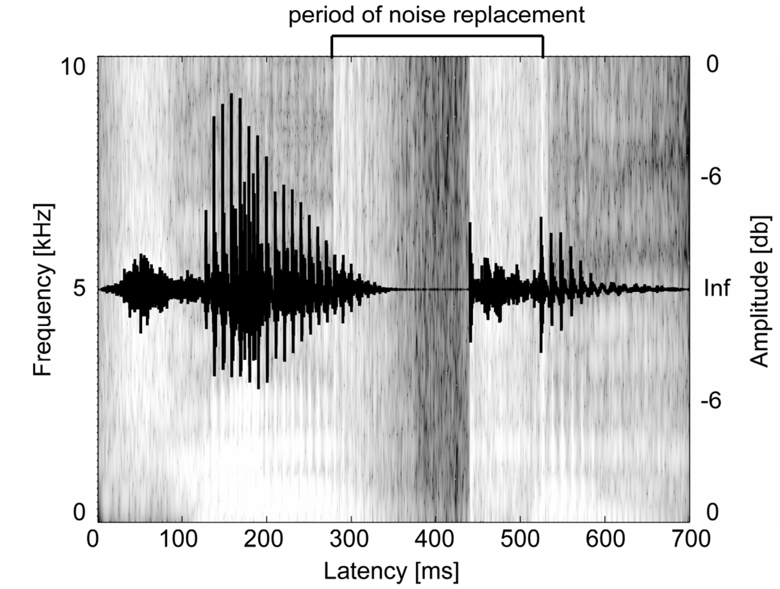
The experimental stimuli were derived from natural recordings of the minimal pair of the German nouns *Falke* (= hawk) and *Falte* (= fold), which were digitized with 44.1 kHz at 16 bit. Both nouns were controlled for frequency in written (*Falke* = 36, *Falte* = 29, (Celex; Baayen et al., 1993)) and spoken German (*Falke* = 0, *Falte* = 1 (Datenbank Gesprochenes Deutsch des Instituts für deutsche Sprache))[3]. For experimental purposes a pair of nouns with maximal similarity was chosen and was equalized further in envelope and second syllable onset using the software package Adobe Audition. Further, the acoustic difference between both nouns was minimized by cross-splicing the first and second syllables with each other which results in two instances of *Falke* and *Falte*. The latency for cross-splicing was chosen such that no co-articulation of /k/ and /t/ on the phoneme /l/ occurred. This latency was identified at 280 ms post stimulus onset in a separate gating test (for more details see also Appendix of study B). To create the noise-replaced items the speech-correlated noise technique (Schroeder, 1968), which flips the sign of half of the sampling points

---

[3] Numbers represent the word frequency = number of occurences of the words in a population of 6.000.000 words in the CELEX database. The Datenbank Gesprochenes Deutsch (http://dsav-wiss.ids-mannheim.de/DSAv/DSAVINFO.HTM) contains over 20 different corpora of different size. The above printed number reflects the overall occurrence of one word in all of those corpora.

chosen at random, was used to create a noise from 280 to 520 ms post stimulus onset that maintains the amplitude of the envelope original but has a flattened spectrum. Noise-overlaid items were produced by adding together the critical portions of the replaced and the original versions point for point. Figure 6 illustrates the oscillogram and spectrogram for one exemplar of (a) noise-overlaid and (b) noise-replaced items. This procedure resulted in three stimulus classes for the experiment: each two noise-overlaid *Falke* (K) and *Falte* (T) items and four noise-replaced

a) K-item (T-item) = noise-overlaid Falke (Falte)



b) #-item = noise-replaced Falke (Falte)

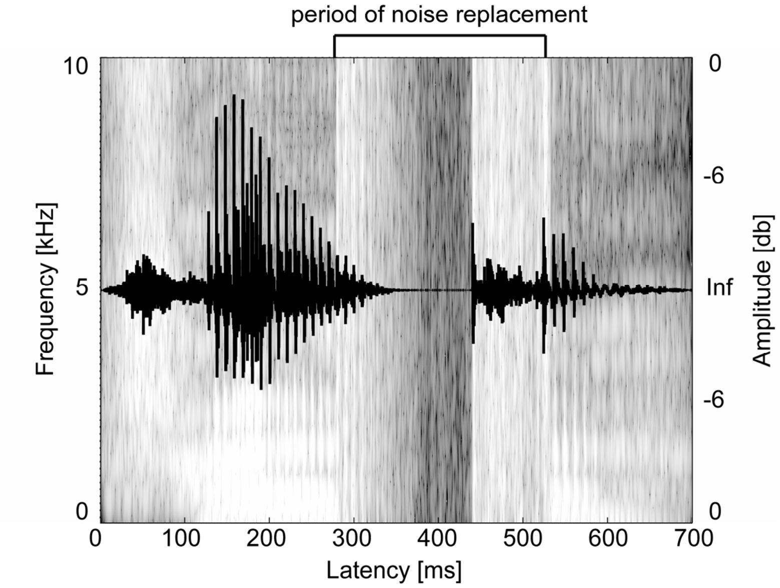


Figure 6. *Stimulus characteristics. Exemplarily the oscillogram and the spectrogram for a noise-overlaid K-item (a) and a derived noise-replaced #-item (b) are shown.*

ambiguous (#) items. According to the assumption of a featurally underspecified mental lexicon (Lahiri and Reetz, 2002) the critical phoneme in the T-items is underspecified for the [coronal] place of articulation while the featural information for K-items with the [dorsal] place of articulation is fully specified. Finally, all stimuli were normalized for peak amplitude and presented in comfortable loudness (approx. 50 dB SPL) via headphones (Sennheiser PMX 60).

Acoustic Stimulation

The present experiment consisted of three experimental blocks of approximately 28 minutes comprising of permanently changing stimulus trains (Fig. 7). One stimulus train consisted of a varying number (from 3 to 8) of K-, T- and #-items. Inside each train of this so called roving standard oddball paradigm (Cowan et al., 1993; Baldeweg et al., 2004) all items belonged to one stimulus class. Within each train we picked out for further analyses (i) each first item to be the **disconfirming** item with respect to the previous train and (ii) every third item to serve as the so called **confirming** stimulus.[4] The number of items per stimulus train as well as the stimulus class of each following train varied randomly through out all three blocks. Overall, this yielded in 12 different conditions (6 disconfirming and 6 confirming) with 120

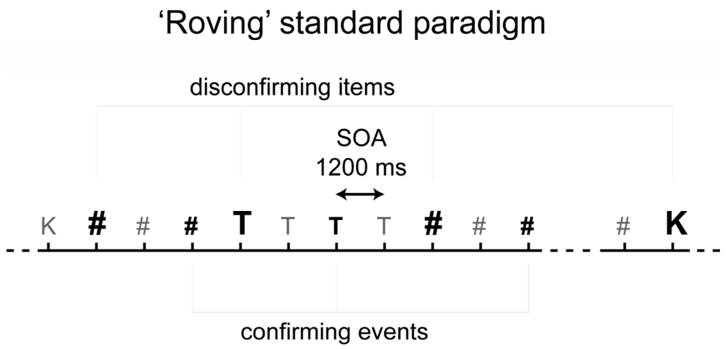


Figure 7. *Illustration of the roving standard stimulation used in study B (and C). The stimulus sequence consisted of repeatedly changing trains of K, T and #-items (indicated by characters) of a variable number (3 to 8 repetitions) of identical items. Every first item of a train served as disconfirming item (large black) and the respectively third item served as confirming item (small black). The EEG data of disconfirming and confirming items were further analysed. SOA, Stimulus onset asynchrony.*

---

[4] Subsequent, the disconfirming items will be marked as item-type$_1$ and the confirming items as item-type$_2$.

occurrences each. Moreover, the 6 disconfirming stimulus classes are characterized by having different predecessors (K_1(t), K_1(#), T_1(k), T_1(#), $\#_1$(k), $\#_1$(t)). The corresponding confirming stimulus classes were labeled respectively.

Participants were seated in an electrically shielded and sound attenuated room. During the experiment the subjects were instructed to ignore all stimuli and watched a silent movie. Before and after the three blocks of passive listening the subjects had to identify the three stimulus classes by pressing a key corresponding to the subjectively heard phoneme at the beginning of the second syllable.

Data acquisition

The EEG (TMS international, Type Porti S/64) was recorded continuously and digitized with 512 Hz. We used an elastic cap (EASY cap) with 62 scalp electrodes at international 10-10 system locations (average reference) and 2 additional electrodes for controlling eye movements below both eyes (see Figure 1). The EEG data were band-filtered from 0.1 to 100 Hz. All impedances were kept below 5 kΩ. The continuous EEG was segmented in epochs from 500 ms prior to 1200 ms post stimulus onset. Using the BESA software package, experimental data was corrected for eye artefacts (Berg and Scherg, 1994) and artefact-flawed epochs were rejected by visual inspection or if epochs exceeded a maximum of 90 µV in amplitude or a gradient of >75 µV.

Data analysis

To analyze the induced spectral changes in GBA (the principle approach was the same as in study A; Hannemann et al., 2007) in the artifact free epochs from -400 ms to 1000 ms of the disconfirming and confirming items, a wavelet analysis using Morlet wavelets was performed. Then, time by frequency energy is averaged across single trials, allowing one to analyze non-phase-locked frequency components (for a description of the Morlet wavelets see General Methods). In order to achieve a good time and frequency resolution wavelets from 10 to 100 Hz in 2 Hz steps were computed. Next the raw wavelet-data were normalized by computing the relative power change for every time by frequency bin compared to the median of the according baseline (-200 to -100 ms prior stimulus onset).

To capture a wide range of cortical sources as well as maintaining a good signal to noise ratio, the mean spectral power of all disconfirming and confirming events was averaged over 6 electrode arrays with 6 electrodes each (Fig. 1). Concerning the lack of exact a priori knowledge of latencies and frequencies which might map process of the phonemic restoration in the gamma band, a similar approach as in study A using permutation tests (Blair and Karniski, 1993) was pursued to compare the differences in spectral power of disconfirming #-items with a K-item as predecessor minus the associated confirming item with the comparable difference having a T-item as predecessor. In the present study these tests were applied to each time-frequency bin from 280 to 1000 ms post stimulus onset for frequencies between 30 and 60 Hz. To make relatively sure that no time-frequency bins passed our criteria by chance, only contiguous bins for at least 30 ms per frequency band which showed a p-value $p<0.01$ (uncorrected) were taken into account for further consideration.

Finally a four-way repeated-measures ANOVA Predecessor (K-item vs. T-item) x Expectation (disconfirming vs. confirming) x Hemisphere (left vs. right) x Position (anterior, medial, posterior) was performed on the time-frequency clusters surviving the initial permutation tests to substantiate our findings. For all analyses involving the factor Position, we checked for the violations of the sphericity assumption using Mauchly’s criterion, and in case of violations report multivariate testing (using Wilks Lamba) instead. Post-hoc test were only applied to time frequency spots that passed the initial permutation tests. These statistical analyses principally comprised a two way repeated-measures ANOVA Predecessor x Expectation and the belonging t-tests to identify the direction of the predicted modulation in induced GBA.

To dissociate the “filling in” of expected lexical information from a pure phonological conflict depending on the specificity of phoneme representations between the particular predecessor and the pivotal disconfirming noise replaced item, the assessed induced GBA was compared with the MMN response (e.g. Naatanen, 2001; Pulvermuller and Shtyrov, 2006) which is sensitive to map phonological conflicts in passive oddball paradigms (Naatanen et al., 1997; Eulitz and Lahiri, 2004; Cornell et al., submitted). Thus, to ensure that the hypothesized induced GBA is not a mere by-product of a MMN elicited by deviant items (= disconfirming items) interrupting a sequence of repeated standard items, we analyzed the evoked potentials (re-referenced to linked mastoids) with a prestimulus baseline of 100 ms recorded at Fz. Again, we examined the mean amplitude in the latency range identified by the permutations test for the induced GBA using the factorial design as described above.

Further, to differentiate the induced brain activity from evoked brain activity in the gamma band range, we also calculated mean amplitudes of the evoked GBA in the same time by frequency windows as those for the induced GBA and analyzed them using the same factorial design. Finally we also analyzed the induced GBA in higher frequency ranges (76-86 Hz) which are known to reflect electromyographic (EMG) activity for facial and head muscles (Cacioppo J.T. et al., 1990) to rule out possible confounds of EMG artefacts (Pulvermuller et al., 1997).

Separate behavioural measures

In addition to the EEG study, two behavioural identification experiments were conducted to gain knowledge about the attentive processing of the noise-replaced items. Twelve subjects (seven female; mean age = 24.5 years, standard deviation [SD] = 3.5 years) participated in each of the experiments. They fulfilled the same criteria as the subjects of the EEG study. In the first experiment the subjects had to identify the stimulus-class for all noise-overlaid and noise-replaced items. Each item was presented six times using the same equipment as for the EEG study. The subjects had to subjectively judge as exact as possible which phoneme has been perceived and respond by pressing the corresponding key on a standard PC keyboard with their right hand.

The experimental design of the second behavioural experiment was made to mimic possible context effects which played a role in the EEG study. Therefore, one experimental trial contained 4 stimuli with the first three items belonging to one item class followed by a fourth item (= target) which could belong to the same or one of the other two item classes. Each of the nine stimulus combinations was presented 24 times which resulted in 216 trials overall. The within trial ISI was 500 ms as in the EEG experiment. After the presentation of the fourth item the subjects had to indicate by button press whether they perceive a K or a T-item at the fourth position as fast and accurately as possible. To analyze the processing of the noise-replaced items the reaction times (RT) were analyzed by means of a mixed-model ANOVA after cropping the lower and upper 10% percentile.

3.2.3 Results

Induced brain responses

Figure 8 shows the induced brain responses in the gamma band range averaged over the six electrode groups (Fig. 1). Depicted are the differences between disconfirming #-items over the respective confirming #-items with a precedent fully specified K-item (= $\#_1$(k) - $\#_2$(k)) compared to those differences with a precedent T-item (= $\#_1$(t) - $\#_2$(t)) where the critical consonant is underspecified for place of articulation. The time-frequency bins which fulfil the permutation test criteria are shown in colour. As Figure 8 depicts, only one extensive cluster over anterior to medial left hemispheric electrode sites passed the criteria and showed a remarkable difference in the 38 - 44 Hz range from around 430 to 490 ms post stimulus onset. The four way repeated-measures ANOVA in this latency range resulted in a significant Predecessor x Expectation x Hemisphere x Position interaction (F(1,15) =

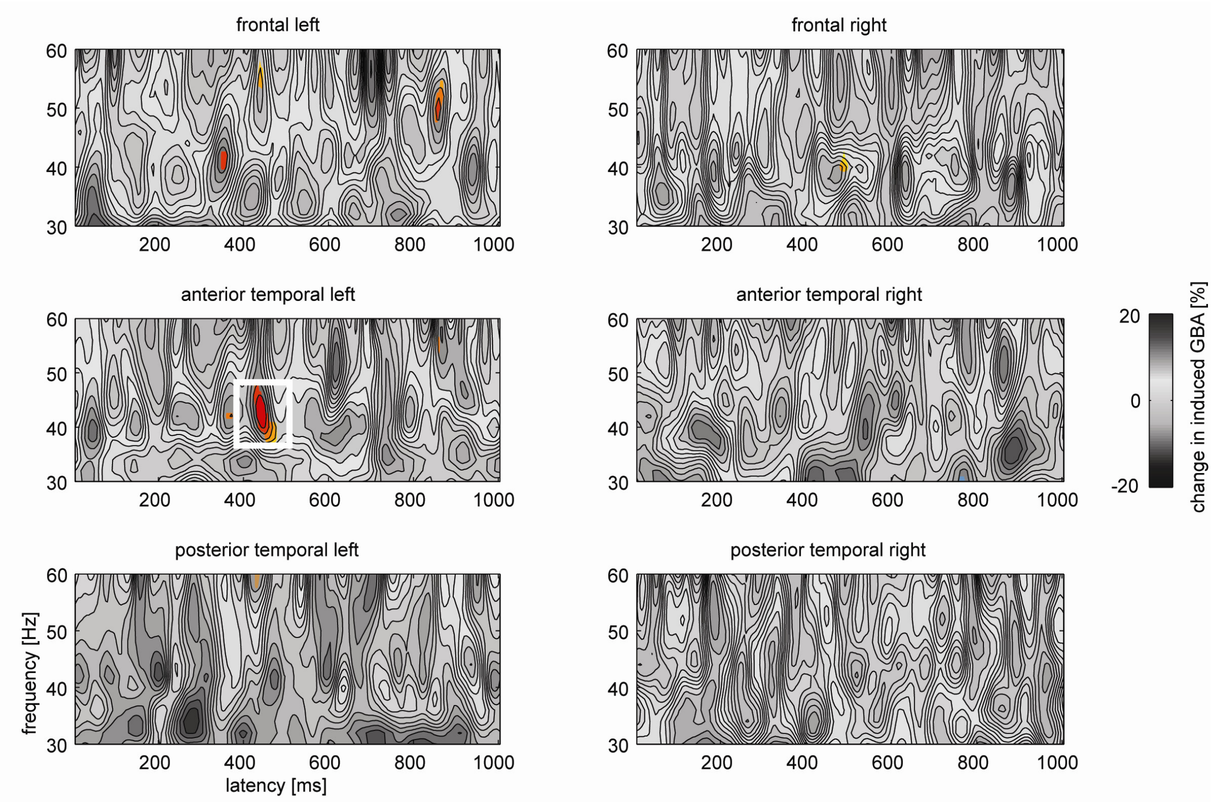


***Figure 8**. Grand mean time frequency (TF) plots of induced brain activity over six brain regions. Depicted are the differences of disconfirming #-items over the associated confirming #-items with a precedent K-item compared to those differences with a precedent T-item. Coloured areas reflect TF bins identified as different by a permutation tests with $p < 0.01$. The framed TF spot further fulfils the continuity criteria and thus constitutes the main finding of this study.*

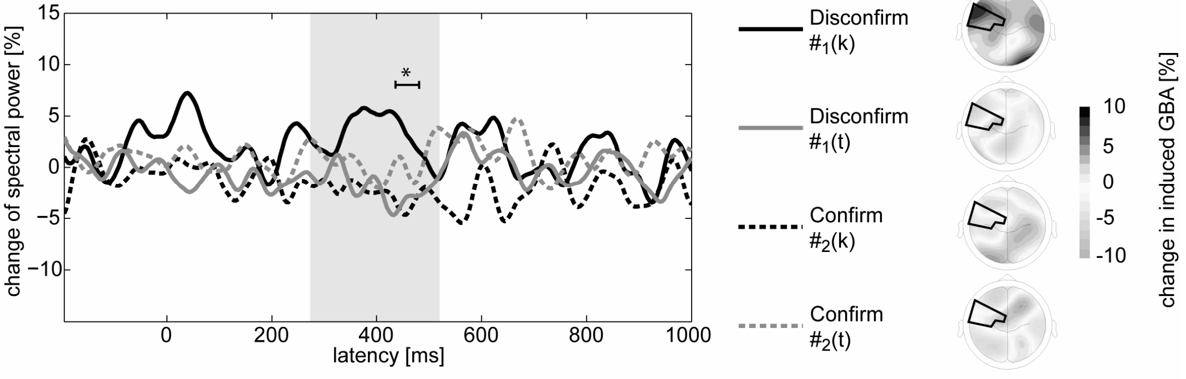
Position	**Hemisphere**	**Expectation created by predecessor**	**Predecessor K-item**		**Predecessor T-item**	
			%GBA change	(± SEM)	%GBA change	(± SEM)
Frontal	Left	Disconfirm	4.29	(2.38)	-3.48	(1.23)
		Confirm	0.29	(1.52)	-2.62	(2.46)
	Right	Disconfirm	2.41	(1.76)	-1.81	(1.70)
		Confirm	-1.81	(1.68)	1.09	(1.66)
Anterior temporal	Left	Disconfirm	**3.69**	**(2.19)**	**-3.59**	**(1.65)**
		Confirm	-3.73	(1.86)	0.26	(1.81)
	Right	Disconfirm	3.77	(3.28)	-1.78	(1.55)
		Confirm	1.82	(1.58)	-2.25	(1.62)
Posterior temporal	Left	Disconfirm	0.51	(1.86)	1.27	(2.00)
		Confirm	1.87	(2.01)	-2.66	(1.93)
	Right	Disconfirm	3.07	(2.36)	-1.49	(1.34)
		Confirm	0.87	(2.52)	-0.64	(1.36)

Table 2. *Mean spectral power for the 430 – 490 ms / 38 – 44 Hz range, averaged across six electrode sites and standard error of mean (SEM) in % change for the noise-replaced #-items. Figures in bold indicate the main result of study B.*

4.09, $p < 0.05$). This result confirms the findings of the permutation tests, in that the prominent modulation of the 38 - 44 Hz spectral power is mainly focused on left lateralized anterior to medial electrode sites (Fig. 8). The changes in induced GBA for the tested time frequency range are summarized in Table 2.

Further post-hoc comparisons were performed for the left anterior temporal spot. The time courses and scalp topographies of the induced changes in the 38 - 44 Hz range over this electrode group for the #-items is shown in the upper part of Figure 9. Only the disconfirming #-items following a K-item ($\#_1(k)$, solid black) with a fully specified critical consonant show a substantial increase compared to the other #-items in the latency range of noise replacement (indicated as grey box), especially in the latency range of 430 to 490 ms as identified by the permutation tests. For this time frequency spot, a two way repeated-measures ANOVA revealed a significant Predecessor x Expectation interaction ($F(1,15) = 18.39$, $p < 0.001$). Further post-hoc comparisons revealed a significant difference between the disconfirming #-items ($t(15) = 2.49$, $p < 0.05$) in favour of a larger value for those #-items which were preceded by a fully specified K-item. No differential modulation was found for the confirming #-items ($t(15) = 1.61$, $p > 0.1$). Additionally post-hoc tests showed significant differences between the disconfirming and the confirming #-items following K-items ($t(15) = 2.64$, $p < 0.05$) as well as T-items ($t(15) = 2.43$, $p < 0.05$). The mean values indicated a positive difference in spectral power if the K-items were in predecessor position and a negative difference if the T-items were in predecessor position. Despite the opposing directionality in the evolution of the spectral power for the #-items following

a) induced GBA in 38 - 44 Hz range



b) evoked GBA in 38 - 44 Hz range

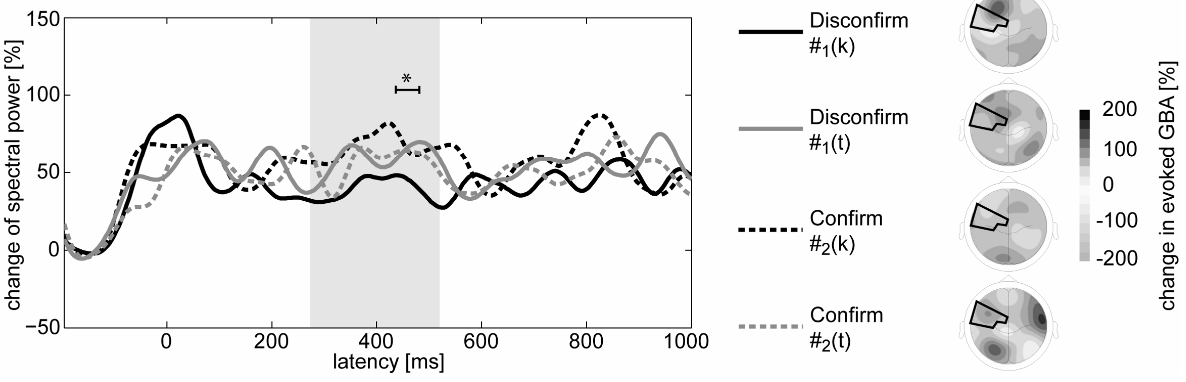


Figure 9. *Depiction of the gamma band activity in the 38 – 44 Hz range. Left: Comparison of the time courses of induced (upper panel) and evoked GBA (lower panel) over left anterior temporal electrode sites. Solid lines represent disconfirming, dashed lines confirming #-items. Black lines picture #-items following K-items and grey lines represent #-items with T-items as predecessor. The underlying grey box pictures the latency range noise-replacement. The starred range always depicts the time course identified by the permutation test. Right: Shown are the topographies of the GBA in the 430 – 490 ms latency range for each #-item respectively. The area of left anterior temporal electrodes sites (cf. Fig. 1) is outlined in black. Depicted latency range is the same as indicated by the starred time course on the left (identified by the permutation tests).*

K and T-items it is important to note, that the main modulation in induced GBA was observed for the disconfirming #-items whereas the confirming #-items did not differ. As Figure 9a indicates, the modulation in 38 – 44 Hz spectral power might last longer than the initial permutation test suggested. For the latency range from 350 to 490 ms the Predecessor x Expectation interaction was also significant ($F(1,15) = 12.25$, $p < 0.001$) with post-hoc t-tests showing significant larger values in spectral power for $\#_1(k)$ compared to $\#_1(t)$ items ($t(15) = 2.96$, $p < 0.01$) and significant differences between $\#_1(k)$ and $\#_2(k)$ items ($t(15) = 2.86$, $p < 0.05$). All other post-hoc analyses revealed no significant differences ($p > 0.2$) for this latency range.

Although the predictions of study B concerning the induced spectral changes were only specific to the “filling in” of expected lexical information as processing step to build up a percept of a phoneme (which are expected to appear first after onset of the noise replacement begins) the time course of the induced 38 – 44 Hz changes in Figure 9a points to another modulation around the onset of the #-items (-50 – 100 ms). However, the corresponding ANOVA showed neither a significant Predecessor x Position interaction ($F(1,15) = 1.87$, $p > 0.1$) nor any main effect (all $F < 1.5$, $p > 0.2$) for the 38 – 44 Hz range and reinforces therefore the non-result of the permutation tests for this latency range.

Analyses of evoked gamma band responses and control for possible EMG confounds

To ascertain that the present results indeed reflect modulations of induced brain activity, we post-hoc analyzed the evoked brain activity for the same time and frequency range. Figure 9 contrasts the time course of the induced and evoked spectral changes in the 38 – 44 Hz range for left anterior medial temporal electrode sites. As exemplified only the induced spectral changes showed a modulation on the disconfirming #-items with larger values for the $\#_1(k)$ items compared to $\#_1(t)$ items whereas the evoked spectral changes exposed no comparable modulation pattern. Statistical analyses analogous to the analyses for the induced brain activity revealed neither a four-way interaction ($F(1,15) < 1$, $p > 0.4$) nor any main effect or interaction for the left anterior temporal electrode sites (all $F < 1$).

Finally, to test for possible EMG artefacts which might be correlated with the induced result in the 38 – 44 Hz range, a four-way ANOVA testing the 76 – 86 Hz range (in which the peak of the spectral density function of muscular contamination could be expected; Cacioppo J.T. et al., 1990) yielded no comparable results for the latency range identified by the permutation tests, especially no Predecessor x Expectation x Hemisphere x Position interaction (F(1,15) = 1.84, p > 0.1).

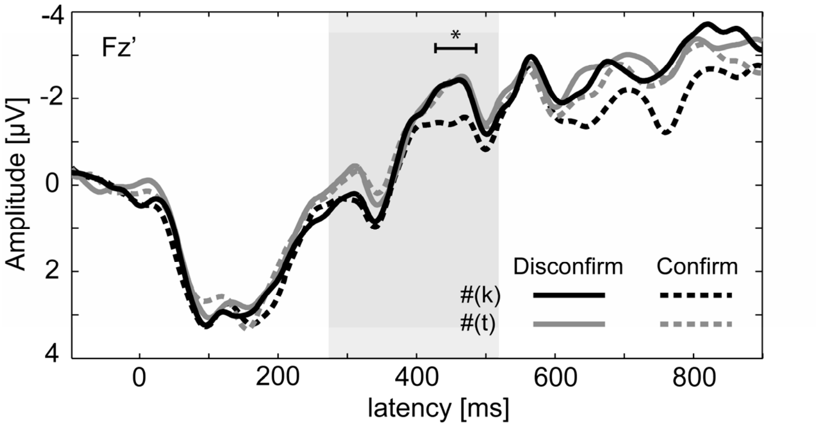


Figure 10. *Evoked potentials at Fz for #-items. Solid lines represent disconfirming and dashed lines confirming #-items. Those #-items following K-items are shown in black and #-items with T-item as predecessor are depicted in grey. The underlying grey box shows the latency range of noise-replacement. The starred range depicts the analyzed time course as identified by the permutation tests for the induced GBA.*

Differentiation of induced brain responses from ERP results at Fz

Figure 10 depicts the re-referenced evoked potentials (ERP) for the disconfirming and confirming #-items following either stimulus trains of fully specified K or underspecified T-items recorded at Fz. In the latency range identified with the permutation tests for the induced brain activity (430 – 490 ms) all #-items except the $\#_2$(k)-items show the same activity pattern. An ANOVA analogous to the induced brain responses showed only a barely significant main effect of Expectation (F(1,15) = 4.64, p < 0.048). The mean values of amplitude point to a stronger negativity for the disconfirming #-items (mean = -2.23 µV) compared to the confirming #-items (mean = -1.79 µV). Notably, statistical analyses revealed neither a Predecessor x Expectation interaction (F(1,15) = 2.44, p > 0.1) nor a main effect of Predecessor (F(1,15) = 1.91, p > 0.1) in this latency range.

		Rate of Keystroke [%]		
		K - key	T – key	Remaining keys
Item-class	K-Items	72.44	23.08	4.48
	T-items	18.18	79.22	2.60
	#-items	27.48	34.82	37.70

Table 3. *Mean identification rates for the K, T and #-items across 12 subjects of the first behavioural control experiment.*

Behavioural measures

As shown in Table 3, the results of the first identification experiment indicate that despite some miscomprehensions the noise-overlaid segments in the K-items and T-items were perceived significantly above chance level as /k/ (72.4 %) or /t/ (79.2 %) respectively. For the #-items the data revealed a tendency to perceive the noise-replaced segment as /t/ (34.8 %) or /k/ (27.5 %) rather than anything else.

The results of the second behavioural experiment are summed up in Table 4. With respect to the influence of the predecessor context on the perception of the #-items, there is a clear preference to interpret the #-items as /t/ perceptions independently of the context. This is reflected in a significant main effect of Percept ($F(1,11) = 17.24$, $p < 0.01$) for the projection rate. However, the reaction time data indicate an influence of the predecessor. Here, subjects go faster for the opposite percept relative to the predecessor. This pattern results in a significant Predecessor x Percept interaction ($F(1,457) = 7.99$, $p < 0.01$).

			K - Percept		T – Percept	
	pre-decessor	disconf. item	RT in [ms]	PR in [%]	RT in [ms]	PR in [%]
Item Sequence	K–Items	#-Item	1056.7	28.5	987.4	71.5
	T-Items	#-Item	991.3	26.9	1037.9	73.1
	#-Items	#-Item	1014.4	33.3	1031.7	66.7

Table 4. *Reaction times (RT) and the projection rate (PR) of #-items onto /k/ and /t/ perceptions in different contexts are summarized for the second behavioural experiment.*

3.2.4 Discussion

To gain a better understanding of how the brain copes with acoustically ambivalent situations the present study was set out to shed light on the brain mechanisms underlying the repairing of fragmentary speech information. Particularly the study investigated the role of lexical specification of phonological details in the mental lexicon and its impact on the phonemic restoration illusion. In order to prevent influences of attention or decision making processes on the phonemic restoration the illusion was investigated by means of a passive oddball paradigm in which the subjects were instructed to ignore the auditory stimuli. To monitor the processing of ambivalent sensory input under the influence of differential top-down mediated expectations of phonemic features study B examined the induced GBA. If the fine structure of phonological information in the mental lexicon may play a significant role in the phonemic restoration, a differential modulation in the induced GBA depending on the specificity of the place of articulation of the phoneme to be restored was hypothesized. The results for the left anterior electrode sites clearly support this assumption. In the latency range of the to-be-expected phoneme for the disconfirming #-items larger values of induced GBA were observed if the expected phoneme was specified for the feature place of articulation (K-item) compared to the underspecified expectation (T-item). These larger values of induced GBA were most pronounced between 430 and 490 ms in the 38 – 44 Hz range. Importantly, there was no differential modulation in induced GBA for the confirming #-items. As Figure 9 illustrates the evoked GBA showed no comparable effects, neither for the disconfirming nor the confirming #-items.

As outlined in study A (Hannemann et al., 2007) a modulation in induced GBA over left anterior temporal electrode sites can be interpreted as a correlate for a match of bottom-up sensory input with lexical expectations which result in resonant neural dynamics (Grossberg, 1999; Grossberg, 2003). The present results for the differential modulation of induced GBA in the restoration illusion, depending on the underlying specificity of phonological features of the expected phoneme, nicely fit in this left anterior temporal framework which merges top-down expectations with sensory input into a perceivable auditory object. According to the featurally underspecified lexicon theory (Lahiri and Reetz, 2002) the critical phoneme in the K-items possesses a full featural specification for the [dorsal] place of articulation while the [coronal] place of articulation for T-items is underspecified. Thus, while the repeatedly presented K-items establish an expectation of a specified place of articulation in the critical phoneme, the T-item cannot build up such specific expectations based on specified featural information in the mental lexicon. The higher induced GBA values for the disconfirming #-items following K-items might therefore reflect the “filling in” or rather merging of the anticipated lexically specified information in the noise replaced part of disconfirming #-items. However, following T-items the expectation does not contain a specified place of articulation and therefore disconfirming #-items induced a significantly smaller GBA. Due to the repeated presentation of #-items before the confirming #-items the perceptual system has built up a non-expectation of phonemes (or an expectation of the underspecified place of articulation as the extended analyses in the noise-replaced latency range of the #-items suggest) which is reflected in the similar values for induced GBA.

To carry over a generated expectation from one item to another and compare it to the actual item short term memory processes described by means of modulated GBA (e.g. Kaiser et al., 2003; Kaiser et al., 2006b; Leiberg et al., 2006; Lenz et al., 2007a) are relevant. Similarly, the present results were observed over anterior temporal brain regions as the topographies in Figure 9 show. Therefore the present modulation of induced GBA is interpreted as correlate for a pre-attentive processing mechanism which accomplishes a “filling in” of lexical expectations in noise-replaced time segments by merging anticipated phonemic information with deficient sensory input. This crucial step towards the formation of an auditory object should later allow a successful perception of speech, particularly because possible effects of co-articulation can be ruled out through cross-splicing the items carefully. Consequently, this pre-attentive modulation in induced GBA is further evidence for lexical top-down support in the phonemic restoration (Samuel, 1997) and generally in the perception of speech in difficult auditory environments.

It is well established that signatures in GBA can differentiate between words and pseudowords (Lutzenberger et al., 1994; Eulitz et al., 1996). As all #-items were acoustically identical (and in a strict sense all pseudowords) this known difference should maximally lead to a main effect of Expectation and does therefore not explain the present results. Thus models favoring strictly bottom-up processes in speech perception (e.g. Norris et al., 2000) or only top-down contribution for perceptual learning (e.g. Norris et al., 2003) can hardly account for the observed differential modulation in induced GBA, especially because all #-items were physically equal and there is no post-perceptual decision making process which might have influenced the GBA.

According to Pulvermuller et al. (Pulvermuller et al., 1996) activity in higher frequency bands contains information about semantic features of words, i.e. it shows differential topographies between verbs and nouns in a lexical decision task. Recently an intracerebral EEG study observed modulations in evoked GBA in a visual semantic decision task (Mainy et al., 2007). Following this argumentation it might be possible, that the observed modulation in induced GBA in the present study is caused by different semantic instead of phonological expectations. As both words which create the expectation for the disconfirming #-items are nouns, were matched for frequency and the observed effect cover different frequency bands this interpretation seems rather unlikely. Nevertheless it can’t be absolutely ruled out that the larger value of induced GBA for $\#_1$(k)-items compared to $\#_1$(t)-items is at least partly due to a differential semantic expectation.

With respect to the findings of Eulitz and Lahiri (2004) the suspending of repeated presentation of fully specified with underspecified items lead to larger phonological conflicts mapped in differential MMNs than vice versa. Thus, if the present modulation in induced GBA in favor to the $\#_1$(k)-items is due to that kind of phonological conflict the MMN should also show a differentiating pattern between the disconfirming expectations of specified and underspecified items. As we found only a general difference between disconfirming and confirming #-items which was independent from the predecessor context (Fig. 10) the present results cannot be explained by variable strength of phonological conflicts, at least in the present latency range of 430 – 490 ms.

The results of the behavioural experiments support and extend our interpretation of the observed gamma band modulation during the processing of the #-items. When attending the stimuli, the pattern of results is different compared to the pre-attentive processing of #-items. Without context, as in the first behavioural experiment, the subjects showed a preference towards perceiving a /t/ over a /k/ and all other possible phonemes. This identification bias toward /t/ was replicated for the projection rates in the second behavioural experiment. Due to the lack of alternatives in this choice task, this bias was even more pronounced. This bias can be interpreted in two ways: (i) The [coronal] place of articulation is regarded as the default place of articulation by phonolgists (cf. Paradis and Prunet, 1991). Therefore subjects showed a preference towards perceiving a /t/ in case of an ambiguous acoustic signal. (ii) It might be also due to the spectral characteristics of the noise replacing the critical consonant, which is spectrally slightly more similar to a /t/ compared to a /k/ (cf. Liberman et al., 1952). Interestingly, reaction time data of the second behavioural experiment indicated context effects. When subjects decided that the actual #-item was the same as the predecessors, the reaction times to these #-items was significantly longer compared to the inexpedient response. The longer RT seems to indicate a more complex decision and evaluation process which is required to align the anticipated phonemic information and the sensory input. Under attentive processing conditions, this RT effect is independent of the specification of featural information in the mental lexicon.

According to that, the modulation in induced GBA in favour of the $\#_1$(k)-items and the prolonged RT enlighten differential aspects of the phonemic restoration illusion. Both describe the matching processes of deficient sensory input and anticipated phonemic information. But, as the behavioural data is generally influenced by external factors, i.e. task formulations etc., the pre-attentive EEG data is free of such influences and thus yield additional insights on the influence of the fine structure of the mental lexicon on this matching process.

Summary

In sum, the current study evinces for the first time a direct correlate for a top-down modulated “filling in” in the phonemic restoration illusion without relying on redundant sentential information. The present induced brain responses of study B again reveal clear evidence for a left lateralized functional network in matching expected lexical information with sketchy sensory input to form a coherent auditory object (Griffiths and Warren, 2004). Further, they demonstrate the influence of the fine structure in the mental lexicon on top-down modulated speech perception processes and are in line with current cortical models of auditory word recognition (e.g. Scott and Johnsrude, 2003). Moreover, the delineated alignment of lexical expectancies with sensory input is in accordance with recent ideas that speech perception is facilitated by internal forward models (e.g. Poeppel et al., 2008). Thus it serves as prerequisite for speech and more generally for conscious object perception (Tallon-Baudry and Bertrand, 1999). Finally the current results of study B experimentally show that the human ability to comprehend speech even under much compromised conditions (i.e. restoring missing phonemes) relies on the interaction of lexical expectancies (i.e. top-down) and the acoustical input and that these interactions can be examined by means of induced GBA.

3.2.5 Appendix of study B: Gating experiment

In order to ensure that possible differences in the articulation of the first syllable or systematic co-articulation of /t/ and /k/ on the phoneme /l/ in the *Falke* (K) and *Falte* (T) items do not influence the results of the following EEG studies, a cross-splicing of the first syllable of the words is necessary. To determine the latency for the cross-splicing such that any influences of co-articulation can be inhibited a separate gating experiment (Marslen-Wilson and Warren, 1994; see also Grosjean, 1996, for an overview over this method) was performed.

Altogether twenty volunteers (nine female; mean age = 24.7 years, standard deviation [SD] = 3.3 years) without neurological or otological diseases participated in this gating experiment. All participants gave their written consent and received class credits or a small financial bonus.

As specified in the method section of study B, the experimental items where based on the natural recordings of the minimal pair of the German nouns *Falte* and *Falke*. The gated stimuli where generated by cutting the original K and T-items into fragments of 240 – 500 ms length, beginning at the original word's onset each. In a sound attenuated room, the items were presented as a sequence of increasingly large fragments, using increments of 20 ms. The final 2 ms of each fragment were progressively attenuated, so that there were no "clicks" at offset. Each gated stimuli was presented 16 times. After every stimulus presentation the subjects had to indicate whether that stimulus represents a K or a T-item by pressing a corresponding key on a standard PC-keyboard. If there were not sure about the identity of the fragment they were required to guess whether the item represents a K or a T-item. The average responses over all subjects to the gated stimuli were illustrated in Figure 11.

By using a binomial test, which is an exact test of the statistical significance of deviations from an expected distribution of observations of two alternatives, each fragment was analyzed whether the single subject designated it significantly as K or T-item. Thus, to be significantly identified, a fragment has to be identified at least 12 times out of the 16 presentations ($p < 0.04$). If a subject identified at least three K or T-item fragments of successive length significantly, the fragment-length of the first identified fragment was qualified as so-called recognition point of the corresponding

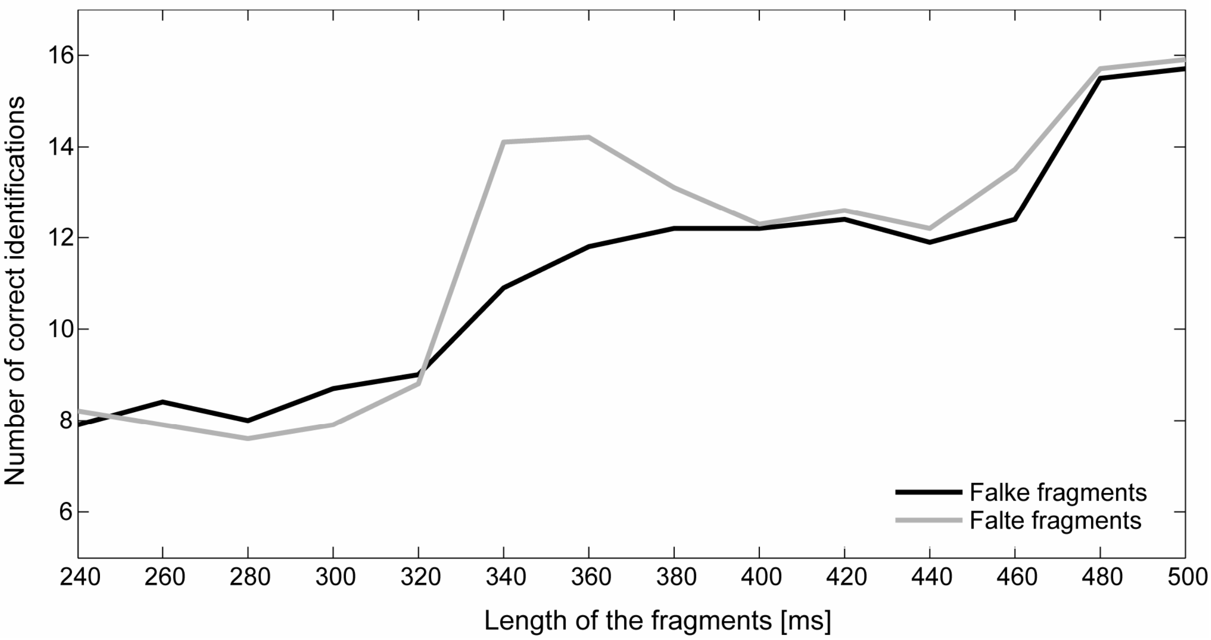


Figure 11. *Illustration of the subject's correct responses to the gated K (black) and T-items (grey) of the gating experiment (max. 16 correct responses for each fragment length were possible).*

item for that subject (W. Nagl, personal communication). Subjecting these latencies to a pared sample t-test, the statistics revealed a significant difference for the identification of K and T-items ($t(19) = 2.16$, $p < 0.05$) with an earlier recognition point for T-items (mean = 370 ms) compared to the K-items (mean = 395 ms).

Concluding on these results, the latency for cross-splicing was chosen 100 ms before the average recognition point of both items. Thus, to prevent any systematic influences of articulation of the first syllable and of the co-articulation of /t/ and /k/ on the phoneme /l/ the K and T-items were cross-spliced at 280 ms post stimulus onset which results in two instances of *Falke* and *Falte* each.

3.3 Study C: Constraints of matching top-down knowledge with deficient sensory input in speech perception.[5]

Models of speech perception relying on Bayesian statistics claim that the successful recognition of speech depends on prior knowledge and actual acoustic information, at least in noisy environmental conditions. Study B examined unification processes in speech perception, i.e. in the phonemic restoration illusion, and suggests that the repairing of defective speech signal is depicted in the gamma band activity.

The present EEG study was set out to expand the knowledge of the general conditions in which repairing of deficient speech signals is possible without relying on broader contexts. Therefore, as compared to study B, a more explicit acoustic difference between expected phoneme and the obliterated speech segment was introduced. Additionally, the specificity of lexical top-down information (specified vs. underspecified phonological information) was modulated and the subject's pre-attentive MMN and GBA responses to obliterated speech segments in German nouns were compared. While the GBA responses evinced no differential pattern, the MMN responses revealed a clear prediction error if the expected phonemic information was not present in the incoming stimulus. Further, the present data corroborate the hypothesis of sparse representations in the mental lexicon (Lahiri and Reetz, 2002) by showing for the first time a differential MMN for word-medial phonological conflicts. Taken together, the current results delineate the detection of unpredicted sensory input from the alignment with pre-existing lexical knowledge by means of differential electrophysiological measures in the perception of speech.

3.3.1 Background

In everyday listening deviations in the speech signal due to mispronunciations as well as interruptions by extraneous sounds cause a big challenge for the speech processing system. A highly useful feature of the speech processing system is its ability to detect and infer unpredicted changes in the sensory input as prerequisite for learning about the environment (Friston, 2005). On the other hand it has been shown that missing acoustic information can be restored by using top-down lexical

---

[5] This section is prepared to be submitted to a peer-reviewed journal (co-authored by C. Eulitz). The author wishes to thank A. Bobrov and C. Massau for their assistance in data collection and artifact rejection.

knowledge (Samuel, 1997; Samuel, 2001; Hannemann et al., 2007; Hannemann and Eulitz, submitted). However, it is not always useful to repair every deviance in the acoustic signal. What are the general conditions that allow for a repair and what conditions prevent it? How does the fine structure of the mental lexicon influence these processes?

Recent models of speech perception based on Bayesian statistics (Friston, 2005; Norris and McQueen, 2008) account for the inherent ambiguity in the input signal due to the noisy environment by assuming that the listener's knowledge of phonemes is represented as likelihood functions. For example, each phoneme might be described in terms of Gaussian distributions over several dimensions, each characterizing a phonemic feature. In this view, word recognition is optimal in the face of ambiguous input when available perceptual evidence is combined with knowledge of the prior probabilities of words. All the more the speech signal is sketchy, the more important the a-priory knowledge becomes in guiding speech perception (Norris and McQueen, 2008).

With respect to the assumption of likelihood functions representing possible instances of a phoneme, the integration of lexical top-down influences and sensory information resulting in resonant states (e.g. Grossberg, 2003) should take place if the characteristics of the bottom-up signal are covered at least partly by the distribution of the expected phonemic features. On the contrary, if there is not sufficient overlap, an error signal indicating the difference between the observed and the predicted input should be used to refine the lexical expectation in order to optimize the speech perception process. According to the idea of predictive coding (Friston, 2005), especial such error signal is generated reflecting any mismatch between the predicted signal and the actual stimulus-generated activity. Corresponding to that account event related potentials, i.e. the MMN, can be understood as transient expressions of this so-called prediction error.

Studies investigating the handling of missing segmental information in speech perception, the so-called phonemic restoration illusion, showed that listeners hear spoken words as intact, even though parts of them have been replaced by an extraneous sound (Warren, 1970; Samuel, 1981a). According to the Grossberg (1999; 2003), the phonemic restoration illusion arises as result of matching lexical top-down expectations with the sketchy sensory input. This proposed top-down lexical influence on phonemic processing (Samuel, 1997; Samuel, 2001) was analyzed in study B by means of modulations in the induced GBA and was interpreted as signature of speech integration and unification processes at the level of working memory (Hagoort, 2005; Bastiaansen and Hagoort, 2006) or rather as a signature of (auditory) object recognition in general (Tallon-Baudry and Bertrand, 1999). The reported gamma band modulation in study B was situated over left anterior temporal electrode sites and hence is similar distributed as the proposed correlate for a match of bottom-up sensory input with lexical expectations (see study A; Hannemann et al., 2007; Hannemann and Eulitz, submitted). Thus, the induced GBA has been shown twice to be a valid measure of lexical guided speech perception and is suitable to describe the restoration of missing speech segments. Subsuming the findings so far, it is hypothesized, that a successful phonemic restoration should be accompanied by a modulation in the induced gamma band and a failure in restoration in a pronounced MMN.

Independently of the acoustic quality of the sensory input, the recognition of speech is a process which requires matching the spoken input with abstract mental representations associated with word candidates. These lexical representations possess abstract units, the phonemes, standing for the word’s subcomponents which are assumed to mediate this mapping. Thus, if one aims to investigate the repair of obliterated speech segments one has to deal with these abstract units, i.e. the phonemes and their features. This task becomes more complicated because the realization of a word or a single phoneme varies with the talker, speech rate, context and several other factors. Unfortunately, the question about the exact nature of that mental representation of language sounds is far from being resolved. However, as not all possible features of phonemes that can be extracted from the acoustic signal are required for recognition, it is assumed that all predictable and nondistinctive information might not necessarily be specified in the mental lexicon (Lahiri and Marslen-Wilson, 1991; Lahiri and Reetz, 2002). This is a key assumption of a featurally underspecified lexicon model (FUL; Lahiri and Reetz, 2002), which serves to explain how the speech perception system economically handles acoustic variance of single words across speakers and contexts beyond the possibility of merely storing all variance (McClelland and Elman, 1986; Bybee, 2001).

The present study was set out to test whether a more explicit difference between acoustical and anticipated phonemic information might prevent the integration of these into one coherent speech percept. Therefore the present study adapted the same experimental design applied in study B but used normal spoken words containing no manipulation instead of the noise-overlaid words and contrasted them with the noise-replaced stimuli. If the more distinct acoustic discrepancy might prevent the repair of the missing speech segments, the noise-replaced items should be accompanied by a distinct mismatch negativity (MMN) depicting the prediction error. The MMN, an automatic change detection response in the brain (Naatanen, 2001; Pulvermuller and Shtyrov, 2006), has been shown to be sensitive amongst others to map phonological conflicts in passive oddball paradigms (Naatanen et al., 1997; Eulitz and Lahiri, 2004; Cornell et al., submitted). With respect to the assumption of the FUL-model (Lahiri and Reetz, 2002), the MMN amplitude should be larger when prediction of a full specification is violated (Eulitz and Lahiri, 2004). On the contrary, if the phonemic restoration take place regardless of the more explicit acoustic difference, we again expected the induced GBA to be larger for the noise-replaced items in case of lexically specified compared to underspecified information in the precursor over left anterior temporal electrode sites as in study B.

To examine the above mentioned dissociation, the phonemic restoration was guided by an experience-setting context in a roving standard oddball design (similar to study B; Cowan et al., 1993; Baldeweg et al., 2004) which was established by one of two words being a minimal pair which preceded a target word containing a noise-replaced segment. All experimental stimuli in this study were derived from natural recordings of the minimal pair of the German nouns *Falte* (= fold; named T-item in the following sections) and *Falke* (= hawk; named K-item). These words setting the context differed in just one segment in no more than one phonological feature characterizing the place of articulation which can be either retrieved from stored knowledge in the mental lexicon (i.e. the [dorsal] place which is called specified in the K-items) or derived from redundant information (i.e. the underspecified [coronal] place in the T-items; Lahiri and Reetz, 2002). Because of the clear intelligibility of the experimental items, also a differential MMN reflecting the phonological conflict between the T and K-items was expected if they were presented consecutively. Here a conflict should be observable if the K-item is followed by a T-item because the extracted feature [coronal] conflicts with the anticipated feature [dorsal]. Because of the underspecification of [coronal] no feature for place of articulation is expected and no conflict should be observable if [dorsal] is extracted from the signal (Lahiri and Reetz, 2002; Eulitz and Lahiri, 2004) which is expected to result in a statistical interaction for the evoked potentials. Summing up the hypotheses, either a modulation in the induced GBA, if integration of expected and deficient acoustic input takes place, or a MMN, if the distinct acoustic discrepancy between the both prevents that integration, in the noise-replaced items was expected.

3.3.2 Materials and Methods

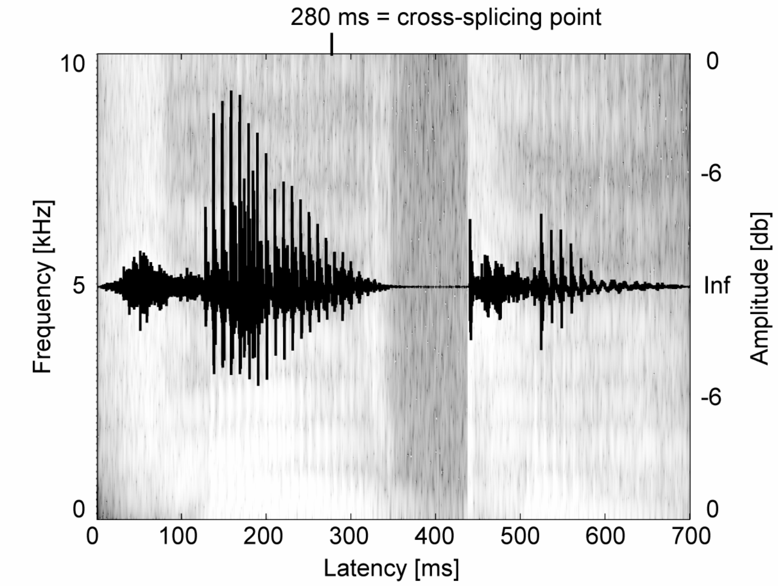
Subjects

Twenty-three right-handed healthy monolingual German-speaking volunteers without neurological or otological diseases participated in this experiment. Because of bad signal-to-noise ratio three subjects had to be excluded and all further analyses were performed for 20 subjects (eleven female; mean age = 23.1 years, standard deviation [SD] = 5.7 years). All participants gave their written consent and received class credits or a small financial bonus.

Stimuli

In this study the experimental stimuli were derived from natural recordings of the minimal pair of the German nouns *Falte* (= fold) and *Falke* (= hawk), which were digitized with 44.1 kHz at 16 bit. Both nouns were controlled for frequency in written (*Falke* = 36, *Falte* = 29, (Celex; Baayen et al., 1993)) and spoken German (*Falke* = 0, *Falte* =1 (Datenbank Gesprochenes Deutsch des Instituts für deutsche Sprache)). For experimental purposes a pair of nouns with maximal similarity was chosen and equalized further in envelope and second syllable onset using the software package Adobe Audition. Additionally, the acoustic difference between both nouns was minimized by cross-splicing the first and second syllables with each other which results in two instances of *Falke* and *Falte*. The latency for cross-splicing was chosen such that no co-articulation of /t/ and /k/ on the phoneme /l/ occurred. This latency was identified at 280 ms post stimulus onset in a separate gating test (see Appendix of study B). To create the noise-replaced items the speech-correlated noise technique (Schroeder, 1968), which flips the sign of half of the sampling points chosen at random, was used to create a noise from 280 to 520 ms post stimulus onset that maintains the amplitude of the envelope original but has a flattened spectrum. Figure 12 illustrates the oscillogram and spectrogram for one exemplar of (a) original and (b) noise-replaced items. This procedure resulted in three stimulus classes for the experiment: each two original *Falte* (T) and *Falke* (K) items and four noise-replaced ambiguous (#) items. According to the claim of a featurally underspecified mental lexicon (Lahiri and Reetz, 2002) the critical phoneme in the T-items is underspecified for the [coronal] place of articulation while the featural information for K-items with the [dorsal] place of articulation is fully specified. Finally, all stimuli were normalized for peak amplitude and presented in comfortable loudness (approx. 50 dB SPL) via headphones (Sennheiser PMX 60).

a) T-item (K-item) = original Falte (Falke)



b) #-item = noise-replaced Falte (Falke)

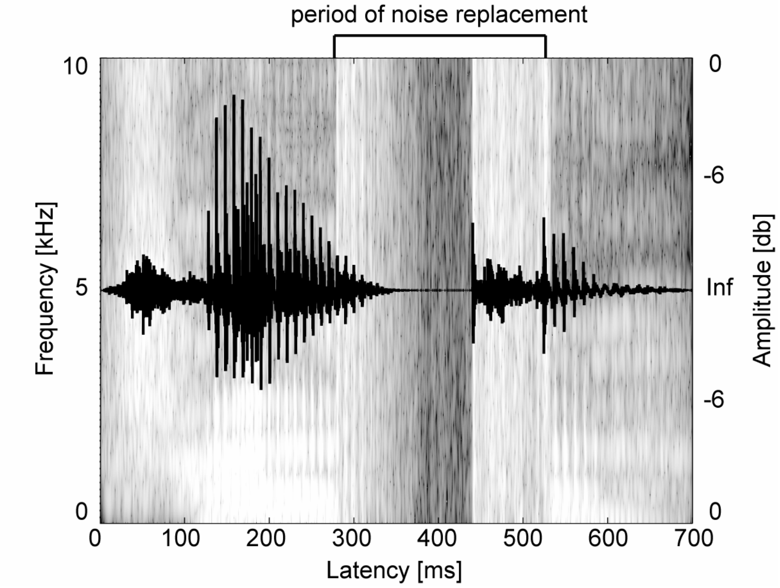


Figure 12. *Stimulus characteristics. Exemplarily the oscillogram and the spectrogram for an original T-item (a) and a derived noise-replaced #-item (b) are shown.*

Acoustic Stimulation

The present study was similar in design as study B and consisted of three experimental blocks of approximately 28 minutes comprising of permanently changing stimulus trains (see study B; Fig. 7). One stimulus train consisted of a varying number (from 3 to 8) of T-, K- and #-items. Inside each train of this so called roving standard oddball paradigm (Cowan et al., 1993; Baldeweg et al., 2004) all items belonged to one stimulus class. Within each train we picked out for further analyses (i) each first item to be the **disconfirming** (or deviant) item with respect to the previous train and (ii) every third item to serve as the so called **confirming** (or standard) stimulus.[6] The number of items per stimulus train as well as the stimulus class of each following train varied randomly throughout all three blocks. Overall, this yielded in 12 different conditions (6 disconfirming and 6 confirming) with 120 occurrences each. Moreover, the 6 disconfirming stimulus classes are characterized by having different predecessors (K_1(t), K_1(#), T_1(k), T_1(#), $\#_1$(k), $\#_1$(t)). The corresponding confirming stimulus classes were named respectively. All participants were seated in a sound attenuated and electrically shielded room. During the experiment the subjects were instructed to ignore all stimuli and watched a silent movie. Before and after the three blocks of passive listening the subjects had to identify the three stimulus classes by pressing a key corresponding to the subjectively heard phoneme at the beginning of the second syllable.

Data acquisition

The electroencephalogram (EEG) recordings were made using a 64-channel system (TMS international, Type Porti S/64) was recorded continuously and digitized with 512 Hz. We used an elastic cap (EASY cap) with 62 scalp electrodes at international 10-10 system locations (average reference) and 2 additional electrodes for controlling eye movements beneath both eyes (see Figure 1). The EEG data were band-filtered from 0.1 to 100 Hz. All impedances were kept below 5 kΩ. The continuous EEG was segmented in epochs from 500 ms prior to 1200 ms post stimulus onset. Using the BESA software package, the experimental data was corrected for eye artefacts (Berg and Scherg, 1994) and artefact-flawed epochs were

---

[6] Subsequent, the disconfirming items will be marked as item-type$_1$ and the confirming items as item-type$_2$.

rejected by visual inspection or if epochs exceeded a maximum of 90 µV in amplitude or a gradient of >75 µV.

Data analysis

To analyze the induced spectral changes in the gamma range in the artefact free epochs of the disconfirming and confirming items, a wavelet analysis identical to that of study B using Morlet wavelets was performed (see General Methods for a description of the Morlet wavelets and Methods section of study B for further details). To test for the integration of anticipated phonemic content and acoustic input the analyses of the induced GBA was limited to left anterior temporal electrode sites which have been shown to depict such unification processes (study A; Hannemann et al., 2007; see also study B). In study B a modulation in the 38 - 44 Hz range from around 430 to 490 ms post stimulus onset illustrating the “filling in” of anticipated information in the phonemic restoration was reported. Therefore the time spectral data in the present study was analyzed in the same latency and frequency range.

Finally a two-way repeated-measures ANOVA Predecessor (T-item vs. K-item) x Expectation (disconfirming vs. confirming) was performed to substantiate the experimental findings. Post-hoc test involving t-tests to identify the direction of the predicted modulation in induced GBA were only applied if the ANOVA reveals a significant interaction.

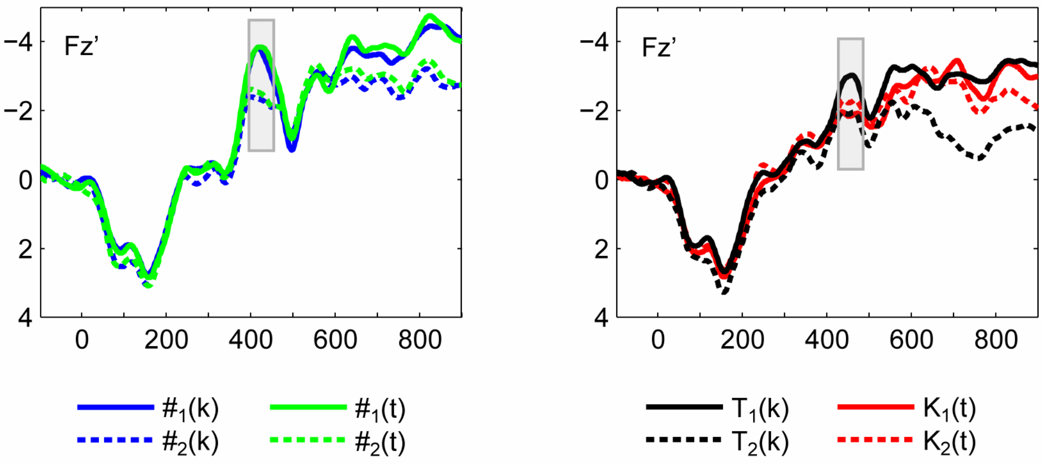
Epochs for the analysis of the ERPs were derived by averaging the same single epochs of the continuous EEG recording after applying a 30 Hz low-pass filter. These epochs lasted from -200 to 1000 ms post stimulus onset. The interval from -100 ms to stimulus onset served as baseline. With respect to the identified latencies of the spectral analyses the ERP amplitudes for the corresponding conditions at Fz position (re-referenced against linked mastoids) were measured as the mean amplitude across 60 ms centred at the mean MMN latency of the grand-average waveform in the latency range from 430 to 490 ms. As a more general amplitude measure the RMS amplitude across all electrodes calculated for the same latency range was used. These parameters were subjected to two-way repeated-measures ANOVAs with the same factorial design described above for the induced GBA.

3.3.3 Results

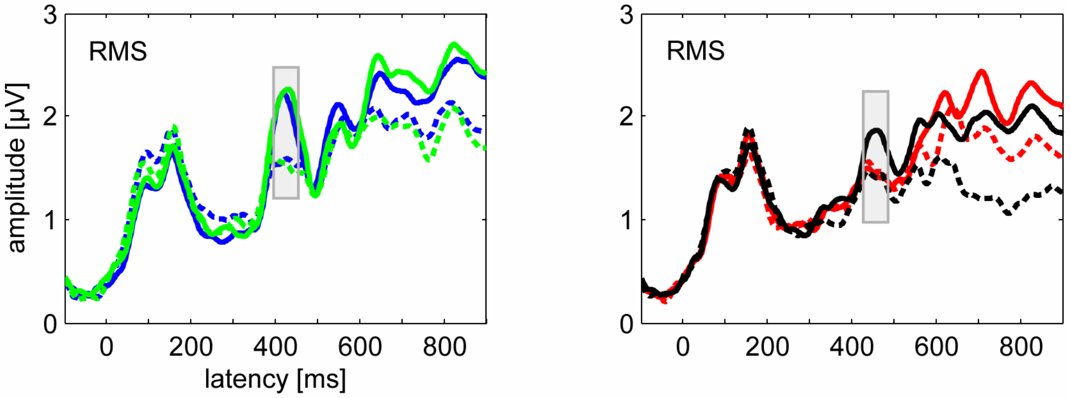
As shown in Figure 13a on the left side, the evoked responses of the disconfirming #-items recorded at Fz show a clear negative deflection in the latency range which was identified for depicting the “filling in” of anticipated specified lexical information in the phonemic restoration illusion (430 – 490 ms; study B) with a negative maximum at 430 ms. On the contrary, the ERPs of the confirming #-items showed no comparable pattern in that latency range. The same pattern can be seen in the RMS amplitudes, calculated across all electrode sites (Fig. 13b). Topographical information about the difference waves of disconfirming minus confirming items is displayed on the lower left side in Figure 13c. The amplitude maps show a typical MMN topography for both conditions (Naatanen, 2001). Because of the similar maps across conditions all analyses were restricted to Fz’ and to the RMS as more general measure for the strength of activation. The grey boxes indicate the time windows for calculating the mean amplitude. Subjecting the mean amplitudes to a two-way repeated measures ANOVA, the statistics revealed a significant main effect for Expectation ($F(1,19) = 15.99$, $p < 0.01$). The difference is due to a stronger negativity for the disconfirming #-items (mean = -3.52 µV) compared to the confirming #-items (mean = -2.32 µV). Notably, statistical analyses revealed neither a Predecessor x Expectation interaction nor a main effect of Predecessor (all $F(1,19) < 1$) in this latency range. The RMS amplitude confirms these results with a significant main effect of Expectation ($F(1,19) = 19.77$, $p < 0.001$) due to a higher RMS for disconfirming (mean = 2.10 µV) compared to confirming #-items (mean = 1.52 µV). Again, neither an interaction nor a main effect of Predecessor (all $F(1,19) < 1$) was revealed. All mean and RMS values for the experimental conditions are summarized in Table 5. With respect to the peak latencies of the MMN, there were no differences between $\#_1$(k) and $\#_1$(t) items ($t(19) = 1.08$, $p > 0.2$).

On the right side of Figure 13a the evoked potentials of the T and K-items with the respectively opponent predecessor recorded at Fz are shown. Around 460 ms the disconfirming T_1(k)-items are more negative compared to the confirming T_2(k)-items and all K-items which depict a pronounced MMN for the T(k) but not the K(t) items. The same pattern of results is also evident for the RMS values and reflected in the topographies of the difference waves (see Figure 13b and 13c). As the K_1(t) items exhibit a slightly more positive deflection compared to the K_2(t) items the topography

a) Event related Potentials recorded at Fz’



b) RMS across all electrodes



c) Topographies of difference waves (disconfirming - confirming item)

Figure 13. *ERPs recorded at Fz’ (upper part), the RMS waveforms (middle part) and the corresponding topographies for the difference waves (MMN, lower part) are shown. The disconfirming $\#_1$-items succeeding T (solid green) and K-items (solid blue) are depicted on the left side and on the right side the disconfirming T_1-items following K-items (solid black) and K_1-items following T-items (solid red) are shown. The corresponding confirming items are depicted in dashed lines. The superimposed grey rectangles illustrate the time windows for calculating the mean MMN and RMS amplitudes as well as the statistics. The topographies for the difference waves (disconfirming minus associated confirming items) in the lower part of the figure are shown as spline interpolated amplitude maps (average reference) with a contour step of 0.2 µV. Blue lines indicate negative potentials, red lines positive potential*

Item	Mean Fz amplitude ± SEM (µV)	Mean RMS amplitude ± SEM (µV)
K_1(t)	- 1.88 ± 0.74	1.43 ± 0.31
K_2(t)	- 2.20 ± 0.84	1.47 ± 0.43
T_1(k)	**- 2.77 ± 0.97**	**1.78 ± 0.41**
T_2(k)	- 1.80 ± 0.77	1.40 ± 0.35
$\#_1$(k)	**- 3.43 ± 1.00**	**2.04 ± 0.49**
$\#_2$(k)	- 2.26 ± 0.98	1.55 ± 0.41
$\#_1$(t)	**- 3.62 ± 0.63**	**2.15 ± 0.35**
$\#_2$(t)	- 2.39 ± 0.79	1.51 ± 0.39

Table 5. *Mean amplitudes for the MMN recorded at Fz and the RMS for the K and T items in reciprocal context and noise-replaced #-items in the K and T context respectively. Figures in bold indicate the main result for the assumed prediction error in the present study.*

in that latency range for the difference wave did not show a classical MMN topography. A two-way repeated measures ANOVA revealed a statistical significant Predecessor x Expectation interaction for both, the mean amplitude at Fz (F(1,19) = 7.99, $p < 0.05$) and for the RMS amplitude (F(1,19) = 4.85, $p < 0.05$). Post-hoc comparisons showed a significant difference for disconfirming and confirming T-items in the K-item context (Fz: t(19) = 2.49, $p < 0.05$; RMS amplitude: t(19) = 2.42, $p < 0.05$) while the K-items in the T-item context revealed no such difference (Fz and RMS: t(19) < 2, $p > 0.1$). Additionally the T and K-items differed at the disconfirming position (Fz: t(19) = 2.92, $p < 0.01$; RMS amplitude: t(19) = 3.21, $p < 0.01$) but did not at the confirming position (Fz and RMS: t(19) < 2, $p > 0.1$). Comparing the peak latencies of the MMN responses at Fz of the combined disconfirming $\#_1$-items with that of the T_1(k) item a significant difference was revealed by the statistics (t(19) = 6.89, $p < 0.001$) indicating an earlier occurrence of the MMN for the $\#_1$-items.

Item	Mean GBA change ± SEM (%) present study	Mean GBA change ± SEM (%) (study B)
$\#_1$(k)	- 1.68 ± 2.14	3.69 ± 2.19
$\#_2$(k)	- 0.14 ± 3.76	-3.59 ± 1.65
$\#_1$(t)	- 0.85 ± 2.45	-3.73 ± 1.86
$\#_2$(t)	1.13 ± 2.29	0.26 ± 1.81

Table 6. *Mean spectral power for the 430 – 490 ms / 38 – 44 Hz range, averaged over left anterior electrode sites and standard error of mean (SEM) in % change for the noise-replaced #-items of the current study. The actual results (left column) are contrasted with the GBA results of Hannemann and Eulitz (submitted) (right column) which have been reflecting the successful repair of the deficient lexically specified speech segment.*

The results of the wavelet analysis for the disconfirming and confirming #-items are depicted in Figure 14. The activity in the gamma band range from 38 – 44 Hz is shown over left anterior electrode sites, which have been shown to map repair processes in the phonemic restoration illusion (see study B). The grey area indicates the time range of 430 – 490 ms which has been discussed to reflect the process of "filling-in" the anticipated phonemic information. As shown in Figure 14, the GBA reveals no marked differential pattern between the experimental conditions. The two-way repeated measures ANOVA for the latency range of 430 – 490 ms revealed neither a Predecessor x Expectation interaction (F(1,19) < 1) nor any main effect (F(1,19) < 2, p > 0.2). The mean values of GBA of the current study as well as results of the previous experiment (see study B; Hannemann and Eulitz, submitted) are summarized in Table 6. Note the differential pattern of GBA changes in the two experiments.

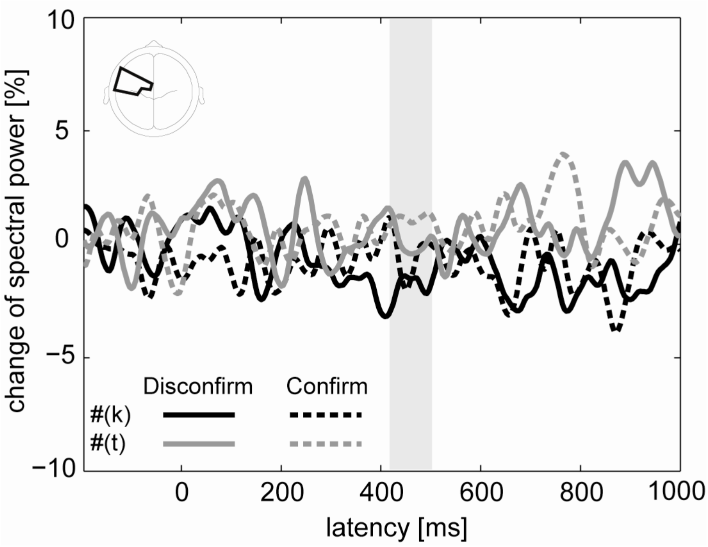


Figure 14. *Illustration of the time course of induced gamma band activity in the 38 – 44 Hz range over left anterior temporal electrode sites (cf. Fig. 1). Solid lines represent disconfirming, dashed lines confirming #-items. Black lines picture #-items following K-items and grey lines represent #-items with T-items as predecessor. The underlying grey box depicts the hypothesized time course (430 – 490 ms) according to which the "filling in" of anticipated lexically specified information in the phonemic restoration should be observable (study B; Hannemann and Eulitz, submitted).*

3.3.4 Discussion

To gain a better understanding of the general conditions that allow or prevent the repairing of deficient speech, the current study was set out to extend previous work, which enlightened brain processes underlying fragmentary speech perception (see study B; Hannemann and Eulitz, submitted). In particular, study C investigated the role of lexical specification of phonological details in the mental lexicon and their impact on the phonemic restoration illusion, especially in contexts of a distinct acoustic difference between actual and anticipated phonemic information. In order to avoid influences of attention or decision making processes on the phonemic restoration the illusion was investigated by means of a passive oddball paradigm in which the subjects were instructed to ignore the auditory stimuli. For the purpose of disentangling possible repair processes from error detection, either a modulation in the induced GBA as a correlate of speech repair processes (as in study B; Hannemann and Eulitz, submitted) or a MMN reflecting the detection of the acoustic discrepancy between the both was hypothesized.

The results of study C clearly demonstrate MMN responses over central frontal electrode sites for disconfirming #-items succeeding T and K-items in the latency range of the to-be-expected phoneme. Admittedly there are no differential MMNs between $\#_1$(k) and $\#_1$(t)-items attributable to the different specification of the to-be-expected phoneme in the respective predecessor observable. Both MMN are equal in size and latency (Figure 13, left side). Apparently the acoustical difference between the noise-replaced segment and the expectation of /t/ or /k/ phonemes is similar. According to Friston (2005) the observed MMN reflects the mismatch between the predicted signal and the actual stimulus-generated activity which can be used to refine the speech perception process. Thus, the pronounced but similar MMNs in the $\#_1$-items are more likely to show a categorical mismatch between speech and no-speech sounds rather than a small difference between expected specified or underspecified phonemic features (Lahiri and Reetz, 2002) or their absence.

However, the results of the present study clearly adduce evidence for a sparse mental lexicon. Looking at the ERPs of the original words a clear MMN can be seen for the T(k)-items but not for the K(t)-items around 460 ms over frontal electrode sites (Figure 13, right side). Thus, in the T(k) case, the repeated presentation of K-items establishes an underlying representation of the [dorsal] place of articulation and therefore conflicts with the extracted surface representation [coronal] of the disconfirming T-item. Reversing the order of appearance of T and K-items, the underlying representation is underspecified for its place feature [coronal] and thus there is no conflict with the surface representation of the [dorsal] place feature extracted from the disconfirming K-items (Lahiri and Reetz, 2002). Thus, the current results are conforming to the pattern of phonological conflicts in vowels (Eulitz and Lahiri, 2004) and consonants (Cornell et al., submitted) mapped in the MMN. Accounts favouring a storage of all acoustic information in the mental lexicon (e.g. McClelland and Elman, 1986; Bybee, 2001) would rather expect a similar pattern in the amplitude of the MMN for both, the T(k) and the K(t) cases. Thus, the present data extend the neurophysiologic evidence speaking for a sparse mental lexicon where only contrastive phonological information is stored (Lahiri and Reetz, 2002).

It has been claimed that the MMN is a compound of the activity of a sensorial and a cognitive comparator mechanism which could be separated in time (e.g. Pulvermuller and Shtyrov, 2006; Maess et al., 2007), especial in pitch perception. The observed latency discrepancy between the MMNs for the disconfirming T(k)-items and #-items in general is in line with this line of research. The MMN for the disconfirming #-items succeeding T and K-items showing a rather categorical mismatch occurs earlier than the more fine grained conflict between the expectancy and the factual occurrence of phonological features.

According to the findings of study B top-down lexical influence on phonemic processing in the phonemic restoration is reflected in modulations in the induced GBA as signature of speech integration and unification processes at the level of working memory (Hannemann and Eulitz, submitted). This pattern was observed over left anterior electrode sites when noise-replaced #-items succeeded noise-overlaid T or K-items. The present results for the #-items did not show the expected nor any other pattern in the induced GBA (Figure 14, Table 6). Apparently the explicit discrepancy between expected and actual acoustic information is too large as the speech processing system is able to integrate the expected phonemic information. Therefore the data show a prediction error depicted in the MMN (Friston, 2005) instead of resonant states depicting the filling in of expected phonemic information in the noise-replaced segment in the phonemic restoration illusion (Grossberg, 2003; Hannemann and Eulitz, submitted). This interpretation is in line with recent models of speech perception based on Bayesian statistics (Norris and McQueen, 2008).

Corresponding to this account the present results can be interpreted as follows: The a-priory assumption generated by virtue of the repeated presentation of T or K-items does by no means map with the characteristics of the actual acoustics. Consequently the aspired alignment with the lexical expectation has to fail.

The phonemic restoration and the so-called continuity illusion, in which a tone with a gap which is masked by broadband noise is perceived as being continuous, are belonging to the much broader phenomenon of auditory induction (Warren et al., 1972; Warren, 1984). In a MMN study it has been shown, that the amplitude of the MMN reflects the perception of an illusory-continuity (Micheyl et al., 2003). They reported larger MMN amplitudes in conditions where the deviants are perceived as continuous and the standards as interrupted or vice versa, than when both were perceived as continuous or both interrupted. Further they claimed that the continuity illusion is well established before a possible prediction error reflected in larger MMN amplitudes is generated, i.e. before 200 ms post-deviance onset. Taking this evidence into account, the present data, which show a pronounced MMN for disconfirming #-items, suggest the failure of the establishing of the restoration of the missing phoneme or at least the reevaluation of the a-priory assumptions to allow for the repairing of the noise-replaced segment at later latencies. However, the present gamma band data do not support the latter assumption. Up to 1000 ms post-stimulus onset there is hardly any modulation observable which might indicate a repair process.

At this point it is important to stress that, albeit an explicit distance between expected and actual phonemic information can prevent the repairing of the missing phoneme, the restoration of the meaning of larger deficient speech units, i.e. words, due to the influence of a broader sentential context, is not implicitly restricted. Issuing the perception of sentence final words with a noise substituted initial phoneme in an EEG study, Sivonen et al. (2006) showed that the semantic processing of the words took place in spite of the substituted onset phoneme. Their result confirmed the whole body of behavioural studies in which listeners report hearing phonemes consistent with sentential or lexical context in locations completely replaced with noise (e.g. Warren, 1970; Samuel, 1997; Samuel, 1981a). However, all these studies can not answer the question if the missing phoneme itself or merely the meaning of the sentence final word was restored. As the study C was set out to prevent attentional and decision making processes to influence the phonemic restoration, it can be concluded, that in classical designs examining the phonemic restoration in sentence context the phoneme itself is not repaired mandatory.

Summary

Summing up, the present data widen the knowledge of the general conditions in which repairing of defective speech signals is possible. In absence of a strictly determining sentential context the integration of expected and actual phonemic information depends largely on the similarity or the overlap of the distribution of the expected phonemic and the actual acoustic features. If this overlap is large, the phonemic restoration is depicted in the induced GBA over left anterior electrode sites (see study B; Hannemann and Eulitz, submitted). On the contrary, if there is only a small or no overlap the current MMN results show that this difference prevents the establishing of a restoration of the missing phoneme. Further the present data extend the electrophysiological evidence for a sparse mental lexicon (Lahiri and Reetz, 2002). To our knowledge they show for the first time a contrastive MMN for word-medial phonological conflicts. Moreover the current results are in line with ideas that speech recognition relies on Bayesian statistics which state that the successful recognition of speech depends on prior knowledge and actual acoustic characteristics (Friston, 2005; Norris and McQueen, 2008).

4 General discussion

In the previous chapters, the data of three studies examining repair processes in speech perception were reported. These studies were unified by the general hypothesis that the perception of deficient speech is possible because lexical top-down expectations and the actual sensory input are integrated in order to obtain a coherent speech percept. Central to all reported studies was the assumption, that the induced GBA is a signature of speech integration and unification processes at the level of working memory (Hagoort, 2005; Bastiaansen and Hagoort, 2006) or rather a signature of (auditory) object recognition (Tallon-Baudry and Bertrand, 1999) and that the MMN reflects the prediction error (Friston, 2005) that is generated, if anticipated information and sensory input mismatches explicitly.

Experiment A confirmed this hypothesis by showing a modulation of induced GBA over left anterior temporal electrode sites as a correlate of the formation of a comprehensible speech percept out of degraded speech information by matching bottom-up sensory input with top-down lexical expectations (Fig. 3). This effect was confirmed and extended in experiment B which revealed a differential modulation in the induced GBA characterizing the phonemic restoration illusion (Fig. 8). This reported modulation, which was again situated over left anterior temporal electrode sites, was shown to depend on the specificity of the place of articulation of the phoneme to be restored in noise-replaced items following a noise-overlaid item. However, instead of revealing a comparable modulation in the gamma band range, experiment C disclosed a prediction error in form of a clear MMN response over frontal electrode sites if noise-replaced items were contrasted with the original items (Fig. 13). The implications of these findings for the understanding of speech repair processes with respect to the initial asked questions are discussed in the following sections.

4.1 Repair processes in speech perception are depicted in gamma band modulations

The first aim of this thesis was to identify the induced GBA as a correlate for the successful recovery of deficient speech under adverse listening conditions. Therefore the induced GBA was analyzed in studies A and B with different experimental designs and stimuli. Because of no exact a priori knowledge of latencies and frequencies which might map these repair processes, permutation tests (Blair and Karniski, 1993), which analyzed the time by frequency data in six electrode arrays (see General Methods, Fig. 1), were applied in both studies. The data of both EEG experiments revealed a modulation in the induced GBA over the left anterior temporal electrode cluster in the 40 Hz range in latency ranges which usually comprise recognition or identification processes in spoken word perception.

Following the argumentation of Hagoort (2005) the neurobiological requirements for this kind of unification of speech signals are those of working memory, which include that lexical building blocks are kept activated for some time while unification operations take place. The cognitive neuroscience literature describes synchronous neuronal activity in the gamma range first of all as correlate of successful representations of coherent single- and multimodal objects as well as memory processes (e.g. Tallon-Baudry and Bertrand, 1999; Pesaran et al., 2002; Busch et al., 2006a; Gruber and Muller, 2006; Leiberg et al., 2006; Gruber et al., 2008). According to the Adaptive Resonance Theory (Grossberg, 1980; Grossberg, 1999) this synchronous neuronal activity is the result of matching bottom-up sensory information with top-down expectations and learned representations, e.g., lexical memory traces. With respect to speech perception, this neuronal synchrony reflects a mechanism by which the brain integrates the different types of information about language, i.e. phonological, morphological, semantic and syntactic information, which is represented in different brain areas (Bastiaansen and Hagoort, 2006).

Corresponding to recent neuroimaging studies using degraded speech stimuli (e.g. Scott et al., 2000; Giraud et al., 2004; Obleser et al., 2007) the intelligibility of degraded speech in passive listening tasks is accompanied by brain activity in inferior frontal as well as anterior and lateral temporal areas. Also an increase in cortical activity in the anterior cingulate as well as superior to middle temporal regions during the perception of degraded but intelligible speech was reported (Giraud et al., 2004). Recently it was speculated, that increased left inferior frontal brain responses reflect top-down efforts to support the perception of speech by the use of semantic information when the perception becomes more difficult due to masking with noise (Scott et al., 2004). When comprehension of degraded speech was likely to succeed due to contextual cues fronto-temporal and prefrontal activations in the left hemisphere were reported (Obleser et al., 2007). According to Hagoort (2005) the left inferior frontal gyrus (IFG) plays an important role in unification processes in speech perception. He claimed that unification operations take place in parallel at the phonological, semantic and syntactic levels of processing.

Taken together, the results of studies A and B clearly evince an anterior temporal framework which merges top-down expectations with sensory input into a perceivable auditory object, especially in the perception of deficient speech input. The observed relationship between unification operations and GBA synchronizations fit well within the larger context of the integrative role that gamma band synchronizations in neuronal networks appears to play in a wider range of cognitive domains and the fact that the same electrode array was identified independently two times further emphasises this claim. Admittedly, it is not possible to infer directly the underlying sources from the surface activity. However, the superior temporal resolution of the EEG adds temporal resolution to the top-down modulated processes seen in the above mentioned neuroimaging studies. In speech perception at least for the integration of semantic and world-knowledge during the interpretation of sentences the benefit of the direct integration of fMRI and oscillatory EEG data have been successfully demonstrated (Hagoort et al., 2004). Hence, to be able to integrate the current findings directly into the framework outlined by the neuroimaging literature, the cortical sources underlying the modulations in induced GBA accompanying restoration and unification processes in the perception of deficient speech signal require further examination. Therefore tying fMRI and EEG measures in speech perception more closely is a stringent next step.

4.2 Successful repairing depends on the compatibility of the actual and the expected input

Second, this thesis started with the notion that it is not optimal for the cognitive system to integrate all possible sensory input with the a-priory expectations. By adapting a similar experimental design as applied in study B, experiment C started examining the general conditions that allow for a repair of obliterated phonemes and what conditions prevent it. Instead of showing a modulation in the induced gamma band range, the results of study C revealed a clear MMN response for the noise-replaced items succeeding normal items with no manipulation. This apparent signature of a prediction error (Friston, 2005) is due to the explicit difference between noisy sensory and expected un-noisy information which prevents the integration of both into a coherent speech percept. Thus, the noise-replaced items did not fit into the likelihood distribution of the expected phoneme and therefore it could not represent a possible instance of the expected phoneme (Norris and McQueen, 2008). The difference between the studies B and C lies in the noise-overlaying of the K and T-item in study B. Albeit the expectancy of the /k/ and /t/ phonemes are the same in both experiments, the anticipated percepts of phonemes in study B also encompass some noise-characteristics. Therefore the difference between the a-priory hypothesis about the next event and the actual sensory input in the roving oddball design is smaller in study B as in study C, which circumvents the integration of both in the latter into a coherent speech percept.

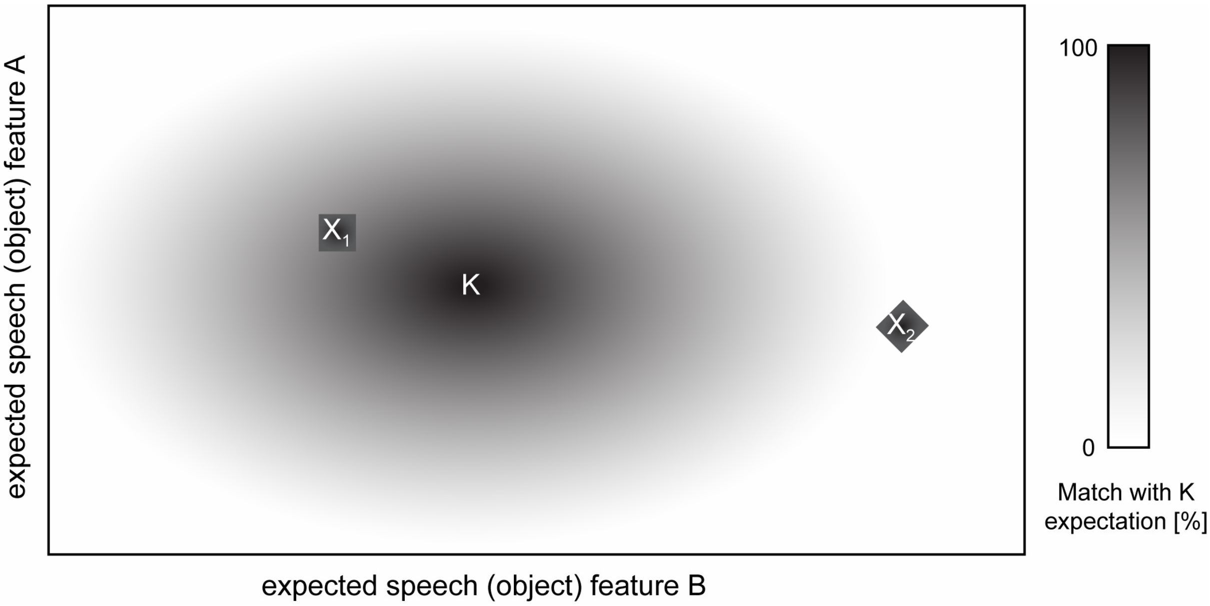


Figure 15. *Illustration of the likelihood distribution of an expected phoneme along two exemplary featural dimensions. Two incoming stimuli, one with partly appropriate characteristics (X_1) and one ill-fitting (X_2) in relation to the expected input are shown.*

The results of the studies B and C are abstractly generalized in Figure 15. Simplified, the likelihood distribution of an expected phoneme K along two exemplarily dimensional characteristics and two acoustic events are shown. Event X_1 partly fits into the distribution of the expected K and event X_2 does not. According to the underlying idea of this thesis the perception of event X_1 leads to observable modulations in induced gamma band responses depicting the merging of a-priory hypothesis about what will be the input and the actual signal. Instead, event X_2 does not fit into the expected range of features and thus reveals a prediction error in form of a MMN. Figure 15 also nicely illustrates the Bayesian approach of speech perception (Norris and McQueen, 2008). Only if the actual features of a sensory event match the expected distribution of features, the a-priory expectation can contribute to the overall speech percept. Otherwise, as consequence of equitation 1 (see General Introduction), the product of the a-priory hypothesis and sensory evidence becomes zero resulting in a zero conditional probability for the expected words or phoneme given the actual sensory evidence. Therefore, the clear difference between actual and expected acoustic features in experiment C must result in a prediction error.

The results of study C also nicely acknowledge early behavioural studies examining the phonemic restoration illusion. These studies reported that the effectiveness of the restoration depends upon the similarity of the replacement sound to the phoneme it replaces and that the replacement of a phoneme with silence inhibits the phonemic restoration (Warren and Obusek, 1971; Layton, 1975; Samuel, 1981b).

According to the Bayesian statistical framework the results of study A can be interpreted as follows: Due to the memorizing of half of the words in the exposure sequence, the lexical memory traces representing these items were strengthened and refreshed. Amongst others, this strengthening leads to a broadening of the likelihood distribution for the expectation of the acoustical and phonemic features. Hence, in the consecutive test sequence the residual features in the degraded speech items were mapped partly by the likelihood distribution of the expected input and therefore were sufficient to activate the strengthened lexical memory traces retained in the working memory.

In sum, the approach of Bayesian statistics to speech perception (e.g. Friston, 2005; Norris and McQueen, 2008) has been proven to be successful to describe a whole range of experimental findings. However, the current data provide only the endpoints of a dimension characterizing the repair and the detection of errors in deficient speech signals. Crucial next steps in exploring these phenomena should encompass a more gradual exploration of the steps between these endpoints, for example by using a linear regression analysis of the EEG data. This analysis has been used, for instance, to discriminate the early visual analysis of the word form from the retrieval of lexico-semantic information in word recognition in the visual domain (Hauk et al., 2006). Thus, up to which dissimilarity between actual and expected input are repair processes observable? Do the electrophysiological parameters change in a linear manner according to this dissimilarity, i.e. does the modulation in the induced gamma band range or the MMN amplitude increase linearly or not? These and other questions have to be answered to be able to draw a comprehensive picture of the perceptual mechanisms that cope with sketchy speech input.

4.3 Influences of the mental lexicon's fine structure on the repair of deficient speech

Because of the enormous amount of variability in the relationship between the acoustic signal and the language it represents, the role of the exact nature of the mental representation of language sounds is far from being resolved. One promising approach to deal with this variability is the idea of a featurally underspecified lexicon (Lahiri and Marslen-Wilson, 1991; Lahiri and Reetz, 2002), which assumes that all predictable and non-distinctive information is not specified in the mental lexicon. To account for the concept of underspecification, the current thesis asked whether or not the repair processes in speech perception, in particular the phonemic restoration, are influenced by the fine structure of the mental lexicon.

Measured in a pre-attentive listening task, the induced gamma band results of experiment B showed a clear modulation which is dependent on the specificity of the expected phoneme. Reported is an increase in induced GBA if the expected phoneme was specified for the feature place of articulation (K-item) compared to the underspecified expectation (T-item). Though, under attentive processing conditions, the evaluation of reaction time data showed no effect of the specification of featural information in the mental lexicon. Pre-attentively, the unification processes seem to provide information founding the basis for further cognitive stages, i.e. decision processes in the phonemic restoration. The higher induced GBA values for the disconfirming #-items following K-items therefore reflect the "filling in" or rather the merging of the anticipated lexically specified information in the noise replaced part of the disconfirming #-items. As the underspecified information serves as the perceptual default (cf. Paradis and Prunet, 1991), no "filling in" of the anticipated information is necessary for the disconfirming #-items succeeding T-items as prerequisite for further processing stages. The latter assumption is verified by the results of the behavioral control experiments reported with the gamma band results of study B. If the subjects had the freedom of choice of all possible phonemes in the identification task, they reported mostly perceiving the /t/, which is underspecified for the place feature. Thus it served perfectly as perceptual default, first of all in the perception of ambiguous acoustic input.

In study C the MMN responses showed a clear prediction error if the expected phonemic information was not present in the actual noise-replaced item. However, this MMN response was not influenced by the differential lexical specificity of the anticipated phoneme. Admittedly, the data for the original words containing no manipulation in experiment C revealed a contrastive MMN for word-medial phonological conflicts according to the predictions derived from the FUL model (Lahiri and Reetz, 2002). Thus, albeit demonstrating that not all processes involved in speech processing necessarily use the amenities of a sparse mental lexicon, these results evince a considerable evidence for the latter. Models or approaches favoring fully specified lexical entries (e.g. McClelland and Elman, 1986; Bybee, 2001) can not account for the present findings. These models neither can explain the differential modulation in the induced GBA in study B nor the contrastive MMN for word-medial phonological conflicts in experiment C. Following this line of argumentation the data rather should reveal similar gamma band and MMN effects in both experiments for both kinds of expectations and their violation. Thus, the findings of experiments B and C favor a sparse mental lexicon and are in line with patterns of phonological conflicts in vowels (Eulitz and Lahiri, 2004) and consonants (Cornell et al., submitted) mapped in the MMN.

However, what is missing in the current thesis is the possibility to generalize the findings to other phonemic features than the place of articulation. Actually the data only account for noise-replaced German stop consonants. Further studies investigating other features in German and other languages are therefore needed to gain a better and extensive understanding of the influence of the specificity of phonemes in the mental lexicon on repair processes in speech perception.

4.4 The auditory induced GBA: a saccadic artifact or a measure reflecting higher cognitive processing?

A vast majority in the cognitive neuroscience literature dealing with synchronous neuronal activity in the gamma band range describes it as correlate of successful representations of coherent single- and multimodal objects as well as memory processes (e.g. Tallon-Baudry and Bertrand, 1999; Busch et al., 2006a; Pesaran et al., 2002; Gruber and Muller, 2006; Leiberg et al., 2006; Gruber et al., 2008). The studies A and B disclosed it as a signature of speech integration and unification processes at the level of working memory (Hagoort, 2005; Bastiaansen and Hagoort, 2006) or rather a signature of (auditory) object recognition in general.

However, a word of caution about high frequency oscillations measured by EEG and MEG are appropriate. By applying a single trial analysis of concomitant EEG and eye tracking in a series of visual perception experiments, Yuval-Greenberg and colleagues (2008) report that the observed modulation induced GBA is tightly time locked to the onsets of involuntary miniature eye movements and concluded that it rather reflects a saccadic “spike potential” than cognitive processing. Their observed effects in the gamma range showed a broadband activity over posterior electrode sites. Concluding on their results, Yuval-Greenberg et al. (2008) claimed that earlier findings of modulations in visual induced GBA are due, at least partly, to properties of micro-saccades. In particular, they argued that the most commonly reported induced GBA around 200–300 ms with a broad frequency spectrum is a product of eye movements, probably of muscular origin, and not a direct measure of neuronal oscillations. Importantly, Yuval-Greenberg and colleagues (2008) did not claim that all scalp recorded induced GBA are due to eye movements (only if they occur in a narrow time window and a broad frequency range) and they do not question the neuronal gamma band synchronization itself. Intracranial neuronal synchronization in the gamma band has been demonstrated more than once directly (e.g. Gray et al., 1989; Fries et al., 2001; Tallon-Baudry et al., 2005) and thus these synchronizations clearly represent an important interactions within networks organized both in space and time in the human brain.

Even though, the raised concerns about the scalp-recorded induced GBA (Yuval-Greenberg et al., 2008) should be taken into account when observing and interpreting oscillatory brain activity. For the following reasons this criticism does, however, not affect the functional interpretation of the present gamma band results. First, studies A and B observed the modulation in induced GBA over left anterior temporal electrode sites and not over occipital brain regions. Second, the all EEG studies did not use the nose-reference which was used by Yuval-Greenberg et al. (2008). This nose reference might be responsible for the projection of the saccadic activity over broad cortical areas, at least partly. Third, the findings of the present thesis are restricted to a relatively small frequency range around 40 Hz. However, the last point might be due to the rather conservative permutation test, which identified the frequency bands depicting the repair processes in the perception of deficient speech. Nevertheless, the analysis of higher frequencies for the sake of controlling for muscular artefacts (Cacioppo J.T. et al., 1990; Pulvermuller et al., 1997) yielded no results as obtained for the 40 Hz frequency range in the studies A and B. Thus, the present data of both studies did not reveal a broad frequency spectrum which is supposed to reflect a rather saccadic origin. The fourth argument against the notion that the present data might be accountable to micro-saccades is the latency of the present findings: all observed effects occur later than 200-300 ms after stimulus onset. However, study A revealed an earlier enhancement in the latency range short before 200 ms, which did not fulfil the criteria of the applied permutation test admittedly. Thus, it might be possible that this rather fragile modulation can be interpreted as artefact despite its anterior temporal topography.

Taken together, it can be concluded that the present induced GBA are not an artefact of saccadic eye movements. Due to the characteristics of the data, this possible influence can be ruled out for both studies reporting signatures in the gamma band for the restoration and the perception of deficient speech over left anterior temporal electrode sites. To be absolutely sure that no influences of micro-saccades contaminate the auditory GBA, future studies could allay these concerns regarding eye movements by using newer methods for the reconstruction of the underlying sources of the oscillatory brain activity (e.g. Dalal et al., 2008), which might be also a more simple alternative to the highly complex parallel measuring of EEG and fMRI data.

4.5 Are there online or offline lexical top-down influences in speech perception?

Taken together, the findings of the current thesis support the view that top-down lexical information facilitates speech perception, especially when the acoustic input is ambiguous or degraded. Autonomous theoretical approaches that concede only top-down influences for learning but refuse the necessity or existence of online feedback in the speech perception process (e.g. Norris et al., 2003; Norris and McQueen, 2008) might be able to explain the results of study A but cannot account for the findings of study B. The latter explicitly demonstrates an online process that integrates anticipated lexical information with the incoming stimulus intermediately. Thus, the “filling in” is more than a reorganisation of memory traces in order to adapt more effectively to future events as claimed by the autonomous models. The present results are therefore in line with several other findings (e.g. Samuel, 1997; Samuel, 2001) that point to interactive processing as a fundamental principle for speech perception as well as for other modalities (McClelland et al., 2006). The interactive view predicts that a-priory lexical information reaches down and reshapes the mental representation of the sound which is actually heard. This is exactly what the findings of study A and B demonstrate. Hence, the current data can be interpreted as depicting the formation of an auditory object (Tallon-Baudry and Bertrand, 1999; Griffiths and Warren, 2004) by delineating resonant states in speech perception (Grossberg, 2003). Further evidence for this claim is revealed by the examination of rhythmic tone sequences. It is known, that the induced GBA persists when expected tones are omitted (Snyder and Large, 2005). Taken together with the findings of the present thesis, it can be concluded that the auditory induced GBA represents a highly valuable tool for investigating resonant states in (illusory) speech perception.

However, if one thinks about the induced GBA as correlate for the reckoning the probability that the actual acoustical input is a realization of a certain phoneme or word (formula 1; $P(Evidence/Hypothesis_i)$) and the a-priory probabilities or initial beliefs about the sensory input (formula 1; $P(Hypothesis_i)$), then the current results match perfectly in the framework of Shortlist B (Norris and McQueen, 2008). Then, the proposed feed forward of information from lower to higher levels may not be questioned. Following this line of interpretation, however, the notion of the functionality of the gamma band itself has to be reconsidered. Does it reflect a cognitive computational process, a state of object representation or both? Unfortunately the current data are not sufficient to answer this question. Future studies addressing this fundamental issue are therefore needed to draw a comprehensive picture about the functionality of the gamma band, not only in relation to speech perception.

4.6 Conclusion

Summing up, the data of the present thesis discern processes of (phonological) unification (Grossberg, 2003; Hagoort, 2005) from the refinement of a-priory hypotheses through prediction error processes (Friston, 2005) in speech perception, particularly in difficult environmental conditions. Further they corroborate the hypothesis of sparse representations in the mental lexicon (Lahiri and Reetz, 2002).

The results proved to be highly analogous to other perceptual domains. Furthermore, they lend plausibility to think of the induced GBA as a correlate of binding the sensory (auditory) input and available lexical information into a coherent perception, i.e. into a recognized phoneme or word. Finally these findings demonstrate, how efficient and economic processing of speech and a-priory information allows the human cognitive system to adapt rapidly to a complex and constant changing environment.

5 References

Andrew C, Pfurtscheller G (1996) Event-related coherence as a tool for studying dynamic interaction of brain regions. Electroencephalogr Clin Neurophysiol 98:144-148.

Axmacher N, Mormann F, Fernandez G, Elger CE, Fell J (2006) Memory formation by neuronal synchronization. Brain Res Rev 52:170-182.

Baayen H, Piepenbrock R, van Rijn H (1993) The CELEX lexical database.

Baldeweg T, Klugman A, Gruzelier J, Hirsch SR (2004) Mismatch negativity potentials and cognitive impairment in schizophrenia. Schizophr Res 69:203-217.

Bar M (2007) The proactive brain: using analogies and associations to generate predictions. Trends Cogn Sci 11:280-289.

Basar E, Basar-Eroglu C, Karakas S, Schurmann M (2001) Gamma, alpha, delta, and theta oscillations govern cognitive processes. Int J Psychophysiol 39:241-248.

Bastiaansen M, Hagoort P (2006) Oscillatory neuronal dynamics during language comprehension. Prog Brain Res 159:179-196.

Berg P, Scherg M (1994) A multiple source approach to the correction of eye artifacts. Electroencephalogr Clin Neurophysiol 90:229-241.

Bertrand O, Bohorquez J, Pernier J (1994) Time-frequency digital filtering based on an invertible wavelet transform: an application to evoked potentials. IEEE Trans Biomed Eng 41:77-88.

Bhattacharya J, Petsche H, Pereda E (2001) Long-range synchrony in the gamma band: role in music perception. J Neurosci 21:6329-6337.

Blair RC, Karniski W (1993) An alternative method for significance testing of waveform difference potentials. Psychophysiology 30:518-524.

Boothroyd A, Nittrouer S (1988) Mathematical treatment of context effects in phoneme and word recognition. J Acoust Soc Am 84:101-114.

Busch NA, Herrmann CS, Muller MM, Lenz D, Gruber T (2006a) A cross-laboratory study of event-related gamma activity in a standard object recognition paradigm. Neuroimage 33:1169-1177.

Busch NA, Schadow J, Frund I, Herrmann CS (2006b) Time-frequency analysis of target detection reveals an early interface between bottom-up and top-down processes in the gamma-band. Neuroimage 29:1106-1116.

Bybee J (2001) Phonology and language use. Cambridge: Cambridge University Press.

Cacioppo J.T., Tassinary L.G., Fridlund A.J. (1990) The skeletomotor system. In: Principles of psychophysiology: physical, social, and inferential elements. (Cacioppo JT, Tassinary L.G., eds), pp 324-385. Cambridge, MA: Cambridge UP.

Cherry EC (1953) Some experiments on the recognition of speech, with one and two ears. J Acoust Soc Am 25:975-979.

Clark HH (1996) Using Language. Cambridge, UK: Cambridge University Press.

Cornell SA, Lahiri A, Eulitz C (submitted) "What you encode is not necessarily what you store": a MMN study of phonological processing.

Cowan N, Winkler I, Teder W, Naatanen R (1993) Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). J Exp Psychol Learn Mem Cogn 19:909-921.

Crone NE, Boatman D, Gordon B, Hao L (2001) Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. Clin Neurophysiol 112:565-582.

Cutler A, Norris D (1979) Monitoring sentence comprehension. In: Sentence processing: Psycholinguistic studies presented to Merrill Garrett. (Cooper WE, Walker ECT, eds), Hillsdale: Erlbaum.

Dalal SS, Guggisberg AG, Edwards E, Sekihara K, Findlay AM, Canolty RT, Berger MS, Knight RT, Barbaro NM, Kirsch HE, Nagarajan SS (2008) Five-dimensional neuroimaging: localization of the time-frequency dynamics of cortical activity. Neuroimage 40:1686-1700.

Darwin C (1874) The Descent of Man, and Selection in Relation to Sex. New York: Hurst & Co.

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

Debener S, Herrmann CS, Kranczioch C, Gembris D, Engel AK (2003) Top-down attentional processing enhances auditory evoked gamma band activity. Neuroreport 14:683-686.

Dehaene S, Changeux JP, Naccache L, Sackur J, Sergent C (2006) Conscious, preconscious, and subliminal processing: a testable taxonomy. Trends Cogn Sci 10:204-211.

Dronkers NF, Wilkins DP, Van VR, Jr., Redfern BB, Jaeger JJ (2004) Lesion analysis of the brain areas involved in language comprehension. Cognition 92:145-177.

Eckhorn R, Frien A, Bauer R, Woelbern T, Kehr H (1993) High frequency (60-90 Hz) oscillations in primary visual cortex of awake monkey. Neuroreport 4:243-246.

Eimas PD, Corbit JD (1973) Selective adaption of linguistic feature detectors. Cognit Psychol 4:99-109.

Endrass T, Mohr B, Pulvermuller F (2004) Enhanced mismatch negativity brain response after binaural word presentation. Eur J Neurosci 19:1653-1660.

Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci 2:704-716.

Engel AK, Singer W (2001) Temporal binding and the neural correlates of sensory awareness. Trends Cogn Sci 5:16-25.

Eulitz C, Eulitz H, Maess B, Cohen R, Pantev C, Elbert T (2000) Magnetic brain activity evoked and induced by visually presented words and nonverbal stimuli. Psychophysiology 37:447-455.

Eulitz C, Lahiri A (2004) Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. J Cogn Neurosci 16:577-583.

Eulitz C, Maess B, Pantev C, Friederici AD, Feige B, Elbert T (1996) Oscillatory neuromagnetic activity induced by language and non-language stimuli. Brain Res Cogn Brain Res 4:121-132.

Fiebach CJ, Gruber T, Supp GG (2005) Neuronal mechanisms of repetition priming in occipitotemporal cortex: spatiotemporal evidence from functional magnetic resonance imaging and electroencephalography. J Neurosci 25:3414-3422.

Friederici AD, Kotz SA (2003) The brain basis of syntactic processes: functional imaging and lesion studies. Neuroimage 20 Suppl 1:S8-17.

Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci 9:474-480.

Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. Science 291:1560-1563.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815-836.

Galambos R (1992) A comparison of certain gamma-band (40 Hz) brain rhythms in cat and man. In: Induced Rhythms in the brain. (Basar E, Bullock T, eds), pp 201-216. Boston: Birkhäuser.

Ganong WF (1980) Phonetic categorization in auditory word perception. J Exp Psychol Hum Percept Perform 6:110-125.

Gaskell MG, Marslen-Wilson W (1997) Integrating form and meaning: A distributed model of speech perception. Language and Cognitive Processes 12:613-656.

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

Goffaux V, Mouraux A, Desmet S, Rossion B (2004) Human non-phase-locked gamma oscillations in experience-based perception of visual scenes. Neurosci Lett 354:14-17.

Gray CM, Konig P, Engel AK, Singer W (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature 338:334-337.

Griffiths TD, Warren JD (2004) What is an auditory object? Nat Rev Neurosci 5:887-892.

Grosjean F (1996) Gating. Language and Cognitive Processes 11:597-604.

Grossberg S (1980) How does a brain build a cognitive code? Psychol Rev 87:1-51.

Grossberg S (1999) The link between brain learning, attention, and consciousness. Conscious Cogn 8:1-44.

Grossberg S (2003) Resonant neural dynamics of speech perception. Journal of Phonetics 31:423-445.

Grossberg S, Boardman I, Cohen M (1997) Neuronal dynamics of variable-rate speech categorization. J Exp Psychol Hum Percept Perform 23:481-503.

Grossberg S, Myers CW (2000) The resonant dynamics of speech perception: interword integration and duration-dependent backward effects. Psychol Rev 107:735-767.

Gruber T, Muller MM (2005) Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. Cereb Cortex 15:109-116.

Gruber T, Muller MM (2006) Oscillatory brain activity in the human EEG during indirect and direct memory tasks. Brain Res 1097:194-204.

Gruber T, Muller MM, Keil A (2002) Modulation of induced gamma band responses in a perceptual learning task in the human EEG. J Cogn Neurosci 14:732-744.

Gruber T, Trujillo-Barreto NJ, Giabbiconi CM, Valdes-Sosa PA, Muller MM (2006) Brain electrical tomography (BET) analysis of induced gamma band responses during a simple object recognition task. Neuroimage 29:888-900.

Gruber T, Tsivilis D, Giabbiconi CM, Muller MM (2008) Induced electroencephalogram oscillations during source memory: familiarity is reflected in the gamma band, recollection in the theta band. J Cogn Neurosci 20:1043-1053.

Gruber T, Tsivilis D, Montaldi D, Muller MM (2004) Induced gamma band responses: an early marker of memory encoding and retrieval. Neuroreport 15:1837-1841.

Haenschel C, Baldeweg T, Croft RJ, Whittington M, Gruzelier J (2000) Gamma and beta frequency oscillations in response to novel auditory stimuli: A comparison of human electroencephalogram (EEG) data with in vitro models. Proc Natl Acad Sci U S A 97:7645-7650.

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

Hagoort P (2008) The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. Philos Trans R Soc Lond B Biol Sci 363:1055-1069.

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

Hagoort P, Van Berkum JJ (2007) Beyond the sentence given. Philos Trans R Soc Lond B Biol Sci 362:801-811.

Hannemann R, Eulitz C (submitted) On the matching of top-down knowledge with sensory input in the perception of ambiguous speech.

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

Hauk O, Davis MH, Ford M, Pulvermuller F, Marslen-Wilson WD (2006) The time course of visual word recognition as revealed by linear regression analysis of ERP data. Neuroimage 30:1383-1400.

Herrmann CS, Lenz D, Junge S, Busch NA, Maess B (2004a) Memory-matches evoke human gamma-responses. BMC Neurosci 5:13.

Herrmann CS, Mecklinger A (2001) Gamma activity in human EEG is related to high-speed memory comparisons during object selective attention. Visual Cognition 8:593-608.

Herrmann CS, Munk MH, Engel AK (2004b) Cognitive functions of gamma-band activity: memory match and utilization. Trends Cogn Sci 8:347-355.

Jackendoff R (2003) Precis of Foundations of language: brain, meaning, grammar, evolution. Behav Brain Sci 26:651-665.

Kaiser J, Buhler M, Lutzenberger W (2004) Magnetoencephalographic gamma-band responses to illusory triangles in humans. Neuroimage 23:551-560.

Kaiser J, Heidegger T, Lutzenberger W (2008a) Behavioral relevance of gamma-band activity for short-term memory-based auditory decision-making. Eur J Neurosci 27:3322-3328.

Kaiser J, Heidegger T, Wibral M, Altmann CF, Lutzenberger W (2008b) Distinct gamma-band components reflect the short-term memory maintenance of different sound lateralization angles. Cereb Cortex 18:2286-2295.

Kaiser J, Hertrich I, Ackermann H, Lutzenberger W (2006a) Gamma-band activity over early sensory areas predicts detection of changes in audiovisual speech stimuli. Neuroimage 30:1376-1382.

Kaiser J, Hertrich I, Ackermann H, Mathiak K, Lutzenberger W (2005) Hearing lips: gamma-band activity during audiovisual speech perception. Cereb Cortex 15:646-653.

Kaiser J, Lennert T, Lutzenberger W (2006b) Dynamics of Oscillatory Activity during Auditory Decision Making. Cereb Cortex.

Kaiser J, Lutzenberger W (2004) Frontal gamma-band activity in magnetoencephalogram during auditory oddball processing. Neuroreport 15:2185-2188.

Kaiser J, Lutzenberger W, Ackermann H, Birbaumer N (2002) Dynamics of gamma-band activity induced by auditory pattern changes in humans. Cereb Cortex 12:212-221.

Kaiser J, Ripper B, Birbaumer N, Lutzenberger W (2003) Dynamics of gamma-band activity in human magnetoencephalogram during auditory pattern working memory. Neuroimage 20:816-827.

Keil A, Muller MM, Gruber T, Wienbruch C, Elbert T (2001) Human large-scale oscillatory brain activity during an operant shaping procedure. Brain Res Cogn Brain Res 12:397-407.

Keselman HJ (1998) Testing treatment effects in repeated measures designs: an update for psychophysiological researchers. Psychophysiology 35:470-478.

Knief A, Schulte M, Bertrand O, Pantev C (2000) The perception of coherent and non-coherent auditory objects: a signature in gamma frequency band. Hear Res 145:161-168.

Kutas M, Federmeier KD (2000) Electrophysiology reveals semantic memory use in language comprehension. Trends Cogn Sci 4:463-470.

Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207:203-205.

Kveraga K, Ghuman AS, Bar M (2007) Top-down predictions in the cognitive brain. Brain Cogn 65:145-168.

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. Hum Brain Mapp 8:194-208.

Ladefoged P, Broadbent DE (1960) Perception of sequence in auditory events. The Quarterly Journal of Experimental Psychology 12:162-170.

Lahiri A, Marslen-Wilson W (1991) The mental representation of lexical form: a phonological approach to the recognition lexicon. Cognition 38:245-294.

Lahiri A, Reetz H (2002) Underspecified Recognition. pp 637-676. Berlin: Mouton.: C. Gussenhoven & N. Warner.

Landau AN, Esterman M, Robertson LC, Bentin S, Prinzmetal W (2007) Different effects of voluntary and involuntary attention on EEG activity in the gamma band. J Neurosci 27:11986-11990.

Layton B (1975) Differential effects of two nonspeech sounds on phonemic restoration. Bulletin of the Psychonomic Society 6:487-479.

Lebedev MA, Messinger A, Kralik JD, Wise SP (2004) Representation of attended versus remembered locations in prefrontal cortex. PLoS Biol 2:e365.

Leiberg S, Kaiser J, Lutzenberger W (2006) Gamma-band activity dissociates between matching and nonmatching stimulus pairs in an auditory delayed matching-to-sample task. Neuroimage 30:1357-1364.

Lenz D, Jeschke M, Schadow J, Naue N, Ohl FW, Herrmann CS (2007a) Human EEG very high frequency oscillations reflect the number of matches with a template in auditory short-term memory. Brain Res.

Lenz D, Schadow J, Thaerig S, Busch NA, Herrmann CS (2007b) What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. Int J Psychophysiol 64:31-38.

Liberman AM, Delattre P, Cooper FS (1952) The role of selected stimulus-variables in the perception of the unvoiced stop consonants. Am J Psychol 65:497-516.

Llinas RR, Ribary U, Jeanmonod D, Kronberg E, Mitra PP (1999) Thalamocortical dysrhythmia: A neurological and neuropsychiatric syndrome characterized by magnetoencephalography. Proc Natl Acad Sci U S A 96:15222-15227.

Lutzenberger W, Pulvermuller F, Birbaumer N (1994) Words and pseudowords elicit distinct patterns of 30-Hz EEG responses in humans. Neurosci Lett 176:115-118.

Maess B, Jacobsen T, Schroger E, Friederici AD (2007) Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. Neuroimage 37:561-571.

Mainy N, Jung J, Baciu M, Kahane P, Schoendorff B, Minotti L, Hoffmann D, Bertrand O, Lachaux JP (2007) Cortical dynamics of word recognition. Hum Brain Mapp.

Marslen-Wilson W, Warren P (1994) Levels of perceptual representation and process in lexical access: words, phonemes, and features. Psychol Rev 101:653-675.

Marslen-Wilson W, Welsh A (1978) Processing interactions and lexical access during word recognition in continuous speech. Cognit Psychol 10:29-63.

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

McClelland JL, Mirman D, Holt LL (2006) Are there interactive processes in speech perception? Trends Cogn Sci 10:363-369.

Mehler J, Sebastian N, Altmann G, Dupoux E, Christophe A, Pallier C (1993) Understanding compressed sentences: the role of rhythm and meaning. Ann N Y Acad Sci 682:272-282.

Melloni L, Molina C, Pena M, Torres D, Singer W, Rodriguez E (2007) Synchronization of neural activity across cortical areas correlates with conscious perception. J Neurosci 27:2858-2865.

Micheyl C, Carlyon RP, Shtyrov Y, Hauk O, Dodson T, Pullvermuller F (2003) The neurophysiological basis of the auditory continuity illusion: a mismatch negativity study. J Cogn Neurosci 15:747-758.

Miller GA, Heise GA, Lichten W (1951) The intelligibility of speech as a function of the context of the test materials. J Exp Psychol 41:329-335.

Miltner WH, Braun C, Arnold M, Witte H, Taub E (1999) Coherence of gamma-band EEG activity as a basis for associative learning. Nature 397:434-436.

Mitra PP, Pesaran B (1999) Analysis of dynamic brain imaging data. Biophys J 76:691-708.

Muller MM, Gruber T (2001) Induced gamma-band responses in the human EEG are related to attentional information processing. Visual Cognition 8:579-592.

Myers EB, Blumstein SE (2008) The neural bases of the lexical effect: an fMRI investigation. Cereb Cortex 18:278-288.

Naatanen R (2001) The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiology 38:1-21.

Naatanen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385:432-434.

Naatanen R, Michie PT (1979) Early selective-attention effects on the evoked potential: a critical review and reinterpretation. Biol Psychol 8:81-136.

Naatanen R, Picton T (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24:375-425.

Naatanen R, Winkler I (1999) The concept of auditory stimulus representation in cognitive neuroscience. Psychol Bull 125:826-859.

Norris D (1994) Short list: a connectionist model of continuous speech recognition. Cognition 52:189-234.

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

Norris D, McQueen JM, Cutler A (2000) Merging information in speech recognition: feedback is never necessary. Behav Brain Sci 23:299-325.

Norris D, McQueen JM, Cutler A (2003) Perceptual learning in speech. Cognit Psychol 47:204-238.

Obleser J, Lahiri A, Eulitz C (2003) Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. Neuroimage 20:1839-1847.

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

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

Osipova D, Takashima A, Oostenveld R, Fernandez G, Maris E, Jensen O (2006) Theta and gamma oscillations predict encoding and retrieval of declarative memory. J Neurosci 26:7523-7531.

Otten M, Nieuwland MS, Van Berkum JJ (2007) Great expectations: specific lexical anticipation influences the processing of spoken language. BMC Neurosci 8:89.

Pantev C (1995) Evoked and induced gamma-band activity of the human cortex. Brain Topogr 7:321-330.

Paradis C, Prunet JF (1991) Phonetics and phonology: Vol.2 The special status of coronals. San Diego, CA: Academic Press.

Pesaran B, Pezaris JS, Sahani M, Mitra PP, Andersen RA (2002) Temporal structure in neuronal activity during working memory in macaque parietal cortex. Nat Neurosci 5:805-811.

Petkov CI, Logothetis NK, Obleser J (in press) Where are the human speech and voice regions and do other animals have anything like them? Neuroscientist.

Pfurtscheller G, Lopes da Silva FH (1999a) Event-related desynchronization. Handbook of Electroencephalography and Clinical Neurophysiology. Amsterdam: Elsevier.

Pfurtscheller G, Lopes da Silva FH (1999b) Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol 110:1842-1857.

Poeppel D, Idsardi WJ, van Wassenhove V (2008) Speech perception at the interface of neurobiology and linguistics. Philos Trans R Soc Lond B Biol Sci 363:1071-1086.

Pulvermuller F (1999) Words in the brain's language. Behav Brain Sci 22:253-279.

Pulvermuller F, Birbaumer N, Lutzenberger W, Mohr B (1997) High-frequency brain activity: its possible role in attention, perception and language processing. Prog Neurobiol 52:427-445.

Pulvermuller F, Keil A, Elbert T (1999) High-frequency brain activity: perception or active memory? Trends Cogn Sci 3:250-252.

Pulvermuller F, Kujala T, Shtyrov Y, Simola J, Tiitinen H, Alku P, Alho K, Martinkauppi S, Ilmoniemi RJ, Naatanen R (2001) Memory traces for words as revealed by the mismatch negativity. Neuroimage 14:607-616.

Pulvermuller F, Preissl H, Lutzenberger W, Birbaumer N (1996) Brain rhythms of language: nouns versus verbs. Eur J Neurosci 8:937-941.

Pulvermuller F, Shtyrov Y (2006) Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. Prog Neurobiol 79:49-71.

Remez RE, Rubin PE, Pisoni DB (1983) Coding of the speech spectrum in three time-varying sinusoids. Ann N Y Acad Sci 405:485-489.

Remez RE, Rubin PE, Pisoni DB, Carrell TD (1981) Speech perception without traditional speech cues. Science 212:947-949.

Rubin P, Turvey MT, van Gelder P (1976) Initial phonemes are detected faster in spoken words than in spoken nonwords. Percept Psychophys 19:384-398.

Rugg MD (1995) Event-related potential studies of human memory. In: The cognitive neurosciences (Michael S.Gazzaniga, ed), pp 1341-1356. Cambridge, MA: MIT Press.

Saberi K, Perrott DR (1999) Cognitive restoration of reversed speech. Nature 398:760.

Samuel AG (1981a) Phonemic restoration: insights from a new methodology. J Exp Psychol Gen 110:474-494.

Samuel AG (1981b) The role of bottom-up confirmation in the phonemic restoration illusion. J Exp Psychol Hum Percept Perform 7:1124-1131.

Samuel AG (1997) Lexical activation produces potent phonemic percepts. Cognit Psychol 32:97-127.

Samuel AG (2001) Knowing a word affects the fundamental perception of the sounds within it. Psychol Sci 12:348-351.

Samuel AG, Kat D (1996) Early levels of analysis of speech. J Exp Psychol Gen 125:28-51.

Samuel AG, Ressler WH (1986) Attention within auditory word perception: insights from the phonemic restoration illusion. J Exp Psychol Hum Percept Perform 12:70-79.

Schneider TR, Debener S, Oostenveld R, Engel AK (2008) Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. Neuroimage 42:1244-1254.

Schroeder MR (1968) Reference signal for signal quality studies. J Acoust Soc Am 44:1735-1736.

Schroger E (1997) On the detection of auditory deviations: a pre-attentive activation model. Psychophysiology 34:245-257.

Schroger E, Tervaniemi M, Huotilainen M (2003) Bottom-up and top-down flows of information within auditory memory: electrophysiological evidence. In: Psychophysics beyond Sensation: Laws and Invariants of Human Cognition. (Kaernbach C, Schroger E, Mueller HJ, eds), pp 391-410. Hillsdale, NJ: Erlbaum.

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123 Pt 12:2400-2406.

Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. Trends Neurosci 26:100-107.

Scott SK, Rosen S, Wickham L, Wise RJ (2004) A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. J Acoust Soc Am 115:813-821.

Scott SK, Wise RJ (2004) The functional neuroanatomy of prelexical processing in speech perception. Cognition 92:13-45.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303-304.

Shtyrov Y, Kujala T, Palva S, Ilmoniemi RJ, Naatanen R (2000) Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. Neuroimage 12:657-663.

Shtyrov Y, Osswald K, Pulvermuller F (2008) Memory traces for spoken words in the brain as revealed by the hemodynamic correlate of the mismatch negativity. Cereb Cortex 18:29-37.

Singer W (1993) Synchronization of cortical activity and its putative role in information processing and learning. Annu Rev Physiol 55:349-374.

Singer W (1999) Neuronal synchrony: a versatile code for the definition of relations? Neuron 24:49-25.

Sinkkonen J, Tiitinen H, Naatanen R (1995) Gabor filters: an informative way for analysing event-related brain activity. J Neurosci Methods 56:99-104.

Sivonen P, Maess B, Lattner S, Friederici AD (2006) Phonemic restoration in a sentence context: evidence from early and late ERP effects. Brain Res 1121:177-189.

Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416:87-90.

Snyder JS, Large EW (2005) Gamma-band activity reflects the metric structure of rhythmic tone sequences. Brain Res Cogn Brain Res 24:117-126.

Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. Trends Cogn Sci 3:151-162.

Tallon-Baudry C, Bertrand O, Delpuech C, Permier J (1997) Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in humans. J Neurosci 17:722-734.

Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J (1996) Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. J Neurosci 16:4240-4249.

Tallon-Baudry C, Bertrand O, Henaff MA, Isnard J, Fischer C (2005) Attention modulates gamma-band oscillations differently in the human lateral occipital cortex and fusiform gyrus. Cereb Cortex 15:654-662.

Tallon-Baudry C, Bertrand O, Peronnet F, Pernier J (1998) Induced gamma-band activity during the delay of a visual short-term memory task in humans. J Neurosci 18:4244-4254.

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC (1995) Integration of visual and linguistic information in spoken language comprehension. Science 268:1632-1634.

Tiitinen H, Sinkkonen J, Reinikainen K, Alho K, Lavikainen J, Naatanen R (1993) Selective attention enhances the auditory 40-Hz transient response in humans. Nature 364:59-60.

Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cognit Psychol 12:97-136.

Warren RM (1970) Perceptual restoration of missing speech sounds. Science 167:392-393.

Warren RM (1984) Perceptual restoration of obliterated sounds. Psychol Bull 96:371-383.

Warren RM, Obusek CJ (1971) Speech perception and phonemic restorations. Percept Psychophys 9:358-363.

Warren RM, Obusek CJ, Ackroff JM (1972) Auditory induction: perceptual synthesis of absent sounds. Science 176:1149-1151.

Warren RM, Riener KR, Bashford JA, Jr., Brubaker BS (1995) Spectral redundancy: intelligibility of sentences heard through narrow spectral slits. Percept Psychophys 57:175-182.

Weisz N, Muller S, Schlee W, Dohrmann K, Hartmann T, Elbert T (2007) The neural code of auditory phantom perception. J Neurosci 27:1479-1484.

Widmann A, Gruber T, Kujala T, Tervaniemi M, Schroger E (2007) Binding symbols and sounds: evidence from event-related oscillatory gamma-band activity. Cereb Cortex 17:2696-2702.

Yordanova J, Kolev V, Demiralp T (1997) The phase-locking of auditory gamma band responses in humans is sensitive to task processing. Neuroreport 8:3999-4004.

Yuval-Greenberg S, Deouell LY (2007) What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. J Neurosci 27:1090-1096.

Yuval-Greenberg S, Tomer O, Keren AS, Nelken I, Deouell LY (2008) Transient induced gamma-band response in EEG as a manifestation of miniature saccades. Neuron 58:429-441.

Zwitserlood P (1989) The locus of the effects of sentential-semantic context in spoken-word processing. Cognition 32:25-64.