available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Top-down knowledge supports the retrieval of lexical information from degraded speech

R. Hannemann[a,*], J. Obleser[b], C. Eulitz[a]

[a]Department of Linguistics, University of Konstanz, Konstanz, Germany
[b]Institute of Cognitive Neuroscience, University College London, London, UK

ARTICLE INFO

Article history:
Accepted 16 March 2007
Available online 28 March 2007

Keywords:
Speech perception
Gammaband activity
Auditory neuroscience

ABSTRACT

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, the present study investigated the modulation of the induced gamma-band activity (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. We conclude 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. Introduction

Successful speech perception is an extraordinary human ability, and the human brain is capable of creating a meaningful speech percept under acoustically adverse conditions. Experimentally, this has been shown using speech embedded in noise (Miller et al., 1951; Boothroyd and Nittrouer, 1988), spectrally degraded or reduced (e.g., Shannon et al., 1995; Warren et al., 1995) or otherwise distorted speech (Saberi and Perrott, 1999). In these cases, successful comprehension cannot be achieved on the basis of sensory information alone.

This study 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

* *Corresponding author*. Department of Linguistics, University of Konstanz, PO Box 174, 78457 Konstanz, Germany.
E-mail address: ronny.hannemann@uni-konstanz.de (R. Hannemann).

0006-8993/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2007.03.069

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 (Engel and Singer, 2001). According to the Adaptive Resonance Theory (ART) (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-band range when comprehension of degraded speech succeeds due to previously strengthened lexical memory traces.

Measured with electro- (EEG) and magnetoencephalogram (MEG), enhanced gamma-band activity (GBA) 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., 2006), 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., 2004; Gruber et al., 2004; Osipova et al., 2006) as well as 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 the present study, 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 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 were 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., 2006). 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., 2007).

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, it 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. We assume 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, we designed an EEG experiment consisting of three experimental sequences. As a baseline measure subjects listened passively to all stimuli in an acoustically degraded format (baseline sequence, see also Experimental procedures and Fig. 1). 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

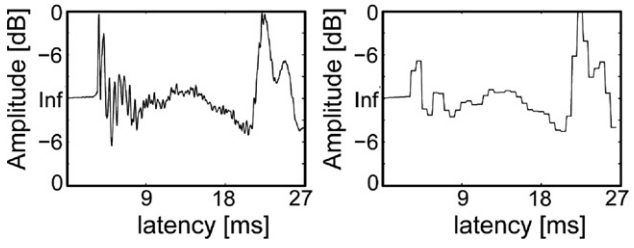

Fig. 1 – Oscillogram of the first 27 ms of the word 'baby' is depicted in original (left) and in degraded form (right).

(Condition = correctly identified or not identified). If our assumption about enhanced induced GBA as a correlate for a successful match holds, we would expect a significant Condition x Sequence interaction. Appropriate time and frequency ranges for those analyses were defined by permutation tests (Blair and Karniski, 1993).

2. Results

2.1. 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 ($F(1,19) = 0.84$; $p < 0.37$).

2.2. Induced brain responses

In absence of a precise presumption about the topography of speech-induced GBA enhancement, we defined six electrode clusters (frontal, anterior temporal and posterior temporal sites; Fig. 2) to capture a wide range of possible cortical sources of brain activity in our analysis. This yielded a 2×2×2×3 factorial design with two levels of Condition (correctly identified–not identified), two levels of Sequence (baseline–test), two levels of Hemisphere (left–right) and three levels of Position (frontal–anterior temporal–posterior temporal). Permutation tests (Blair and Karniski, 1993), which were applied to all bins of 17–48 Hz to reveal exact time-frequency (TF) bins during which the expected enhancement is evident (see Experimental procedures), identified one cluster of time-frequency (TF) bins over anterior temporal electrode sites in the left hemisphere that passed all criteria for further statistical analyses (First, only bins in the 17–48 Hz range showing an enhancement of at least one standard deviation compared to the baseline activity for the correctly identified items in the test sequence; second, the difference between test sequence and baseline sequence for the correctly identified versus not identified comparison was required to be significant at a $p < 0.01$, uncorrected; third, at least four contiguous bins).

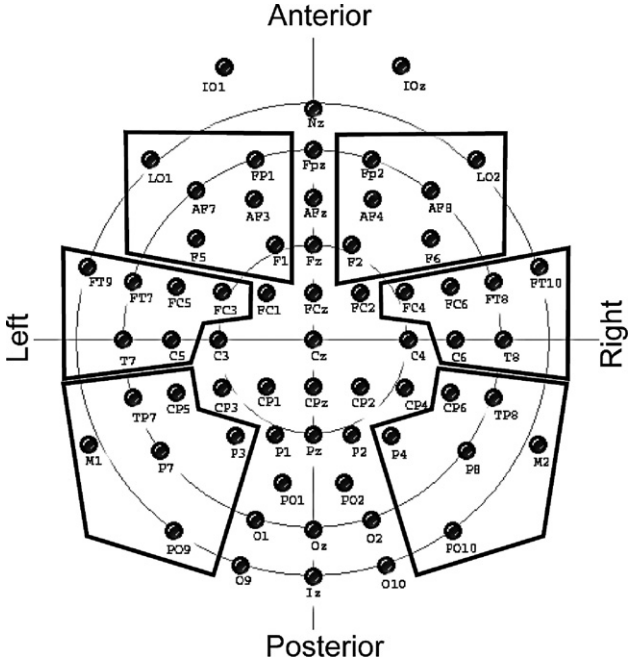

Fig. 2 – Electrode montage and groupings used for statistical analyses are shown.

The time-frequency plots in Fig. 3 depict the differences of correctly identified over not identified items in the test sequence compared to the baseline sequence averaged into electrode groupings (see Fig. 2). 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 to 360 ms.

Fig. 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.

A four-way repeated-measures ANOVA for this time and frequency range revealed a significant Condition × Sequence × Hemisphere × Position Interaction, (Wilks Lambda $F(1,18) = 9.94$, $p < 0.001$) 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 × Sequence interaction for the change in spectral power, $F(1,19) = 17.72$, $p < 0.001$ for left anterior temporal electrode sites. For these electrode sites the mean spectral power is depicted in Fig. 5. Further post-hoc comparisons showed a significant main effect for Sequence for correctly identified items, $F(1,19) = 9.25$; $p < 0.007$, 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

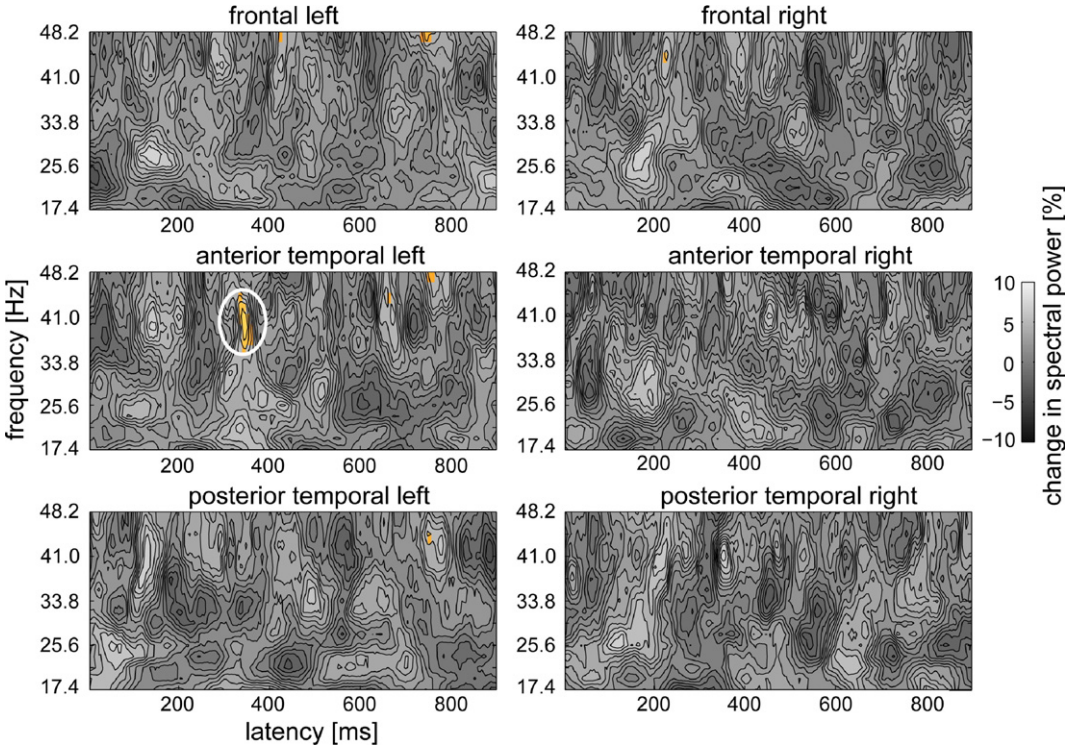

Fig. 3 – Grand mean TF plots of percentual change in induced brain activity over six brain regions (see Fig. 2 for electrode specifications). Depicted is the difference of correctly identified minus not identified items in the test sequence relative to the corresponding difference of the baseline sequence. 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 encircled in white between 330 and 360 ms fulfills all criteria for further statistical analyses and constitutes the main finding of the study.

Condition in the test sequence, $F(1,19)=7.44$; $p<0.013$ as well as (barely significant) in the baseline sequence, $F(1,19)=4.44$, $p<0.049$. 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 does not show a significant main effect.

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 Fig. 4a also points to an earlier enhancement in the latency range between 160 and 200 ms. Repeated-measures ANOVA confirmed that this effect is also elicited over left anterior temporal sites (Condition × Sequence × Hemisphere × Condition $F(2,38)=5.77$, $p<0.007$) and shows a response pattern that closely resembles the ensuing robust GBA effect (Sequence × Condition interaction $F(1,19)=10.68$, $p<0.004$). 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.

2.3. *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. Fig. 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, Fig. 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 ($p<0.093$). Fig. 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 × Sequence interaction or main effect of Condition, which could be 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 ($F(1,19)=6.12$; $p<0.02$).

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 the gamma-band range are not solely due to

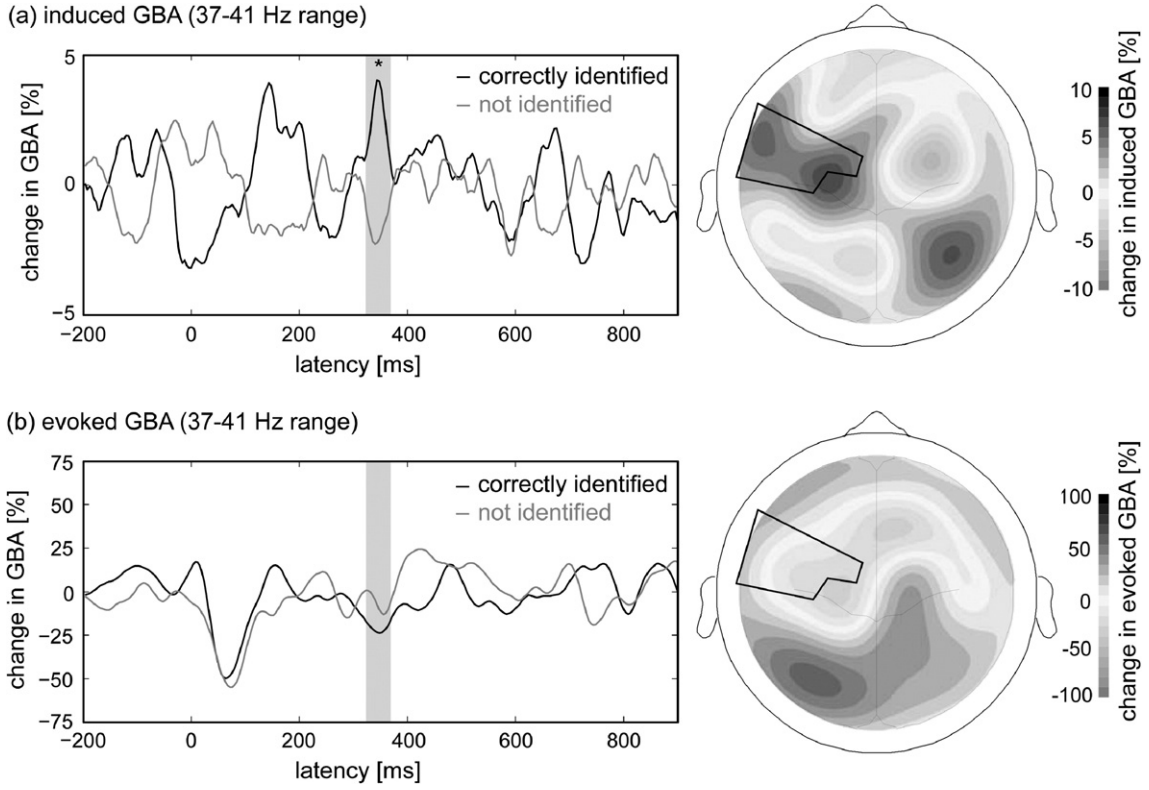

Fig. 4 – Depiction of gamma band activity in the 38–41 Hz range. Left: Comparison of the time courses of induced (upper panel) and evoked GBA (lower panel) over left anterior temporal electrode sites. Black lines represent the correctly identified items (test minus baseline sequence) and gray lines show 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 electrode sites (cf. Fig. 2) 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.

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 ($p < 0.25$) 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 et al., 1990) yielded no results as obtained for the 38–41 Hz frequency

Table 1 – Mean spectral power for the 38–41 Hz / 330–360 ms range, averaged across six electrode sites and standard error of mean (SEM) in % change

Position	Hemisphere	Sequence	Correctly identified		Not identified	
			%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)

Figures in bold type indicate the main result of this study.

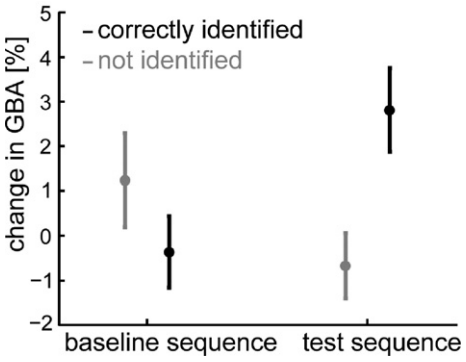

Fig. 5 – Mean values and standard error of mean of the change in induced GBA for 38–41 Hz range in 330–360 ms latency range for left anterior temporal electrode sites are shown.

range (specifically no Condition x Sequence x Hemisphere x Position interaction $F < 1$).

3. Discussion

This study 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. We found one marked increase in induced gamma power at left anterior temporal electrode sites 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., 2006) as well as for the auditory domain by recent results of Lenz et al. (2007) 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. (2007) 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 topographical distribution we observed, 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., 2006). In our study, 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. 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 explanations, 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) intelligibility of degraded speech stimuli is accompanied by brain activity in inferior frontal as well as anterior and lateral temporal areas in passive listening tasks. Giraud and colleagues (Giraud et al., 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 to 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 counter-balanced 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.

3.1. *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. Our findings 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.

4. Experimental procedures

4.1. *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-min pre-test and received the same reimbursement. Informed consent was obtained from all participants. All 35 subjects were monolingual Germans.

4.2. *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 bisyllablic, 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. 1). 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).

4.3. *Experimental design*

The experiment 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, we 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 button press and 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.

4.4. *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. 2 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. This method forms a good compromise between frequency and time resolution (Sinkkonen et al., 1995). It provides a time-varying magnitude of the signal in each frequency band, leading to time-frequency (TF) representations of the signal. Time by frequency energy is averaged across single trials, allowing one to analyze non-phase-locked frequency components. This method is described in detail elsewhere (e.g., Tallon-Baudry et al., 1997). 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. 2). 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-band 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 for further analyses. Further statistical analyses were performed only on time-frequency bins for which the permutation test showed 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 did not form a cluster of at least 4 bins in the temporally smoothed data set before we calculated any ANOVA. Fig. 3 shows the result of the permutation tests on the temporary data after the described data reductions, with significant clusters in color.

To substantiate our findings, a four-way repeated-measures ANOVA Condition (correctly identified vs. not identified) × Sequence (baseline vs. test) × Hemisphere (left vs. right) × Position (anterior, medial, posterior) was performed on the original (unsmoothed) data for all time-frequency regions clusters surviving the initial selection process (Fig. 3). 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 (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 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 that had or had not occurred in non-degraded form in the exposure sequence, 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.

Acknowledgments

The current research was supported by grants of the German Science Foundation to C.E. (SFB 471) and the Leibniz Preis awarded to A. Lahiri. J.O. is currently funded through a Welcome Senior Research Fellowship of Sophie K. Scott and receives additional support from the Landesstiftung Baden-Württemberg. We further wish to thank Barbara Awiszus and Christiane Wolf for their help in EEG data acquisition and Andreas Keil for helpful comments.

REFERENCES

Berg, P., Scherg, M., 1994. A multiple source approach to the correction of eye artifacts. Electroencephalography and Clinical Neurophysiology 90 (3), 229–241.

Blair, R.C., Karniski, W., 1993. An alternative method for significance testing of waveform difference potentials Psychophysiology 30 (5), 518–524.

Boothroyd, A., Nittrouer, S., 1988. Mathematical treatment of context effects in phoneme and word recognition. Journal of the Acoustical Society of America 84 (1), 101–114.

Busch, N.A., Herrmann, C.S., Muller, M.M., Lenz, D., Gruber, T., 2006. A cross-laboratory study of event-related gamma activity in a standard object recognition paradigm. Neuroimage 33 (4), 1169–1177.

Cacioppo, J.T., Tassinary, L.G., Fridlund, A.J., 1990. The skeletomotor system. In: Cacioppo, J.T., Tassinary, L.G. (Eds.), Principles of Psychophysiology: Physical, Social, and Inferential Elements. Cambridge UP, Cambridge, MA, pp. 324–385.

Crone, N.E., Boatman, D., Gordon, B., Hao, L., 2001. Induced electrocorticographic gamma activity during auditory perception. Clinical Neurophysiology 112 (4), 565–582.

Debener, S., Herrmann, C.S., Kranczioch, C., Gembris, D., Engel, A.K., 2003. Top-down attentional processing enhances auditory evoked gamma band activity. Neuroreport 14 (5), 683–686.

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 (3), 243–246.

Engel, A.K., Singer, W., 2001. Temporal binding and the neural correlates of sensory awareness. Trends in Cognitive Sciences 5 (1), 16–25.

Eulitz, C., Maess, B., Pantev, C., Friederici, A.D., Feige, B., Elbert, T., 1996. Oscillatory neuromagnetic activity induced by language and non-language stimuli. Brain Research. Cognitive Brain Research 4 (2), 121–132.

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 (4), 447–455.

Friederici, A.D., Kotz, S.A., 2003. The brain basis of syntactic processes: functional imaging and lesion studies. Neuroimage 20 (Suppl 1), S8–S17.

Giraud, A.L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M.O., Preibisch, C., Kleinschmidt, A., 2004. Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. Cerebral Cortex 14 (3), 247–255.

Goffaux, V., Mouraux, A., Desmet, S., Rossion, B., 2004. Human non-phase-locked gamma oscillations in experience-based perception of visual scenes. Neuroscience Letters 354 (1), 14–17.

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

Grossberg, S., 1999. The link between brain learning, attention, and consciousness. Consciousness and Cognition 8 (1), 1–44.

Gruber, T., Muller, M.M., Keil, A., 2002. Modulation of induced gamma band responses in a perceptual learning task in the human EEG. Journal of Cognitive Neuroscience 14 (5), 732–744.

Gruber, T., Tsivilis, D., Montaldi, D., Muller, M.M., 2004. Induced gamma band responses: an early marker of memory encoding and retrieval. Neuroreport 15 (11), 1837–1841.

Haenschel, C., Baldeweg, T., Croft, R.J., 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. Proceedings of the National Academy of Sciences of the United States of America 97 (13), 7645–7650.

Herrmann, C.S., 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, C.S., Lenz, D., Junge, S., Busch, N.A., Maess, B., 2004. Memory-matches evoke human gamma-responses. BMC Neuroscience 5, 13.

Kaiser, J., Lutzenberger, W., 2004. Frontal gamma-band activity in magnetoencephalogram during auditory oddball processing. Neuroreport 15 (14), 2185–2188.

Kaiser, J., Lutzenberger, W., Ackermann, H., Birbaumer, N., 2002. Dynamics of gamma-band activity induced by auditory pattern changes in humans. Cerebral Cortex 12 (2), 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 (2), 816–827.

Kaiser, J., Lennert, T., Lutzenberger, W., 2006. Dynamics of oscillatory activity during auditory decision making. Cerebral Cortex (Epub ahead of print).

Knief, A., Schulte, M., Bertrand, O., Pantev, C., 2000. The perception of coherent and non-coherent auditory objects: a signature in gamma frequency band. Hearing Research 145, 162–168.

Lebedev, M.A., Messinger, A., Kralik, J.D., Wise, S.P., 2004. Representation of attended versus remembered locations in prefrontal cortex. PLoS Biology 2 (11), e365.

Lenz, D., Schadow, J., Thaerig, S., Busch, N.A., Herrmann, C.S., 2007. What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. International Journal of Psychophysiology 64 (1), 31–38.

Lutzenberger, W., Pulvermuller, F., Birbaumer, N., 1994. Words and pseudowords elicit distinct patterns of 30-Hz EEG responses in humans. Neuroscience Letters 176 (1), 115–118.

Miller, G.A., Heise, G.A., Lichten, W., 1951. The intelligibility of

speech as a function of the context of the test materials. Journal of Experimental Psychology 41 (5), 329–335.
Muller, M.M., Gruber, T., 2001. Induced gamma-band responses in the human EEG are related to attentional information processing. Visual Cognition 8, 579–592.
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 (3), 1839–1847.
Obleser, J., Wise, R.J., Dresner, M.A., Scott, S.K., 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. Journal of Neuroscience 27 (9), 2283–2289.
Osipova, D., Takashima, A., Oostenveld, R., Fernandez, G., Maris, E., Jensen, O., 2006. Theta and gamma oscillations predict encoding and retrieval of declarative memory. Journal of Neuroscience 26 (28), 7523–7531.
Pantev, C., 1995. Evoked and induced gamma-band activity of the human cortex. Brain Topography 7 (4), 321–330.
Pulvermuller, F., 1999. Words in the brain's language. Behavioral and Brain Sciences 22 (2), 253–279.
Pulvermuller, F., Birbaumer, N., Lutzenberger, W., Mohr, B., 1997. High-frequency brain activity: its possible role in attention, perception and language processing. Progress in Neurobiology 52 (5), 427–445.
Rugg, M.D., 1995. Event-related potential studies of human memory. In: Gazzaniga, Michael S. (Ed.), The cognitive neurosciences. MIT Press, Cambridge, MA, pp. 1341–1356.
Saberi, K., Perrott, D.R., 1999. Cognitive restoration of reversed speech. Nature 398 (6730), 760.
Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends in Neurosciences 26 (2), 100–107.
Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123 (Pt 12), 2400–2406.
Scott, S.K., Rosen, S., Wickham, L., Wise, R.J., 2004. A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. Journal of the Acoustical Society of America 115 (2), 813–821.
Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270 (5234), 303–304.
Sinkkonen, J., Tiitinen, H., Naatanen, R., 1995. Gabor filters: an informative way for analysing event-related brain activity. Journal of Neuroscience Methods 56 (1), 99–104.
Tallon-Baudry, C., Bertrand, O., 1999. Oscillatory gamma activity in humans and its role in object representation. Trends in Cognitive Sciences 3 (4), 151–162.
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. Journal of Neuroscience 16 (13), 4240–4249.
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. Journal of Neuroscience 17 (2), 722–734.
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. Journal of Neuroscience 18 (11), 4244–4254.
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 (6432), 59–60.
Warren, R.M., Riener, K.R., Bashford Jr., J.A., Brubaker, B.S., 1995. Spectral redundancy: intelligibility of sentences heard through narrow spectral slits. Perception and Psychophysics 57 (2), 175–182.
Yordanova, J., Kolev, V., Demiralp, T., 1997. The phase-locking of auditory gamma band responses in humans is sensitive to task processing. Neuroreport 8 (18), 3999–4004.