

Multiple brain signatures of integration in the comprehension of degraded speech

Jonas Obleser^{a,*}, Sonja A. Kotz^b

^a Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

^b Research Group "Neurocognition of Rhythm in Communication", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Article history:

Received 17 August 2010
Revised 26 November 2010
Accepted 6 December 2010
Available online 21 December 2010

Keywords:

EEG
Auditory
Speech comprehension
Semantic processing
N100
N400
Gamma

ABSTRACT

When listening to speech under adverse conditions, expectancies resulting from semantic context can have a strong impact on comprehension. Here we ask how minimal variations in semantic context (cloze probability) affect the unfolding comprehension of acoustically degraded speech. Three main results are observed in the brain electric response. First, auditory evoked responses to a degraded sentence's onset (N100) correlate with participants' comprehension scores, but are generally more vigorous for more degraded sentences. Second, a pronounced N400 in response to low-cloze sentence-final words, reflecting the integration *effort* of words into context, increases linearly with improving speech intelligibility. Conversely, transient enhancement in Gamma band power (γ , ~40–70 Hz) during high-cloze sentence-final words (~600 ms) reflects top-down-*facilitated* integration. This γ -band effect also varies parametrically with signal quality. Third, a negative correlation of N100 amplitude at sentence onset and the later γ -band response is found in moderately degraded speech. This reflects two partly distinct neural strategies when dealing with moderately degraded speech; a more "bottom-up," resource-allocating, and effortful versus a more "top-down," associative and facilitatory strategy. Results also emphasize the non-redundant contributions of phase-locked (evoked) and non-phase-locked (induced) oscillatory brain dynamics in auditory EEG.

© 2010 Elsevier Inc. All rights reserved.

Introduction

The influence of semantic context in listening to, analysing and comprehending speech signals is well known. However, the mechanisms behind context-aided speech comprehension have remained surprisingly opaque. This holds especially true for speech comprehension in noisy environments where semantic context is thought to be highly relevant (e.g. Miller et al., 1951; Stickney and Assmann, 2001).

Increasing efforts are made to study the functional neuroanatomy behind this complex sensory–perceptive–cognitive process of speech comprehension in a more integrative approach. Various recent psycholinguistic (Shinn-Cunningham and Wang, 2008; van Linden and Vroomen, 2007) and neuroscientific (Davis et al., 2007; Davis and Johnsrude, 2007; Hannemann et al., 2007; Obleser et al., 2007; Sivonen et al., 2006; van Atteveldt et al., 2007) studies of speech-receptive function were devoted to those factors in speech communication that lie beyond the bottom-up mechanisms of the auditory pathways in analysing the signal itself. Besides semantic meaning in the narrow sense, a whole family of influences facilitate and disambiguate speech comprehension, especially under acoustically compromised conditions, amongst them talker familiarity (e.g., Eisner and McQueen, 2005); prior experience and knowledge (e.g., Hanne-mann et al., 2007); and prosodic and emotional cues (e.g., Gandour

et al., 2004). The main focus of this investigation is on the potential relationship of early, sensory-driven processes, as reflected in early EEG components (N100) and later cognitive–linguistic integration processes, as reflected in the sentence-final N400 and the Gamma-band response.

In the present study, we aimed at narrowing down the possible top-down influences and constrained the sentence context to a minimal and constant one. We employed the well-studied measure of cloze probability. Originally developed as a measure of text readability (Taylor, 1953), it is an empirical measure for lexical expectancy of a given word in a context. Variations of it have been adopted into important studies of semantic processing, mostly in N400 electroencephalography (EEG) designs (Connolly et al., 1995; Federmeier et al., 2007; Gunter et al., 2000; Kutas and Hillyard, 1984). A stronger N400 in response to the sentence-final key word is elicited by a low-cloze sentence such as "she weighs the flour" compared to an (otherwise identical) high-cloze version such as "she sifts the flour". This N400 effect is commonly interpreted to reflect higher cognitive effort to integrate the low-cloze (higher lexical competition) elements of the sentence. A previous study by Aydelott and colleagues had utilised degradation of the sentence context (but not the target word) by 1-kHz-low-pass filtering, while presenting the sentence word by word with 200-ms intervals (Aydelott et al., 2006). The study delivered important initial confirmation that acoustic degradation of the context influences the magnitude of the integration-related N400: No proper N400-like response to incongruent target words built up when the preceding sentence material had been strongly low-pass

* Corresponding author. Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany.

E-mail address: obleser@cbs.mpg.de (J. Obleser).

filtered. However, the current study goes a step further by using a parametric (i.e., more than twofold) variation of the signal degradation; by utilizing naturally spoken sentences; and by taking a comprehensive look at various EEG signatures (i.e., the evoked or phase-locked N100 and N400 potentials, and the induced or non-phase-locked Gamma-band response, see next paragraph).

Especially at the point of sentence meaning integration (i.e., the sentence-final keyword), we expect to observe distinct influences of semantic expectancy constraints on degraded speech in the stimulus-correlated EEG: The N400, as described above, is unanimously interpreted as a signature of integration effort; we therefore expect an enhanced N400 at low-expectancy words. As outlined below, we also expect facilitation of integration (an enhancement for high-expectancy words) being reflected in a modulation of Gamma-band (γ , >30 Hz) oscillations.

Apart of the N400 literature, a whole line of evidence points to an increased effect (i.e., higher amplitudes in response to highly predictable sentence endings) for Gamma-band (γ , >30 Hz) oscillations. Many studies on the top-down formation of percepts have focussed on this most prominent electrophysiological signature in the visual domain (e.g., Gruber et al., 2002; Tallon-Baudry and Bertrand, 1999), in multisensory integration (e.g., Schneider et al., 2008), but also in the auditory (e.g., Lenz et al., 2007), and the speech domain (Hannemann et al., 2007; Shahin et al., 2009). All of these indicate that facilitation of percept formation (“gestalt”) through memory representations (“adaptive resonance,” Grossberg, 2009) is accompanied by comparably focal bursts of enhanced Gamma-band synchrony (for review see e.g., Fries, 2009). This synchrony surfaces as power enhancements (compared to selected baseline periods) in the Gamma-band range when analysing the EEG data for non-phase-locked oscillations.

In order to allow for a precise spectro-temporal control over the speech degradation levels applied, we chose noise-band vocoding (e.g., Shannon et al., 1995) as a manipulation technique. It allows for exact control over the spectral detail being presented (and hence the speech intelligibility) in arbitrary levels (for review see Shannon et al., 2004). It was originally designed to mimic the spectro-temporal characteristics of the input a cochlear implant carrier would receive (Faulkner et al., 2001). Published evidence on the EEG responses to noise-vocoding in normal-hearing listeners has been lacking so far. Unlike fMRI, EEG offers to potentially reveal (and decouple the time courses of) partly differential mechanisms that underlie speech comprehension, such as more “bottom-up,” effortful resource-allocation versus rather “top-down” associative facilitation.

The current experiment utilises a simple 2×3 factorial design: first, simple sentences were varied in the strength of the semantic expectancy coupling within that sentence (cloze probability). Second, these sentences were subjected to multiple levels of speech degradation in order to disentangle the (possibly interactive) mechanisms of *degraded signal quality* (a well-controlled sensory-acoustic parameter) and *cloze probability* (a well-controlled cognitive-linguistic parameter) on deflections of the auditory event-related potential.

In an early time range (approximately 100 ms after sound onset), effects of the degree of speech degradation are expected on the N100 component which has been shown to index early abstraction and percept-formation stages (Krumbholz et al., 2003; Naatanen, 2001; Obleser et al., 2006), especially in the context of an active comprehension task (for task effects on the N100/N100m see Bonte et al., 2009; Obleser et al., 2004a; Poeppel et al., 1996). Also based on a recent study finding enhanced N100 amplitudes in response to degraded sound (Miettinen et al., 2010), we expect the following: the more thorough the degradation of the signal, the more neural effort is likely be allocated to encoding the acoustic signal and mapping it onto known phonological and lexical categories, leading to an enhanced N100 amplitude measured on the scalp.

For the typical locus of semantic integration effort (N400, 300–500 ms post onset of the critical word in the sentence context), we

expect an enhanced N400 to the sentence-final word in low-cloze probability sentences; however, the influence of various levels of speech degradation on the build-up of the sentence context is expected to modulate the N400 (cf. Aydelott et al., 2006).

Lastly, we have good reason to expect enhanced synchronisation in the Gamma band frequency range for facilitated integration of meaning for high-cloze more than for low-cloze sentences, but its dependency on acoustic signal quality and its possible interaction with other signatures of the speech-evoked EEG will be of particular interest.

Materials and methods

Participants

Thirty participants (15 females; age range 19–32 years) took part in this experiment. All were native speakers of German, had normal hearing and no history of neurological or language-related problems. They had no prior experience with noise-vocoded speech and had not taken part in any of the pilot experiments or a recent functional MRI study (Obleser and Kotz, 2010) using the same sentence material. Participants received financial compensation of 8 € per hour.

Stimulus material

Stimuli were recordings of spoken German sentences, all consisting of pronoun (Er or Sie), verb and object, in present tense. Every sentence with the framing constituents pronoun and object was used in two versions, incorporating either a verb low in cloze probability (e.g., Sie wiegt das Mehl [she weighs the flour]) or high in cloze probability (e.g., Sie siebt das Mehl [she sifts the flour]). Previous to its use in Obleser and Kotz (2010), the material had been developed, tested and used in a visual-stimulation experiment on cloze probability (Gunter et al., 2000). It consists of 40 base (to give another example, [he ... the beer]) sentences in a low (e.g., [...sees...]) and a high (e.g., [...drinks...]) cloze probability version, respectively. Gunter et al., 2000, also showed that the sentences have empirical cloze probabilities of $15 \pm 5\%$ (mean \pm standard deviation; maximum 24) for the low-cloze versions and $74 \pm 13\%$ (minimum 56) for the high-cloze versions.

For the present study, one male and one female trained speaker recorded the sentence material (originally sampled at a 44.1 kHz rate, and recorded in a sound-proof booth). Offline editing included resampling to 22.05 kHz, cutting at zero-crossings before and after each sentence, and root mean square normalization of amplitudes. From each final audio file (40 base sentences \times 2 cloze levels \times 2 speakers; 160 recordings), various spectrally degraded versions were created using a Matlab™-based noise-band vocoding algorithm. Originally developed to mimic degraded speech input in cochlear implantation, noise vocoding is an effective technique to manipulate the spectral detail while preserving the temporal envelope of the speech signal and render it more or less intelligible in a graded and controlled fashion, depending on the number of bands used, with more bands yielding a more intelligible speech signal (Shannon et al., 1995; for estimates on spectral detail actually provided by cochlear implants see Fishman et al., 1997; Zeng and Galvin, 1999). A substantial amount of data has been recently gathered in brain imaging studies employing this technique (Davis and Johnsrude, 2003; Obleser et al., 2007; Scott et al., 2000, 2006; Warren et al., 2006), but—as of yet—no EEG studies are available.

Own pilot data acquired in a different participant sample identified 4-band vocoded speech to yield a significant gain in comprehension when a sentence was presented in its high- rather than its low-cloze version (Obleser and Kotz, 2010). We chose to use a pass band of up to 9 kHz that covers most of the critical spectrum of speech (cf. Davis et al., 2005; Obleser et al., 2007, 2008). We then noise-band-vocoded

all sentence files into three different versions, a 1-band version (yielding essentially signal-correlated noise with no spectral detail whatsoever), a 4-bands version and a spectrally very detailed (i.e. not very degraded) 16-bands version.

Experimental procedures

Task

During EEG recording, participants performed a comprehension-rating task to keep them in an alert listening mode without forcing their attention on particular linguistic or acoustic aspects of the stimulus material. Participants were required to listen attentively to the sentences and to indicate by way of a four-way button system how comprehensible this trial's sentence had been. Indicating (or rating)

comprehension comes close to natural listening to speech, and the rating technique shows remarkable consistency with actual recognition scores (Fig. 1; see Davis and Johnsrude, 2003, where rating and recognition scores within participants showed a correlation score of .98; see also Obleser et al., 2008). Fig. 1 illustrates the task during EEG recordings. Three seconds after each stimulus (with sentences' duration varying naturally up to a maximal duration of 1.4 s), a Question mark occurred on the screen prompting for the participant's button press. The question mark disappeared after the button press, or after 2 s in case of a no button press. The response period was followed by a 2-s period (indicated by an eye symbol present for 500 ms) during which participants were instructed to blink if necessary. Then, with a random stimulus onset asynchrony of 500–1500 ms, the next trial began. Button-to-value assignment was

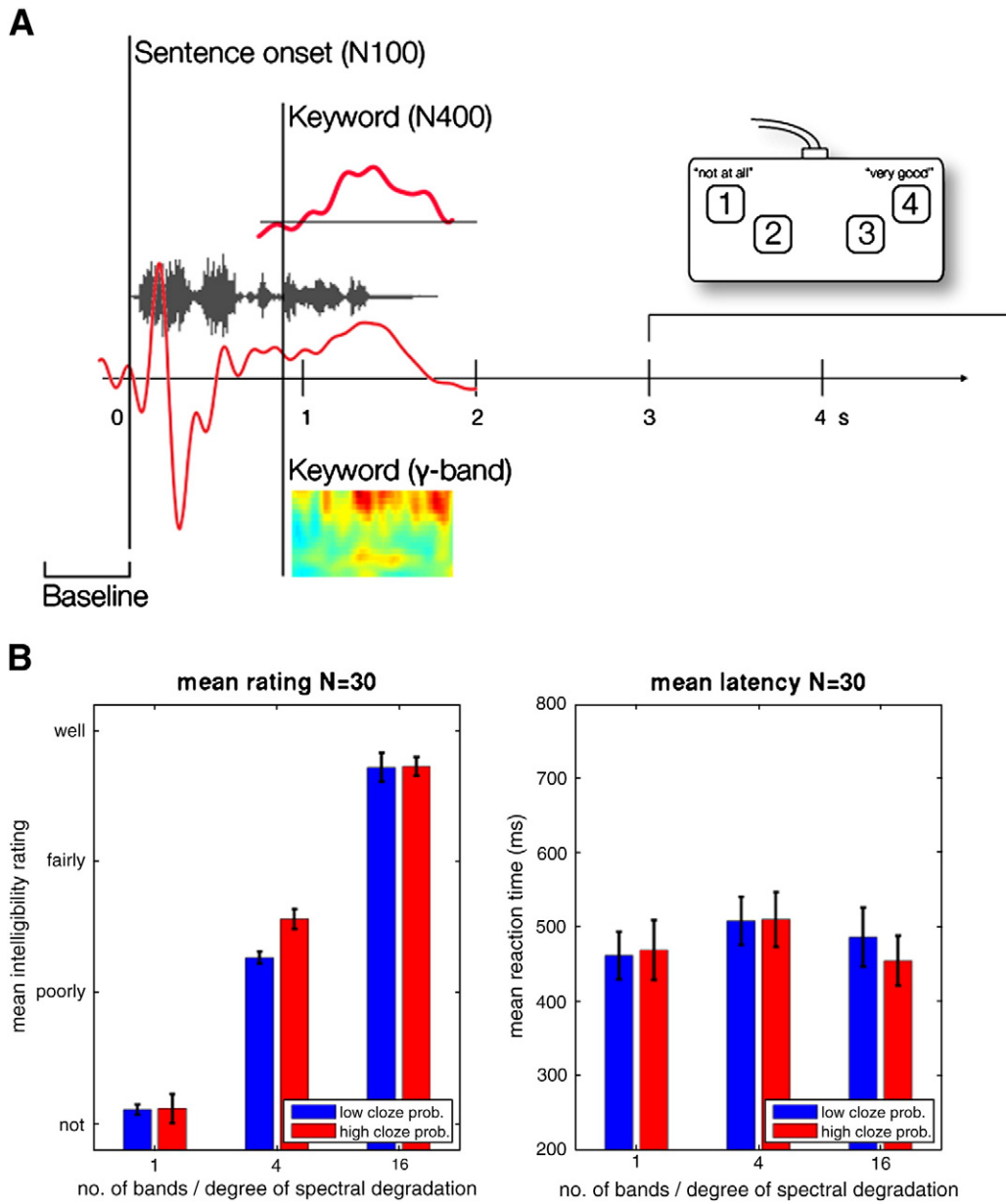


Fig. 1. A. Schematic Display of Experiment and Analyses. An exemplary time course of a trial is given, with indication of the three major analyses: N100 phase-locked to sentence or sound onset; N400 phase-locked to onset of the sentence-final keyword; Changes in the induced (non-phase-locked) Gamma (γ) band power during or following the sentence-final keyword. Three seconds after trial onset, the participant was prompted for a behavioural (rating) response of comprehension. B. Behavioural results. Mean ratings (left panel) and mean reaction times (right panel; both ± 1 standard error of the mean) are shown, split in levels of degradation (number of bands) and semantic expectancy (high cloze, red bars; low cloze, blue bars).

counterbalanced across subjects (with four buttons in total and two buttons per hand, the “very intelligible” response was either mapped to the leftmost or the rightmost digit).

Electroencephalogram acquisition

The electroencephalogram (EEG) was recorded from 64 Ag–AgCl electrodes mounted on a custom-made cap (Electro-Cap International) according to the modified expanded 10–20 system. Signals were recorded continuously with a passband of DC to 200 Hz and digitized at a sampling rate of 500 Hz, referenced against the left mastoid. Bipolar horizontal and vertical EOGs were recorded for artefact rejection purposes. Electrode resistance was kept under 5 k Ω .

After a brief (10-trial) familiarization period, the actual experiment was started. Subjects were required to listen attentively to noise-vocoded sentences, which were pseudo-randomly drawn from one of the six conditions (high or low cloze probability, see above, in either 1-, 4-, or 16-band noise-vocoded format. 40 trials of each condition amounted to a total of 240 trials, lasting approximately 40 minutes.

Data analysis

Data were re-referenced offline to linked mastoids, and a 0.01-Hz highpass finite impulse response (FIR) filter was applied. To analyse early event-related responses (N100) and to characterise the overall waveform morphology, ERPs were averaged for epochs of 2000 ms including a 200-ms pre-stimulus baseline. Trials contaminated by EOG or movement artefacts were rejected when exceeding a threshold of 40 μ V. A zero-phase shift 15-Hz low-pass Butterworth filter was applied to the evoked potential waveforms.

The N400 analysis was based on similarly processed averages, which were time-locked to the sentence-final keyword phrase of each stimulus sentence instead. More specifically, the acoustically less variant beginning of the keyword phrase “[the flour]” rather than the keyword “[flour]” directly was used for triggering the ERP.

All effects of interest were most evident over the midline electrodes, and an exploratory region of interest analyses comparing left and right electrode effects did not yield evidence for lateralisation on any effect whatsoever on the scalp level. Therefore, statistical results for the evoked (phase-locked) potentials are reported for a midline region of interest including channels Fz, FCz, Cz, CPz, Pz and POz. Averaged data were submitted to a 3 \times 2 repeated measures ANOVA with factors degradation (1-, 4-, or 16-band; degrees of freedom were Greenhouse–Geiser-corrected where adequate) and cloze probability (low, high).

First, to account for the experimental nature of this study, consecutive analyses of variance (ANOVA) were performed on the whole average epoch in 50-ms time windows. Based on these systematic statistical tests and close visual inspection, time windows of interest were defined and analysed for further statistical significance of mean ERP amplitude differences. Lastly, correlations of the participants' comprehension ratings with their individual ERP amplitudes in the N100 and N400 time windows were quantified.

Analysis of non-phased locked (induced) activity

As outlined in the introduction, we also analysed the non-phase-locked brain oscillations in response to the sentence-final keyword. To this end, we used all condition- and subject-specific epochs (1-s time window post sentence-final keyword onset, and a 1-s baseline window pre sentence onset; derived from the artefact-rejected and high-pass filtered continuous data, re-referenced to an average reference) and analysed them using Fieldtrip (<http://www.ru.nl/fcdonders/fieldtrip>; an open source Matlab toolbox for EEG and MEG

data analysis developed at the F.C. Donders Centre for Cognitive Neuroimaging). This processing step was performed for 28 of 30 subjects; two subjects had to be excluded due to too few trials being left after artefact rejection for robust time–frequency estimates.

We used an implemented time–frequency analysis approach using Morlet wavelets (Tallon-Baudry and Bertrand, 1999; Tallon-Baudry et al., 1997), with which the time series were convolved (using a standard factor for time/frequency-resolution of m of 7; cf. e.g., Gruber et al., 2002; Hannemann et al., 2007). Wavelet-based approaches for estimating time–frequency representations of EEG data form a good compromise between frequency and time resolution and are especially suited for detecting induced high-frequency oscillations that may occur during brief periods of time.

We focused on relative changes of Gamma (γ) power during the sentence-final keyword (250 to 750 ms after keyword onset; thus comparable to the time range of the described N400 response in the evoked EEG analysis) and for the frequency range of 30 to 80 Hz, since we had a clear *a priori* hypothesis on γ -band activity as a top-down modulated EEG parameter from previous experiments using degraded input (both visual, Gruber et al., 2002, and auditory, Hannemann et al., 2007).

Average time–frequency representations per participant and condition were thus gained, reflecting relative changes in (de-) synchronisation, in response to the sentence-final keyword, relative to an equally long silent pre-sentence (–500 to 0 ms) baseline.

For a comprehensive statistical assessment in the 2 \times 3 design, we calculated the individual differences of high greater low cloze conditions in each degradation level (1-, 4- and 16-band) and submitted these to a parametric *t*-contrast in a massive permutation test (as outlined in Maris and Oostenveld, 2007; 1000 iterations). This procedure checked for time–frequency–electrode clusters (clusters of at least three adjacent electrodes) that would show increasing effects in the high–low cloze difference, as the signal varied in intelligibility (1- to 4- to 16-band speech). The permutation test, although in this case including time bins from 250 to 750 ms across a broad range of γ -band frequency bins (30–80 Hz) as well as across 46 channels (covering the left and right scalp), effectively controls for multiple comparisons error at the cluster level, ensuring a type I error probability smaller than 0.05. The resulting test statistic is a cluster T_{sum} value, summing the *t* values within a time–frequency–electrode cluster (Maris and Oostenveld, 2007; for application see e.g. Medendorp et al., 2007; Schlee et al., 2009). A second permutation test on the same channels and time windows was also run, taking into account only a higher register of the γ -band (60–80 Hz), which has been increasingly found relevant in perceptual tasks (e.g., Van Der Werf et al., 2008; Wyart and Tallon-Baudry, 2008).

Results

Comprehension rating data

The rating data of all 30 participants was analysed and yielded a clear interaction pattern of factors cloze probability and intelligibility/degradation ($F(1.8, 50.8) = 45.9, p < 0.0001$, Greenhouse–Geiser corrected). At intermediate degradation (4-band vocoding), high-cloze sentences were rated as more comprehensible than low-cloze analogues ($t(29) = -4.06, p < 0.001$). There was also a strong overall effect of degradation on perceived comprehensibility ($F(1.6, 45) = 1262.7, p < 0.0001$, Greenhouse–Geiser corrected), confirming previous accounts of good correspondence between actual and self-perceived comprehension (Davis and Johnsrude, 2003; Obleser et al., 2008).

Event-related responses to the sentence onset

As is evident in the example in Fig. 1 as well as in the condition-specific responses in Fig. 2, all conditions elicited a typical N1-P2-

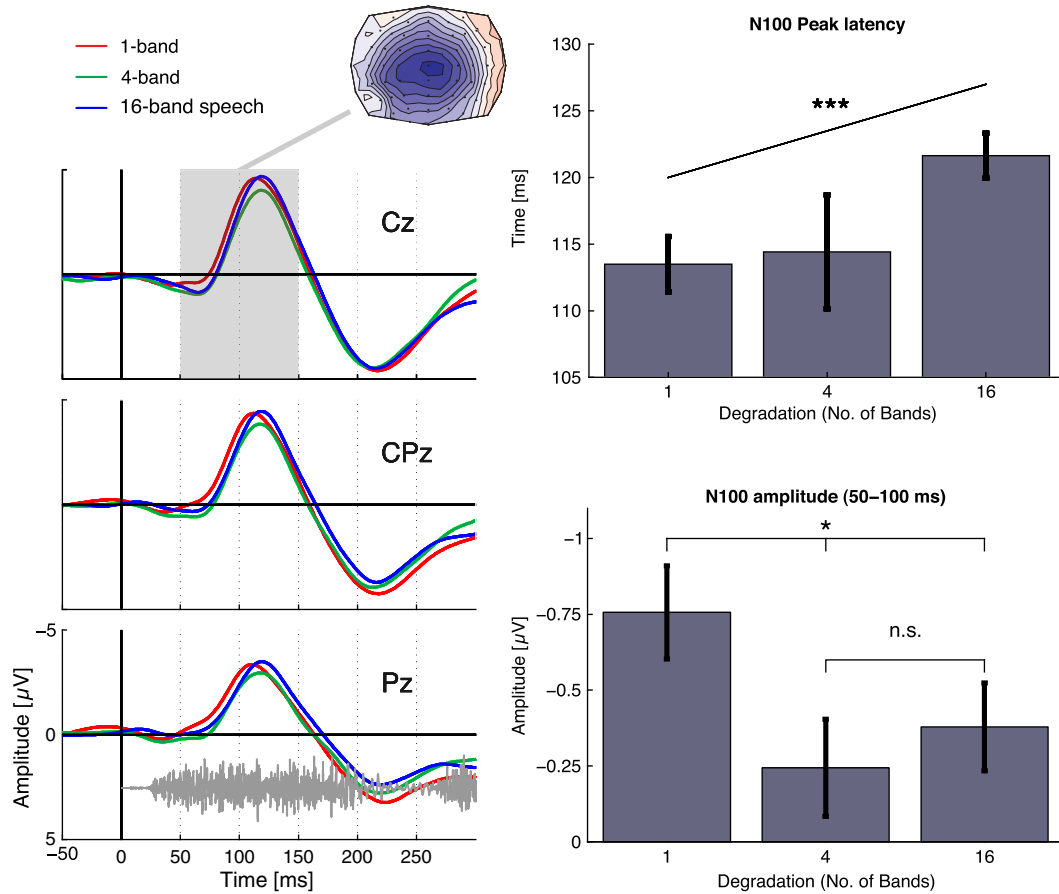


Fig. 2. Evoked potential phase-locked to sentence onset. The left panel shows a scalp topography (N100 time range, in response to 1-band speech, $\pm 5 \mu\text{V}$) and the respective time courses, for midline channels Cz, and Pz, split in levels of degradation (number of bands). Right panels: Parametric N400 effects. The top panel show the peak latency of the N100, split in levels of degradation; linear trend $***p < 0.001$. The bottom panel shows the early N100 amplitude (mean 50–100 ms over midline electrodes Fz to POz ± 1 standard error of the mean), split in levels of degradation (number of bands). $*p < 0.05$.

complex followed by a negative-going, N400-like deflection (see specific averages to sentence-final keywords below).

Statistical analyses showed a strong main effect of degradation on N100 amplitude in the 50–100 ms and 100–150 ms time windows, with respective $F(2,58)$ values of 5.71 and 5.54, both $p < 0.01$.

Mainly the earlier and more vigorous N100 amplitudes in response to highly degraded 1-band vocoded speech were driving this effect (Fig. 2). A peak latency analysis over the same midline electrodes confirmed the linear trend in N100 peak latency ($F(1,27) = 54.05$, $p < 0.0001$), with 1-band speech eliciting the earliest and 16-band speech the latest N100 peak (Fig. 2).

Notably, the degradation sensitivity in the N100 time window was inversely correlated with participant's rating behaviour: Those participants who showed a more pronounced N100 amplitude sensitivity to levels of degradation (i.e., difference between 1-band and 16-band speech) also showed wider differences in comprehension ratings after each trial (i.e., they more fully exploited the range of available rating levels). Pearson's correlation coefficient was $r = 0.44$, $p < 0.02$ (Fig. 5).

A strong oddball-like positive response to 1-band speech was also observed at posterior to occipital channels in a later time range (350 to 700 ms, $41.7 > F(2,58) > 27.8$, all $p < 0.0001$), most likely elicited by the perceptual pop-out of highly unintelligible speech (~30% of all trials). This did not yield comparable correlations with individual rating behaviour.

Event-related responses to the sentence-final keyword

The event-related signal perturbations in response to the sentence-final keyword are illustrated in Fig. 3. A negative-going

response with parieto-central distribution and a peak at about 400 ms is the most prominent deflection, matching a prototypical N400 response. As for the changes in statistical significance for main effects of cloze probability and signal degradation, as well as their interaction, both factors contribute significantly to the negative perturbation in the time windows preceding the N400 (main effect of number of bands peaking at 250–300 ms, $F(1,8,53.4) = 8.7$, $p < 0.001$, Greenhouse–Geiser corrected; main effect of cloze peaking at 350–400 ms, $F(1,29) = 7.92$, $p < 0.002$), the interaction of both factors attains significance exactly in the N400 time window (400–450 ms; (Cloze \times Number of bands interaction $F(1.9,55.7) = 4.37$, $p < 0.02$, Greenhouse–Geiser corrected).

As Fig. 3 also shows, the significant interaction results from the expectancy-driven N400 (stronger responses for low-cloze probability keywords) increasing with better signal quality and is strongest at the best-quality, 16-band level.

The increasing low-high cloze differentiation in the N400 amplitude with increasing signal intelligibility was also evident in post-hoc paired t -tests per degradation level. There was no low vs. high cloze difference in 1-band speech whatsoever ($t(29) < 1$, $p > 0.50$). In 4-band speech, a rise in N400 amplitude was observed for both cloze levels, with a strong trend for an augmented amplitude in low cloze ($t(29) = -1.97$, $p = 0.06$). At 16-bands, the expected cloze-driven N400 effect surfaced, as low cloze sentence-final keywords elicited a substantial N400 negativity and led to a highly significant cloze difference ($t(29) = -3.26$, $p < 0.003$). Inspecting the 3-tiered pattern of N400 responses across degradation levels, low cloze keywords led to monotonically rising N400 responses with better signal quality, whereas high cloze keywords invoked an inverted u-shaped pattern.

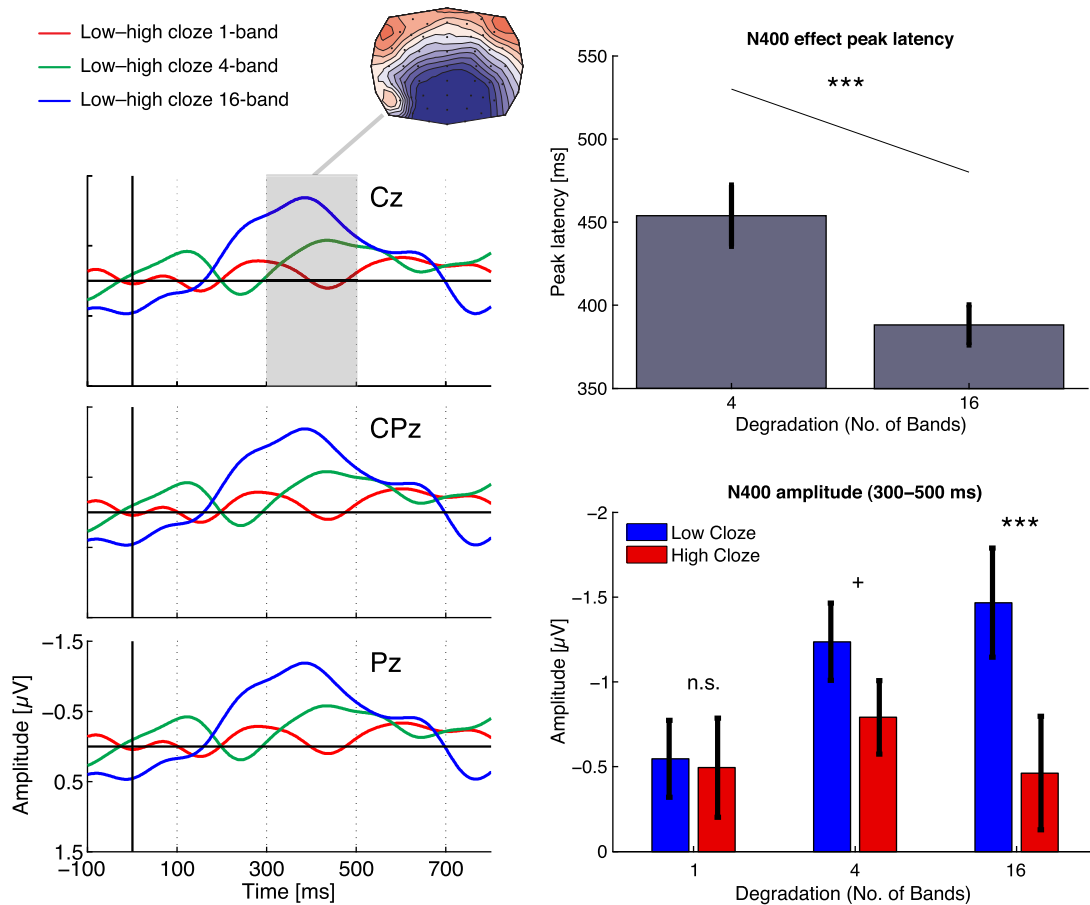


Fig. 3. Evoked potential phase-locked to sentence-final keyword. The left panel shows a scalp topography of the cloze effect (low–high cloze; 300–500 ms, in response to 16-band speech, $\pm 3 \mu\text{V}$). The right panel shows the respective time course of the cloze effect (low–high cloze), for midline channels Cz, CPz, and Oz, split in levels of degradation (number of bands). Note gradual build-up of an N400 with increasing signal quality. Right panels: Parametric N400 effects. The top panel shows the peak latency of the N400 cloze effect (low–high cloze conditions), split in the two levels of degradation where an N400 effect was observed; *** $p < 0.001$. The bottom panel shows the N400 amplitude (mean 300–500 ms over midline electrodes Fz to POz ± 1 standard error of the mean), split in levels of degradation (number of bands) and semantic expectancy (high cloze, red bars; low cloze, blue bars). *** $p < 0.001$; + $p < 0.1$.

Where present (4-band and 16-band signals), the N400 effect (i.e., low cloze eliciting more negative potentials than high cloze) was not only stronger in amplitude but peaked also earlier for more intelligible speech. This effect of degradation on the peak latency of the N400 cloze effect ($t(29) = 3.85$, $p < 0.0001$) is also shown in Fig. 3 (upper right panel).

Non-phase-locked activity

We followed up our evoked (i.e., phase-locked) ERP analysis with an additional search for potential non-phase-locked changes in brain oscillatory power on the sentence-final keyword. We especially had the hypotheses that (i) non-phase-locked brain activity should yield complementary effects not observable in the classic event-related, that is phase-locked analysis reported above (e.g., Makeig et al., 2004; Tallon-Baudry and Bertrand, 1999), and that (ii) non-phase locked brain activity in the Gamma-band range should indicate facilitated comprehension processes (e.g., Hannemann et al., 2007; Lenz et al., 2007)—thus, more prominent event-related synchronisations were expected for high- than for low-cloze sentence keywords.

Visual inspection of the high vs. low cloze differences at the various levels of signal degradation showed an incremental enhancement of Gamma power for the high versus low cloze keyword differences in a broad time–frequency range. At intermediate (4-band) and at sufficient (16-band) speech signal quality, recurring time–frequency patches of Gamma enhancement in the 40-Hz range

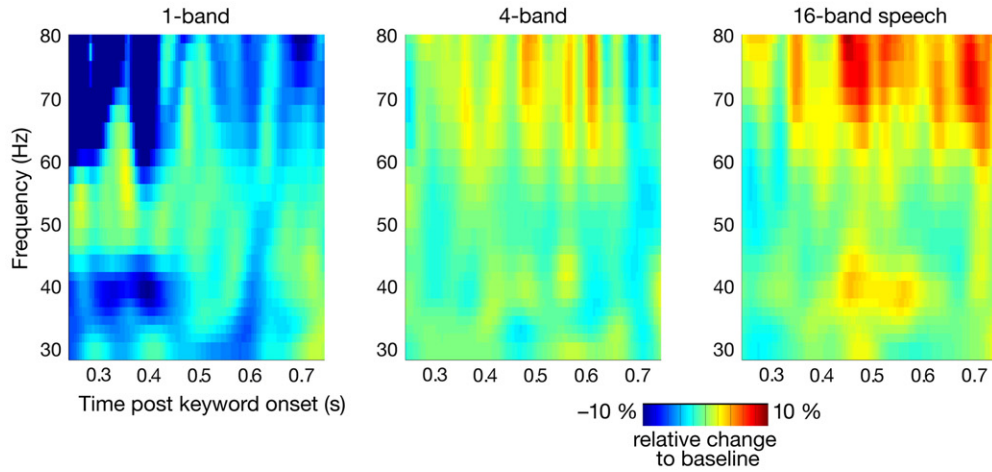
as well as in the 60–80 Hz range were observable in the time–frequency plots (Fig. 4, top panels).

We thus ran permutation tests for significant time–frequency–electrode clusters (see *Methods*) and specifically tested for a parametric increase of in γ -band power for the high–low cloze difference, as we moved from 1- to 4- to 16-band speech.

This test yielded the most robust time–frequency–electrode cluster in the 60–80 Hz gamma frequency range. It was broadly distributed over left fronto-temporal electrodes in the 550–650-ms time range (sum of T values in cluster $T_{\text{sum}} = 730.8$, $df = 27$, $p = 0.014$). It thus directly ensued the N400 time window (Fig. 4B; electrode positions belonging to the significant cluster are shown in bold).

The time–frequency plots in Fig. 4A tentatively also hinted at a comparable effect in a slightly lower 40-Hz γ -band register. We repeated the permutation test for the 30–50 Hz register, which yielded the cluster at the same locations (left and right fronto-temporal electrodes), but it was only significant at a trend level (cluster $T_{\text{sum}} = 389.5$, $p = 0.088$). When testing most conservatively over the entire γ -band (30–80 Hz) in a third permutation test, the cluster peaking around 70 Hz was still found significant (cluster $T_{\text{sum}} = 775$, $p = 0.04$). Taken together, the time–frequency–channel area encompassing 60–80 Hz activity in the 600-ms time window over left fronto-temporal electrodes carries the most robust signatures of the expectancy–intelligibility dependency, while the lower γ -band appears more variable across participants and conditions, yielding reduced statistical significance.

A High > Low Cloze: Increase with signal quality (all channels)



B Parametric increase of High > Low Cloze with signal quality (550–650 ms)

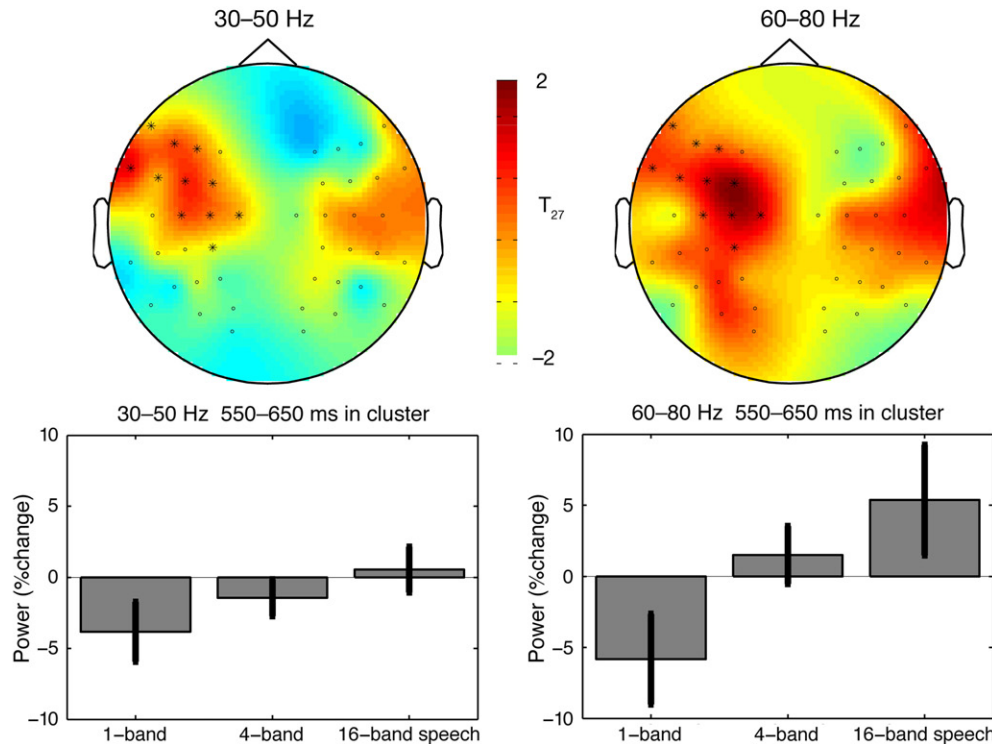


Fig. 4. Effect of semantic expectancy and degradation on gamma (γ) band activity. **A.** Time–frequency plot (average across all channels). Gamma band power is colour-coded ($\pm 10\%$ change relative to baseline; see text for details), panels from left to right show the high–low cloze effect for 1-band, 4-band and 16-band speech. **B.** Scalp Topographies and Bar graphs of the Parametric Gamma effects. T ($df=27$) values from the massed permutation test are plotted, separately for a low-Gamma (30–50 Hz) and a high-Gamma (60–80 Hz) register. Note the left- and right fronto-central peaks in statistical significance. Highlighted channels in the left panel indicate the location of the significant channel–time–frequency-cluster (see text). The bar graphs below each scalp represent the parametric power change ($\pm 10\%$) in the 600-ms time range within the electrode cluster highlighted in the left scalp topography above. Note the monotonic increase in Gamma power in both registers for the high–low cloze effect, going from 1- to 4- to 16-band speech.

Discussion

The human brain is able to resolve the semantic relations within sentences, and it will derive meaning even when the acoustic quality is less than ideal. However, the mechanisms and hierarchies of acoustic processing and semantic processing are unclear. In the current study, we set out to further disentangle the interaction of speech signal processing and semantic expectancy processing, by analysing the time course of the concomitant neural events using EEG. By analysing a whole cascade of brain electric signatures (early N100

to sentence onset; N400 and induced Gamma-band oscillations to sentence-final keywords; plus behavioural responses), we were also able to test whether signatures of early sensory processing (N100) are directly linked to the ensuing allocation of further resources in the semantic analysis and the processing of top-down expectancies.

Four main results shall be summarised here: Firstly, the N100 response was affected only by varying levels of speech degradation. It peaks earlier and appears augmented in amplitude for unintelligible (1-band) speech. The individual “dynamic range” of the N100 amplitude modulation (parameterised as the amplitude difference

between 1- and 16-band N100 responses) was positively correlated with the individual “dynamic range” or range of comprehension ratings that listeners expressed after each trial.

Secondly, an enhanced integration effort in highly predictable sentences that is reflected by augmented N400 amplitude clearly depends on the level of speech degradation. The additional computations required to process verb–argument relations as in “weigh–flour” compared to highly expected “sift–flour” (i.e., the cloze probability effect) linearly increased in amplitude as the signal quality improved. This interaction of acoustics-level and semantics-level cues attained significance only in the N400 time window (350–450 ms post keyword onset).

Thirdly, the opposite effect (a vigorous response to high-cloze speech) was observed in the domain of non-phase-locked brain activity, in the Gamma-band range, at a comparably late stage of word processing (600 ms post onset, that is, following the N400 time window). This is in line with previous reports on top-down mediated gains in Gamma-band amplitude for comprehension of (degraded) speech (Hannemann et al., 2007; Shahin et al., 2009), and it relates closest to previous studies on top-down effects onto speech comprehension (Obleser et al., 2007; see also schematic summary of data in Fig. 6).

Lastly, a significant correlation of the individual expectancy-driven change in Gamma-band power at sentence ending with the very early N100 amplitude change at sentence onset was found. At 16-band speech (i.e., where the most vigorous Gamma-band change was elicited), a negative correlation across 28 participants between these remote EEG signatures could be substantiated.

As the schematic summary of findings in Fig. 6 illustrates, sentence comprehension under adverse listening conditions yields at least three distinct EEG signatures. All three reflect distinct effects of signal degradation and prove the strengths of temporally sensitive EEG in studying sentence comprehension.

The more vigorous N100 deflection for less intelligible speech is the earliest significant effect observed. The sentence-initial N100 is not influenced by the semantic manipulation (which occurs only at the sentence-final word), but it indexes an initial allocation of resources and formation of a sensory memory trace (e.g., Gage and Roberts, 2000; Obleser et al., 2004a; Schroger et al., 2003). The observed stronger activation through the unfamiliar spectro-temporal patterns of 1-band speech underlines this interpretation: In a speech context (subjects attempted to comprehend all trials and rated them afterwards), the allocation of processing resources and the (possibly

futile) attempt to form and maintain a sensory memory trace should be strongest in response to the least familiar stimuli, here 1-band speech.

Two important correlations observed here speak to this link of comprehension and the N100 amplitude. First, those participants who exploited the full range of possible rating levels more exhaustively (i.e., showed more “dynamic range” in their ratings) also showed a stronger modulation or dynamic range in their N100 amplitudes in response to very degraded and highly intelligible speech. Second, an overall stronger sentence-initial N100 (reflected here by a stronger N100 to highly intelligible speech) was a good predictor of a weaker Gamma-band effect (in the 40–Hz range) in sentence-final semantic integration. As argued below, strong N100 amplitudes may reflect high effort and initial resource allocation. The negative correlation of the N100 amplitude at 16-band (i.e., degraded yet well-intelligible) speech with later Gamma power at the final keyword may suggest that strong initial N100s versus strong Gamma-band responses at integration reflect two partly distinct neural strategies of listeners dealing with moderately degraded speech; a more “bottom-up,” resource-allocating, and effortful strategy (which may not necessarily be successful; Obleser and Eisner, 2009; Pisoni, 2000) versus a more top-down, associative, or facilitatory strategy (Engel et al., 2001; Hannemann et al., 2007; Herrmann et al., 2009).

While such distinction remains subject to further testing in bigger subject samples, the observed correlations (Fig. 5) suggest that the early N100 amplitude is a good predictor of behaviour as well as of ensuing neural processing stages occurring later. Note that a comparable link of early and late integration stages could not be established for N100 and N400 amplitude. This provides further evidence for the complimentary information on language processing that is offered by “classical” N400-type analyses and oscillatory EEG analysis.

One could argue that the N400 effect, reflecting stronger responses to low cloze keywords, and the ensuing Gamma-band oscillations, reflecting stronger responses to high cloze keywords, are just two views on the same neural phenomenon. This is unlikely for various reasons: First, their time window of occurrence is not identical. Second, the topographically broadly distributed N400 stands in contrast to the more focal and somewhat more anterior Gamma-band bursts. Lastly, while a moderate negative correlation between these two markers is observed (which is in line with our effort versus facilitation argument which is mainly based on the N100–Gamma-band anticorrelation), it does not attain significance: Individual N100

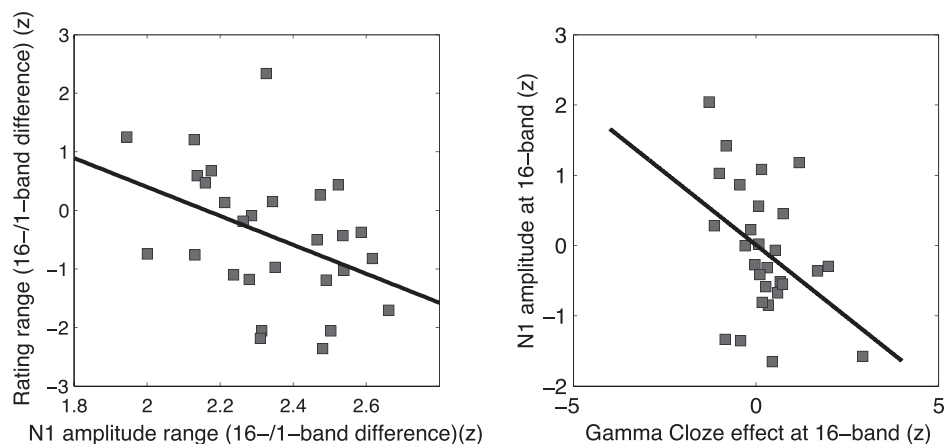


Fig. 5. Correlations of Ratings, N100 amplitude and Gamma-band power. The left panel illustrates the correlation between the “dynamic range” of participants’ N100 amplitudes (individual mean amplitude differences between 16- and 1-band signals; z-normalised) and the “dynamic range” of comprehension ratings (individual rating score differences between 16- and 1-band signals; z-normalised); Pearson’s $r = 0.44$, $p < 0.02$. The right panel illustrates the negative correlation between individual N100 amplitude (z-normalised) at sentence-onset in response to moderately degraded speech (16 bands) and individuals’ Gamma-band power enhancement at the sentence-final keyword (also z-normalised, also in response to 16-band speech); Pearson’s $r = -0.41$, $p > 0.03$.

amplitudes in response to 16-band speech can explain three times more of the observed Gamma-band variance ($r^2 \sim 18\%$) than N400 amplitudes can ($r^2 \sim 6\%$).

The N400 is arguably the best-studied of all language-related evoked potential components, and debate over its functional significance is comparably sparse (see [Kutas and Federmeier, 2000](#) for review). A higher effort in integrating a given word into a sentence context elicits a stronger N400 response. The integration process is less effortful given a high cloze relationship between preceding verb (or context, for that matter) and keyword, which then elicits only a weak N400 ([Kutas and Hillyard, 1984](#)). Our results in the N400 time range on the sentence-final keyword are good confirmation of this effect (cf. [Aydelott et al., 2006](#)), obtained in naturally spoken, auditorily presented sentences.

However, our data add an additional aspect to this picture: they provide evidence that the semantic integration process, which is facilitated by high expectancy, is “gated” by signal quality (vocoding level) in a parametric fashion ([Fig. 3](#)). The N400-reduction effect (low expectancy eliciting stronger negativity than high expectancy) is absent in 1-band speech; present yet weak in 4-band speech; and it is prominent and peaks earliest in 16-band speech signal quality. The statistical interaction of factors cloze and signal quality in the N400 time range confirms this ([Fig. 3](#), right bottom panel).

Note that the concept of neural effort here has different implications in the sentence-initial N100 and in the keyword-triggered N400: As argued in detail above, the vigorous N100 amplitudes in response to more degraded speech are taken to reflect more effortful *signal-dependent* computations being devoted to attempts of encoding the unfamiliar and spectrally degraded speech signal. In a new set of data on word comprehension, this observation has been replicated using 4 different levels of vocoding: An according monotonic decrease in N100 amplitude with increasing signal quality has been observed ([Obleser et al., in prep.](#)). However, the higher effort of semantic integration, which is commonly assumed to be reflected in more vigorous N400 amplitudes, is distinct from this: To reiterate, effortful semantic computation is more evident in less degraded signals, that is, only when the signal quality allows for the semantic manipulations to exert their effect on comprehension.

How do our Gamma-band findings relate to current views on the functional role of human Gamma oscillations? The reported parametric Gamma-band enhancement is in keeping with previous results and sheds light on the dependency of bottom-up acoustic–sensory analysis with ensuing, top-down-mediated language processes. In one of the very few previous studies on top-down-mediated comprehension of degraded speech ([Hannemann et al., 2007](#)), a Gamma-band enhancement with highly comparable scalp distribution was found. Statistics on the current data surpassed the threshold at a left-anterior cluster of channels ([Fig. 4B](#)). The current effect is observed slightly later in time. However, it was elicited here in a sentence context rather than in a simple two-syllable word context.

The role of transient gamma-band enhancement in perception has received wide attention across modalities and species (for review see [Fries, 2009](#); [Tallon-Baudry and Bertrand, 1999](#)). As outlined in the introduction, the current design is a particularly fitting test case for the hypothesis that enhanced gamma-band oscillations are a neural correlate of top-down-facilitated language perception: From the (parametrically degraded) speech signal itself, the listener could extract either comparably low (low cloze) or comparably high (high-cloze) levels of a semantic relation between verb and object in the sentence; when a high-cloze relation is present, this should facilitate the processing of the sentence-final keyword, and, intertwined with this, the access of sentence-level meaning. Thus, the gamma-band oscillation change in the current manipulation further fosters a functional role for Gamma-band oscillations in semantic processing. The observed Gamma-band enhancement also shows the expected topography (fronto-temporo-central channels, more pronounced on

the left than on the right) and time–frequency distribution (500–600 ms post sentence-final keyword; peaking in the expected ~ 40 -Hz as well as at in the ~ 70 -Hz range).

Two previous studies further corroborate the semantic facilitation interpretation of the present Gamma-band effect: Firstly, [Schneider and colleagues](#) demonstrated in a cross-modal priming paradigm an enhanced Gamma response, peaking at 40–50 Hz and ranging from 30 to about 70 Hz, to auditory targets that formed a semantically congruent context with visual primes. The scalp distribution of these effects also yielded a fronto-temporal topography (cf. [Fig. 4](#); [Schneider et al., 2008](#)). Secondly, [Shahin and colleagues](#) report a Gamma-band (centred at about 40 Hz) enhancement in a semantic compared to a simple voice task performed on auditorily presented words, which exhibited also a left fronto-central scalp distribution ([Shahin et al., 2009](#)).

As an aside, we are confident that these data are a robust signature of a perceptual–cognitive effect in the Gamma-band range, as the data were thoroughly checked for possible confounds that are known to affect measured γ -band power. Most importantly, any artefactual or non-cognitive source for the Gamma-band modulation would be required to explain the complex high-greater-low cloze modulation and in particular its parametric change across three different levels of acoustic degradation. This can be ruled out in an auditory EEG setup like the current one. Also, oscillations in the Theta-band (θ , 4–7 Hz; see [Bastiaansen and Hagoort, 2006](#)), upon which we had not based a specific hypotheses about top-down facilitation, have been shown to be closely linked to retrieval from semantic memory, and, more generally, the faster Gamma bursts has been shown to be depend on the phase of slower Theta (e.g., [Quilichini et al., 2010](#)); here, a post-hoc analysis of power changes in the Theta-band revealed a monotonic power increase that was weaker than yet concomitant to the Gamma-band changes.

To our knowledge, our study is the first to demonstrate the parametric dependence of induced, top-down facilitated gamma-band enhancement on auditory signal quality. As outlined for the N400 above, the facilitatory Gamma-band effect also depends parametrically on the signal quality of speech. The parametric dependence of all brain electric signatures on signal quality is an additional insight offered by these data. [Fig. 6](#) shows a schematic summary of all parametric signal-quality dependencies observed.

By analysing three distinct EEG measures (N100, N400, gamma-band oscillations) and by checking for possible subject-to-subject correlations between them, the current data open various venues to tie the results to the existing body of literature on degraded speech processing on the one hand, and semantic processing on the other.

Given the amount of studies which have previously researched the likely source generators of the N100 component (mostly based on the spatially advantageous MEG; e.g., [Krubholz et al., 2003](#); [Obleser et al., 2004b](#); [Scherg et al., 1990](#)), a tentative tie is suggested between the enhanced N100 amplitude observed in response to more strongly degraded speech and the recent functional neuroimaging data on an enhanced activation in posterior superior temporal cortex (planum temporale) to incomprehensible speech or language conditions (for review see [Price, 2010](#); [Obleser et al., in preparation](#); [Obleser and Eisner, 2009](#)). The N100 is likely to receive strong generating contributions from the planum temporale ([Scherg et al., 1990](#)), which is also assigned an important intermediate role in processing complex ([Griffiths and Warren, 2002](#)) or unpredictable ([Overath et al., 2007](#)) acoustic stimuli. A re-analysis of the [Obleser and Kotz, 2010](#) data also showed that it is the planum temporale that shows a tight inverse relationship to subject's comprehension performance of noise-vocoded simple sentences, which is in line with the current N100 results ([Figs. 5 and 6](#)). We thus suggest a nexus of neuroanatomy, neural functioning, and behaviour (planum temporale; N100; comprehension) in the perception of degraded speech that is open to further experimental inquiry.

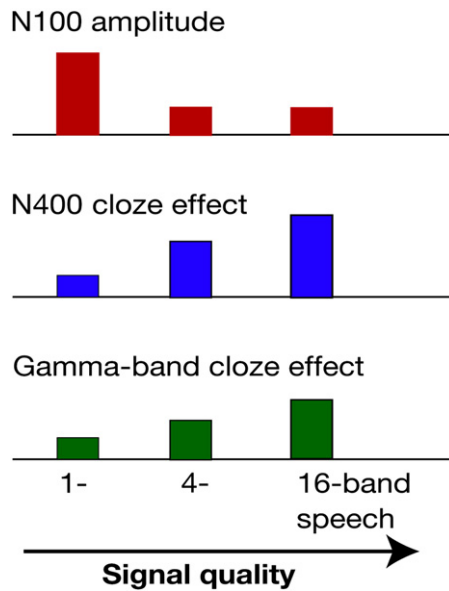


Fig. 6. Schematic summary of results. All three main analysis targets (N100 amplitude; low > high cloze at the N400; high > low in the Gamma band oscillations) are displayed in a schematic fashion, as a function of signal quality (increasing from 1 to 4 to 16 bands, left to right). Note the correspondence to the correlation of low N100 and high Gamma (also shown in Fig. 5), and the complimentary information contained in the N400 (low > high cloze) and the Gamma band (high > low cloze).

The current study sacrificed spatial fidelity and a close link to neuroanatomy for the temporal fidelity offered by EEG. This was beneficial in two ways. It allowed us to disentangle certain stages of processing as the sentence progresses, and it allowed analysing the oscillatory dynamics of the ongoing EEG signal during speech comprehension. The participants tested were not identical with the ones who had undergone functional MR scanning while listening to these materials earlier (Obleser and Kotz, 2010); this leaves us only to speculate about links between functional neuroanatomy and the observed EEG correlates at this time. We have argued for a nexus of posterior superior temporal gyrus activation, the N100 modulation observed here, and the overt comprehension behaviour shown by both subject samples. In the N400 time range, it is of note that the parametric “gating” of the N400 (i.e. the cloze difference the N400 reflects being dependent on signal quality) resembles the activation pattern of the left inferior frontal gyrus (IFG; Brodmann’s area 44) we had observed in the fMRI study. While inferior–frontal contributions to the generation of the N400 have been suggested (Maess et al., 2006), it is not very likely that the N400 is generated frontally. It rather receives strong contributions from superior temporal cortex (Halgren, 2004; Uusvuori et al., 2008; Van Petten and Luka, 2006), which was also found in the previous fMRI study (bilateral posterior superior temporal sulcus activation accompanied low more than high cloze sentences). In sum, the conjunction of the fMRI and the current EEG study suggests an analogy (rather than identity) between the processes reflected in the IFG BOLD response and processes supposedly reflecting effortful integration in the N400 time window. For the Gamma-band, least is known in terms of neural generator structures at least in the framework of auditory and language studies, and improved source localisation of the Gamma-band effects reported here is needed to foster function–neuroanatomy links, for example to the angular gyrus, a brain structure repeatedly suggested by own and other’s fMRI data to play a facilitatory role in access or manipulation of semantic information (Eisner et al., 2010; Obleser and Kotz, 2010; Obleser et al., 2007; Raettig and Kotz, 2008; for review see Binder et al., 2009).

A testable hypothesis that follows from this brief EEG–fMRI synopsis would be a fronto-temporal loop, where the “bottom-up,” resource- and effortful processes reported here (initial N100; N400 at integration) here should covary more closely with posterior-temporal brain structures (especially PT, and middle temporal gyrus; see Lau et al., 2008), whereas putative facilitatory processes (as reflected by high-frequency oscillatory mechanisms Fries, 2009) should covary more closely with prefrontal and parietal structures.

It would also allow to test whether there are individual differences in which of these strategies is chosen when confronted with degraded speech (as suggested by the across-participants negative correlation of N100 and Gamma-band in our data), and whether this affects the relative contributions in a fronto-temporal network.

Conclusions

To the best of our knowledge, the current study is the first to take a comprehensive look at human brain electric signatures of speech comprehension, while parametrically varying both speech signal quality (“bottom-up”) as well as semantic constraints (“top-down”). Those two parametric variations allowed us to sample the early N100 response, the late N400 response, the non-phase-locked changes in the late Gamma-band power, as well as participant’s post-trial behaviour for separate as well as joint effects of these manipulations. Fig. 6 gives a simplified overview over the results. The reported sensitivity of the early N100 to degradation and its correlation with late, post-trial comprehension reports tie this component closer to strategies of effortful coping with degraded speech. The N400, showing the expected effect of semantic expectancy as reflected by cloze probability, also varied as a function of degradation. We conclude that the computations of semantic integration as indexed by the augmented N400 are “gated” by sufficient signal quality. Complementary to this and in an ensuing time window, the changes in Gamma-band power exhibited the same dependence on signal quality, yet showed an opposite effect; likely to reflect the facilitated integration of the sentence-final keyword into sentential meaning. Lastly, interesting venues for further research are opened by the observed negative correlation at moderate signal degradation: While the N100 appeared generally reduced for less degraded speech, its (lower) amplitude was a significant predictor of later facilitatory processes as indexed by the (higher) Gamma-band response. We tentatively suggest that this is a good indication for opponent processes in dealing with degraded speech, namely resource-intensive bottom-up versus context-exploiting, facilitatory top-down ones.

Acknowledgments

The Max Planck Society (J.O., S.K.) and the Deutsche Forschungsgemeinschaft (S.K.) supported this research. Thom Gunter kindly provided the pre-tested sentence material that formed the basis of our stimuli, and Stuart Rosen (University College London) provided the original code snippets for noise-vocoding. Mathias Barthel and Beatrice Neumann helped edit the audio material, and Conny Schmidt helped acquire the EEG data. Two anonymous reviewers helped substantially improve this manuscript.

References

- Aydelott, J., Dick, F., Mills, D.L., 2006. Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology* 43, 454–464.
- Bastiaansen, M., Hagoort, P., 2006. Oscillatory neuronal dynamics during language comprehension. *Prog. Brain Res.* 159, 179–196.
- Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796.

- Bonte, M., Valente, G., Formisano, E., 2009. Dynamic and task-dependent encoding of speech and voice by phase reorganization of cortical oscillations. *J. Neurosci.* 29, 1699–1706.
- Connolly, J.F., Phillips, N.A., Forbes, K.A., 1995. The effects of phonological and semantic features of sentence-ending words on visual event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* 94, 276–287.
- Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431.
- Davis, M.H., Johnsrude, I.S., 2007. Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear. Res.* 229, 132–147.
- Davis, M.H., Johnsrude, I.S., 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.
- Davis, M.H., Coleman, M.R., Absalom, A.R., Rodd, J.M., Johnsrude, I.S., Matta, B.F., Owen, A.M., Menon, D.K., 2007. Dissociating speech perception and comprehension at reduced levels of awareness. *Proc. Natl Acad. Sci. USA* 104, 16032–16037.
- Eisner, F., McQueen, J.M., 2005. The specificity of perceptual learning in speech processing. *Percept. Psychophys.* 67, 224–238.
- Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., Scott, S.K., 2010. Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *J. Neurosci.* 30, 7179–7186.
- Engel, A.K., Fries, P., Singer, W., 2001. Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* 2, 704–716.
- Faulkner, A., Rosen, S., Wilkinson, L., 2001. Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant speech processor. *Ear Hear.* 22, 431–438.
- Federmeier, K.D., Wlotko, E.W., De Ochoa-Dewald, E., Kutas, M., 2007. Multiple effects of sentential constraint on word processing. *Brain Res.* 1146, 75–84.
- Fishman, K.E., Shannon, R.V., Slattery, W.H., 1997. Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *J. Speech Lang. Hear. Res.* 40, 1201–1215.
- Fries, P., 2009. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* 32, 209–224.
- Gage, N.M., Roberts, T.P., 2000. Temporal integration: reflections in the M100 of the auditory evoked field. *NeuroReport* 11, 2723–2726.
- Gandour, J., Tong, Y., Wong, D., Talavage, T., Dziedzic, M., Xu, Y., Li, X., Lowe, M., 2004. Hemispheric roles in the perception of speech prosody. *Neuroimage* 23, 344–357.
- Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. *Trends Neurosci.* 25, 348–353.
- Grossberg, S., 2009. Cortical and subcortical predictive dynamics and learning during perception, cognition, emotion and action. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1223–1234.
- Gruber, T., Müller, M.M., Keil, A., 2002. Modulation of induced gamma band responses in a perceptual learning task in the human EEG. *J. Cogn. Neurosci.* 14, 732–744.
- Gunter, T.C., Friederici, A.D., Schriefers, H., 2000. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci.* 12, 556–568.
- Halgren, E., 2004. How can intracranial recordings assist MEG source localization? *Neuro. Clin. Neurophysiol.* 2004, 86.
- Hannemann, R., Obleser, J., Eulitz, C., 2007. Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Res.* 1153, 134–143.
- Herrmann, C.S., Frund, I., Lenz, D., 2009. Human gamma-band activity: a review on cognitive and behavioral correlates and network models. *Neurosci. Biobehav. Rev.* 34 (7), 981–992.
- Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C., Lutkenhoner, B., 2003. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cereb. Cortex* 13, 765–772.
- Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470.
- Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163.
- Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933.
- Lenz, D., Schadow, J., Thaeig, S., Busch, N.A., Herrmann, C.S., 2007. What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. *Int. J. Psychophysiol.* 64, 31–38.
- Maess, B., Herrmann, C.S., Hahne, A., Nakamura, A., Friederici, A.D., 2006. Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing. *Brain Res.* 1096, 163–172.
- Makeig, S., Debener, S., Onton, J., Delorme, A., 2004. Mining event-related brain dynamics. *Trends Cogn. Sci.* 8, 204–210.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190.
- Medendorp, W.P., Kramer, G.F., Jensen, O., Oostenveld, R., Schoffelen, J.M., Fries, P., 2007. Oscillatory activity in human parietal and occipital cortex shows hemispheric lateralization and memory effects in a delayed double-step saccade task. *Cereb. Cortex* 17, 2364–2374.
- Mietinen, I., Tiitinen, H., Alku, P., May, P.J., 2010. Sensitivity of the human auditory cortex to acoustic degradation of speech and non-speech sounds. *BMC Neurosci.* 11, 24.
- Miller, G.A., Heise, G.A., Lichten, W., 1951. The intelligibility of speech as a function of the context of the test materials. *J. Exp. Psychol.* 41, 329–325.
- 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.
- Obleser, J., Eisner, F., 2009. Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn. Sci.* 13, 14–19.
- Obleser, J., Kotz, S.A., 2010. Expectancy constraints in degraded speech modulate the speech comprehension network. *Cereb. Cortex* 20, 633–640.
- Obleser, J., Albert, T., Eulitz, C., 2004a. Attentional influences on functional mapping of speech sounds in human auditory cortex. *BMC Neurosci.* 5, 24.
- Obleser, J., Lahiri, A., Eulitz, C., 2004b. Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J. Cogn. Neurosci.* 16, 31–39.
- Obleser, J., Scott, S.K., Eulitz, C., 2006. Now you hear it, now you don't: transient traces of consonants and their nonspeech analogues in the human brain. *Cereb. Cortex* 16, 1069–1076.
- Obleser, J., Wise, R.J., Dresner, M.A., Scott, S.K., 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. *J. Neurosci.* 27, 2283–2289.
- Obleser, J., Eisner, F., Kotz, S.A., 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* 28, 8116–8123.
- Overath, T., Cusack, R., Kumar, S., von Kriegstein, K., Warren, J.D., Grube, M., Carlyon, R.P., Griffiths, T.D., 2007. An information theoretic characterisation of auditory encoding. *PLoS Biol.* 5, e288.
- Pisoni, D.B., 2000. Cognitive factors and cochlear implants: some thoughts on perception, learning, and memory in speech perception. *Ear Hear.* 21, 70–78.
- Poeppel, D., Yellin, E., Phillips, C., Roberts, T.P., Rowley, H.A., Wexler, K., Marantz, A., 1996. Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Brain Res. Cogn. Brain Res.* 4, 231–242.
- Price, C.J., 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. NY Acad. Sci.* 1191, 62–88.
- Quilichini, P., Sirota, A., Buzsáki, G., 2010. Intrinsic circuit organization and theta-gamma oscillation dynamics in the entorhinal cortex of the rat. *J. Neurosci.* 30, 11128–11142.
- Raettig, T., Kotz, S.A., 2008. Auditory processing of different types of pseudo-words: an event-related fMRI study. *Neuroimage* 39, 1420–1428.
- Scherg, M., Vajsar, J., Picton, T.W., 1990. A source analysis of the late human auditory evoked field. *J. Cogn. Neurosci.* 1, 336–355.
- Schlee, W., Mueller, N., Hartmann, T., Keil, J., Lorenz, I., Weisz, N., 2009. Mapping cortical hubs in tinnitus. *BMC Biol.* 7, 80.
- Schneider, T.R., Debener, S., Oostenveld, R., Engel, A.K., 2008. Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *Neuroimage* 42, 1244–1254.
- Schroger, E., Tervaniemi, M., Huotilainen, M., 2003. Bottom-up and top-down flows of information within auditory memory: electrophysiological evidence. In: Kaernbach, C., Schröger, E., Müller, H.J. (Eds.), *Psychophysics beyond Sensation: Laws and Invariants of Human Cognition*. Erlbaum, Hillsdale, NJ, pp. 391–410.
- 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, 2400–2406.
- Scott, S.K., Rosen, S., Lang, H., Wise, R.J., 2006. Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. *J. Acoust. Soc. Am.* 120, 1075–1083.
- Shahin, A.J., Picton, T.W., Miller, L.M., 2009. Brain oscillations during semantic evaluation of speech. *Brain Cogn.* 70, 259–266.
- Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. *Science* 270, 303–304.
- Shannon, R.V., Fu, Q.J., Galvin III, J., 2004. The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngol. Suppl.* 552, 50–54.
- Shinn-Cunningham, B.G., Wang, D., 2008. Influences of auditory object formation on phonemic restoration. *J. Acoust. Soc. Am.* 123, 295–301.
- Sivonen, P., Maess, B., Lattner, S., Friederici, A.D., 2006. Phonemic restoration in a sentence context: evidence from early and late ERP effects. *Brain Res.* 1121, 177–189.
- Stickney, G.S., Assmann, P.F., 2001. Acoustic and linguistic factors in the perception of bandpass-filtered speech. *J. Acoust. Soc. Am.* 109, 1157–1165.
- 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., Delpeuch, C., Perrier, J., 1997. Oscillatory gamma-band (30–70 Hz) activity induced by a visual search task in humans. *J. Neurosci.* 17, 722–734.
- Taylor, W.L., 1953. Cloze procedure: a new tool for measuring readability. *Journalism Quarterly* 30, 414–433.
- Uusvuori, J., Parviainen, T., Inkinen, M., Salmelin, R., 2008. Spatiotemporal interaction between sound form and meaning during spoken word perception. *Cereb. Cortex* 18, 456–466.
- van Atteveldt, N.M., Formisano, E., Goebel, R., Blomert, L., 2007. Top-down task effects override automatic multisensory responses to letter–sound pairs in auditory association cortex. *Neuroimage* 36, 1345–1360.
- Van Der Werf, J., Jensen, O., Fries, P., Medendorp, W.P., 2008. Gamma-band activity in human posterior parietal cortex encodes the motor goal during delayed prosaccades and antisaccades. *J. Neurosci.* 28, 8397–8405.
- van Linden, S., Vroomen, J., 2007. Recalibration of phonetic categories by lipread speech versus lexical information. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1483–1494.
- Van Petten, C., Luka, B.J., 2006. Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain Lang.* 97, 279–293.
- Warren, J.D., Scott, S.K., Price, C.J., Griffiths, T.D., 2006. Human brain mechanisms for the early analysis of voices. *Neuroimage* 31, 1389–1397.
- Wyart, V., Tallon-Baudry, C., 2008. Neural dissociation between visual awareness and spatial attention. *J. Neurosci.* 28, 2667–2679.
- Zeng, F.G., Galvin III, J.J., 1999. Amplitude mapping and phoneme recognition in cochlear implant listeners. *Ear Hear.* 20, 60–74.