Suppressed Alpha Oscillations Predict Intelligibility of Speech and its Acoustic Details

Jonas Obleser[1] and Nathan Weisz[2,3]

[1]Max Planck Research Group "Auditory Cognition", Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany, [2]Department of Psychology and [3]Zukunftskolleg, University of Konstanz, 78457 Konstanz, Germany

Address correspondence to Jonas Obleser, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany. Email: obleser@cbs.mpg.de.

Modulations of human alpha oscillations (8–13 Hz) accompany many cognitive processes, but their functional role in auditory perception has proven elusive: Do oscillatory dynamics of alpha reflect acoustic details of the speech signal and are they indicative of comprehension success? Acoustically presented words were degraded in acoustic envelope and spectrum in an orthogonal design, and electroencephalogram responses in the frequency domain were analyzed in 24 participants, who rated word comprehensibility after each trial. First, the alpha power suppression during and after a degraded word depended monotonically on spectral and, to a lesser extent, envelope detail. The magnitude of this alpha suppression exhibited an additional and independent influence on later comprehension ratings. Second, source localization of alpha suppression yielded superior parietal, prefrontal, as well as anterior temporal brain areas. Third, multivariate classification of the time–frequency pattern across participants showed that patterns of late posterior alpha power allowed best for above-chance classification of word intelligibility. Results suggest that both magnitude and topography of late alpha suppression in response to single words can indicate a listener's sensitivity to acoustic features and the ability to comprehend speech under adverse listening conditions.

Keywords: alpha, auditory processing, degraded speech, EEG, neural oscillations, noise-vocoding

Introduction

The seeming ease with which listeners comprehend speech contrasts with the sparsely understood neural mechanisms enabling it. The functional neuroanatomy important for speech perception and comprehension in humans has been mapped out increasingly over the past decade (for review, see Hickok and Poeppel 2007; Rauschecker and Scott 2009). In parallel but separate from this, oscillatory rhythms in the human brain have gained interest mainly in vision and attention research (e.g., Engel et al. 2001; Jensen et al. 2007; Klimesch et al. 2007; Lakatos et al. 2008; Schroeder et al. 2008; Fries 2009). Unfortunately, there is much less known about oscillatory perturbations in auditory and speech perception (for review, see Weisz et al. 2011). Speech is a prime test case for studying any neural mechanism of perception and comprehension, as it notoriously happens in acoustically compromised circumstances that range in severity from phone lines and noisy environments to degraded hearing and cochlear implants.

In the current study, we explore the degradation of spectrotemporal acoustic features in the speech signal (Rosen 1992; Shannon et al. 1995), how they affect speech comprehension, and which oscillatory brain rhythms respond to it. In particular, we foresee a role for alpha rhythms (approximately 8–13 Hz) in speech comprehension: Alpha activity is receiving renewed interest and has been tied, mostly in visual experiments, to changing task demands and cognitive effort. In particular, alpha power is seen decreasing in brain areas required to actively process information, for example, in response to words at the point when access of the mental lexicon or the word meaning is likely to happen (for review, see Klimesch et al. 2007).

A parsimonious framework for changes in alpha power and their role in cognitive processing has been suggested by Jensen and colleagues (Osipova et al. 2008; Jensen and Mazaheri 2010), who see the general function of alpha oscillations in "gating by inhibition." In this framework, high-amplitude alpha oscillations are suited to gate or suppress spontaneous or task-irrelevant higher frequency (i.e., Gamma) oscillations in neural networks; similarly, a relative "silencing" of alpha oscillations takes away this inhibitory gating and allows for Gamma oscillations to occur. In line with this assumption, the few studies targeting Gamma oscillations in speech and language perception report increased Gamma power when access of the mental lexicon (in words) or integration of meaning (in sentences or cross-modally) becomes facilitated (e.g., Hannemann et al. 2007; Schneider et al. 2008; Shahin et al. 2009; Obleser and Kotz 2011).

Advantageously, the functional inhibition framework offers a link between the easily discernible and reliably estimable changes in alpha (8–13 Hz) power and the cognitive operations triggered by speech input. In this experiment, we test whether parametric degradations of acoustic detail known to affect intelligibility of speech (Obleser et al. 2008) parametrically drive the power of alpha oscillations in the human electroencephalogram (EEG). Using the noise-vocoding technique (Shannon et al. 1995; mostly used to mimic cochlear implantation and often used in imaging studies, for example, Davis and Johnsrude 2003; Scott et al. 2006; Obleser et al. 2008), we can parametrically vary fine temporal (i.e., spectral) and coarse temporal (i.e., envelope) detail available to the listener, which both will directly affect speech intelligibility (e.g., Rosen 1992; Xu et al. 2005; Lorenzi et al. 2006).

We specifically hypothesize that the degree of alpha suppression during and after auditory presented words should increase with increasing intelligibility of these words; as outlined above, this would reflect less need for functional inhibition and less effortful speech processing (Klimesch et al. 2007; Jensen and Mazaheri 2010; Weisz et al. 2011). Furthermore, we will also be able to test potential differential efficacy of spectral and envelope detail to modulate neural oscillations. A multivariate pattern classifier trained and tested on the time–frequency representations of degraded words will allow us assessing the predictive power of oscillatory changes on speech intelligibility.

Methods and Materials

Participants

Twenty-four healthy participants (12 female; mean 23.8 years of age ± 2.2 years standard deviation [SD]) took part in this study. All were monolingual speakers of German, had normal hearing and no history of neurological or language-related problems. They were naïve toward noise-vocoded speech. Participants received a financial compensation of 15 €. The procedure was approved of by the local ethics committee and in accordance with the declaration of Helsinki.

Stimuli

Stimuli were randomly drawn from a 560-item pool of recordings of spoken German mono-, bi-, and trisyllabic nouns (Kotz et al. 2002; Obleser et al. 2008). Words were recorded in a soundproof chamber by a trained female speaker and digitized at a 44.1 kHz sampling rate. Off-line editing included downsampling to 22.05 kHz, cutting at zero-crossings before and after each word, and normalization of root mean squared amplitude. The large pool of word stimuli allowed us to entirely avoid repetition of word items for any given participant.

From each word's final audio file, parametrically degraded versions were created using a Matlab-based noise band–vocoding algorithm. Noise vocoding is an effective technique to manipulate the spectral or fine temporal detail while preserving the amplitude envelope of the speech signal (Shannon et al. 1995). This renders the signal more or less intelligible in a graded and controlled way, depending on the number of bands used and more bands yielding a more intelligible speech signal. The technique has been used widely in behavioral and brain imaging studies before (Scott et al. 2000; Faulkner et al. 2001; Davis and Johnsrude 2003; Scott et al. 2006; Obleser et al. 2007). Usually, in noise vocoding, the spectral degradation is varied by specifying the number of spectral bands extracted, and the extracted amplitude envelopes in each band are carefully smoothed (low-pass filtered) by a value that is unlikely to affect intelligibility, for example, 250 Hz, as the relevant temporal envelope perturbations that contribute to intelligibility in speech mainly lie below 20 Hz (Xu et al. 2005).

For this study, however, we also systematically varied the coarse temporal features of the speech signal by smoothing the amplitude envelopes (Fig. 1A) over a range previously identified to have a distinct influence on intelligibility (for usage of this approach in a functional magnetic resonance imaging [fMRI] study, see Obleser et al. 2008). It is noteworthy that, with this additional dimension of signal processing, noise vocoding enables an orthogonal manipulation of the spectral and envelope variations in any given signal. This would be hard to achieve with other techniques because it allows splitting the signal into an arbitrary number of frequency bands and then smoothing the excitatory envelopes of each band with an arbitrary low-pass filter cutoff. For example, simply low-pass filtering a speech signal (as it has been used previously in studies of speech intelligibility) would remove fast perturbations from the envelope but of course it would also affect the overall spectral content available. As shown in Figure 1A, this is not the case in the orthogonal manipulations devised here.

Procedure

During EEG acquisition, participants were seated in front of a black screen computer monitor in a dimly lit and soundproof EEG cabine, wearing Sennheiser HD-280 headphones. Sounds were presented at 60 dB(A) sound pressure level.

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 stimuli words and to indicate by way of a 4-way button system how comprehensible a given trial's word stimulus had been. This rating technique shows remarkable consistency with actual recognition scores (see also Davis and Johnsrude 2003, where rating and recognition scores within participants showed a correlation score of 0.98) and was also used with this material in an fMRI experiment before (Obleser et al. 2008). Other own EEG experiments showed that the

rating task is able to accurately replicate known comprehension effects (Obleser and Kotz 2011; Fig. 1B).

A single trial was timed as follows: Three seconds after each stimulus (with spoken word stimuli's duration varying naturally; average file length was 0.6 ± 0.12 s SD), a question mark occurred on the screen prompting for the participant's button press. The question mark disappeared after the button press (time-out after 2 s). The response period was followed by a 2-s period (indicated by an eye symbol present for 0.5 s) during which participants were instructed to blink if necessary. Then, with a random stimulus onset asynchrony of 0.5–1.5 s, the next trial began.

Button-to-value assignment was counterbalanced across subjects (with 4 buttons in total and 2 buttons per hand, the "very intelligible" response was either mapped to the leftmost or the rightmost digit). After a brief (15-trial) familiarization period, the actual experiment was started which was broken down in 3 runs, yielding a total experimental time of about half an hour.

In total, 320 trials of interest were acquired (trials for one additional condition with monaural stimulation were also acquired, see the design of Obleser et al. 2008; these trials were not part of this analysis, however). That is, for every of the 16 conditions in the spectral × envelope design (Fig. 1), 20 trials were presented. Note that in principle all analyses were designed to find parametric effects of the 2 manipulated stimulus dimensions, "spectral detail" and "envelope detail." Thus, all analyses were run on the marginal averages over the design's cells, and 8 marginal conditions resulted from averaging, each with a maximum of 80 averages and each reflecting a level of spectral detail (2, 4, 8, or 16 bands in vocoding) or envelope detail (2, 4, 8, or 16 Hz low-pass filtering applied to the vocoding bands).

### Electroencephalographic Recording and Data Analysis

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

All data were analyzed using the Fieldtrip software (http://www.ru.nl/fcdonders/fieldtrip; Oostenveld et al. 2011), an open source Matlab toolbox for EEG and magnetoencephalogram (MEG) data analysis developed at the F. C. Donders Centre for Cognitive Neuro-imaging as well as custom Matlab (Mathworks) scripts. After each recording session, individual electrode positions were tracked using a Polhemus FASTRAK electromagnetic motion tracker (Polhemus, Colchester, VT, USA).

Off-line analysis followed the protocol of defining the data samples of interest (1 s pre- to 2 s poststimulus onset); re-referencing all data to average reference; applying a high-pass filter of 0.3 Hz (including 2-s padding of all trials before filtering); automatically rejecting trials affected by ocular or muscle artefacts, using the fieldtrip-implemented, z score–based rejection routines. On average, less than 25% of all trials and in no subject or condition more than 48% per subject and condition were rejected; each condition- and subject-specific average contained on average 53.4 trials. A 2 × 4 repeated measures analysis of variance on rejection rates with factors degradation level (1–4) and spectral/envelope showed no significant effect whatsoever; all $F < 1$, ensuring the absence of systematic signal to noise ratio differences between conditions).

The resulting clean data were submitted to a time–frequency analysis of "induced" or non–phase-locked changes in brain oscillations. For the time–frequency analysis, we used the Fieldtrip-implemented version of the Wavelet approach using Morlet wavelets (Tallon-Baudry et al. 1997; Tallon-Baudry and Bertrand 1999), with which the time series were convolved. Wavelet-based approaches to estimating time–frequency representations of EEG data form a good compromise between frequency and time resolution (here, a constant resolution factor $m = f/\sigma_f$ of 7 was used). We convolved the signal with the wavelets in the frequency domain from −0.5 to 1 s in 20-ms steps and from 2 to 48 Hz in 2-Hz steps. The resulting power estimates in each time–frequency bin poststimulus onset were transformed into estimates of relative

A 4 × 4 Stimulus Manipulation
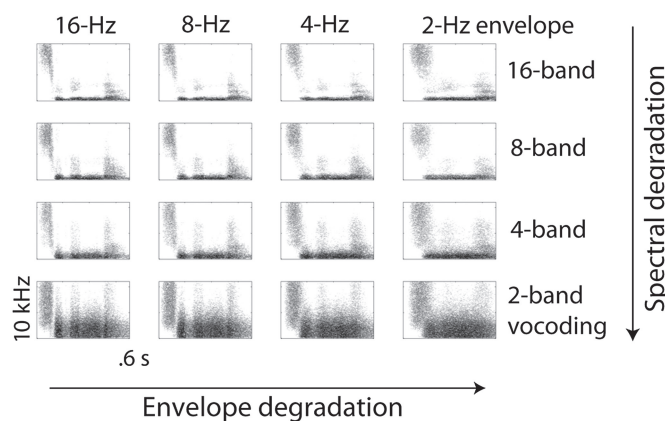
B Comprehension rating
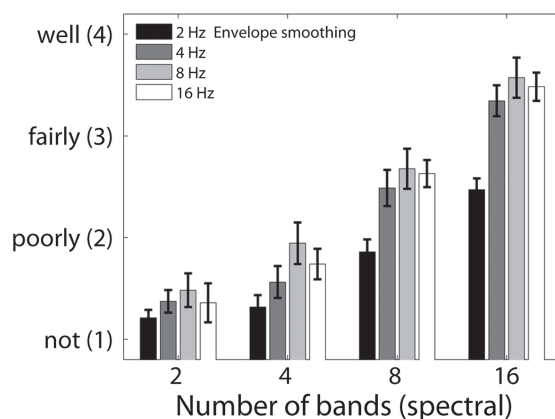
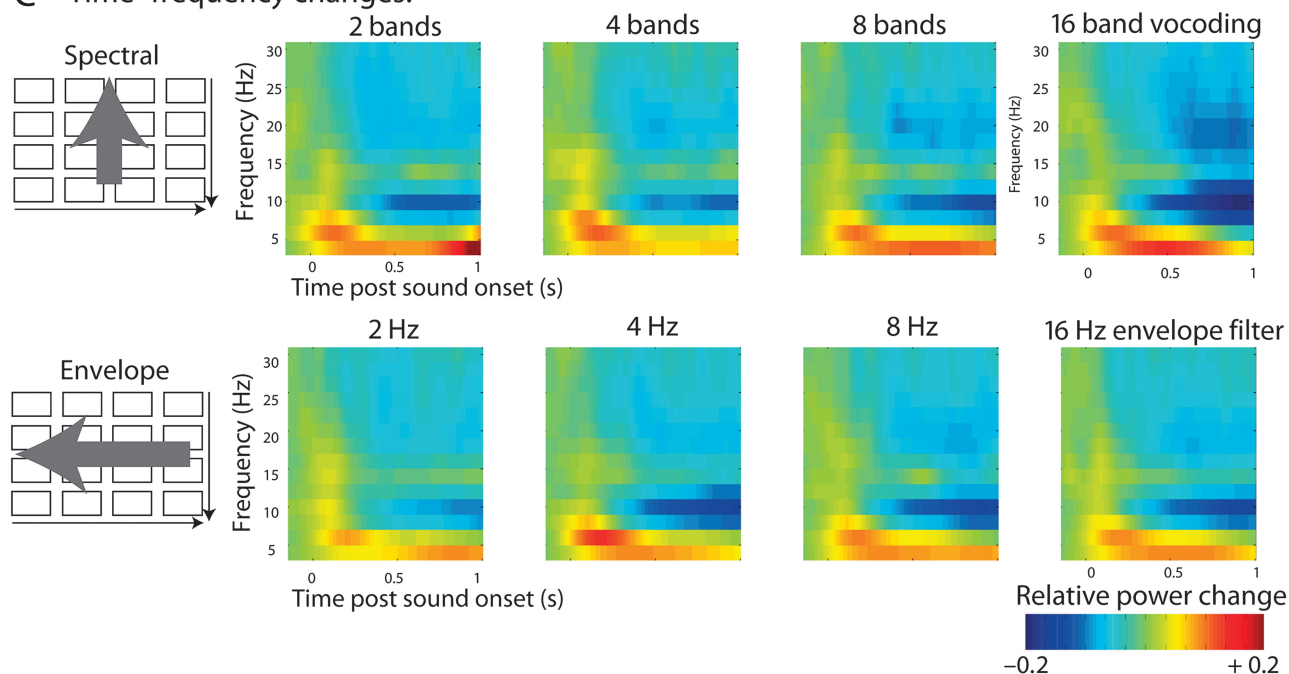
C Time–frequency changes:

Figure 1. (*A*) Stimulus design of spectral and envelope variation. Arrows indicate the 2 dimensions along which the signal was degraded. (*B*) Comprehension rating results during EEG. Note the 2 main effects of spectral and envelope degradation. (*C*) Time–frequency power changes poststimulus onset per conditions. Spectral (top row) and envelope (bottom row) manipulations affect mainly the magnitude of late alpha suppression. The panels reflect averages over the posterior–central channels that were part of the statistical clusters shown in Figure 2.

power change compared with power estimates in a baseline window (–0.5 to –0.1 s). Average time-frequency representations per participant and condition were thus gained, reflecting "relative" changes in [split it as "(de-)syn - chronization"] in response to a given condition, relative to an equally long silent presentence (–500 to 0 ms) baseline. The ensuing parametrical statistical tests (see below) were run using the absolute power estimates again, that is, on the potentially more sensitive direct comparisons of the different conditions (Maris and Oostenveld 2007).

Statistical Analysis

Subject- and condition-specific averages of the time-frequency representations were submitted to parametric 2-sided regression *t*-tests, embedded in a massed permutation test (as outlined in Maris and Oostenveld 2007; 1000 iterations). In essence, this procedure checks for time-frequency-electrode clusters (here: clusters of at least

3 adjacent electrodes in size) that show parametric effects of either a power "decrease" or an "increase" covarying with the manipulated stimulus dimension. The permutation tests included all bins from 0.1 to 0.9 s across a broad range of frequencies (2–48 Hz) as well as across 46 channels (covering symmetrically the left and right scalp, leaving out the midline and most eccentric electrodes). Such a test effectively controls for an inflated 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).

Source Localization

Sources of alpha activity were also localized using an adaptive spatial filter (Dynamic Imaging of Coherent Sources, DICS; Gross et al. 2001) in the frequency domain. The DICS technique is based on the

cross-spectral density matrix, which was obtained in every trial and condition by applying a multitaper fast fourier transform estimate of the time windows and frequencies of interest (Here yielding estimates centered at 11 Hz with a 3 Hz spectral smoothing and a window length of 400 ms centered at 750 ms poststimulus onset). A realistically shaped 3-layer boundary elements model of the brain was used, adjusted to the individual EEG electrode locations gathered (which were warped off-line using the Fieldtrip function "ft_electroderealign," applying a rigid-body transformation) and a standard template MRI. The resulting volume in each individual was then divided into a grid with a 1-cm resolution, and the lead field was calculated for each grid point. Using the cross-spectral density matrices and the individual lead field, a spatial filter was constructed for each grid point, and the spatial distribution of power was estimated for each condition in each subject. A common filter was constructed from all baseline and posttrigger segments (i.e., based on the cross-spectral density matrices of the combined conditions). Subject- and condition-specific solutions were spatially normalized to Montreal Neurological Institute (MNI) space and averaged across subjects, for display on an MNI template (using SPM8). Figure 4 shows the result of cluster-based statistical tests (essentially the same tests as used for the electrode-level data before) that yielded voxel clusters for covariation of source power with spectral and envelope degradation, respectively. This was mainly done for illustration purposes, and unlike the tests for channel-time-frequency clusters outlined above, no strict cluster-level significance testing was applied. T values are plotted on a standard MR template, and MNI coordinates mentioned in-text refer to brain structures that showed local maxima of activation.

Multivariate Classification

As an additional stage in data analysis, we sought to analyse the information contained in local channel-time-frequency patterns rather than significant power differences. Adopting a protocol from functional MRI research ("searchlight" approach; Kriegeskorte et al. 2006; Kriegeskorte and Bandettini 2007), a linear classifier (support vector machine [SVM], using the "libsvm" matlab toolbox, v2.89) was trained on decoding certain stimulus attributes (see below) from local time-frequency patterns. Many studies in cognitive neuroscience have recently reported accurate classification performance on functional MRI voxel pattern using an SVM classifier (e.g., Haynes and Rees 2005; Formisano et al. 2008), and SVM is one of the most widely used classification approach across research fields. For our main classification question (accurate classification of high vs. low intelligibility using the very salient spectral or number of bands variation), a feature vector was obtained using the power change estimates from a set of time-frequency bins (see searchlight approach below) as feature values.

In brief, a linear SVM separates training data points x for 2 different given labels (e.g, "intelligible," 8- and 16-band speech vs. "unintelligible," 2- and 4-band speech) by fitting a hyperplane $w^T x + b = 0$ defined by the weight vector w and an offset b. The classification performance (accuracy) was tested using a leave-one-out cross validation across participants' data sets: The classifier was trained on $n - 1$ data sets, while an nth subject's data set was left out for later testing the classifier in "predicting" the labels from the local time-frequency activation pattern, ensuring strict independence of training and test data. Classification accuracies were obtained by comparing the predicted labels with actual data labels and averaged across the 16 leave-one-out iterations afterward, resulting in a mean classification accuracy value per channel-time-frequency bin.

Importantly, we chose a multivariate so-called searchlight approach to assess the local discriminative pattern over the entire channel-time-frequency space measured (Kriegeskorte et al. 2006): Multivariate pattern classifications were conducted for each position in the channel-time-frequency space, with the searchlight feature vector containing the power estimates for that time-frequency bin and a defined group of its closest neighbors. Here, a searchlight radius of 2 adjacent bins in time and frequency and averaging across up to 5 neighboring channels was selected (maximal distance 5 cm; analogous to the neighbor selection used in the cluster statistics above). This yielded about 12 time-frequency bins per searchlight position. Thus, any significant bin shown in the figures will represent a robust local

pattern of on average 12 nearest-neighbor bins (averaged across up to 5 adjacent channels). "Robustness" was ensured by constructing boot-strapped ($n = 1000$) confidence intervals (CIs) for all bins' mean accuracy and thresholding the maps to only show those time-frequency bins whose lower 95% confidence limit of mean accuracy did not cover the 50% chance level.

As an additional control for a possible inflated alpha error due to multiple comparisons (which is often neglected when using CIs; Benjamini and Yekutieli 2005), we used a procedure suggested in analogy to the established false discovery rate (FDR; e.g., Genovese et al. 2002), called "false coverage-statement rate" (FCR). In brief, we "selected" those time-frequency bins whose 95% CI for accuracy did not cover the 50% (chance) level in a first pass (see above). In a second correcting pass, we (re-)constructed FCR-corrected CIs for these select bins at a level of $1 - R \times q/m$, where R is the number of selected bin at the first pass, m is the total number of time-frequency bins tested, and q is the tolerated rate for false coverage statements, here 0.05 (Benjamini and Yekutieli 2005). Effectively, this yielded FCR-corrected bin-wise confidence limits at $\alpha \sim 0.004$ rather than 0.05. The FCR correction procedure as well as the across-subjects leave-one-out-validation is described in detail in Obleser et al. (2010).

Results

The behavioral rating results gathered after each EEG trial accurately reproduce the known effect of more acoustic detail improving word comprehension (Fig. 1B). Both manipulations affected the comprehension rating (spectral detail, 2–16 bands of vocoding: $F_{1.72,39.52} = 1114$, $P < 0.0001$; envelope detail, 2–16 Hz low-pass filtering of the vocoding envelopes: $F_{2.62,60.20} = 980.1$, $P < 0.0001$). Also in line with previous tests using these materials (see Obleser et al. 2008), the interaction proved also significant ($F_{5.76,132.42} = 972.7$, $P < 0.0001$) and is best explained by the "primacy" of spectral detail: With low-spectral detail (2 bands), comprehension ratings were on average never better than approximately 1.5 on a 1–4/"not"–"well" scale, irrespective of envelope detail in the signal; with high-spectral detail (16 bands), ratings never dropped below 2.5 instead (Fig 1B).

Our EEG analyses focused on the induced (i.e., not strictly phase locked but time locked) changes in power of alpha frequency oscillations, depending on the 2 dimensions of acoustic detail that we manipulated (spectral or fine temporal detail vs. envelope or coarse temporal detail) and the 4 levels of degradation along each dimension.

Overall, the time-frequency grand averages shown in Figure 1C illustrate a pattern that is typical for auditory stimulation: At first, strong initial enhancement or synchronization of power (relative to the average of the baseline period) in lower theta and alpha frequencies occurs. It extends up into the lower Gamma-band range and is most prominent at parietocentral channels of the scalp (around Cz). It is also strongly reflecting the mainly phase-locked parts of the signal (i.e., the evoked potential one observes when averaging in the time domain). From approximately 300 ms post sound onset on, this initial enhancement is followed by a pronounced and temporally as well as spatially wide-spread suppression or resynchronization that has its peak in the alpha frequency range but extends to the beta range and lasts up to 1 s after sound onset.

Time–Frequency Clusters Covarying With Acoustic Detail

The first statistical analysis sought for monotonic changes in power covarying with the increase of spectral detail. Positive time–frequency–channel clusters resulting from this test would indicate increases of power as the signal contains more spectral

or fine structure detail, that is, as the signal changes from 2- over 4- and 8-band to 16-band vocoding. Negative clusters would accordingly indicate decreases or suppression of power.

As shown in Figure 2A (upper row), a significant cluster of power suppression was found over mostly left parietocentral channels and later extending also to left frontal channels; it extended from about 500–900 ms post word onset and peaked in the alpha band (8–13 Hz) range ($P < 0.0001$). The bar graph in the upper right panel illustrates the monotonic decrease in alpha power (relative to the prestimulus baseline) within a representative channel of this cluster (TP7) with increasing spectral detail (and hence, increasing comprehension of speech, cf. the behavioral results in Figure 1B).

Testing for clusters of monotonic change with the increase of envelope detail yielded 2 negative clusters ($P < 0.003$ and $P < 0.03$). Both reflected a parietocentral peak in the alpha band and also began after 500 ms and broadened across the scalp. It is of note that the much stronger, first negative cluster of envelope detail changes has its peak over right parietocentral clusters. However, the later broadening to the left scalp (reflected statistically by the second cluster) as well as the absence of a clear-cut hemisphere × manipulation interaction supports best the general conclusion that parietocentral alpha suppression is monotonically dependent on the amount of acoustic detail. The strength differences of the clusters (Fig. 2A) are in line with the behavioral data (Fig. 1B), which all put forward a stronger leverage of spectral detail on both, alpha suppression and subjective comprehension ratings.
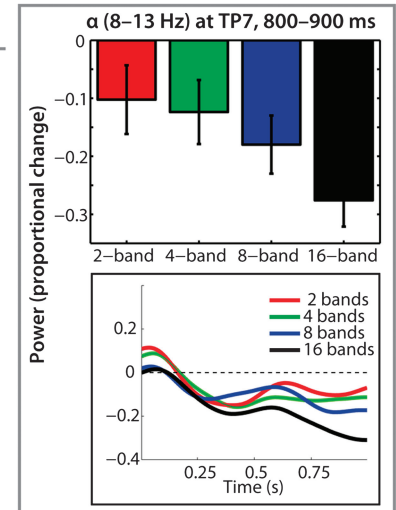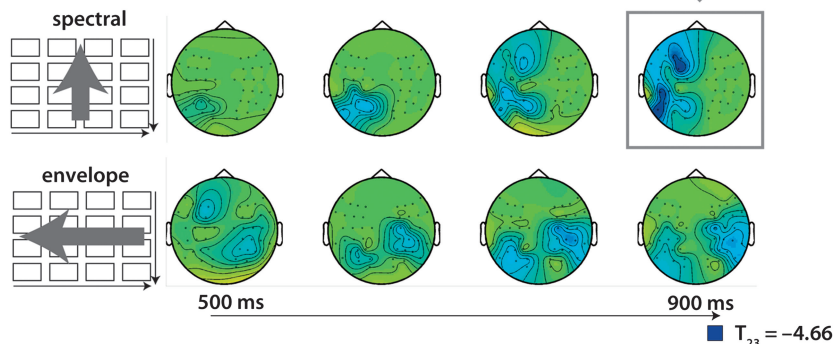
In a similar vein, the only significant positive cluster (i.e., a monotonic increase in power with more acoustic detail being present) was also observed with increasing spectral but not envelope detail: Power was enhanced in the theta frequency (4–7 Hz) range from about 300 to 500 ms ($P < 0.02$). This is shown in Figure 2B and the lower left bar graph. Note that this effect preceded the alpha suppression effects in time.

Influence of Alpha Power on Comprehension Ratings

Across conditions and participants, there was a notable correlation of the extent of posterior–central alpha power suppression and subjective comprehension ratings expressed after the trial by participants (Pearson's $r = -0.30$, $P < 0.01$; when using Spearman's rank correlation to better account for the ordinal scaling of the rating data, $r = -0.29$, $P < 0.01$).

To properly account for the between- and within-subject sources of variance, a mixed regression model was run, with ratings as the dependent measure and alpha power as well as spectral degradation as predictors, plus subject as a random term. This confirmed the strong influence of alpha power onto the rating ($F_{1,92} = 337$, $P < 10^{-4}$). Figure 3 illustrates this in the left panel, showing a scatter plot of all condition- and participant-specific alpha power estimates (average of 800–900 ms over channels P5/P6, PO3/PO4, and PO7/PO8) versus the average comprehension ratings in these conditions and participants. Additionally, the model confirmed that the actual spectral degradation level (2–16 bands) has a very strong

A Alpha (8–13 Hz) desynchronization

spectral

envelope

500 ms   900 ms

$T_{23} = -4.66$

α (8–13 Hz) at TP7, 800–900 ms

Power (proportional change)

2–band 4–band 8–band 16–band

2 bands
4 bands
8 bands
16 bands

0.25   0.5   0.75
Time (s)

B Theta (4–7 Hz) synchronization

spectral

300 ms   700 ms

$T_{23} = 4.66$

θ (4–7 Hz) at Fz, 300–500 ms

Power (proportional change)
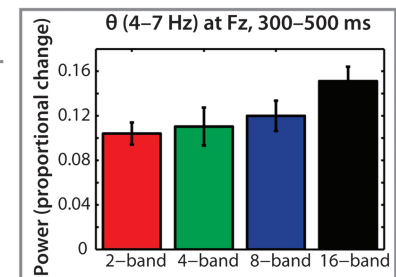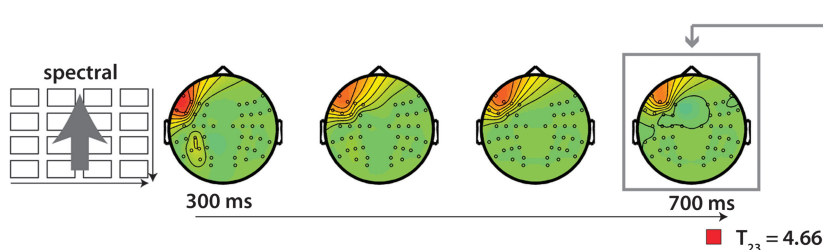
2–band 4–band 8–band 16–band

Figure 2. Time–frequency–electrode cluster statistics. (A) Alpha power changed as a function of both spectral (top row) and envelope (middle row) detail, as also indicated by the bar graph on the right. Over time (500–900 ms shown), stronger decreases in alpha power occurred for more acoustic detail being present. (B) Theta power at left frontotemporal channels increased in the 300–500 ms time range as a function of spectral detail. See Results for details.
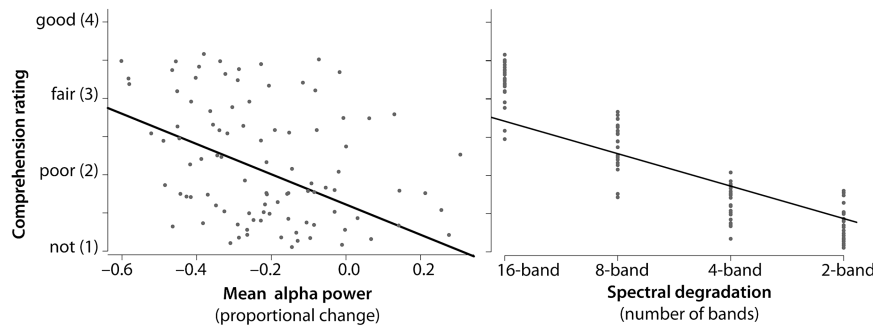
Figure 3. Influence of alpha power and spectral detail on comprehension. A negative correlation of alpha late power at posterior channels and ratings of comprehension was observed (left panel, $P < 0.01$). The right panel further illustrates the influence of spectral detail onto these rating. Mixed regression models (accounting for within- and between-subject variance, see Results) indicated that alpha power and spectral detail exhibit independent influences on the comprehension ratings.

influence on the rating ($F_{1,92} = 909.7$, $P < 10^{-4}$; shown in the right panel of Fig. 3). Importantly, however, the interaction term of spectral degradation and alpha power was not significant ($P = 0.23$)—this indicates that, while spectral degradation drives the late alpha suppression, both had an independent strong influence on post-trial rating behavior.

Lastly, late alpha power was also negatively correlated to the power of preceding theta, indicating an inverse relationship of theta power and ensuing alpha suppression in these data (Pearson's $r = -0.27$, $P < 0.01$). Overall, the alpha suppression effect appears most tightly coupled to comprehension ratings.

Source Localization of Alpha Power Changes

In order to infer tentatively on the most likely source localization of the late alpha suppression, data from the time–frequency window of interest were submitted to a DICS beamformer routine (centered at 750 ms post word onset and 11 Hz; see Materials and Methods for details). The source space representations of relative alpha change compared with baseline were then also tested for voxel clusters covarying with the monotonic increase in spectral and envelope detail, respectively. The results are shown in Figure 4. Voxels co-varying most strongly with both spectral and envelope detail manipulations were located in bilateral superior parietal cortex (MNI peak coordinates [26, –60, 58] in the right and [–21, –34, 71] in the left hemisphere), extending also into the right occipital cortex; in the left inferior prefrontal cortex (–47, 39, 2); and bilaterally in the anterior superior temporal cortex (left temporal pole, [–46, 21, –26], right superior temporal gyrus/ planum polare [56, 2, –6]). When running separate source analyses for covariation with spectral and with envelope degradation (not shown), it can be shown that the strongest source peaks ($|t_{23}| \geqslant 2$) are driven by the spectral changes; envelope detail changes, which had a lesser influence on comprehension also contribute only $|t_{23}|$ values lower than 2.

Multivariate Searchlight Classification of Intelligibility from Time–Frequency Data

Classification accuracies in the across-subjects leave-one-out classification procedure performed across time–frequency bins (averaged across up to 5 adjacent electrodes) confirmed that the alpha frequency range (approximately 8–13 Hz) in the peri-, and poststimulus time window at posterior electrodes contained characteristic information about the 2 broad categories of speech intelligibility (high spectral detail in 8-

and 16-band speech vs. low-spectral detail in 2- and 4-band speech); allowing for significant above-chance classification. The multiplot of time, frequency, and electrodes in Figure 5 has been masked to show only time–frequency bins whose accuracy CIs (corrected for multiple comparisons; see Materials and Methods) do not cover chance level (50%). It is evident that the highest classification accuracies (classifying data of an nth subject after training with $n - 1$ independent data sets) appear in the alpha frequency range at 600–900 ms post word onset particularly at posterior sites. The scalp topography of classification accuracies in this time–frequency rage further illustrates this (Fig. 5, bottom panel).

Task-Dependent Versus Task-Independent Aspects of Alpha Suppression

Lastly, we aimed to quantify how much of the late alpha suppression effect and its dependence on acoustic detail might have been induced by the comprehension-rating task that had been performed following each trial. To this end, we reinvited 2 participants (S5 and S23; both had initially produced strong parametric alpha effects) after a period of more than 1 year. They underwent the identical experiment, however, instead of engaging in an active comprehension-rating task they were only asked to "listen attentively to the more or less comprehensible words" and were prompted afterward with on-screen digits 1–4 (instead of a question mark) to press the respective button. All this was performed entirely in keeping with the original experiment's trial timing, number of trials, recording, and analysis parameters.

Figure 6 illustrates the effect of task presence versus absence on the alpha effect for increasing spectral detail, based on 2 sessions (task presence vs. absence) from 2 participants. Using data from the 2 respective sessions of these 2 subjects and calculating F-statistics at the source level over single trials, we performed 2 tests on the source localizations of the respective late alpha-band power: First, which voxels would show an effect of spectral detail increase (see main result), irrespective of task? Active voxels in this test would reflect task-independent effects of spectral detail. It turned out that this effect had contributions from superior parietal cortex bilaterally as well as superior temporal cortex also in this small sample of only 2 subjects ($F_{3,837} \geqslant 2.6$). This corroborated the group result reported here as well as an earlier single-subject report (Weisz et al. 2011).

Second, however, we tested which voxels would show an effect of task presence versus absence, irrespective of spectral
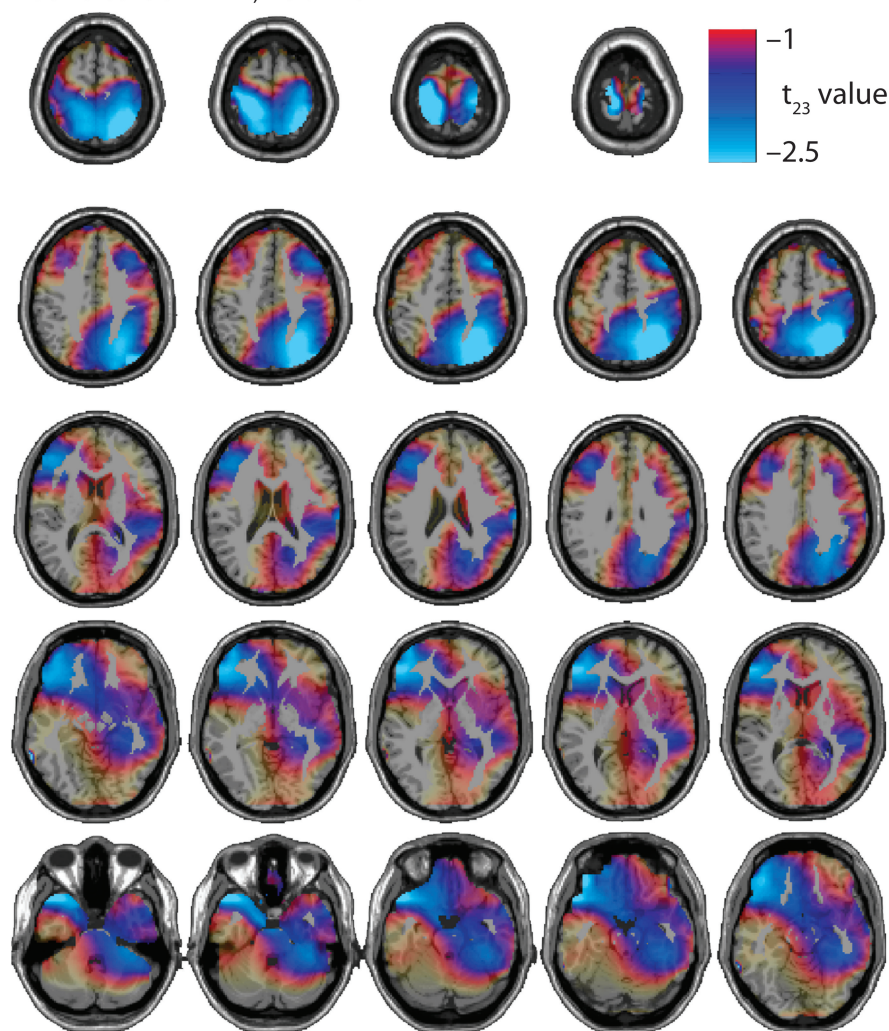
Figure 4. Source localization statistics on the alpha suppression effect. The panels show source localization of the acoustic degradation effect on late alpha power (source-level regression T statistic) plotted onto axial slices ($z = -34$ to 80) of a standard T_1-weighted MR image. It confirms the strong contribution of superior parietal cortex but also highlights a right dorsolateral prefrontal source and bilateral anterior temporal lobe sources. See text for details.

detail. This test result yielded very strong activation in attention- and task-monitoring regions in the anterior cingulate as well as the posterior cingulate/cuneal cortex (Fig. 6, activation shown in red; $F_{1,835} \geqslant 8$). These areas exhibited stronger alpha modulation when an active task was following the trial. In sum, and despite some overlap mainly in right parietoocciptial cortex (shown in green), the effect of task has an underlying source topography that is distinct from the spectral degradation effect that forms the main focus of our study.

Discussion

Acoustic degradations of the spectral and envelope detail are able to severely compromise the intelligibility of speech. The present study sought to establish a potential link between oscillatory power in alpha (8–13 Hz) power of the human scalp potential and the (dis-)ability to comprehend speech

despite such acoustic degradation. To this end, we analyzed the time–frequency oscillatory changes occurring in auditory presentation of parametrically degraded words.

Main findings can be summarized as follows: First, the strong alpha power suppression observed from about 500 ms post word onset on depended parametrically on spectral and envelope detail. The fewer acoustic detail was present in a stimulus (i.e., the less likely that the stimulus would be comprehended), the relatively more alpha power was observable over posterior scalp electrodes.

Second, spectral and envelope degradation differed in their specific leverage on alpha power and on comprehension, with spectral changes having a more potent influence on essentially all parameters tested: In line with 2 previous experiments using the same manipulations (Obleser et al. 2008), spectral degradation had a stronger leverage on comprehension ratings. Concomitantly, changes in spectral detail were most closely tied to alpha power change at left scalp electrode sites (and
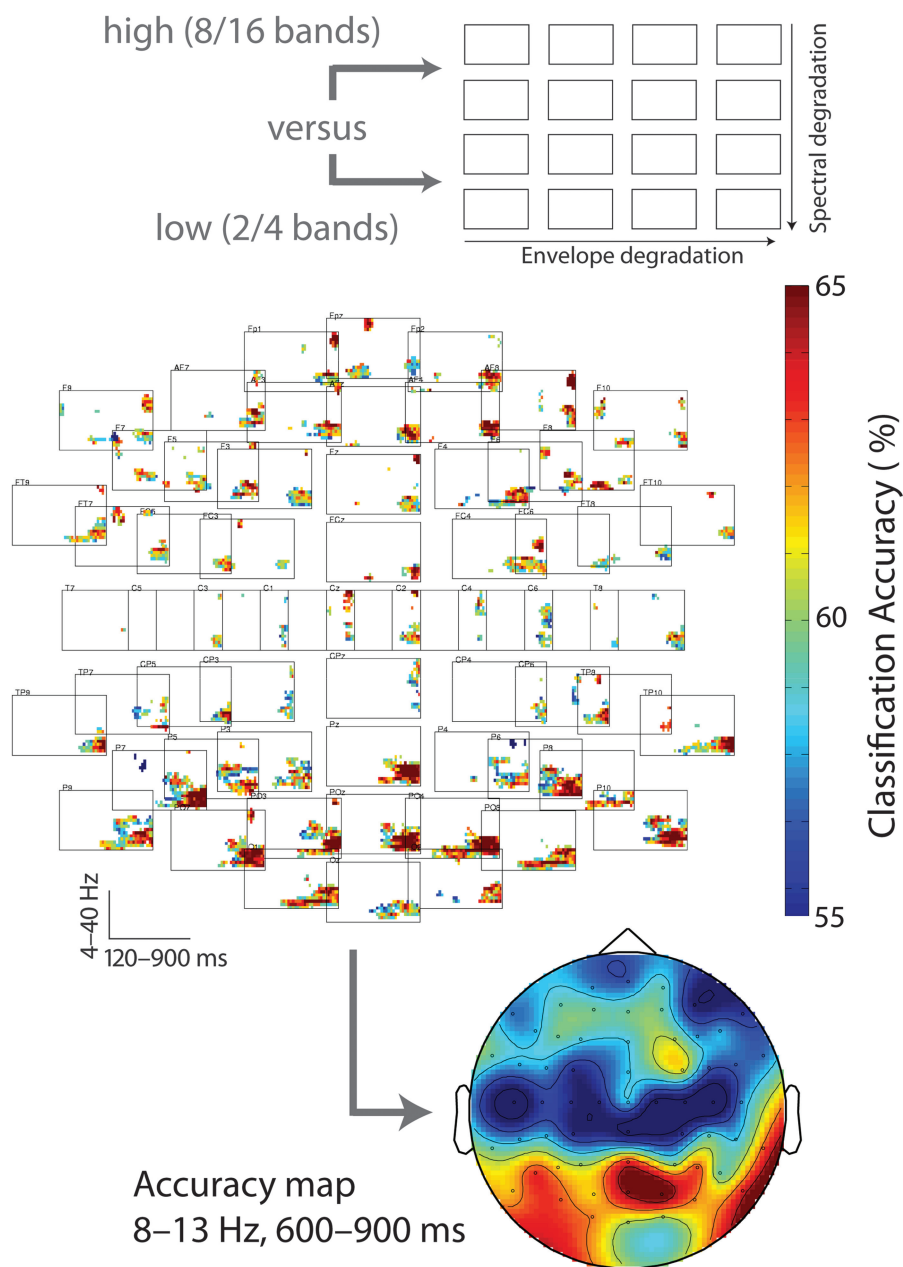
Figure 5. Results of multivariate pattern classification. The channel–time–frequency array in the middle plots significant above-chance bins that allow classifying hardly intelligible (2- and 4-band speech) from likely intelligible (8- and 16-band speech) speech (see schematic display in top panel). Only clusters with a multiple-comparisons-corrected accuracy confidence limit that does not cover chance level (50%) and a cluster extent of at least 6 adjacent bins are plotted. The bottom panel shows a scalp map for the top-accuracy time–frequency area, the late (600–900 ms) alpha (8–13 Hz) range. See Materials and Methods and Results for details.

superior parietal cortical generators), while envelope detail was slightly less tight-coupled to alpha power changes, which appeared first more right lateralised and later bilaterally. Changes in spectral detail were additionally yielding a theta power increase, with most intelligible words yielding strongest theta power over left electrodes. (This preceded the alpha power decrease and was, across conditions and participants, negatively correlated with it.)

Third, when training and testing a machine learning algorithm on the electrode-time-frequency patterns from independent participant data sets, it were the late alpha power changes at posterior electrodes that classified best patterns associated with trials likely to be comprehended from patterns of trials more unlikely to be comprehended.

Alpha and Theta Oscillations in the Comprehension of Speech

The data presented fill a missing piece in at least 2 puzzles, which are mostly studied separately: the neural mechanisms of speech perception and comprehension and the role of neural oscillations in cognition. As for the former, only comparably few studies have exploited the temporal precision of M/EEG
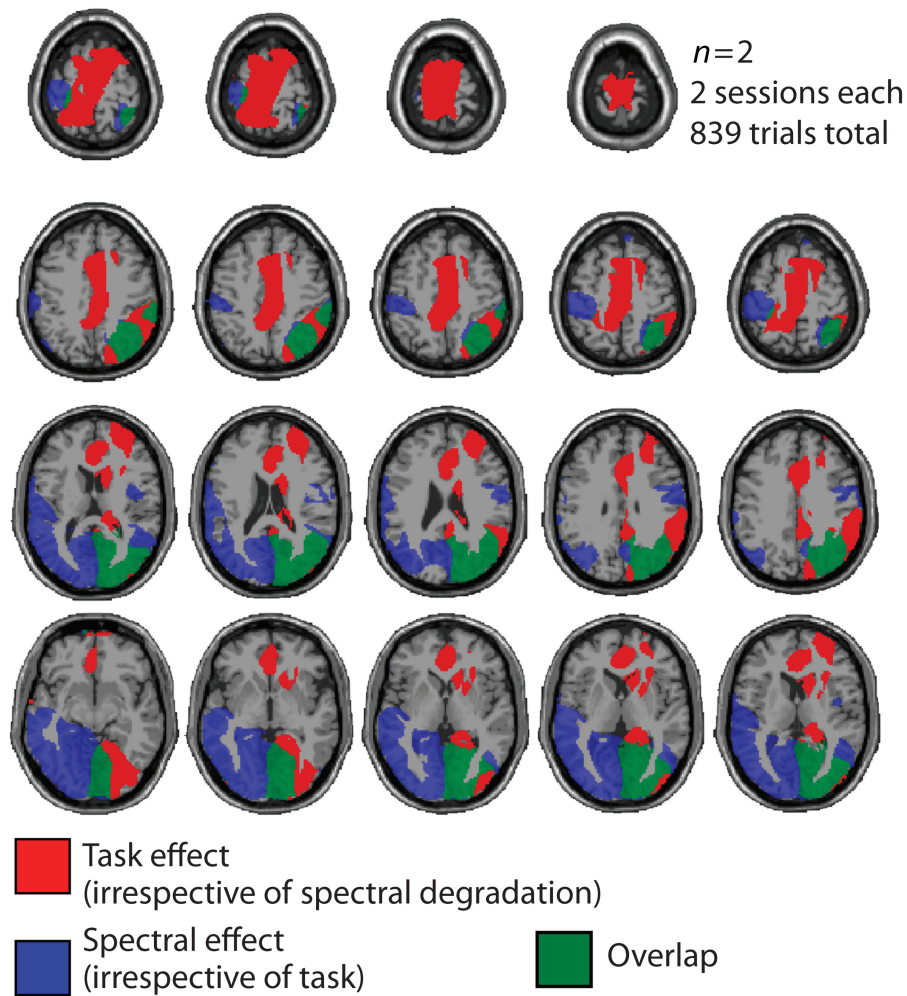
Figure 6. Task in of the spectral degradation effect. For $n = 2$ subjects and $k = 2$ sessions (with and without the active comprehension task), the effect of task versus no task (in red), the effect of spectral degradation (in blue), and their overlap (in green) are shown. Axial slices ($z = -34$ to 80) of a standard T_1-weighted MR image are shown. The active task enhances the observed alpha power suppression (source-level F-statistic for task vs. no task over trials; shown in red). However, irrespective of task, the main finding of alpha suppression as a function of spectral detail replicates (source-level F-statistic with a linear contrast for spectral detail over trials, shown in blue). Overlap is strongest in right superior parietooccipital cortex.

for studying oscillatory dynamics in auditory perception (e.g., Weisz et al. 2007; Kerlin et al. 2010) and speech comprehension (e.g., Luo and Poeppel 2007; Shahin et al. 2009; Obleser and Kotz 2011), and considerable emphasis has been put on the gross functional neuroanatomy of speech comprehension instead. As for the latter, increasing amounts of data are gathering that put forward analogous roles of alpha and other oscillations in audition as described in much greater detail already for other modalities (for a review on audition and alpha, see Weisz et al. 2011). The current study is to our knowledge the first that directly tests spectrotemporal specifics of speech intelligibility and their respective imprints on alpha oscillations.

Two particular findings deserve elaboration here: Presence or absence of the spectral detail of speech signals, often referred to as "temporal fine structure" (Rosen 1992; Gilbert and Lorenzi 2010), proved to be the critical and determining parameter that affected not only comprehension most directly but also almost all brain parameters reported in the present study. Presence or absence of the slower temporal, or "envelope," detail was much less effective in driving comprehension as well as concomitant brain parameters. This is

particularly notable, as the slow temporal amplitude envelope of speech receives a lot of attention in current models of speech perception (Luo and Poeppel 2007; Chandrasekaran et al. 2009), and it is the one of the few speech parameters for which simple yet testable models of neural processing exist: The envelope of speech is mainly driven by syllable rate, usually 3–7 Hz, which happens to be a theta rhythm. Thus, phase locking of the cerebral theta rhythm to the envelope of speech is a first potent mechanism for "syncing" the receiving brain to the emitted signal (Luo and Poeppel 2007; Lakatos et al. 2008; Schroeder et al. 2008; Ghitza and Greenberg 2009).

The present results complicate this view somewhat: With respect to behavior, they again demonstrate the fact that envelope cues alone are certainly helpful but not sufficient for speech comprehension (e.g., Shannon et al. 1995; Xu et al. 2005), and that speech without any envelope cues can be intelligible as well (Lorenzi et al. 2006). Instead, with respect to neural oscillations, they show that all tested parameters (degree of alpha suppression, but also theta enhancement) depend more on spectral than on envelope detail. This does not detract from a possible role for theta–envelope phase locking

in speech processing, though, as our study used single word rather than longer utterances and thus does not allow valid conclusions on theta phase-locking behavior over longer periods of time. A tentative hypothesis for further studies could be that (1) spectral cues are ultimately more essential than envelope cues for intelligibility (i.e., the auditory system's ability to match speech sound features onto meaning; see evidence cited above), and that (2) on the utterance or discourse level, theta phase locking to the envelope should vary as a function of intelligibility, that is, spectral cues being present that render the speech signal intelligible in the first place (Ahissar et al. 2001; M. H. Davis, personal communication).

The Functional Network of Alpha Suppression in Degraded Speech

The relative late occurrence of the observed alpha differences is not very surprising, given the word length of on average 600 ms and the time that slow alpha modulations can take to develop (200–300 ms). The time point of maximal alpha suppression nevertheless speaks to a high-hierarchical, integrative rather than a very low-hierarchical, sensory processing stage.

A closer look at the brain areas implied to be involved in this alpha oscillatory network also can be interpreted in favor of such a late processing stage of degraded speech input, spanning across wide-spread brain areas: The peak of alpha suppression originates in the superior parietal cortex, which has been linked closely to executive and attention processes in many previous studies (e.g., Buschman and Miller 2007; Medendorp et al. 2007; Koenigs et al. 2009) and which belongs to the "dorsal attention network" (Sadaghiani et al. 2010). A capture of the attention system by auditory stimulus features has been demonstrated, in a much more low-level beep/flash design, for example, by Foxe et al. 1998. The strong modulation of alpha power in these regions might reflect a differential (dis-) engagement of the parietooccipital (primary visual) attention system depending on speech intelligibility.

However, the localization results also indicate subpeaks of alpha suppression in far "downstream" auditory processing areas (Rauschecker and Scott 2009; Obleser et al. 2011) in the anterior temporal cortex and inferior frontal cortex: In particular, the peak in right anterior superior temporal gyrus collocates well with a hemodynamic (blood oxygenation, blood oxygen level–dependent [BOLD]) increase reported for these kinds of spectrotemporal degradation variations reported in Obleser et al. (2008). Given the overriding evidence for negative correlations of alpha power and BOLD increase (e.g., Laufs et al. 2006; Sadaghiani et al. 2010), this provides a plausible link of the current EEG results to previous BOLD data on the same materials and task. The inferior frontal alpha suppression peak also aligns well with reports of increased inferior frontal cortex BOLD activation in processing degraded speech (e.g., Obleser and Kotz 2010; Davis et al. 2011).

In sum, the network suggested by the temporal and neuroanatomical characteristics of the alpha suppression effect is the following: Higher levels of intelligibility not only trigger activation in downstream auditory processing areas but they also strongly modulate a posterior attentional network. The following section will lay out possible mechanisms of how degraded speech triggers differential needs for attention and functional inhibition in more detail.

Adverse Listening and the Role of Functional Inhibition

In terms of neural mechanisms, how can the observed alpha power changes in adverse listening situations be interpreted? The strongest suppression for the most intelligible condition could be considered the "default" state in speech comprehension that has been observed in other studies before (e.g., Shahin et al. 2009). We would like to argue that the relative decline in alpha power is likely to reflect an increase in mental operations performed on the speech signal, thus, more-attentive active cognitive processing. This is what follows directly from a view that associates high alpha power in broadly distributed neural networks with functional inhibition of more local high-frequency oscillations, which in turn are thought to reflect neural computations (Fries 2009; Jensen and Mazaheri 2010). A post hoc look at gamma power in the present data yielded a trend-level increase of gamma power within the clusters identified as alpha suppression clusters. This is in favor of this interpretation (cf. Osipova et al. 2008), but any gamma power difference might have been partly canceled out by the acoustically somewhat variable stimuli (e.g., varying naturally in exact length, see Materials and Methods), while the alpha effects might have been more robust against such variance at the stimulus level.

Viewed from a different angle, however, an alternative hypothesis would be that the relative "lack" of such alpha suppression in severely degraded speech reflects neural oscillators that keep the alpha power high in order to "gate out" erroneous and misleading activations in language- and meaning-related areas. This view can also be derived from the functional inhibition framework: Relative increases of alpha power have been reported during working memory retention (e.g., Jensen et al. 2002; Leiberg et al. 2006), most likely reflecting inhibitory control over items in memory (for review, see Klimesch et al. 2007). Also, it is very likely that listening to degraded speech taxes the cognitive resources of working memory (Pisoni 2000) and selective attention (Shinn-Cunningham and Best 2008). Thus, at this stage, it cannot be entirely ruled out that the observed alpha modulations do reflect relative "increases" in alpha power for more degraded stimuli, reflecting the enhanced need for executive control.

The stronger need for executive control in more severely degraded speech is not to indicate a simple task- or response-driven process: recall that the retest with 2 participants in an almost passive task-free setup did affect the "overall" degree of alpha suppression, but it showed the same "relative" suppression of alpha power dependent on acoustic detail. Thus, the alpha modulation observed is not entirely explained by task- or response-driven processes; it rather seems to depend to large extents on stimulus-driven influences.

Conclusions

The data presented bear relevance to auditory neuroscience, as there are so few previous human studies on neural oscillations during speech comprehension. More importantly though, they provide a new, additional neural parameter that can be employed in further challenging scientific questions: Namely, alpha power changes accompanying the perception of single words and reflecting, to satisfying degrees, a listener's ability and effort to comprehend speech. Alpha power changes have the advantage of being very prominent in the human EEG power spectrum; for the present design, we were also able to

run the classifier on single trials from a single subject (data not shown), yielding also alpha power in the peri-, and poststimulus time range over posterior channels as the most accurate electrode–time–frequency site to tell low-spectral detail/intelligibility trials from high-spectral detail/intelligibility trials. This, plus its monotonic dependency on spectral detail and concomitant comprehension ratings, renders late alpha suppression a possibly valuable parameter when studying populations where M/EEG recording is the only neuroimaging technique of choice (particularly in cochlear implant carriers who are not eligible for functional MRI), as well as populations where comprehension tasks and experiments optimized for signal/noise ratio are not always feasible (hearing-impaired children or elderly participants).

In conclusion, the present data suggest a major role for large-scale alpha-frequency oscillatory networks in comprehending speech under adverse conditions and coping with degradations of its major acoustic feature dimensions. The results also show that comparably late stages of peri- and poststimulus alpha activity are surprisingly informative on the level of acoustic detail and the degree to which listeners are able to utilize it for speech comprehension.

Funding

Notes

References

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci U S A. 98:13367-13372.

Benjamini Y, Yekutieli D. 2005. False discovery rate-adjusted multiple confidence intervals for selected parameters. J Am Stat Assoc. 100:71-81.

Buschman TJ, Miller EK. 2007. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science. 315:1860-1862.

Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA. 2009. The natural statistics of audiovisual speech. PLoS Comput Biol. 5:e1000436.

Davis MH, Ford MA, Kherif F, Johnsrude IS. Forthcoming 2011. Does semantic context benefit speech understanding through "top-down" processes? Evidence from time-resolved sparse fMRI. J Cogn Neurosci. doi:10.1162/jocn_a_00084.

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

Engel AK, 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.

Formisano E, De Martino F, Bonte M, Goebel R. 2008. "Who" is saying "what"? Brain-based decoding of human voice and speech. Science. 322:970-973.

Foxe JJ, Simpson GV, Ahlfors SP. 1998. Parieto-occipital approximately 10 Hz activity reflects anticipatory state of visual attention mechanisms. Neuroreport. 9:3929-3933.

Fries P. 2009. Neuronal gamma-band synchronization as a fundamental process in cortical computation. Annu Rev Neurosci. 32:209-224.

Genovese CR, Lazar NA, Nichols T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage. 15:870-878.

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

Gilbert G, Lorenzi C. 2010. Role of spectral and temporal cues in restoring missing speech information. J Acoust Soc Am. 128:EL294-EL299.

Gross J, Kujala J, Hamalainen M, Timmermann L, Schnitzler A, Salmelin R. 2001. Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc Natl Acad Sci U S A. 98:694-699.

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

Haynes JD, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci. 8:686-691.

Hickok G, Poeppel D. 2007. The cortical organization of speech processing. Nat Rev Neurosci. 8:393-402.

Jensen O, Gelfand J, Kounios J, Lisman JE. 2002. Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. Cereb Cortex. 12:877-882.

Jensen O, Kaiser J, Lachaux JP. 2007. Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci. 30:317-324.

Jensen O, Mazaheri A. 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci. 4:12.

Kerlin JR, Shahin AJ, Miller LM. 2010. Attentional gain control of ongoing cortical speech representations in a "cocktail party". J Neurosci. 30:620-628.

Klimesch W, Sauseng P, Hanslmayr S. 2007. EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res Rev. 53:63-88.

Koenigs M, Barbey AK, Postle BR, Grafman J. 2009. Superior parietal cortex is critical for the manipulation of information in working memory. J Neurosci. 29:14980-14986.

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

Kriegeskorte N, Bandettini P. 2007. Analyzing for information, not activation, to exploit high-resolution fMRI. Neuroimage. 38:649-662.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. Proc Natl Acad Sci U S A. 103:3863-3868.

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE. 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. Science. 320:110-113.

Laufs H, Holt JL, Elfont R, Krams M, Paul JS, Krakow K, Kleinschmidt A. 2006. Where the BOLD signal goes when alpha EEG leaves. Neuroimage. 31:1408-1418.

Leiberg S, Lutzenberger W, Kaiser J. 2006. Effects of memory load on cortical oscillatory activity during auditory pattern working memory. Brain Res. 1120:131-140.

Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BC. 2006. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad Sci U S A. 103:18866-18869.

Luo H, Poeppel D. 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 54:1001-1010.

Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG-and MEG-data. J Neurosci Methods. 164:177-190.

Medendorp WP, Kramer GF, Jensen O, Oostenveld R, Schoffelen JM, 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.

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

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

Obleser J, Kotz SA. 2011. Multiple brain signatures of integration in the comprehension of degraded speech. Neuroimage. 55:713-723.

Obleser J, Leaver AM, Van Meter J, Rauschecker JP. 2010. Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. Front Psychol. 1:232.

Obleser J, Meyer L, Friederici AD. 2011. Dynamic assignment of neural resources in auditory comprehension of complex sentences. Neuroimage. 56:2310-2320.

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

Oostenveld R, Fries P, Maris E, Schoffelen JM. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci. 2011: 156869.

Osipova D, Hermes D, Jensen O. 2008. Gamma power is phase-locked to posterior alpha activity. PLoS One. 3:e3990.

Pisoni DB. 2000. Cognitive factors and cochlear implants: some thoughts on perception, learning, and memory in speech perception. Ear Hear. 21:70-78.

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

Rosen S. 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci. 336:367-373.

Sadaghiani S, Scheeringa R, Lehongre K, Morillon B, Giraud AL, Kleinschmidt A. 2010. Intrinsic connectivity networks, alpha oscillations, and tonic alertness: a simultaneous electroencephalography/functional magnetic resonance imaging study. J Neurosci. 30:10243-10250.

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

Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A. 2008. Neuronal oscillations and visual amplification of speech. Trends Cogn Sci. 12:106-113.

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

Scott SK, Rosen S, Lang H, Wise RJ. 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 AJ, Picton TW, Miller LM. 2009. Brain oscillations during semantic evaluation of speech. Brain Cogn. 70:259-266.

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

Shinn-Cunningham BG, Best V. 2008. Selective attention in normal and impaired hearing. Trends Amplif. 12:283-299.

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

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

Weisz N, Hartmann T, Müller N, Obleser J. 2011. Alpha rhythms in audition: cognitive and clinical perspectives. Front Psychol. 2:73.

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

Xu L, Thompson CS, Pfingst BE. 2005. Relative contributions of spectral and temporal cues for phoneme recognition. J Acoust Soc Am. 117:3255-3267.