

ORIGINAL ARTICLE

The Human Neural Alpha Response to Speech is a Proxy of Attentional Control

Malte Wöstmann, Sung-Joo Lim and Jonas Obleser

Department of Psychology, University of Lübeck, Maria-Goeppert Straße 9a, 23562 Lübeck, Germany

Address correspondence to Malte Wöstmann and Jonas Obleser, Department of Psychology, University of Lübeck, Maria-Goeppert Straße 9a, 23562 Lübeck, Germany. Email: malte.woestmann@uni-luebeck.de (M.W.); jonas.obleser@uni-luebeck.de (J.O.).

Abstract

Human alpha (~10 Hz) oscillatory power is a prominent neural marker of cognitive effort. When listeners attempt to process and retain acoustically degraded speech, alpha power enhances. It is unclear whether these alpha modulations reflect the degree of acoustic degradation per se or the degradation-driven demand to a listener's attentional control. Using an irrelevant-speech paradigm and measuring the electroencephalogram (EEG), the current experiment demonstrates that the neural alpha response to speech is a surprisingly clear proxy of top-down control, entirely driven by the listening goals of attending versus ignoring degraded speech. While ($n = 23$) listeners retained the serial order of 9 to-be-recalled digits, one to-be-ignored sentence was presented. Distractibility of the to-be-ignored sentence parametrically varied in acoustic detail (noise-vocoding), with more acoustic detail of distracting speech increasingly disrupting listeners' serial memory recall. Where previous studies had observed decreases in parietal and auditory alpha power with more acoustic detail (of target speech), alpha power here showed the opposite pattern and increased with more acoustic detail in the speech distractor. In sum, the neural alpha response reflects almost exclusively a listener's goal, which is decisive for whether more acoustic detail facilitates comprehension (of attended speech) or enhances distraction (of ignored speech).

Key words: acoustic detail, alpha oscillation, attention, speech, working memory

Introduction

Selective attention allows neural enhancement of relevant information while simultaneously filtering out irrelevant distractors (Serences and Kastner 2014), thereby promoting successful maintenance of relevant information in memory. Neurally, the demand on attention and memory is reflected in modulations of oscillatory power, specifically in the alpha band (~10 Hz) in the magneto-/electroencephalogram (M/EEG). A higher degree of associated attention demand at task boosts the extent of alpha power enhancement (for reviews, see Jensen and Mazaheri 2010; Foxe and Snyder 2011). Across various sensory modalities, alpha power has been implicated in functional inhibition of task-irrelevant distraction (somatosensation: Haegens et al. 2011; vision: Snyder and Foxe 2010; audition: Wöstmann, Hermann, et al. 2016).

The domain of human speech comprehension provides an excellent case for examining the neural basis of selective

attention and working memory. Speech signals are often degraded by a mixture of irrelevant, interfering sound sources that increase the demand on attention and memory (Shinn-Cunningham 2008; Wilsch and Obleser 2016). Therefore, successful listening requires neural enhancement of target sounds at the expense of distractors (Fritz et al. 2007). To decrease distractor interference, listeners must deploy selective attention to target sounds during perceptual encoding and during working memory maintenance.

The link between neural alpha oscillatory power and speech processing has been well established. For instance, acoustic degradation of target speech, detrimental to comprehension, evokes an alpha power increase (Fig. 1B; Obleser and Weisz 2012; Obleser et al. 2012; Becker et al. 2013; Scharinger et al. 2014; Wöstmann, Hermann, et al. 2015; McMahan et al. 2016). Based on these findings, high alpha power was considered as a neural indicator of high attention and memory demand from

degraded speech, often referred to as “listening effort”. But do these neural responses reflect the degree of acoustic degradation per se, or rather the degradation-driven demand on attention and memory?

Unlike previous studies that showed more acoustic detail in task-relevant target speech to reduce cognitive effort (on memory: [Obleser et al. 2012](#); on attention: [Wöstmann, Herrmann, et al. 2015](#)), our recent behavioral study suggested the opposite effect of better acoustics on cognitive effort ([Wöstmann and Obleser 2016](#)): More acoustic detail in the to-be-ignored speech segments did increase distraction. In other words, listeners were more susceptible to distraction from task-irrelevant speech with higher acoustic detail. These results suggest that listening effort depends entirely on a listener’s behavioral goals and not on stimulus acoustics per se. Here we use EEG to demonstrate that the neural alpha response reflects the effort to ignore a speech distractor, that is, attentional control.

During goal-directed behavior, attention strongly interacts with working memory ([Awh et al. 2006](#)). Evidence for this interaction comes from “retro-cue” experiments, which demonstrate that attention to items in memory improves item recall ([Oberauer and Hein 2012](#); [Lim et al. 2015](#)). We presume that the interplay between attention and memory is strong in degraded speech processing as well: Speech comprehension requires active maintenance of target speech (implemented in the phonological loop of working memory; [Baddeley 1992](#)) and the suppression of neural activity unrelated to target speech (implemented through attention to target speech and away from distractors). We therefore adapted an established paradigm that interfaces attention and working memory, the so-called irrelevant-speech paradigm ([Colle and Welsh 1976](#)). In this paradigm, we acoustically degraded the speech distractor (using noise-vocoding, [Fig. 1A](#); [Rosen et al. 1999](#)) while leaving the target speech signal acoustically intact. Importantly, speech is known to be more distracting if it contains more acoustic detail ([Ellermeier et al. 2015](#); [Wöstmann and Obleser 2016](#)). If neural responses in the alpha band are driven mainly by stimulus acoustics, we expect a similar decrease in alpha power irrespective of whether more acoustic detail is preserved in target or distractor speech. Instead, the present data will demonstrate how the listening goal (here, ignoring speech) drives the neural oscillatory response. Hence, a parametric increase of alpha

power with more acoustic detail of the distractor is reported ([Fig. 1B](#)).

Materials and Methods

Participants

23 German native speakers (12 females; mean age: 24.5 years, SD: 3 years) took part in this study. Data of one additional participant were excluded from all analyses due to a high number of noise-contaminated trials in the EEG. Participants gave informed consent and were financially compensated for their participation. None reported any history of neurological or mental health concerns. Experimental procedures were approved by the ethics committee of the University of Lübeck.

Speech Materials

Full detail on the speech materials and the irrelevant-speech task can be found in [Wöstmann and Obleser \(2016\)](#). Recordings of German spoken digits from 1 to 9 (female voice, average duration: 0.6 s; range: 0.5–0.7 s) from a previous study ([Obleser et al. 2012](#)) served as to-be-attended target speech stimuli.

For the task-irrelevant speech distractor, we used a German version of the speech-in-noise (SPIN) sentences ([Erb et al. 2012](#)) adopted from [Kalikow et al. \(1977\)](#). For the present study we used 180 different sentences (5–8 words, average duration: 2.1 s), which were spoken by the same female voice as the digits. Half of the task-irrelevant sentences had highly predictable final words (e.g., “She covers the bed with fresh sheets” translated from German; where “sheets” is highly predictable from the sentence content) whereas the final word predictability of the remaining sentences was low (e.g., “We are very happy about the sheets” where “sheets” is not predictable from the sentence content). Since the predictability manipulation was found to have no effect in a previous behavioral study ([Wöstmann and Obleser 2016](#)) it was not further analyzed here.

Acoustic degradation was only applied to the speech distractor. To manipulate the acoustic detail, the distractor sentences were spectrally degraded using 1, 4, or 32 frequency channels (ch) for noise-vocoding. In detail, the speech materials were divided into 1, 4, or 32 frequency channels using a

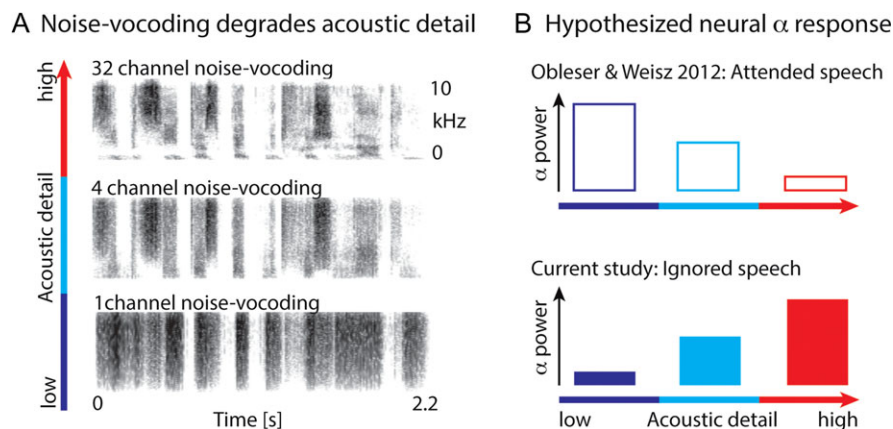


Figure 1. Hypothesized study outcomes. (A) Noise-vocoding with a higher number of frequency channels increases the acoustic (spectral) detail of speech (for details, see Materials and Methods). (B) The goal of this study was to demonstrate that the neural alpha (α) response to speech is under control of a listener’s focus of attention. That is, previous research has shown alpha power decreases when more acoustic detail facilitates comprehension of attended speech (empty bars; e.g., [Obleser and Weisz 2012](#)). But in the current study, we expected that alpha power would instead increase as more acoustic detail aggravates the distraction of ignored speech (filled bars).

bank of bandpass filters (sixth order Butterworth filters spanning frequencies 70–9000 Hz; logarithmic spacing of filter center frequencies according to Greenwood 1990). For each frequency channel, the slow amplitude fluctuation (i.e., the temporal envelope) was extracted (using half-wave rectification and low-pass filtering at 500 Hz; second order Butterworth filter) and multiplied with bandpass filtered noise (white noise filtered with the bandpass filter of the respective frequency channel). Finally, the processed frequency channels were summed up. In essence, the noise-vocoding technique parametrically degrades the spectral content of the acoustic signal (i.e., the fine structure; see Wöstmann, Schröger, et al. 2015) but leaves the temporal information largely intact (Shannon et al. 1995). For a complete list of sentences and further details on the noise-vocoding procedure, see Erb et al. (2012).

Figure 1A shows speech spectrograms for 1, 4, and 32 ch noise-vocoded speech. A higher number of frequency channels results in higher spectral detail and thus enhances intelligibility (Faulkner et al. 2001; also for the particular sentence materials used in the present study: Hartwigsen et al. 2015).

Loudness Adjustments

Target digits were equalized to the same root-mean squared (rms) sound amplitude (−30 dB Full Scale, FS). The sound intensity of differentially noise-vocoded speech distractors (1, 4, 32 ch) was adjusted in order to equalize the perceived loudness. In detail, we found in a previous unpublished behavioral experiment that listeners' perceived loudness ratings for individual spoken words increased with fewer frequency channels used for noise-vocoding ($n = 16$ participants; 18–30 years; native German speakers; no self-reported hearing problems). Thus, in the present study 1, 4, and 32 ch noise-vocoded speech stimuli were accordingly adjusted to 28.5, 31.4, and 32.8 dB FS, respectively. Note our loudness adjustments were relatively weak and that even a much stronger loudness manipulation of distractor speech of 75 versus 60 dB A in an irrelevant-speech task was found to have no effect on behavioral performance (Ellermeier and Hellbrück 1998).

Task Design and Procedure

We used an adapted version of the irrelevant-speech paradigm (Colle and Welsh 1976; Jones and Morris 1992). On each trial, participants listened to the German digits from 1 to 9 presented in random order. Spoken digits had an onset-to-onset delay of 0.75 s, resulting in an average duration of digit presentation of 6.6 s (depending on the duration of the final digit; Fig. 2A). The digit presentation was followed by a 5-s memory retention period during which one randomly selected task-irrelevant distractor sentence was presented. The sentence was presented on average 1.435 s (randomly jittered from 1.035 to 1.835 s) after the offset of the last digit. During the presentation of speech stimuli, participants fixated on a cross in the center of the screen. Speech stimuli were presented through Sennheiser HD-25 headphones at a comfortable level of ~65 dB A.

After the memory retention period, participants saw a number pad with the digits from 1 to 9 on the computer screen (Fig. 2A). In order to prevent participants' motor preparation for a particular behavioral response, the digits in the number pad were randomly arranged on each trial. Participants used the computer mouse to select the digits in their order of presentation. After each selection of an individual digit, the digit disappeared from the number pad. After the selection of all 9 digits

from the number pad, an additional mouse click was required to start the next trial.

Prior to the experiment, participants were instructed to internally rehearse the spoken digits in their order of presentation during the retention period in order to keep memory decay low. Participants were instructed to not close their eyes and to not speak the digits out loudly during a trial. Participants performed approximately 10–20 practice trials to become familiarized with the task.

In the experiment, each participant completed 180 trials, 60 per noise-vocoding condition (1, 4, and 32 ch). The trial order was fully randomized. The entire experiment took approximately 1'10" to complete. The experiment was divided in 4 blocks of 45 trials each. Participants took a short break in-between every 2 blocks.

Behavioral Data Analysis

To quantify participants' accuracy in the serial recall of digits from memory, we considered digits recalled at their respective position of presentation as "correct", and all remaining responses as "incorrect". We calculated the proportion of correctly recalled digits for the 3 noise-vocoding conditions at individual digit positions (Fig. 2B). For statistical analysis, we averaged the proportion correct data across the 9 digit positions for the 3 noise-vocoding conditions. Important for proportion data, the assumption of normality was not violated in any of the 3 conditions (Shapiro–Wilk test; all $P_s > 0.5$). We submitted the proportion data to a repeated-measures ANOVA with noise-vocoding (1, 4, and 32 ch) as a within-subject factor, followed by post hoc dependent-samples t-tests.

For repeated-measures ANOVAs we report Greenhouse-Geisser (GG) epsilon (ϵ) and GG-corrected P -values in case of significant violation of the sphericity assumption (Mauchly's test; $P < 0.05$).

EEG Recording and Preprocessing

The EEG from 64 active scalp electrodes (Ag/Ag-Cl) was recorded (ActiChamp, Brain Products, München, Germany) at a sampling rate of 1000 Hz, with a DC–280 Hz bandwidth, against a left mastoid reference (channel TP9). All electrode impedances were kept below 5 kOhm. To ensure equivalent placement of the EEG cap across participants, the vertex electrode (Cz) was placed at 50% of the distance between inion and nasion and between left and right ear lobes.

Offline, the continuous EEG data were filtered (0.3-Hz high-pass finite impulse response (FIR) filter, zero-phase lag, order 5574; 180-Hz low-pass FIR filter, zero-phase lag, order 100). The continuous data were segmented into epochs relative to the onset of the first digit (−2 to 16 s). An independent component analysis (ICA) was used to detect and reject components corresponding to eye blinks, saccadic eye movements, muscle activity and heartbeats. On average 48% (SD: 8%) of components were rejected. Finally, trials in which an individual EEG channel exceeded a range of 200 microvolts were rejected (on average 17% of trials, SD: 16%). The proportion of rejected trials did not differ across noise-vocoding conditions (repeated-measures ANOVA; $F_{2, 44} = 0.26$; $P = 0.77$; $\eta^2_p = 0.01$). The EEG data were re-referenced to the average of all EEG channels (average reference). For the EEG data analyses we used the Fieldtrip toolbox (version 2012-12-16; Oostenveld et al. 2011) for Matlab (R2013b) and customized Matlab scripts.

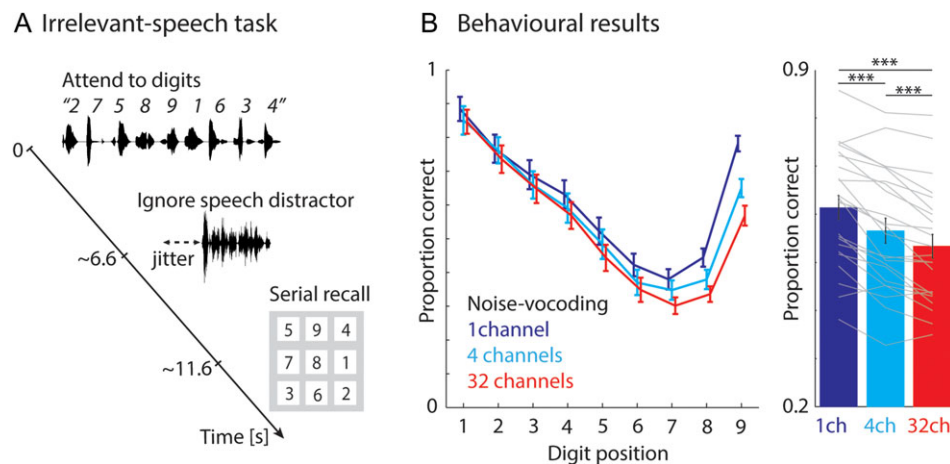


Figure 2. Irrelevant-speech task and behavioral results. (A) Design of the irrelevant-speech task. On each trial, participants attended to the spoken digits from 1 to 9 presented in random order. Their task was to keep the serial order of digits in memory during the presentation of one task-irrelevant distractor sentence, which had to be ignored. In the end of each trial, participants had the task to select the digits in the order of presentation from a visually presented randomly ordered number pad. (B, left) Colored lines show average proportions of correctly recalled digits as a function of the digit position. (B, right) Bars indicate the average proportion correct averaged across all (9) digit positions. Thin gray lines show the average proportion correct for individual participants (which decreased for 22 of 23 participants for high compared to low acoustic detail, 32 ch vs. 1 ch). Errorbars show ± 1 between-subjects standard error of the mean (SEM); *** $P < 0.001$.

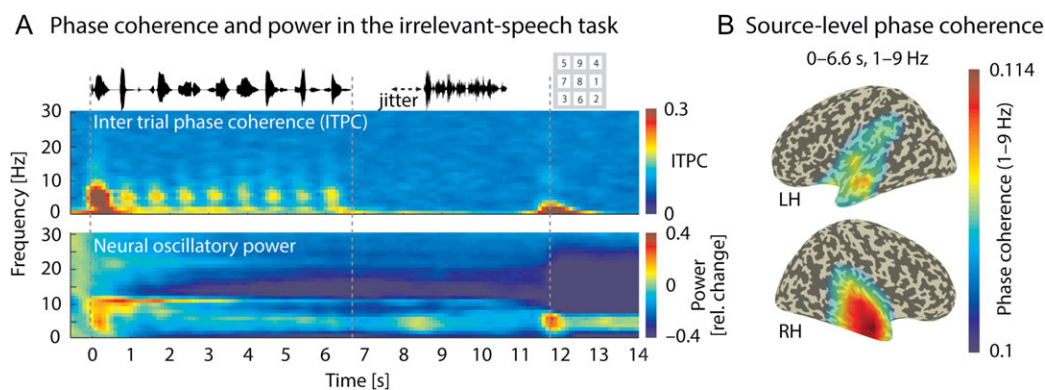


Figure 3. Neural oscillatory activity in the irrelevant-speech task. (A) Inter-trial phase coherence (ITPC, top), and oscillatory power (bottom) averaged across all (64) EEG electrodes in the irrelevant-speech task. Low-frequency ITPC was high at the onsets of spoken digits and at the onset of the visually presented number pad (at -11.6 s). (B) In order to estimate the neural sources of auditory activation, we localized low-frequency ITPC during the encoding of spoken digits (0–6.6 s, 1–9 Hz). Only the highest 25% of ITPC values are shown (i.e., $ITPC > 0.1025$) on the inflated brain surfaces. Low-frequency auditory phase-locking was strongest in bilateral auditory cortex regions in the temporal lobes.

Time–Frequency Analysis of Oscillatory Power

To obtain time–frequency representations of single-trial EEG data, we calculated the complex Fourier coefficients for a moving time window (fixed length of 1 s; hanning taper; moving in steps of 0.1 s through the trial). Fourier coefficients were obtained for frequencies 1–30 Hz with a frequency resolution of 1 Hz.

We performed 2 time–frequency analyses, one for the whole trial duration (digit encoding, retention, recall) and the other for the memory retention period time-locked to the onset of the speech distractor. For the whole trial analysis (Fig. 3) Fourier coefficients were calculated from -2 to 14 s relative to the onset of the first spoken digit. Representations of oscillatory power were obtained by squaring the magnitude of the complex Fourier coefficients. The grand-average power representation was calculated through averaging across $n = 23$ participants. Prior to visualization, the grand-average power representation was baseline-corrected through calculation of the relative change in oscillatory power with respect to a pre-stimulus baseline (-0.5 to 0 s).

In addition to oscillatory power we analyzed inter-trial phase coherence (ITPC) on the sensor and on the source level (Fig. 3). The rationale of this analysis was to demonstrate the fidelity of our EEG source localization, which is known to be limited in comparison to MEG (Leahy et al. 1998). ITPC for the whole trial duration was calculated through division of the complex Fourier coefficients by their magnitudes, followed by averaging across trials and calculation of the magnitude of the resulting complex value.

For the analysis of oscillatory power in the retention period (Fig. 4), Fourier coefficients were calculated from -2 to 7 s relative to the onset of the speech distractor. Single-trial power was obtained (squared magnitude of the complex Fourier coefficients) and normalized by calculating the relative power change with respect to the average power across all trials (in all conditions) in the time window -0.5 to 0 s (relative to the onset of the speech distractor).

Our major hypothesis was that more acoustic detail of the speech distractor would induce an increase in oscillatory power

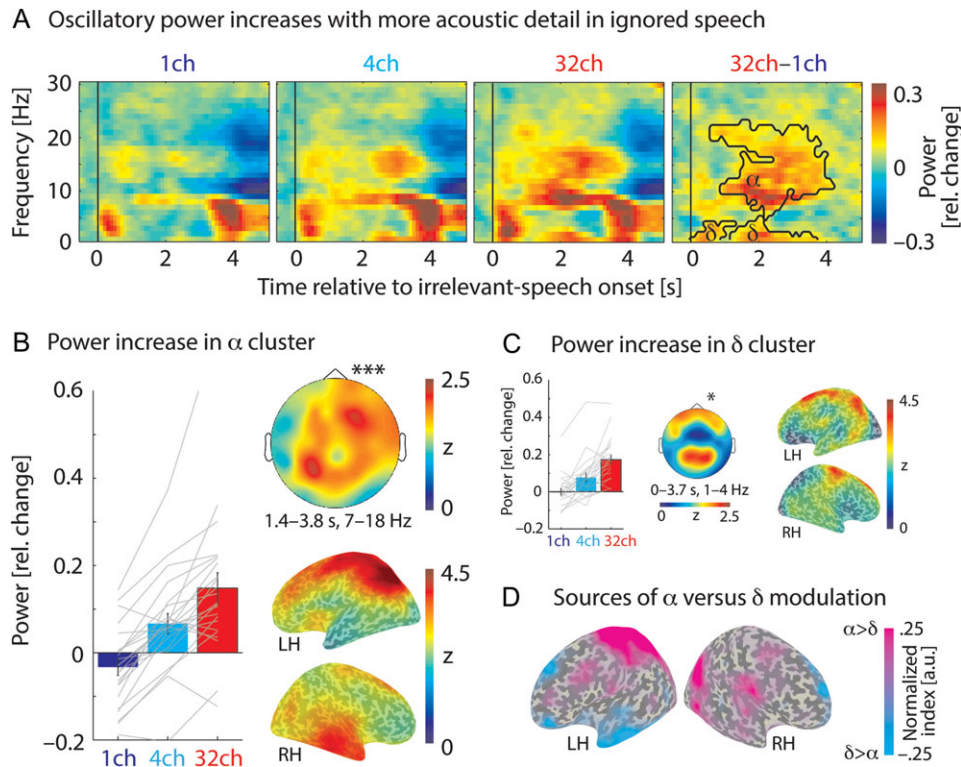


Figure 4. Acoustic detail of distracting speech increases neural oscillatory power. (A) Time–frequency representations of oscillatory power relative to speech distractor onset (0 s) for 3 noise-vocoding conditions (1, 4, and 32 ch) averaged across all (64) EEG electrodes. The rightmost time–frequency representation shows the power difference between the conditions with highest and lowest acoustic detail (32 ch–1 ch). Outlines indicate clusters showing a significant power increase with more acoustic detail of the speech distractor in the delta (and lower theta) frequency range (δ ; earlier cluster $P = 0.049$, later cluster $P = 0.004$) and in the alpha (and lower beta) frequency range (α ; $P < 0.001$). (B) Bars show the average power in the significant alpha cluster. Errorbars show ± 1 between-subjects SEM. Thin gray lines indicate individual participants' power values averaged within the cluster. Topographic maps and source localizations are shown for the strongest power modulation within the cluster in the alpha frequency range (1.4–3.8 s, 7–18 Hz); * $P < 0.05$, *** $P < 0.001$. (C) Same as B for the significant cluster in the delta frequency band. (D) Sources of alpha and delta power modulations with more acoustic detail (32 ch–1 ch) were normalized (i.e., scaled between 0 and 1) and contrasted using an index: $(\text{norm. } \alpha \text{ mod.} - \text{norm. } \delta \text{ mod.}) / (\text{norm. } \alpha \text{ mod.} + \text{norm. } \delta \text{ mod.})$. Alpha power modulation was relatively stronger in parietal and right auditory cortex regions, but delta power modulation was stronger in frontal and left temporal regions.

in the alpha frequency band (~10 Hz). To test this hypothesis, we applied a similar 2-level statistical analysis used in previous studies (Obleser et al. 2012; Wöstmann, Herrmann, et al. 2015). On the single-participant level, we used an independent-samples regression t-test to regress power values of individual trials on linearly increasing, zero-centered predictor values (–1, 0, 1), corresponding to the 3 acoustic conditions (1, 4, or 32 ch noise-vocoding). Linearly (instead of logarithmically) spaced predictor values were used, as logarithmic spacing of stimulus frequencies relates to linear spacing in auditory perception (Attneave and Olson 1971). This resulted in one time–frequency–electrode matrix for each participant, containing t-values for the (linear) increase/decrease in power with respect to acoustic detail. In order to account for slight differences in the number of trials between conditions and participants, t-values were transformed to z-values.

On the group level, the distribution of z-values across frequencies (1–30 Hz) and time (0–5 s relative to the onset of the speech distractor) was tested against zero using a cluster-based permutation one-sample t-test (Maris and Oostenveld 2007). This test clusters t-values of adjacent bins in time–frequency–electrode space (minimum cluster size: 3 adjacent electrodes) and compares the summed t-statistic of the observed cluster against 10 000 randomly drawn clusters from the same data with permuted condition labels. The P-value of a cluster corresponds to the proportion of Monte Carlo

iterations in which the summed t-statistic of the observed cluster is exceeded.

Source Localization of Oscillatory Activity

We performed 2 source localizations. First, in order to obtain the neural sources of auditory activation, we calculated low-frequency (1–9 Hz) ITPC on the source level during the encoding of digits (applying a similar procedure used in Wöstmann, Herrmann, et al. 2016). In detail, a standard headmodel (Boundary Element Method, BEM; 3-shell) was used to calculate leadfields for a grid of 1 cm resolution. We applied the Dynamic Imaging of Coherent Sources (DICS) beamformer approach (Gross et al. 2001) implemented in Fieldtrip. For each participant, we calculated an adaptive spatial filter from the leadfield and the cross-spectral density of Fourier transforms centered at 5 Hz with ± 4 Hz spectral smoothing (resulting in a frequency range of 1–9 Hz) in the time interval 0–6.6 s relative to the onset of the first digit. This filter was applied to single-trial Fourier transforms (1–9 Hz, frequency resolution: 0.15 Hz). ITPC at each grid point and for each frequency was calculated and averaged across frequencies.

Second, we determined neural sources of the significant clusters that exhibited oscillatory power increases with more acoustic detail of the speech distractor. Based on the entire memory retention period duration (–0.5 to 4 s relative to speech

distractor onset) across all trials (in all experimental conditions), we calculated 2 spatially adaptive filters for each participant: first, for all trials in all experimental conditions in the delta band ($2\text{ Hz} \pm 2\text{ Hz}$ spectral smoothing) and second, for the alpha frequency range ($12.5\text{ Hz} \pm 5.5\text{ Hz}$ spectral smoothing). These filters were then used to localize oscillatory power independently for the acoustically most distinct conditions (1 and 32 ch) in the time interval 0–3.7 s for the delta frequency band (1–4 Hz) and 1.4–3.8 s for the alpha frequency band (7–18 Hz). Furthermore, for both frequency ranges, oscillatory power in the –0.5–0 s baseline interval across all experimental conditions was localized and used to calculate the relative change in oscillatory power for each grid point in the same manner as on the sensor level (see above). We calculated the power difference (32 ch–1 ch) for each grid point and participant and tested these difference values against zero (using uncorrected one-sample *t*-tests). Finally, the obtained *t*-values were transformed to *z*-values.

In order to visualize the contrast of the significant alpha and delta power modulations on the source level, we calculated a normalized index (Fig. 4D). In detail, each participant's power increase with more acoustic detail (32 ch–1 ch) on the source level was extracted separately for the significant alpha and delta cluster. We normalized each participant's alpha and delta power modulations separately by scaling them between 0 and 1 across all grid points and calculated the normalized difference between alpha and delta modulations at each grid point as $(\text{norm. alpha mod.} - \text{norm. delta mod.}) / (\text{norm. alpha mod.} + \text{norm. delta mod.})$.

For visualization of source localizations, ITPC, *z*-values, and the normalized index were mapped onto a standard inflated brain surface (corresponding to the SPM anatomical template in MNI coordinates).

Brain-Behavior Correlations

We investigated linear relationships between the modulation of behavior and EEG responses with more acoustic detail of task-irrelevant speech. To this end, we contrasted neural oscillatory power in the significant clusters for the 2 most extreme acoustic conditions (32 ch–1 ch). In order to ensure that a fixed change in the proportion correct weighs relatively more when overall performance is high (i.e., relatively close to ceiling), proportion correct data were first transformed to rationalized arcsine units (rau; Studebaker 1985) and then contrasted between acoustic conditions (32 ch–1 ch). To assess the strength of the relationship between behavioral and neural modulations, we used Pearson correlation if both variables entered in the correlation analysis could be assumed to be normally distributed (Shapiro–Wilk test; $P > 0.05$) and Spearman correlation otherwise.

Effect Sizes

Effect size measures are reported as partial eta-squared (η^2_p) for repeated-measures ANOVAs, and *r*-equivalent (bound between 0 and 1; Rosenthal and Rubin 2003) for dependent-samples *t*-tests. For each significant cluster obtained from cluster-based permutation tests we averaged *r*-values of the multiple *t*-tests for all time–frequency–electrode bins belonging to the cluster into a composite cluster-effect size *R* (Wöstmann, Herrmann, et al. 2015).

Results

More Acoustic Detail of Distracting Speech Decreases Memory Recall

Figure 2B shows participants' task performance, quantified as the proportion of correctly recalled digits in each position. As it is typical for any serial memory task such as the irrelevant-speech task, the proportion of correctly recalled digits was high for digits presented early during encoding (primacy effect) and for the final digit (position 9; recency effect).

Importantly, the average proportion of correctly recalled digits parametrically decreased with more acoustic detail of task-irrelevant speech (average proportion correct, 1 ch: 0.61, 4 ch: 0.57, 32 ch: 0.53; Fig. 2B right), which resulted in a significant main effect of acoustic detail (Greenhouse-Geisser $\epsilon = 0.8$; $F_{2, 44} = 45.9$; $P < 0.001$; $\eta^2_p = 0.68$). Post hoc pairwise comparisons revealed a significantly lower proportion correct for 32 ch versus 4 ch ($t_{22} = 5.02$; $P < 0.001$, $r = 0.73$), for 32 ch versus 1 ch ($t_{22} = 7.87$; $P < 0.001$, $r = 0.86$), and for 4 ch versus 1 ch ($t_{22} = 5.84$; $P < 0.001$, $r = 0.78$). These results show that the speech distractor disrupts memory stronger if it is acoustically more intact (in agreement with Wöstmann and Obleser 2016).

Oscillatory Activity During the Irrelevant-Speech Task

In order to understand modulations of oscillatory activity in a particular time interval (here: during the processing of distracting speech) it is necessary to first inspect oscillatory activity over the entire duration of a trial. Figure 3A illustrates average oscillatory activity throughout the whole trial period (across all experimental conditions and all participants). Low-frequency neural oscillations were strongly phase-locked (i.e., ITPC was high) at the onsets of spoken digits (Fig. 3A, top). As expected for auditory activation, 1–9 Hz ITPC was localized to bilateral auditory cortex regions in the temporal lobes (Fig. 3B), which demonstrates the fidelity of our EEG source localization.

Relative to a pre-trial baseline (–0.5 to 0 s), oscillatory power in the alpha frequency band (~10 Hz) was enhanced initially during the encoding of spoken digits and decreased later during the trial (Fig. 3A, bottom). Power in the beta band (~15–20 Hz) was decreased throughout the trial duration; particularly after the onset of the visual number pad when participants made their behavioral response. Due to the temporal onset jitter of the speech distractor, no solid distractor-evoked activity is seen in ITPC or power. For an analysis of the event-related potential (ERP) in our irrelevant-speech task, see the Supplementary Information.

Oscillatory Power Increases with Acoustically Richer Distracting Speech

A cluster-based permutation test revealed 3 time–frequency–electrode clusters showing significant increases in oscillatory power with more acoustic detail of the speech distractor (Fig. 4A). Most importantly and in line with our hypothesis, one highly significant cluster exhibited a power increase in the alpha (α) and also in the lower beta frequency range with more acoustic detail of the speech distractor (time window: 0.6–4.2 s; frequencies: 5–23 Hz, 60 of 64 electrodes; $P < 0.001$; $R = 0.54$; Fig. 4B). The alpha/beta power increase in this cluster was localized to parietal cortical regions (in the left hemisphere) and also to auditory cortical regions (in the right hemisphere).

Furthermore, 2 clusters were observed in the delta (δ) and lower theta frequency range; an earlier cluster (time

window: 0–1.3 s; frequencies: 1–4 Hz, 28 of 64 electrodes; $P = 0.049$; $R = 0.52$) and a later cluster (time window: 1.2–3.7 s; frequencies: 1–5 Hz, 60 of 64 electrodes; $P = 0.004$; $R = 0.56$). Since these 2 clusters were in close time–frequency–electrode proximity, they were collapsed into one combined cluster (i.e., the union of the 2 clusters) for further analyses. Source localization revealed that delta/theta power in left parietal and bilateral fronto-central cortical regions increased with more acoustic detail of the speech distractor (Fig. 4C). Besides oscillatory power, also the posterior negativity in the event-related potential (ERP) was modulated by more acoustic detail in distracting speech (see Supplementary Information).

Due to the dominance of the alpha/beta cluster by frequencies in the alpha band (7–13 Hz), and of the delta/theta cluster by frequencies in the delta band (1–3 Hz) these clusters will be referred to as “alpha cluster” and “delta cluster”, respectively.

Alpha Versus Delta Power Modulation and Relation to Behavior

Figure 4D contrasts the anatomical origins of alpha versus delta power increases with more acoustic detail in the speech distractor. Alpha power modulation was clearly dominated by left parietal cortex and also auditory regions in the right hemisphere (pink). Delta power, on the contrary, was modulated relatively stronger in frontal and left temporal cortex regions (light blue). Interestingly, despite their distinct spectral and anatomical origins, these alpha and delta power modulations do share some variance (correlation of alpha power change with more acoustic detail, 32 ch–1 ch, versus respective delta power change; $r_{\text{Spearman}} = 0.408$; $P = 0.054$). This indicates a tendency for listeners who increase their alpha power stronger in response to more acoustically intact distracting speech also to show a stronger delta power increase.

In previous studies, varying acoustic detail of target speech had been shown to modulate neural alpha power but also behavioral performance; moreover, these neural and behavioral modulations were significantly correlated (Obleser and Weisz 2012; Wöstmann, Herrmann, et al. 2015). Here, we asked whether a similar correlation would be observed for an acoustically degraded speech distractor, which participants had to ignore. We thus tested whether the behavioral performance decrease with more acoustic detail would correlate with the delta and alpha power increases with more acoustic detail. Modulations of behavioral performance and neural oscillatory power were quantified as the differences (i.e., 32 ch–1 ch) in the respective measures (the *rau*-transformed proportion correct, and oscillatory power in the significant clusters). The correlation was neither significant for the delta cluster ($r_{\text{Spearman}} = -0.122$; $P = 0.579$) nor for the alpha cluster ($r_{\text{Spearman}} = -0.25$; $P = 0.247$).

Significant brain-behavior correlation might well be present in the data, but in time–frequency–electrode regions different from those clusters showing a significant power increase with more acoustic detail (i.e., different from our significant clusters in Fig. 4). To test this, we conducted a potentially more robust within-subject analysis to relate single-trial performance to single-trial oscillatory power. To this end, each participant’s single-trial time–frequency representations of power in the retention period were regressed on single-trial performance scores (ranging between 0 and 9 correctly recalled digits; *z*-transformed), separately for the 3 acoustic detail conditions. On the group level, we tested these single-subject statistics against zero using 3 dependent-samples cluster-based permutation tests,

one for each acoustic detail level. However, no significant clusters (cluster closest to statistical significance: $P = 0.053$) were found, and no simple, direct relation of oscillatory power during distraction to resulting behavior was observable.

Discussion

We asked in how far neural alpha and other oscillatory dynamics during auditory attention are being controlled by a listener’s top-down goals. The present study examined the nature of a well-established human electrophysiological response to acoustically degraded sound—the increase of ~10 Hz alpha oscillatory power. We predicted that alpha power increases should hinge entirely on increases in the degree of task detriment—that is, distraction—and not acoustic degradation per se. In fact, alpha power (and also low-frequency delta/theta power) increased, rather than decreased, with more acoustic detail in a speech distractor. This complete pattern reversal of parametric alpha power change demonstrates that the neural alpha response to degraded speech depends largely on a listener’s goal to attend versus to ignore speech.

More Acoustic Detail of Task-Irrelevant Speech Increases Distraction

Listeners’ serial recall of digits from memory was impaired when distracting speech with more acoustic detail was presented during memory retention (Fig. 2B). This agrees with 2 previous studies, which used noise-vocoded speech distractors in the irrelevant-speech task as well (Ellermeier et al. 2015; Wöstmann and Obleser 2016). The distraction likely originates on the level of a listener’s attentive processing of target speech in working memory.

During target speech retention, participants actively rehearse the serial order of digits in working memory (Baddeley 1992). This requires that attention be directed to digits in memory and away from the speech distractor. When the speech distractor is acoustically more intact it might draw attention away from items in memory, which eventually leads to a partial occupation of the limited working memory capacity by the distractor. Furthermore, our speech materials are of higher intelligibility when more acoustic detail is preserved in noise-vocoding (Hartwigsen et al. 2015), which might increase interference by the semantic content of the speech distractor (i.e., informational masking; e.g., Schneider et al. 2007).

It has been argued that more acoustic detail of a sound distractor impedes the serial rehearsal in working memory, irrespective of any attentional capture from the distractor (Jones and Morris 1992). Evidence for this distractor-driven memory demand comes also from a positron emission tomography (PET) study, which has shown that memory load and distracting speech both modulate neural activity in working memory areas in dorsolateral prefrontal cortex (Gisselgard et al. 2004). As we will argue below, our EEG results indicate that distracting speech with more acoustic detail increases cognitive demands on attention and working memory.

Focus of Attention Governs the Neural Alpha Response to Degraded Speech

The most important finding of this study was that more distracting speech of higher acoustic detail induced an increase in neural alpha power (Fig. 4A, B). This result is the exact opposite of previous electrophysiological studies, which found an alpha

power decrease with more acoustic detail, however, of a target speech signal (Obleser and Weisz 2012; Obleser et al. 2012; Becker et al. 2013; Scharinger et al. 2014; Wöstmann, Herrmann, et al. 2015; McMahan et al. 2016). Thus, as expected, the modulation of neural alpha power is not fully determined by the mere stimulus acoustics. Rather, increasing alpha power indicates high demands on attention, which can be triggered by less acoustic detail of target speech as well as by more acoustic detail of distractor speech (Fig. 1).

The power increase with more acoustic detail was not restricted to the alpha frequency band but spread to the lower beta band up to 22 Hz (Fig. 4A). Alpha and beta power desynchronize in response to speech (Shahin et al. 2009; Obleser and Weisz 2012), which might reflect speech encoding. Furthermore, beta power correlates with motor activity preceding a behavioral response (Pfurtscheller et al. 2003; Tzagarakis et al. 2010). However, our participants were naïve to the spatial arrangement of digits in the number pad, which was randomly determined on each trial. Hence, the preparation for a mouse movement to a particular digit in the number pad was impossible before the number pad was presented. We consider it thus likely that the alpha power modulation leaked into the beta band and that our alpha and beta modulations reflect the same cognitive process.

The oscillatory alpha and beta power increase with more acoustic detail was localized in parietal cortex regions in the left hemisphere and in auditory cortex regions in the right temporal lobe (Fig. 4B,D). Parietal cortex is a common source of high alpha power during attention, likely indicating inhibitory control processing in a supramodal parietal attention network (Banerjee et al. 2011). Higher parietal alpha power in our study suggests increasing effort to direct the focus of attention to target speech in memory when distraction by acoustically intact task-irrelevant speech is high.

In a recent MEG study we found modulations of alpha power not just in parietal but also in auditory cortex regions, which indicated participants' ignoring versus attending of one of two concurrent speech streams (Wöstmann, Herrmann, et al. 2016). The present alpha power increase with more acoustic detail of the speech distractor was particularly strong in right-hemispheric auditory cortex regions, which agrees with previous findings of stronger right- versus left-hemispheric auditory alpha modulations in spatial attention tasks (Müller and Weisz 2012; Weisz et al. 2014). Although our task did not require spatial attention, the alpha modulation in right auditory cortex conforms with the stronger involvement of right versus left auditory cortex in the attentional coding of auditory space (Zatorre and Penhune 2001). Functionally, higher auditory alpha power indicates stronger suppression of neural activity in auditory regions (de Pestors et al. 2016) which is thought to facilitate an attentional "filtering out" of distracting, task-irrelevant input (Strauß et al. 2014; Wöstmann, Herrmann, et al. 2016).

In a related study in the visual modality, Bonnefond and Jensen (2012) observed an increase in visual alpha power in occipital cortex regions prior to the onset of a strong versus weak visual distractor. They concluded that high alpha power protects working memory content against visual distraction. Applying the same reasoning to the present study, high auditory alpha power would inhibit neural processing of the auditory distractor in order to protect items in working memory.

It is of note, however, that the present alpha effect, as in previous studies (Banerjee et al. 2011; Obleser and Weisz 2012; Obleser et al. 2012; Wöstmann, Herrmann, et al. 2016) had

contributions not only from auditory cortical, but more dominantly from parietal areas. Parietal cortex is part of the dorsal attention network (DAT; Sadaghiani et al. 2010), which is thought to reduce the manifestation of alpha oscillations in task-related sensory cortex regions (Sadaghiani and Kleinschmidt 2016) when allocating selective attention. In the present study, however, participants had the task to ignore the task-irrelevant distractor, that is, to inhibit attentional allocation. Relative inhibition of the DAT through high alpha power might thus be a means to neurally facilitate ignoring.

It is also noteworthy that we did not find a linear relationship of participants' behavior and their modulation of oscillatory power during distraction in the present study. The direction of the (non-significant) brain-behavior correlations in the present data is in accordance with prior studies (i.e., stronger modulation of behavior with acoustic detail coincides with stronger modulation of oscillatory power). However, there are crucial conceptual differences between our study and prior studies that have reported such a brain-behavior correlation.

First, previous studies often did not use task accuracy to establish a brain-behavior correlation, but rather other behavioral measures that reflect a participant's perceptual certainty, such as confidence ratings (Wöstmann, Herrmann, et al. 2015), speech comprehension ratings (Obleser and Weisz 2012), or response times (Bonnefond and Jensen 2012). Thus, the number of correctly recalled digits might not be sensitive enough to establish a reliable brain-behavior correlation.

Second and more interestingly, our experimental manipulation of acoustic detail deviates from most speech-in-noise studies in that it was applied solely to the to-be-ignored speech distractor (instead of to-be-attended target speech). It is conceivable that in the present paradigm, where behavior depends on the recall of attended digits but the neural response on acoustic detail of ignored speech, the brain-behavior relationship is thus considerably more indirect compared to most paradigms where both task performance and the neural response directly relate to an attended stimulus.

Distracting Speech Modulates Low-Frequency Delta and Theta Power

In addition to alpha power, listeners' delta and lower theta power (1–5 Hz) also increased parametrically with more acoustic detail of distracting speech (Fig. 4C). This finding is in line with Obleser and Weisz (2012), who found a left frontal theta power (4–7 Hz) increase with more acoustic detail of individual target words. Thus, in contrast to the neural alpha response, which entirely reverses for more acoustic detail of target versus distractor speech (see above), low-frequency power shows a similar parametric increase with more acoustic detail of target speech (Fig. 2B in Obleser and Weisz 2012) and distractor speech (current study; Fig. 4C).

During target speech processing, low-frequency oscillations phase-lock to the acoustic input, a phenomenon called "speech tracking" (for an overview, see Wöstmann, Fiedler, et al. 2016). Importantly, the neural tracking of target speech against background noise relatively increases when more acoustic detail is preserved in noise-vocoding (Ding et al. 2013; Kong et al. 2015) and when acoustically degraded speech is intelligible (as compared to unintelligible spectrally rotated speech; Peelle et al. 2013). It has recently been shown that top-down delta/theta signals, particularly in the left hemisphere, modulate the coupling of auditory cortical responses to target speech (Park et al. 2015). In our irrelevant-speech task it should be performance-beneficial

to reduce the auditory cortical coupling to distractor speech. Frontal and left temporal delta/theta power modulations might thus indicate a listener's need to limit auditory cortical coupling to distracting speech of higher acoustic detail.

Alternatively, it has been shown that frontal theta power correlates positively with increasing load in working memory (i.e., with a higher number of to-be-retained memory items; e.g., [Jensen and Tesche 2002](#)). In agreement with this view, previous EEG studies found evidence for theta power modulations during serial memory processing in the irrelevant-speech task ([Weisz and Schlittmeier 2006](#); [Schlittmeier et al. 2011](#)). Recall, however, that the modulations in delta/theta shared only about 20% of their variance with the modulations in alpha/beta ($r = 0.408$; $P = 0.054$), which points to partly distinct underlying neural processes.

Of direct relevance to such a potential distinction of alpha/beta and delta/theta mechanisms, [Hsieh et al. \(2011\)](#) demonstrated a dissociation between alpha and theta power modulations in a visual working memory task: While parieto-occipital alpha power increased during the maintenance of items, frontal theta power increased when participants retained their temporal order. If we transfer these results to our irrelevant-speech task, theta power might reflect the serial rehearsal of digits in memory, whereas alpha power might indicate a listener's attentional focus on working memory items to reduce distractor interference (for a similar dissociation of theta and alpha power, see [Roux and Uhlhaas 2014](#)). Higher delta/theta and alpha power might thus reflect higher demand on "memory of" and "attention to" target speech items, respectively.

Conclusions

It is a well-established finding that neural oscillatory ~10-Hz alpha power decreases with more acoustic detail when listeners aim to comprehend speech. We demonstrate here that a listener's goals are capable of entirely reversing this alpha response: Parietal and auditory alpha power parametrically increase instead when more acoustic detail renders task-irrelevant speech more effortful to ignore. We conclude that the neural alpha response is not governed by speech acoustics per se but by the acoustics-driven demand on a listener's attention and memory systems.

Supplementary Material

Supplementary data is available at *Cerebral Cortex* online.

Funding

This research was supported by an ERC grant (ERC-CoG-2014 AUDADAPT) to Jonas Obleser, as well as funds by the Volkswagen foundation (BIT-CHAT).

Notes

We thank Lea-Maria Schmitt and Franziska Scharata for help with EEG data acquisition. *Conflict of Interest*: None declared.

References

Attneave F, Olson RK. 1971. Pitch as a medium: a new approach to psychophysical scaling. *Am J Psychol.* 84(2):147–166.

Awh E, Vogel EK, Oh SH. 2006. Interactions between attention and working memory. *Neuroscience.* 139(1):201–208. doi:10.1016/j.neuroscience.2005.08.023.

Baddeley A. 1992. Working memory. *Science.* 255(5044):556–559.

Banerjee S, Snyder AC, Molholm S, Foxe JJ. 2011. Oscillatory alpha-band mechanisms and the deployment of spatial attention to anticipated auditory and visual target locations: supra-modal or sensory-specific control mechanisms? *J Neurosci.* 31(27):9923–9932. doi:10.1523/JNEUROSCI.4660-10.2011.

Becker R, Pefkou M, Michel CM, Hervais-Adelman AG. 2013. Left temporal alpha-band activity reflects single word intelligibility. *Front Syst Neurosci.* 7:121. doi:10.3389/fnsys.2013.00121.

Bonnefond M, Jensen O. 2012. Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Curr Biol.* 22(20):1969–1974. doi:10.1016/j.cub.2012.08.029.

Colle HA, Welsh A. 1976. Acoustic masking in primary memory. *J Verbal Learning Verbal Behav.* 15(1):17–31.

de Pestras A, Coon WG, Brunner P, Gunduz A, Ritaccio AL, Brunet NM, de Weerd P, Roberts MJ, Oostenveld R, Fries P, et al. 2016. Alpha power indexes task-related networks on large and small scales: a multimodal ECoG study in humans and a non-human primate. *Neuroimage.* 134:122–131. doi:10.1016/j.neuroimage.2016.03.074.

Ding N, Chatterjee M, Simon JZ. 2013. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage.* 88 C:41–46. doi:10.1016/j.neuroimage.2013.10.054.

Ellermeier W, Hellbrück J. 1998. Is level irrelevant in "irrelevant speech"? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *J Exp Psychol Hum Percept Perform.* 24(5):1406–1414.

Ellermeier W, Kattner F, Ueda K, Doumoto K, Nakajima Y. 2015. Memory disruption by irrelevant noise-vocoded speech: effects of native language and the number of frequency bands. *J Acoust Soc Am.* 138(3):1561–1569. doi:10.1121/1.4928954.

Erb J, Henry MJ, Eisner F, Obleser J. 2012. Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychologia.* 50(9):2154–2164. doi:10.1016/j.neuropsychologia.2012.05.013.

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(5):431–438.

Foxe JJ, Snyder AC. 2011. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front Psychol.* 2:154. doi:10.3389/fpsyg.2011.00154.

Fritz JB, Elhilali M, David SV, Shamma SA. 2007. Auditory attention—focusing the searchlight on sound. *Curr Opin Neurobiol.* 17(4):437–455. doi:10.1016/j.conb.2007.07.011.

Gisselgard J, Petersson KM, Ingvar M. 2004. The irrelevant speech effect and working memory load. *Neuroimage.* 22(3):1107–1116. doi:10.1016/j.neuroimage.2004.02.031.

Greenwood DD. 1990. A cochlear frequency-position function for several species—29 years later. *J Acoust Soc Am.* 87(6):2592–2605.

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 USA.* 98(2):694–699. doi:10.1073/pnas.98.2.694.

Haegens S, Handel BF, Jensen O. 2011. Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. *J Neurosci.* 31(14):5197–5204. doi:10.1523/JNEUROSCI.5199-10.2011.

Hartwigsen G, Golombek T, Obleser J. 2015. Repetitive transcranial magnetic stimulation over left angular gyrus modulates the predictability gain in degraded speech comprehension. *Cortex.* doi:10.1016/j.cortex.2014.08.027.

- Hsieh LT, Ekstrom AD, Ranganath C. 2011. Neural oscillations associated with item and temporal order maintenance in working memory. *J Neurosci.* 31(30):10803–10810. doi:10.1523/JNEUROSCI.0828-11.2011.
- Jensen O, Mazaheri A. 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front Hum Neurosci.* 4:186. doi:10.3389/fnhum.2010.00186.
- Jensen O, Tesche CD. 2002. Frontal theta activity in humans increases with memory load in a working memory task. *Eur J Neurosci.* 15(8):1395–1399.
- Jones DM, Morris N. 1992. Irrelevant speech and serial recall: implications for theories of attention and working memory. *Scand J Psychol.* 33(3):212–229.
- Kalikow DN, Stevens KN, Elliott LL. 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am.* 61(5):1337–1351.
- Kong YY, Somarowthu A, Ding N. 2015. Effects of spectral degradation on attentional modulation of cortical auditory responses to continuous speech. *J Assoc Res Otolaryngol.* 16(6):783–796. doi:10.1007/s10162-015-0540-x.
- Leahy RM, Mosher JC, Spencer ME, Huang MX, Lewine JD. 1998. A study of dipole localization accuracy for MEG and EEG using a human skull phantom. *Electroencephalogr Clin Neurophysiol.* 107(2):159–173.
- Lim SJ, Wöstmann M, Obleser J. 2015. Selective attention to auditory memory neurally enhances perceptual precision. *J Neurosci.* 35(49):16094–16104. doi:10.1523/JNEUROSCI.2674-15.2015.
- Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods.* 164(1):177–190. doi:10.1016/j.jneumeth.2007.03.024.
- McMahon CM, Boisvert I, de Lissa P, Granger L, Ibrahim R, Lo CY, Miles K, Graham PL. 2016. Monitoring alpha oscillations and pupil dilation across a performance-intensity function. *Front Psychol.* 7:745. doi:10.3389/fpsyg.2016.00745.
- Müller N, Weisz N. 2012. Lateralized auditory cortical alpha band activity and interregional connectivity pattern reflect anticipation of target sounds. *Cereb Cortex.* 22(7):1604–1613. doi:10.1093/cercor/bhr232.
- Oberauer K, Hein L. 2012. Attention to information in working memory. *Curr Dir Psychol Sci.* 21(3):164–169. doi:10.1177/0963721412444727.
- Obleser J, Weisz N. 2012. Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cereb Cortex.* 22(11):2466–2477. doi:10.1093/cercor/bhr325.
- Obleser J, Wöstmann M, Hellbernd N, Wilsch A, Maess B. 2012. Adverse listening conditions and memory load drive a common alpha oscillatory network. *J Neurosci.* 32(36):12376–12383. doi:10.1523/JNEUROSCI.4908-11.2012.
- 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. doi:10.1155/2011/156869.
- Park H, Ince RA, Schyns PG, Thut G, Gross J. 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr Biol.* 25(12):1649–1653. doi:10.1016/j.cub.2015.04.049.
- Peelle JE, Gross J, Davis MH. 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb Cortex.* 23(6):1378–1387. doi:10.1093/cercor/bhs118.
- Pfurtscheller G, Graimann B, Huggins JE, Levine SP, Schuh LA. 2003. Spatiotemporal patterns of beta desynchronization and gamma synchronization in corticographic data during self-paced movement. *Clin Neurophysiol.* 114(7):1226–1236.
- Rosen S, Faulkner A, Wilkinson L. 1999. Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *J Acoust Soc Am.* 106(6):3629–3636.
- Rosenthal R, Rubin DB. 2003. *r* equivalent: a simple effect size indicator. *Psychol Methods.* 8(4):492–496. doi:10.1037/1082-989X.8.4.492.
- Roux F, Uhlhaas PJ. 2014. Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends Cogn Sci.* 18(1):16–25. doi:10.1016/j.tics.2013.10.010.
- Sadaghiani S, Kleinschmidt A. 2016. Brain networks and alpha-oscillations: structural and functional foundations of cognitive control. *Trends Cogn Sci.* 20(11):805–817. doi:10.1016/j.tics.2016.09.004.
- 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(30):10243–10250. doi:10.1523/JNEUROSCI.1004-10.2010.
- Scharinger M, Herrmann B, Nierhaus T, Obleser J. 2014. Simultaneous EEG-fMRI brain signatures of auditory cue utilization. *Front Neurosci.* 8:137. doi:10.3389/fnins.2014.00137.
- Schlittmeier SJ, Weisz N, Bertrand O. 2011. What characterizes changing-state speech in affecting short-term memory? An EEG study on the irrelevant sound effect. *Psychophysiology.* 48(12):1669–1680. doi:10.1111/j.1469-8986.2011.01263.x.
- Schneider BA, Li L, Daneman M. 2007. How competing speech interferes with speech comprehension in everyday listening situations. *J Am Acad Audiol.* 18(7):559–572.
- Serences JT, Kastner S. 2014. A multi-level account of selective attention. In: Nobre AC, Kastner S, editors. *The Oxford handbook of attention.* New York: Oxford University Press.
- Shahin AJ, Picton TW, Miller LM. 2009. Brain oscillations during semantic evaluation of speech. *Brain Cogn.* 70(3):259–266. doi:10.1016/j.bandc.2009.02.008.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. 1995. Speech recognition with primarily temporal cues. *Science.* 270(5234):303–304.
- Shinn-Cunningham BG. 2008. Object-based auditory and visual attention. *Trends Cogn Sci.* 12(5):182–186. doi:10.1016/j.tics.2008.02.003.
- Snyder AC, Foxe JJ. 2010. Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. *J Neurosci.* 30(11):4024–4032. doi:10.1523/JNEUROSCI.5684-09.2010.
- Strauß A, Wöstmann M, Obleser J. 2014. Cortical alpha oscillations as a tool for auditory selective inhibition. *Front Hum Neurosci.* 8:350. doi:10.3389/fnhum.2014.00350.
- Studebaker GA. 1985. A “rationalized” arcsine transform. *J Speech Hear Res.* 28(3):455–462.
- Tzagarakis C, Ince NF, Leuthold AC, Pellizzer G. 2010. Beta-band activity during motor planning reflects response uncertainty. *J Neurosci.* 30(34):11270–11277. doi:10.1523/JNEUROSCI.6026-09.2010.
- Weisz N, Müller N, Jatzev S, Bertrand O. 2014. Oscillatory alpha modulations in right auditory regions reflect the validity of acoustic cues in an auditory spatial attention task. *Cereb Cortex.* 24(10):2579–2590. doi:10.1093/cercor/bht113.
- Weisz N, Schlittmeier SJ. 2006. Detrimental effects of irrelevant speech on serial recall of visual items are reflected in

- reduced visual N1 and reduced theta activity. *Cereb Cortex*. 16(8):1097–1105. doi:10.1093/cercor/bhj051.
- Wilsch A, Obleser J. 2016. What works in auditory working memory? A neural oscillations perspective. *Brain Res*. 1640 (Pt B):193–207. doi:10.1016/j.brainres.2015.10.054.
- Wöstmann M, Fiedler L, Obleser J. 2016. Tracking the signal, cracking the code: Speech and speech comprehension in non-invasive human electrophysiology. *Lang Cogn Neurosci*. doi:10.1080/23273798.2016.1262051, forthcoming.
- Wöstmann M, Herrmann B, Maess B, Obleser J. 2016. Spatiotemporal dynamics of auditory attention synchronize with speech. *Proc Natl Acad Sci USA*. doi:10.1073/pnas.1523357113.
- Wöstmann M, Herrmann B, Wilsch A, Obleser J. 2015. Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *J Neurosci*. 35(4): 1458–1467. doi:10.1523/JNEUROSCI.3250-14.2015.
- Wöstmann M, Obleser J. 2016. Acoustic detail but not predictability of task-irrelevant speech disrupts working memory. *Front Hum Neurosci*. 10(538). 10.3389/fnhum.2016.00538.
- Wöstmann M, Schröger E, Obleser J. 2015. Acoustic detail guides attention allocation in a selective listening task. *J Cogn Neurosci*. 27(5):988–1000. doi:10.1162/jocn_a_00761.
- Zatorre RJ, Penhune VB. 2001. Spatial localization after excision of human auditory cortex. *J Neurosci*. 21(16):6321–6328.