Induced Alpha And Beta Electroencephalographic Rhythms Covary With Single-Trial Speech Intelligibility In Competition

Vibha Viswanathan [1,*], Hari M. Bharadwaj [2], Michael G. Heinz [3], Barbara G. Shinn-Cunningham [4]

[1] Department of Neurological Surgery, University of Pittsburgh, Pitttsburgh, PA 15260. [2] Department of Communication Science and Disorders, University of Pittsburgh, Pitttsburgh, PA 15260. [3] Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN 47907. [4] Neuroscience Institute, Carnegie Mellon University, Pitttsburgh, PA 15213.

*Correspondence: viv46@pitt.edu

Acknowledgments

This study was supported by grants from the National Institutes of Health [Grant Nos. F31DC017381 (to V.V.), T32DC11499 (to V.V.), F32DC020649 (to V.V.), R01DC009838 (to M.G.H.), R01DC015989 (to H.M.B.), R01DC013825 (to B.G.S.-C.), and R01DC019126 (to B.G.S.-C.)] and from Action on Hearing Loss [Grant No. G72 (to M.G.H.)].

1. Abstract

Neurophysiological studies suggest that intrinsic brain oscillations influence sensory processing, especially of rhythmic stimuli like speech. Prior work suggests that brain rhythms may mediate perceptual grouping and selective attention to speech amidst competing sound, as well as more linguistic aspects of speech processing like predictive coding. However, we know of no prior studies that have directly tested, at the single-trial level, whether brain oscillations relate to speech-in-noise outcomes. Here, we combined electroencephalography while simultaneously measuring intelligibility of spoken sentences amidst two different interfering sounds: multi-talker babble or speech-shaped noise. We find that induced parieto-occipital alpha (7–15 Hz; thought to modulate attentional focus) and frontal beta (13–30 Hz; associated with speech-motor predictive coding) oscillations covary with trial-wise percent-correct scores; importantly, alpha and beta power provide significant independent contributions to predicting single-trial behavioral outcomes. Moreover, we observed large individual differences in the across-trial distribution of alpha and beta power as well as in the alpha-to-beta power ratio. This raises the possibility that listeners employed different task strategies: some may have relied heavily on focusing attention, while others on predictive coding. These results can inform models of speech processing and guide noninvasive measures to index different neural processes that together support complex listening.

2. Introduction

Understanding speech in the presence of interfering sounds—e.g., competing talkers or other sources of background noise—is a difficult perceptual task that our brains solve everyday (Cherry, 1953). However, the neural mechanisms facilitating “cocktail-party” listening remain poorly understood. Neurophysiological studies in humans [using electro- (EEG) and magneto-encephalography (MEG) as well as invasive intracranial recordings] and other animal species suggest that brain rhythms (Buzsáki and Draguhn, 2004) in different frequency bands may mediate sensory processing (Schroeder and Lakatos, 2009). This, in turn, may facilitate speech understanding in competition. For instance, in a mixture of competing sources, brain oscillations in the low-frequency delta (1–3 Hz) and theta (3–7 Hz) bands preferentially phase-lock to the slow temporal fluctuations (i.e., envelopes) in attended speech (Ding and Simon, 2012; O’sullivan et al., 2015), while power fluctuations in the low-gamma (30–70 Hz; Viswanathan et al., 2019) and high-gamma (70–120 Hz; Mesgarani and Chang, 2012; Golumbic et al., 2013) bands selectively synchronize to attended versus ignored speech envelopes. This target-speech envelope phase-locking in the brain may aid listeners in selectively processing a target speech source in an acoustic mixture, thereby influencing speech intelligibility across different everyday listening conditions (Viswanathan et al., 2021).

In addition to phase-locked (or evoked) neuronal oscillations, induced brain rhythms have also been implicated in cocktail-party listening. For instance, focused auditory attention leads to increase in the power of the alpha (7–15 Hz) rhythm in parieto-occipital areas, which is specifically thought to be a hallmark of neuronal mechanisms related to suppression of sensory distractors (Foxe and Snyder, 2011; Strauß et al., 2014) like visual input (Adrian, 1944). During auditory spatial selective attention, parieto-occipital alpha power becomes lateralized: alpha power increases contralateral to the hemifield of distracting sounds (i.e., ipsilateral to the hemifield of focus; Banerjee et al., 2011; Deng et al., 2019; Deng et al., 2020). This alpha lateralization has been reported to predict individual differences in spoken-digit identification when listeners hear a mixture of spatially separated sources (Love et al., 2020). Moreover, even for tasks that involve spatially co-localized speech and distractor sources, prior studies report positive correlation between the overall magnitude (versus lateralization) of alpha power in centro-parietal EEG channels and speech comprehension across signal-to-noise ratios (SNRs; for spoken sentences; Hall et al., 2019) and across individuals (for spoken digits; Alhanbali et al., 2022).

In contrast to induced alpha, which has been implicated in auditory attention, the beta (13–30 Hz) rhythm is thought to support maintenance of the current sensorimotor state (Engel and Fries, 2010) and speech-motor predictive coding (Arnal and Giraud, 2012; Lewis and Bastiaansen, 2015). More generally, motor-theory accounts of speech recognition posit that sensorimotor integration between fronto-motor areas controlling articulation (e.g., inferior frontal gyrus and premotor cortex) and temporal-parietal cortical areas implicated in phonetic category representation mediates top-down sensory prediction to modulate and stabilize speech representation (Hickok et al., 2011; Skipper et al., 2017; Pulvermüller, 2018), especially in adverse listening conditions such as in background noise (Adank et al., 2012; Du et al., 2014). In line with this notion, frontal beta power correlates with sensory prediction precision in vocoded word identification (Cope et al., 2017), with auditory cortical entrainment to continuous speech

(Park et al., 2015; Keitel et al., 2017), and with comprehension for time-compressed speech sentences (Pefkou et al., 2017). Moreover, across individuals, beta-band synchrony between premotor and temporal-parietal cortical regions correlates positively with syllable identification in noise (Alho et al., 2014).

Despite the prior literature linking alpha and beta rhythms to speech processing, we know of no prior studies that tested whether trial-to-trial variations in the overall magnitude of induced parieto-occipital alpha power and frontal beta power relate to trial-wise speech intelligibility when competing sounds are present. The present study explored this question using human EEG and simultaneous speech intelligibility measurements of spoken sentences under masking. Because different competing sounds could drive different degrees of demand on selective attention versus contextual prediction (e.g., attentional demand—and hence alpha power—may be greater when the masker is a competing speech stream or multi-talker babble versus stationary noise; Shinn-Cunningham, 2008), we used two different maskers in this study: multi-talker babble and speech-shaped stationary noise. We examined the extent to which the overall magnitude of induced oscillatory power in different frequency bands relates to speech intelligibility in each masking condition on a trial-by-trial basis. Building on our previous work, where we related phase-locked neural responses to speech understanding in different listening conditions in the same dataset (Viswanathan et al., 2021), here we focused on induced brain activity. Specifically, we examined frequency bands in which the prior literature reports induced responses to speech. Because we sought to quantify induced activity on a single-trial level, we focused on alpha and beta rhythms here (rather than higher-frequency gamma band activity) due to the relatively greater SNR of alpha and beta in EEG measurements (Buzsáki et al., 2012; Viswanathan et al., 2019).

3. Materials and Methods

The stimuli, participants, experimental design, and hardware used in the current study are described in detail in the materials and methods in Viswanathan et al. (2021). Below, we describe each briefly.

3.1. Stimulus generation

Target speech that listeners were instructed to attend were Harvard/Institute of Electrical and Electronics Engineers (IEEE) sentences (Rothauser, 1969) spoken in a female voice and recorded as part of the PN/NC corpus (McCloy et al., 2013). Stimuli were created for two different speech-in-noise experimental conditions, as described below.

1. Condition 1: Speech in babble (SiB). Speech was added to spectrally matched four-talker babble at -2 dB SNR. The long-term spectra of the target speech sentences were adjusted to match the average (across instances) long-term spectrum of four-talker babble. In creating each SiB stimulus, a babble sample was randomly selected from a list comprising 72 different four-talker babble maskers obtained from the QuickSIN corpus (Killion et al., 2004).

2. Condition 2: Speech in speech-shaped stationary noise (SiSSN). Speech was added to spectrally matched stationary Gaussian noise, i.e., speech-shaped stationary noise, at -5 dB

SNR. The long-term spectra of the target speech sentences and that of stationary noise were adjusted to match the average (across instances) long-term spectrum of four-talker babble. A different realization of stationary noise was used for each SiSSN stimulus.

The particular SNRs used for SiB and SiSSN yielded average speech intelligibility values close to 50% (Figure 1; also see Viswanathan et al., 2021), which helped avoid floor and ceiling effects when quantifying percent-correct scores on a trial-by-trial basis. Note that the speech and masker sources in each acoustic mixture were co-localized (presented diotically) for all stimuli.

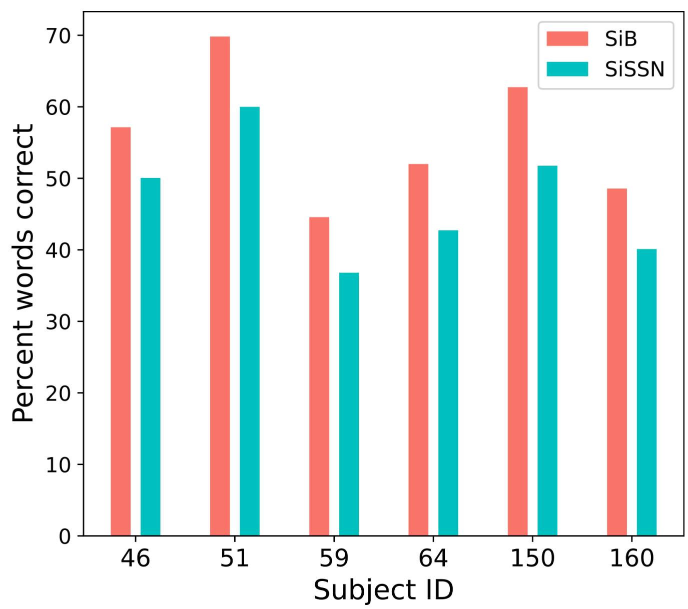

Figure 1. Percent words correct as a function of subject and experimental condition (SiB versus SiSSN). Data shown are pooled over trials.

3.2. Participants

Data from six human subjects (one male, five female) aged 19–31 years were analyzed for this study. All subjects were native speakers of North American English, had pure-tone hearing thresholds better than 20 dB hearing level in both ears at standard audiometric frequencies between 250 Hz and 8 kHz, and reported no history of neurological disorders. All subjects also had distortion-product and click-evoked otoacoustic emissions within the normal range (Gorga et al., 1993) as well as normal tympanograms. Subjects provided informed consent in accordance with protocols established at Purdue University. Data were collected from each subject over the course of one or two visits (with a total visit time of approximately 5 hours).

3.3. Experimental design

Each subject performed 175 trials of speech intelligibility testing for each of the two experimental conditions, with a distinct target sentence in every trial. In total, 1050 trials were collected for each experimental condition across the subject cohort. The different experimental conditions were intermingled across trials.

Thirty-two-channel EEG was measured as subjects performed the speech identification task. The target speech sentences were presented at a sound level of 72 dB sound pressure level (SPL), while the level of the background was set to obtain the desired SNR.

Subjects were instructed that they would be listening for a woman's voice speaking a sentence in each trial and that at the end of the trial they would have to verbally repeat the sentence back to the experimenter sitting beside them. They were told that it would be the same woman's voice every time but that the type and level of background noise would vary across trials. They were also told that the noise would start first in each trial with the target woman's voice starting approximately one second later. They were encouraged to guess as many words as they could if they heard a sentence only partially.

At the beginning of each trial, subjects were presented with a visual cue that read "stay still and listen now" in red font. The audio stimulus started playing one second afterward. In every trial, the background noise started first, while the target speech started 1.25 seconds later to allow sufficient time to cue the subjects' attention to the stimulus. The target was at least 2.5 seconds long. After the target sentence ended, the background noise continued for a short, variable amount of time. Two hundred ms after the noise ended, subjects were presented with a different visual cue that read "repeat now" in green font, cueing them to report the target sentence. This delayed response design avoided motor artifacts and speech-motor preparation signals from contributing to the EEG recorded during listening. For each trial, intelligibility was scored on five pre-determined keywords (which excluded articles and prepositions) in the target sentence and then converted to a percent-correct score (whose value in each trial was either 0, 20, 40, 60, 80, or 100). Before the actual EEG experiment, subjects performed a short training demo task, which used the same listening conditions and target speaker's voice as the actual experiment but a different set of Harvard/IEEE target sentences from the main experiment.

3.4. Hardware

The entire experiment was conducted in a sound-treated booth. A personal desktop computer controlled all aspects of the experiment, including triggering sound delivery and storing data. Special-purpose sound-control hardware (System 3 real-time signal processing system, including digital- to-analog conversion and amplification; Tucker Davis Technologies, Alachua, FL) presented audio through insert earphones (ER-2; Etymotic, Elk Grove Village, IL) coupled to foam ear tips. The earphones were custom shielded by wrapping the transducers in layers of magnetic shielding tape made from an amorphous cobalt alloy (MCF5; YSHIELD GmbH & Co., Ruhstorf, Germany) and then placing them in 3-mm-thick aluminum enclosures to attenuate electromagnetic interference. The signal cables driving the transducers were shielded with braided metallic Techflex (Techflex, Sparta, NJ). All shielding layers were grounded to the chassis of the digital-to-analog (D/A) converter. The absence of measurable electromagnetic artifact was verified by running intense click stimuli through the transducers with the transducers positioned in the same location relative to the EEG cap as actual measurements but with foam tips left outside the ear. All audio signals were digitized at a sampling rate of 48.828 kHz. The EEG signals were recorded at a sampling rate of 4.096 kHz using a BioSemi (Amsterdam, The Netherlands)

ActiveTwo system. Recordings were done with 32 cephalic electrodes and two additional earlobe electrodes.

3.5. EEG processing

All six subjects were able to stay still during the presentation of the sentences and respond on cue. EEG signals were preprocessed by re-referencing channel data to the average of the two earlobe reference electrodes. Then, the signal space projection method was used to construct spatial filters to remove eye blink and saccade artifacts (Uusitalo and Ilmoniemi, 1997). Finally, the broadband EEG was bandpass filtered between 1 and 400 Hz and parceled into epochs, each of which corresponded to a distinct trial. Each epoch started 0.5 seconds before the "stay still and listen now" cue of the corresponding trial, and ended when the shortest target sentence ended; thus the total duration of each epoch was 5.25 seconds.

For each subject and experimental condition, the EEG response spectrogram in each epoch was calculated using a Slepian-tapered complex exponential wavelet (Slepian, 1978; Thomson, 1982). Five cycles were used to estimate each time-frequency bin with a time-full-bandwidth product of 2.

Based on our observation of clear induced brain oscillations in the alpha (7–15 Hz) and beta (13–30 Hz) bands (Figure 2), we derived scalp topomaps in each of these two bands for the pre-stimulus (the one-second-long time period between the "stay still and listen now" cue and the stimulus presentation) and during-stimulus (the 3.75-seconds-long time period between the start of stimulus presentation and end of the target sentence) periods; this was done by averaging the response spectrogram over all band-specific frequencies, epochs, subjects, experimental conditions, and time samples in the corresponding period.

To obtain overall measures of alpha power in each trial, the spectrogram in the corresponding epoch was averaged in the alpha band over parieto-occipital channels (A9, A10, A11, A12, A13, A14, A15, A16, A17, A18, A19, A20, A21, and A22; based on Figure 3A topomap). This was done separately over the pre-stimulus period and the during-stimulus period. An overall measure of trial-specific pre- and during-stimulus beta-band power was obtained using a similar approach, but by using frontal channels (A1, A2, A3, A4, A5, A6, A25, A26, A27, A28, A29, A30, and A31) instead (based on Figure 3B topomap). Since alpha power in the pre- and during-stimulus periods were strongly correlated (Figure 4A), we combined the pre- and during-stimulus power by averaging before performing further analysis. Similarly, since the beta power in the pre- and during-stimulus periods were correlated (Figure 4B), the average power across pre- and during-stimulus periods was used for all further analysis.

3.6. Statistical analysis

We tested whether greater alpha power is associated with higher percent-correct score across different trials. A linear model was built with the different trials across subjects as observations. The log of alpha power was taken as the response variable (to ensure response data are normally distributed; Thomson and Chave, 1991) while percent-correct score and condition (SiB versus SiSSN) were predictors. Percent correct was treated as an ordered factor variable (with six levels:

0, 20, 40, 60, 80, and 100) and condition as a factor variable (with two levels). Statistical tests used ANOVA (Type II tests with the F statistic) along with post-hoc t-tests. The same approach was used to test whether greater beta power is associated with higher percent-correct score across trials.

We performed several posthoc analyses to quantitatively assess the precise relationship between percent-correct score and alpha or beta power. First, we examined the individual terms in the omnibus model to assess the respective contributions of the linear, quadratic, cubic, and fourth order terms in the model. The linear term had the largest contribution to the overall main effect in both the alpha and beta models. To understand the linear term further, we performed pairwise t-tests for percent-correct scores going from 0 to 20, 20 to 40, 40 to 60, etc. (i.e., successive difference contrast coding).

To test whether alpha and beta power are correlated across trials, we used a linear mixed-effects model to account for any random effect of subject. Log beta power in different trials across subjects was the response, and log alpha power, condition (factor variable with two levels), and subject were predictors. Anova (Type II Wald F tests with Kenward-Roger degree of freedom; Kenward and Roger, 1997) was used for statistical testing.

Behavioral outcomes across different trials may be influenced by both top-down selective attention and contextual prediction. To explore this possibility directly, we examined response error patterns, computing the histogram (across trials and subjects) of number of words correct per sentence separately for each experimental condition. If either selective attention or predictive coding—or a combination of the two—influence trial-wise behavioral outcome, performance across the different words should be correlated within a trial. We tested for this effect separately for each experimental condition by comparing the histogram of number of words correct per sentence against the distribution under the null hypothesis of independent outcomes across different words. Under the null hypothesis, the performance on any particular word in a sentence (i.e., whether or not the word was reported correctly) has a Bernouilli distribution with parameter p = average proportion correct score for the particular condition. Moreover, the probability of reporting correctly x words out of a total of 5 words per sentence is binomial with parameters $n = 5$, and p = average proportion correct score for the condition. Assuming independent outcomes across different sentences, the probability that M sentences out of a total of 1050 sentences per condition (pooled over all six subjects; each subject performed 175 sentences per condition) had x words correct is also binomial, with parameters $n = 1050$ and p = probability that x words per sentence are correct. We compared this final probability distribution (modeling the null distribution of independent word outcomes in each condition) with the histogram of number of words correct per sentence. Specifically, we generated p-values describing, for each experimental condition, the likelihood of observing the actual correlation across words within each trial, assuming that performance on the words was truly independent.

We used a multinomial linear regression model to test whether beta power contributes additionally to predicting percent-correct score over the contribution of alpha power alone, and vice-versa (i.e., whether alpha power contributes additional predictive power over that contributed by beta power alone). The percent-correct score in different trials across subjects was the response (treated as an ordered factor variable with six levels: 0, 20, 40, 60, 80, and 100); the predictors were log alpha power, log beta power, and condition (factor variable with two levels). Likelihood-ratio Type II tests were used for statistical testing by calculating the deviance (i.e., -2 times log likelihood-ratio) and comparing it to a chi-squared distribution (Wilks, 1938).

3.7. Software accessibility

Stimulus presentation was controlled using custom MATLAB (The MathWorks, Inc., Natick, MA) routines. EEG preprocessing was performed using the open-source software tools MNE-PYTHON (Gramfort et al., 2014) and SNAPsoftware (Bharadwaj, 2018). All further data analyses were performed using custom software in PYTHON (Python Software Foundation, Wilmington, DE). Statistical analyses were performed using R (R Core Team; www.R-project.org). Visualizations used the colorblind-friendly Colorbrewer (Harrower and Brewer, 2003) and Color Universal Design (Ichihara et al., 2008) colormap palettes. Copies of all custom code can be obtained from the authors.

4. Results

We wished to quantify induced brain oscillations in different canonical frequency bands (Buzsáki and Draguhn, 2004) on a trial-by-trial basis and relate those to speech intelligibility, also measured on a trial-by-trial basis. For this, we first computed the EEG response spectrogram in each trial (see Materials and Methods: EEG processing) to examine induced (versus evoked) oscillatory activity in different frequency bands. Figure 2A shows the average EEG spectrogram over the 32 EEG channels, and all trials, subjects, and experimental conditions. Figure 2B shows the average EEG spectrum obtained by computing the mean over time of the spectrogram shown in Figure 2A. As seen in Figure 2, induced brain oscillations are clearly visible in the alpha (7–15 Hz) and beta (13–30 Hz) bands. Because we did not see any induced activity outside the alpha and beta frequency ranges, we restricted all further analyses in the current study to just the alpha and beta bands.

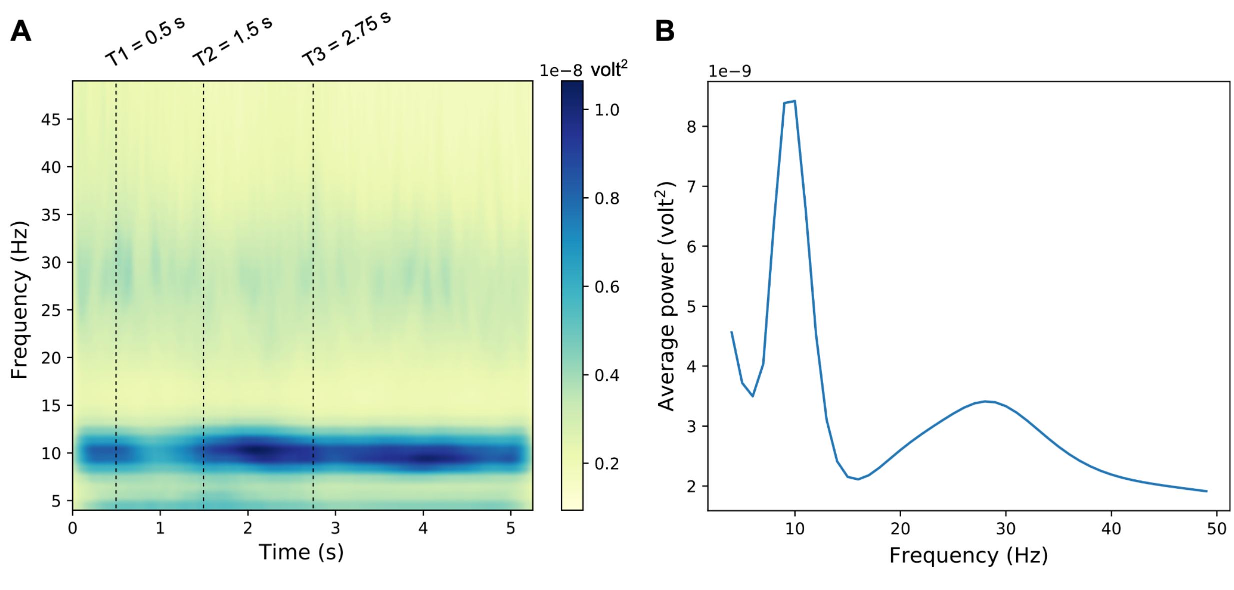

Figure 2. Average EEG response spectrogram (A) and spectrum (B). The spectrogram and spectrum shown are averaged over the 32 EEG channels, and all trials, subjects, and experimental conditions. Note that the time before T1 corresponds to baseline. At T1, the "stay still and listen now" visual cue was shown. At T2, the audio stimulus started playing. At T3, presentation of the target speech sentence started; target presentation lasted until at least 5.25 s (and was longer for the longer sentences).

To better understand the neural sources of these alpha and beta induced oscillations, we computed their scalp topography (see Materials and Methods: EEG processing). Figures 3A and 3B show average (over band-specific frequencies, trials, subjects, experimental conditions, and time samples) scalp topomaps for the alpha and beta bands, respectively; the topomaps are plotted separately for the pre- and during-stimulus periods so as to be able to visualize any differences in the contributions of preparatory rhythmic activity and stimulus-induced oscillatory activity (Foxe and Snyder, 2011). Results suggest that the strongest alpha power occurs in the parieto-occipital channels (Figure 3A) and the strongest beta power occurs in the frontal channels (Figure 3B).

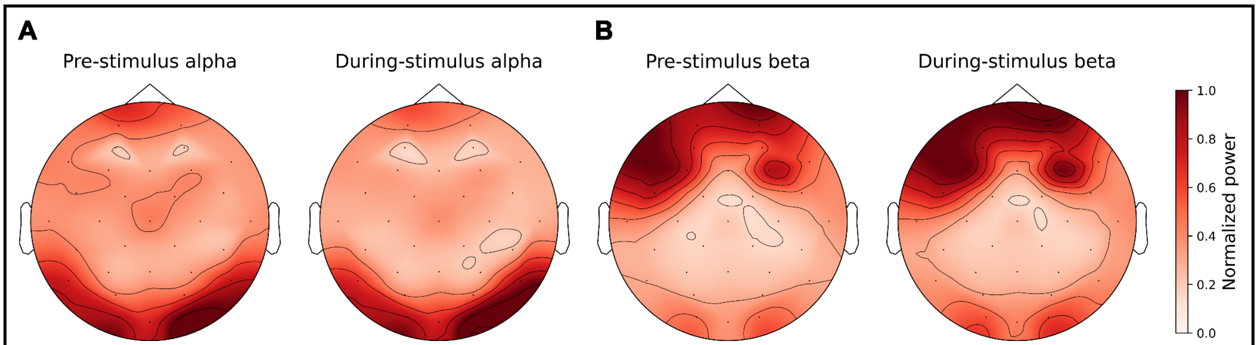

Figure 3. Average scalp topography maps for the induced oscillatory power in the alpha (A) and beta (B) bands, shown separately for the pre- and during-stimulus time periods. The topomaps are averaged over band-specific frequencies, trials, subjects, experimental conditions, and time samples.

Based on Figure 3 scalp topomaps, we used parieto-occipital EEG channels to derive an overall measure of alpha power for each trial from the EEG response spectrogram; we did this separately for the pre- and during-stimulus periods (see Materials and Methods: EEG processing). Similarly, we derived an overall measure of trial-specific pre- and during-stimulus beta power, but by using frontal EEG channels instead. Figures 4A and 4B plot the power in different trials across subjects and conditions in the pre-stimulus period against the corresponding values in the during-stimulus period for the alpha and beta bands, respectively. Across trials, pre- and during-stimulus induced oscillation power was significantly correlated for both alpha ($R^2 = 0.6292$, $p < 2e-16$) and beta ($R^2 = 0.9151$, $p < 2e-16$). For beta, the during-stimulus power was roughly equal to the pre-stimulus power (in Figure 4B, data fall along the identity line); however, for alpha, the during-stimulus power was consistently higher than the pre-stimulus power (in Figure 4A, data fall above the diagonal), but the pre-stimulus value nonetheless predicted the during-stimulus alpha power value. We were interested in whether the difference in alpha (and beta) power across trials was related to behavioral performance and wished to quantify the level of alpha (and beta) in each trial relative to the other trials. Therefore, we averaged the pre- and during-stimulus periods together for all further analyses.

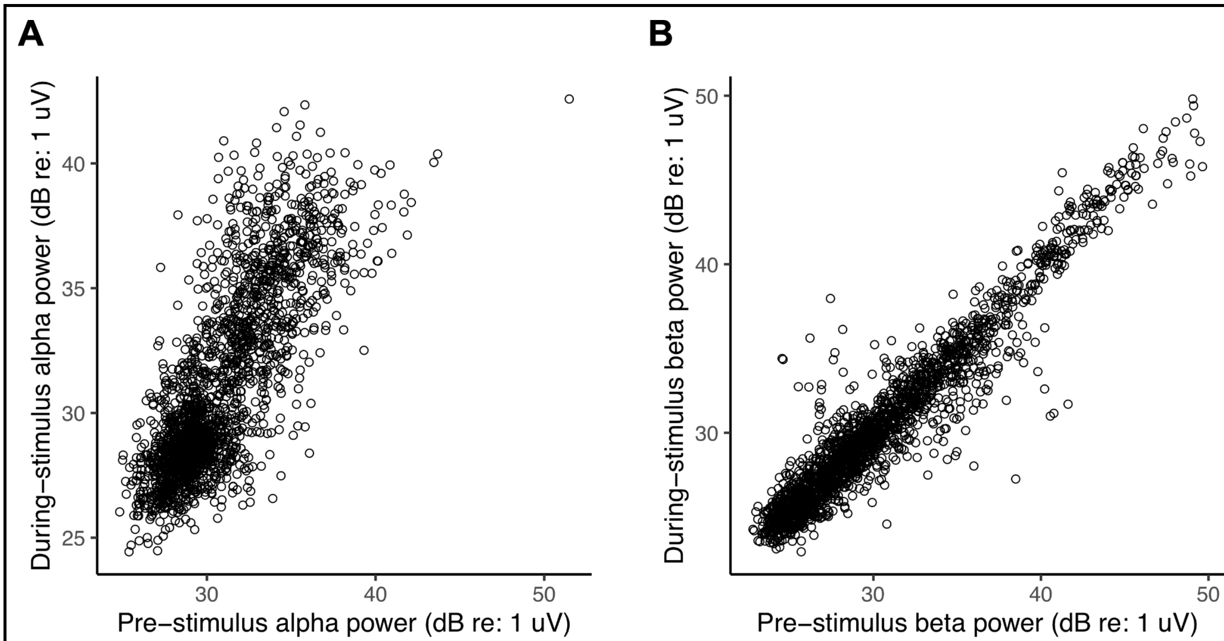

Figure 4. Overall parieto-occipital alpha (A) and frontal beta (B) power in different trials (across subjects and conditions) in the pre-stimulus period against the corresponding values in the during-stimulus period.

Across trials, we compared percent-correct score with the corresponding alpha and beta power. Figure 5A shows alpha power versus percent-correct score in different trials across subjects, separately for each condition. Alpha power covaried significantly with percent-correct score within condition ($F = 4.7789$, $p = 0.0002397$; see Materials and Methods: Statistical analysis for details). Figure 5B shows beta power versus percent-correct score in different trials across subjects, separately for each condition. Beta power too covaried significantly with percent-correct score within condition ($F = 6.4915$, $p = 5.346e\text{-}06$). Posthoc analyses revealed that the largest contribution to the main effect of score on alpha and beta power came from the linear term, indicating that alpha and beta power increased with score ($T = 4.216$, $p = 2.59e\text{-}05$ for alpha; $T = 4.173$, $p = 3.13e\text{-}05$ for beta). The quadratic term was also significant in predicting alpha power, but carried a negative coefficient ($T = -2.521$, $p = 0.0118$) in line with the plateauing of alpha power with increasing score seen in Figure 5A. In the model for beta power, the cubic term was also significant ($T = 3.003$, $p = 0.00271$), in line with the U-shaped trend seen over the 20–100% range of scores in Figure 5B. We conducted pairwise t-tests to compare the changes in alpha and beta power for each step increase in percent-correct score. These sequential-difference-contrast analyses showed that alpha and beta power increased when percent-correct score increased from 0 to 20 ($T = 1.994$, $p = 0.0463$ for alpha; $T = 4.249$, $p = 2.24e\text{-}05$ for beta); however, for the successive steps (20 to 40, 40 to 60, etc.), the increase in power was not significant for either alpha or beta. Figure 5A also suggests that for a given percent-correct score, alpha power is greater for speech in stationary noise than in babble; however, statistical testing did not reveal a significant effect of experimental condition (SiB versus SiSSN) on alpha power.

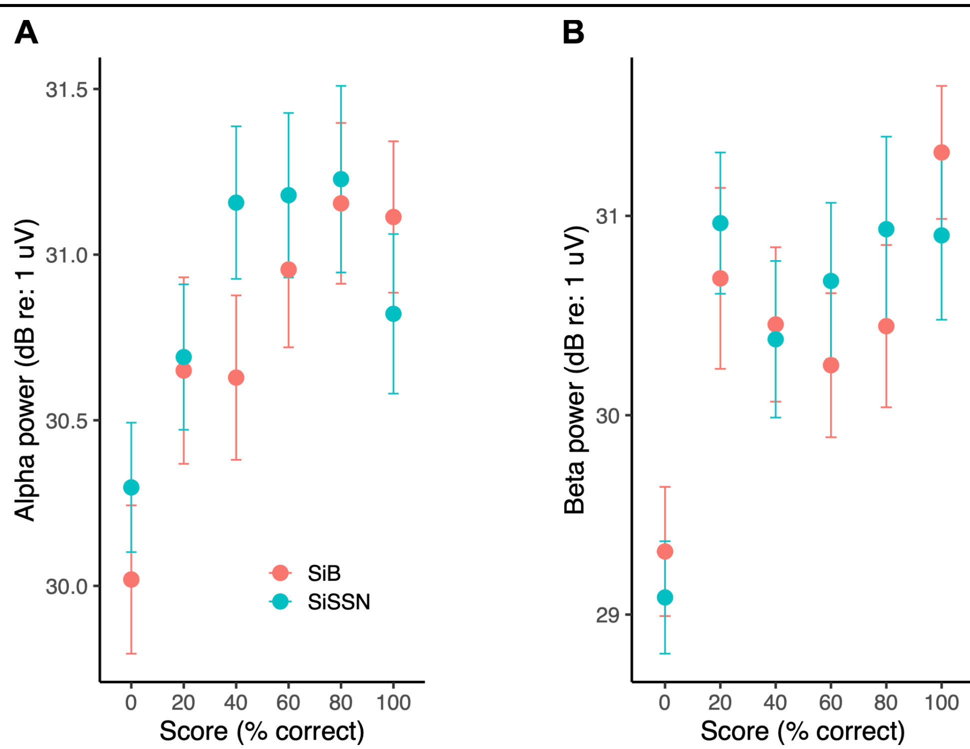

Figure 5. Alpha (A) and beta (B) power [mean and standard error of the mean (STE)] versus percent-correct score in different trials across subjects. Data are shown separately for each experimental condition (SiB versus SiSSN).

Because alpha and beta power were both related to percent-correct score on a given trial, we directly compared alpha and beta power on a trial-by-trial basis. Figure 6A plots beta power versus alpha power in different trials across subjects and conditions, with data from each subject shown in a different color. We found significant correlation between alpha and beta power across trials (F = 175.7166, p < 2e-16), even after accounting for the random effect of subject (i.e., we see significant correlation even within subject). Another interesting observation from Figure 6A is that there appear to be individual differences in the overall magnitude of alpha and beta power across trials, as well as in the distribution of the alpha-to-beta power ratio across trials. The latter individual differences are quantified in Figure 6B. While some subjects (e.g., subject 46) show an alpha-to-beta power ratio greater than 1 across all trials, others (e.g., subject 64) show an alpha-to-beta power ratio less than 1 in most trials. This result raises the possibility that the listening strategy used may differ across individuals, with some listeners using focused attention to perform the task and others relying more on predictive coding.

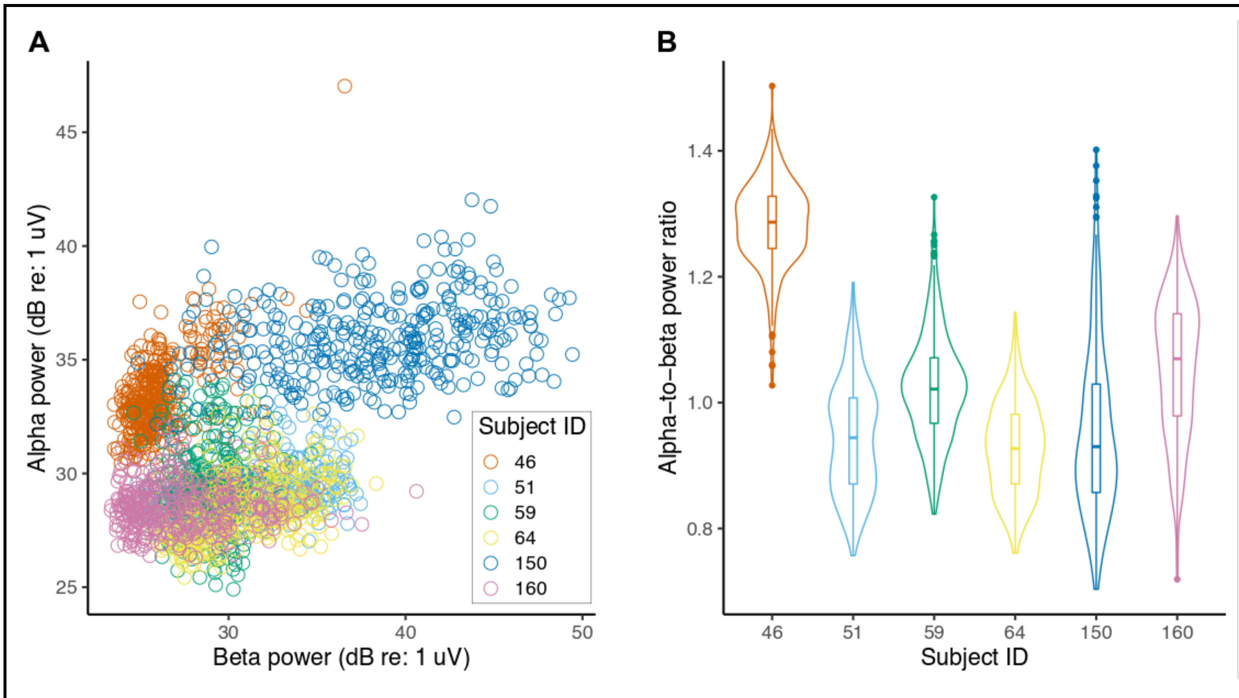

Figure 6. (A) Beta power versus alpha power in different trials (across subjects and conditions). The different colors in the plot correspond to different subjects. (B) The distribution of alpha-to-beta power ratio across trials for each subject is shown as a violin plot. The median (horizontal bar), 50% confidence limits (box), and 95% confidence limits (whiskers) of each distribution are also shown.

To directly test whether the behavioral outcome across different trials may be influenced by selective attention or contextual prediction, we plotted behavioral response error patterns. Figure 7 shows the histogram (across trials and subjects) of number of words correct per sentence, separately for each experimental condition. These data show that the probability with which subjects get 0, 1, 2, 3, 4, or 5 words correct per sentence is significantly different from the expected distribution under the null hypothesis of independent word outcomes (p-values shown in Figure 7; see Materials and Methods: Statistical analysis for details). Thus, there is a high correlation in performance across the different words in a trial in both experimental conditions

(SiB and SiSSN), which suggests that either top-down selective attention or predictive coding, or both, influence the behavioral outcome in each trial.

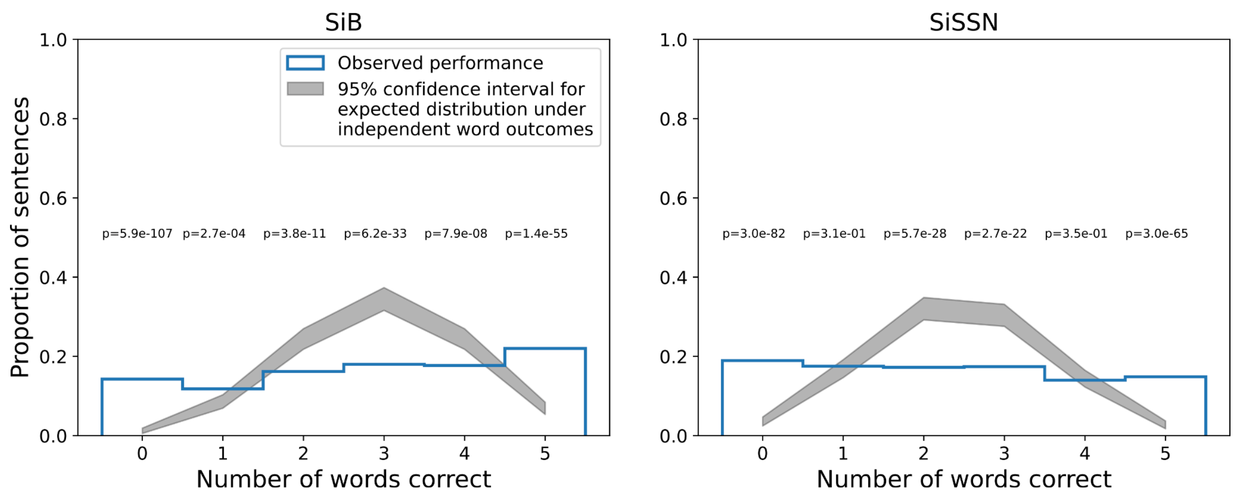

Figure 7. Normalized histogram (across trials and subjects) of number of words correct per sentence for SiB and SiSSN conditions (blue unfilled bars). Shown also is the 95% confidence interval for the expected distribution under the null hypothesis of independent word outcomes (gray patch). P-values for the observed data under the null hypothesis are also indicated.

Although alpha and beta power were correlated across trials (Figure 6A), beta power contributed significant additional predictive power to predict within-condition percent-correct score over the contribution of alpha power alone (deviance chi-square for the full model versus the same model without beta = 25.209, $p = 0.000127$), and vice-versa (deviance chi-square for the full model versus the same model without alpha = 15.475, $p = 0.008515$). Thus, not only do induced oscillations in both alpha and beta bands relate to speech intelligibility in noise on a trial-by-trial basis within condition, but crucially, alpha and beta power each make significant independent contributions to predicting trial outcome (i.e., trial-wise speech intelligibility). This result is quantified in Figure 8.

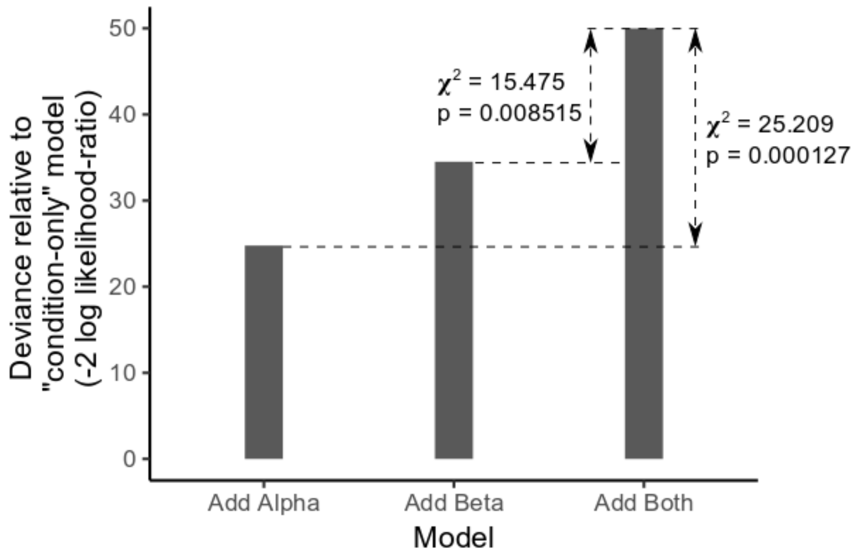

Figure 8. Likelihood ratio deviances (relative to a base model with condition as the only predictor) for three different multinomial models of percent-correct score: (1) model with condition and alpha power as predictors, (2) model with condition and beta power as predictors, and (3) model with condition, alpha power, and beta power as predictors. The deviance chi-square statistic and corresponding p-value are indicated for the comparison between models (1) and (3), and separately also for the comparison between models (2) and (3).

5. Discussion

Using human EEG with simultaneous speech intelligibility measurements in different masking conditions (speech in multi-talker babble, and speech in speech-shaped noise) in the present study, we found that induced brain oscillations in the alpha and beta bands relate to speech intelligibility in competition on a trial-by-trial basis. Specifically, we found that the overall magnitudes (averaged over the pre- and during-stimulus periods) of alpha power in parieto-occipital EEG channels and beta power in frontal channels significantly covary with, and importantly independently contribute to, single-trial speech intelligibility in our speech-in-noise tasks. These results are consistent with the posited role of the parieto-occipital alpha rhythm in auditory selective attention (Foxe and Snyder, 2011; Strauß et al., 2014; Banerjee et al., 2011; Deng et al., 2019; Deng et al., 2020) and the frontal beta rhythm in speech-motor predictive coding (Engel and Fries, 2010; Arnal and Giraud, 2012; Lewis and Bastiaansen, 2015) that is thought to stabilize speech representation in adverse listening conditions (Hickok et al., 2011; Skipper et al., 2017; Pulvermüller, 2018). The interpretation that some combination of top-down selective attention and predictive coding influences the behavioral outcome in each trial is also supported by behavioral response error patterns in our data (Figure 7).

Our results are in line with prior reports of a positive correlation between alpha power in parietal EEG channels and speech intelligibility in noise (e.g., across SNRs as quantified in the during-stimulus period by Hall et al., 2019, and across individuals as quantified in the pre-stimulus period by Alhanbali et al., 2022). However, at least at first glance, our results appear to be at odds with other reports (e.g., by Obleser and Weisz, 2012 and Becker et al., 2013, who used noise-vocoded speech in quiet, and Dimitrijevic et al., 2017, who used digits in noise) that better comprehension is associated with alpha suppression (rather than a power increase) in the late during-stimulus period in temporal brain regions and central EEG channels. This discrepancy may be explained in part by the existence of multiple neural generators of task-related alpha (i.e., alpha power in the parieto-occipital and central EEG channels may reflect two different mechanisms of alpha; Deng et al., 2020). Moreover, some of these studies presented speech in quiet rather than with simultaneous competing sounds, which could evoke different mechanisms (Obleser and Weisz, 2012 and Becker et al., 2013).

Foxe and Snyder (2011) distinguish between parieto-occipital alpha seen in an unaroused state (e.g., when visual stimuli are ignored; Adrian, 1944) and that seen in selective attention across different stimuli (especially spatial selective attention, where alpha power is lateralized according to the hemifield of focus; Banerjee et al., 2011; Deng et al., 2019; Deng et al., 2020). In the present study, the target speech and masker sources were both presented diotically rather than spatially separated; thus, even though it required selective attention, our task did not involve any spatial focus of attention. It may be that the alpha in the current study, which covaries with trial-wise speech intelligibility, reflects an overall suppression of the visual scene and focus of auditory attention, rather than a mechanism specific to stimulus selection. Another possibility is that there may be a common mechanism in play across the parieto-occipital alpha seen in the two cases. Indeed, the frontoparietal attention network becomes active during spatial attention and working memory for auditory stimuli as well as for visual inputs, even though many earlier studies assume it is strictly a visuospatial processing network (Corbetta and Shulman, 2002; Michalka et al., 2015; Michalka et al., 2016; Noyce et al., 2017; Noyce et al., 2022). Thus, future studies should disambiguate between the different mechanisms by which the alpha rhythm may mediate suppression of sensory distractors (Foxe and Snyder, 2011; Strauß et al., 2014), especially for co-localized sources like those used in the current study.

Our current results show that trial-by-trial variations in alpha and beta power are correlated, even within subject (Figure 6A; statistics given in Results). Prior studies have also reported that oscillatory activity within the alpha and beta bands are correlated (Carlqvist et al., 2005), even though they may represent distinct functions. Despite being correlated, alpha and beta power each provide significant independent contributions to predicting single-trial percent-correct score (Figure 8). Moreover, there are individual differences in the overall magnitude of alpha and beta power as well as in the alpha-to-beta power ratio across trials (Figure 6). It may be that those individuals with larger alpha-to-beta ratios relied more on focused attention to perform the task whereas those with smaller alpha-to-beta ratios relied more on predictive coding. In both cases, trial-wise performance improved as either alpha or beta power increased (Figure 5). Our behavioral measurements cannot disambiguate between the different listening strategies of using focused attention versus relying on predictive coding. However, measuring confusion patterns can inform which listening strategy an individual subject used (e.g., when a subject made an error, whether they reported a word from the competing stream or a new, contextually suitable word; Ruggles et al., 2011). Future experiments can directly test whether a higher alpha-to-beta power ratio is associated with greater reliance on focused attention versus robust prediction, and whether a lower alpha-to-beta power ratio is associated with the opposite.

References

Noyce, A. L., Cestero, N., Michalka, S. W., Shinn-Cunningham, B. G., & Somers, D. C. (2017). Sensory-biased and multiple-demand processing in human lateral frontal cortex. Journal of Neuroscience, 37(36), 8755-8766.

Noyce, A. L., Lefco, R. W., Brissenden, J. A., Tobyne, S. M., Shinn-Cunningham, B. G., & Somers, D. C. (2022). Extended frontal networks for visual and auditory working memory. Cerebral Cortex, 32(4), 855-869.

Adank, P., Davis, M. H., & Hagoort, P. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. Neuropsychologia, 50(1), 77-84.

Adrian, E. D. (1944). Brain rhythms. Nature, 153, 360-362.

Alhanbali, S., Munro, K. J., Dawes, P., Perugia, E., & Millman, R. E. (2022). Associations between pre-stimulus alpha power, hearing level and performance in a digits-in-noise task. International Journal of Audiology, 61(3), 197-204.

Alho, J., Lin, F. H., Sato, M., Tiitinen, H., Sams, M., & Jääskeläinen, I. P. (2014). Enhanced neural synchrony between left auditory and premotor cortex is associated with successful phonetic categorization. Frontiers in Psychology, 5, 394.

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. Trends in cognitive sciences, 16(7), 390-398.

Banerjee, S., Snyder, A. C., Molholm, S., & Foxe, J. J. (2011). Oscillatory alpha-band mechanisms and the deployment of spatial attention to anticipated auditory and visual target locations: supramodal or sensory-specific control mechanisms?. Journal of Neuroscience, 31(27), 9923-9932.

Becker, R., Pefkou, M., Michel, C. M., & Hervais-Adelman, A. G. (2013). Left temporal alpha-band activity reflects single word intelligibility. Frontiers in systems neuroscience, 7, 121.

Bharadwaj, H. M. (2018). “SNAPsoftware/ANLffr: Software tools for electrophysiology from the Systems Neuroscience of Auditory Perception Lab,” https://github.com/SNAPsoftware/ANLffr (Last viewed September 14, 2021).

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. science, 304(5679), 1926-1929.

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. Nature reviews neuroscience, 13(6), 407-420.

Carlqvist, H., Nikulin, V. V., Strömberg, J. O., & Brismar, T. (2005). Amplitude and phase relationship between alpha and beta oscillations in the human electroencephalogram. Medical and Biological Engineering and Computing, 43(5), 599-607.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. The Journal of the acoustical society of America, 25(5), 975-979.

Cope, T. E., Sohoglu, E., Sedley, W., Patterson, K., Jones, P. S., Wiggins, J., ... & Rowe, J. B. (2017). Evidence for causal top-down frontal contributions to predictive processes in speech perception. Nature Communications, 8(1), 1-16.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nature reviews neuroscience, 3(3), 201-215.

Deng, Y., Reinhart, R. M., Choi, I., & Shinn-Cunningham, B. G. (2019). Causal links between parietal alpha activity and spatial auditory attention. Elife, 8.

Deng, Y., Choi, I., & Shinn-Cunningham, B. (2020). Topographic specificity of alpha power during auditory spatial attention. Neuroimage, 207, 116360.

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2017). Cortical alpha oscillations predict speech intelligibility. Frontiers in human neuroscience, 11, 88.

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. Proceedings of the National Academy of Sciences, 109(29), 11854-11859.

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proceedings of the National Academy of Sciences, 111(19), 7126-7131.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations—signalling the status quo?. Current opinion in neurobiology, 20(2), 156-165.

Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. Frontiers in psychology, 2, 154.

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., ... & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. Neuron, 77(5), 980-991.

Gorga, M. P., Neely, S. T., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., ... & Jesteadt, W. (1993). A comparison of transient-evoked and distortion product otoacoustic emissions in normal-hearing and hearing-impaired subjects. The Journal of the Acoustical Society of America, 94(5), 2639-2648.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. Neuroimage, 86, 446-460.

Harrower, M., & Brewer, C. A. (2003). ColorBrewer. org: an online tool for selecting colour schemes for maps. The Cartographic Journal, 40(1), 27-37.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. Neuron, 69(3), 407-422.

Ichihara, Y. G., Okabe, M., Iga, K., Tanaka, Y., Musha, K., & Ito, K. (2008, January). Color universal design: the selection of four easily distinguishable colors for all color vision types. In Color Imaging XIII: Processing, Hardcopy, and Applications(Vol. 6807, pp. 206-213). SPIE.

Keitel, A., Ince, R. A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. Neuroimage, 147, 32-42.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. Biometrics, 983-997.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. The Journal of the Acoustical Society of America, 116(4), 2395-2405.

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. Cortex, 68, 155-168.

Love, J., Shinn-Cunningham, B., & Bharadwaj, H. (2020). Endogenous brain oscillations in the 10–20 Hz range during auditory spatial attention. The Journal of the Acoustical Society of America, 148(4), 2468-2468.

McCloy, D., Souza, P., Wright, R., Haywood, J., Gehani, N., and Rudolph, S. (2013). "The PN/NC Corpus. Version 1.0," http://depts.washington.edu/phonlab/resources/pnnc (Last viewed March 17, 2017).

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. Nature, 485(7397), 233-236.

Michalka, S. W., Kong, L., Rosen, M. L., Shinn-Cunningham, B. G., & Somers, D. C. (2015). Short-term memory for space and time flexibly recruit complementary sensory-biased frontal lobe attention networks. Neuron, 87(4), 882-892.

Michalka, S. W., Rosen, M. L., Kong, L., Shinn-Cunningham, B. G., & Somers, D. C. (2016). Auditory spatial coding flexibly recruits anterior, but not posterior, visuotopic parietal cortex. Cerebral Cortex, 26(3), 1302-1308.

Obleser, J., & Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cerebral cortex, 22(11), 2466-2477.

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. Journal of Neuroscience, 32(36), 12376-12383.

O'sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., ... & Lalor, E. C. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. Cerebral cortex, 25(7), 1697-1706.

Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Current Biology, 25(12), 1649-1653.

Pefkou, M., Arnal, L. H., Fontolan, L., & Giraud, A. L. (2017). θ-Band and β-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. Journal of Neuroscience, 37(33), 7930-7938.

Pulvermüller, F. (2018). Neural reuse of action perception circuits for language, concepts and communication. Progress in neurobiology, 160, 1-44.

Rothauser, E. H. (1969). IEEE recommended practice for speech quality measurements. IEEE Trans. on Audio and Electroacoustics, 17, 225-246.

Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. Proceedings of the National Academy of Sciences, 108(37), 15516-15521.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. Trends in neurosciences, 32(1), 9-18.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. Trends in cognitive sciences, 12(5), 182-186.

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. Brain and language, 164, 77-105.

Slepian, D. (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty V: the discrete case", Bell Systems Tech. J, 57(5), 1371-1430.

Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. Frontiers in human neuroscience, 8, 350.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. Proceedings of the IEEE, 70(9), 1055-1096.

Thomson, D. J., & Chave, A. D. (1991). Advances in spectrum analysis and array processing. Jackknifed error estimates for spectra, coherences, and transfer functions. Advances in spectrum analysis and array processing, 58-113.

Uusitalo, M. A., & Ilmoniemi, R. J. (1997). Signal-space projection method for separating MEG or EEG into components. Medical and biological engineering and computing, 35(2), 135-140.

Viswanathan, V., Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2019). Electroencephalographic signatures of the neural representation of speech during selective attention. Eneuro, 6(5).

Viswanathan, V., Bharadwaj, H. M., Shinn-Cunningham, B. G., & Heinz, M. G. (2021). Modulation masking and fine structure shape neural envelope coding to predict speech intelligibility across diverse listening conditions. The Journal of the Acoustical Society of America, 150(3), 2230-2244.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. The annals of mathematical statistics, 9(1), 60-62.