NeuroImage 263 (2022) 119627

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Brainstem speech encoding is dynamically shaped online by fluctuations in cortical α state

Jesyin Lai[a,b,c], Caitlin N. Price[a,b,d], Gavin M. Bidelman[a,b,e,f,*]

[a] Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
[b] School of Communication Sciences and Disorders, University of Memphis, Memphis, TN, USA
[c] Diagnostic Imaging Department, St. Jude Children's Research Hospital, Memphis, TN, USA
[d] Department of Audiology and Speech Pathology, University of Arkansas for Medical Sciences, Little Rock, AR, USA
[e] Department of Speech, Language and Hearing Sciences, Indiana University, 2631 East Discovery Parkway, Bloomington, IN 47408, USA
[f] Program in Neuroscience, Indiana University, 1101 E 10th St, Bloomington, IN 47405, USA

ARTICLE INFO

Keywords:
EEG
Cortical α
Frequency-following responses
Attention and arousal

ABSTRACT

Experimental evidence in animals demonstrates cortical neurons innervate subcortex bilaterally to tune brainstem auditory coding. Yet, the role of the descending (corticofugal) auditory system in modulating earlier sound processing in humans during speech perception remains unclear. Here, we measured EEG activity as listeners performed speech identification tasks in different noise backgrounds designed to tax perceptual and attentional processing. We hypothesized brainstem speech coding might be tied to attention and arousal states (indexed by cortical α power) that actively modulate the interplay of brainstem-cortical signal processing. When speech-evoked brainstem frequency-following responses (FFRs) were categorized according to cortical α states, we found low α FFRs in noise were weaker, correlated positively with behavioral response times, and were more "decodable" via neural classifiers. Our data provide new evidence for online corticofugal interplay in humans and establish that brainstem sensory representations are continuously yoked to (i.e., modulated by) the ebb and flow of cortical states to dynamically update perceptual processing.

1. Introduction

Central auditory processing of target sounds is influenced by multiple factors, such as background noise (Billings et al., 2020; Binkhamis et al., 2019; Song et al., 2011), task engagement (Saderi et al., 2021; Shaheen et al., 2021), attention (Price and Bidelman, 2021; Saiz-Alía et al., 2019), and arousal state (Mai et al., 2019; Saderi et al., 2021). These factors may interact to modulate sound processing differentially. For instance, in complex listening environments, attention aids the selection of behaviorally-relevant inputs over irrelevant background noise to prioritize target cues for robust speech-in-noise (SIN) understanding (Price and Moncrieff, 2021). Attention is defined as a top-down cognitive process that alerts and orients listeners to focus on environmental and external stimuli (Petersen and Posner, 2012). There is ample experimental evidence in animals that cortical neurons innervate subcortex (i.e., inferior colliculus) (Bajo and Moore, 2005; Beyerl, 1978) to provide top-down fine-tuning of subcortical auditory neural coding (Atiani et al., 2009; Gao and Suga, 2000; Suga and Ma, 2003). Attention has also been shown to influence all stages of human auditory processing from the inner ear to cortex (Galbraith et al., 2003; Giard et al., 1994; Lukas, 1980; Picton and Hillyard, 1974; Rinne et al., 2008).

Still, while attentional modulation of auditory cortical activity is well established (Alho et al., 2014; Alho and Vorobyev, 2007; Picton and Hillyard, 1974), attention effects on auditory brainstem activity in humans remains highly controversial (Dunlop et al., 1965; Galbraith and Kane, 1993; Picton et al., 1971; Varghese et al., 2015). For instance, brainstem responses recorded during speech perception tasks typically fail to vary with listening state (Varghese et al., 2015), despite concomitant changes in cortex. In recent work, we used frequency-following responses (FFRs)–a scalp-recorded potential reflecting the brain's spectrotemporal encoding of sound–to directly investigate attentional influences on auditory brainstem processing during speech listening tasks. Although a mixture of sources of the hearing pathway contribute to the scalp-recorded FFR (Bidelman, 2015; 2018; Sohmer and Pratt, 1977; Sohmer et al., 1977), its short latency [6-8 ms (Galbraith et al., 2000)], correspondence with intracranial potentials (Smith et al., 1975), and lesion data in both cats (Davis and Britt, 1984; Smith et al., 1975) and humans (Sohmer et al., 1977) indicate the brainstem inferior colliculus

* Corresponding author at: Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, IN, 47408, USA.
E-mail addresses: jesyin.v@gmail.com (J. Lai), gbidel@indiana.edu (G.M. Bidelman).

https://doi.org/10.1016/j.neuroimage.2022.119627.
Received 15 August 2022; Accepted 12 September 2022
Available online 16 September 2022.

is the primary generator of the response. When recorded via EEG, the human FFR is dominantly generated by the brainstem (and sometimes even more peripheral auditory nuclei) across the majority of frequencies in speech (Bidelman, 2015; Bidelman and Momtaz, 2021). Using source-reconstructed FFRs to further isolate responses of subcortical origin, we demonstrated attention enhances brainstem neural coding to aid noise-degraded speech perception (Price and Bidelman, 2021). We also found attentional enhancements in neural signaling between sub- and neocortical levels with stronger functional connectivity in the top-down (corticofugal) direction during challenging listening conditions. While these data provided neuroimaging evidence for efferent influences on brainstem auditory processing in humans, they did not speak to the *temporal dynamics* of auditory efferent system function. The allocation of attention is itself dynamic, waxing and waning over time (Helfrich et al., 2018; Kucyi et al., 2017). Presumably, changes in brainstem sound processing from corticofugal feedback occur online, as cortex dynamically shifts functional states dependent on stimulus or task demands (e.g., listener engagement, arousal, attention). Demonstrating FFRs depend on concurrent cortical activation would establish new and direct evidence that the corticofugal system dynamically shapes subcortical auditory processing in humans in real-time.

Besides attention, top-down pathways also regulate arousal (Krone et al., 2017). While attention is usually directed towards external stimuli and involves the allocation of processing resources to relevant stimuli (Coull, 1998), arousal is related more to the internal state of physiological reactivity of the subject (Cohen, 2014; Eysenck, 1982; Robbins and Everitt, 1995), such as sleep, wakefulness, or excitement. When attending to sounds and actively engaging in a task, one's internal arousal state is not static but fluctuates. Hence, we posited speech sound representations, as indexed via FFRs, might simultaneously track with trial-by-trial variations in cortical arousal state. Such findings would demonstrate online functional effects of the corticofugal pathway in the human auditory system.

In this study, we measured source-reconstructed brainstem and cortical neuroelectric electroencephalogram (EEG) activity as human listeners performed SIN tasks aimed to tax perceptual and attentional processing. The EEG task was designed to contain clean (no background noise) vs. noise condition and active vs. passive condition. We hypothesized top-down arousal effects would be more prominent when participants were actively performing the speech task. The passive condition was included as a negative control where little to no top-down effects were expected. On the other hand, the noise condition was included to exaggerate arousal effects since participants might be consciously suppressing background noise or segregating targets from distractors. Since the EEG task was relatively easy under the clean condition, arousal effects were expected to be less prominent than in the noise condition. In brief, these four conditions allowed us to study potential graded differences in arousal effects. In addition, brainstem speech processing was measured via FFRs. FFRs represent sustained, phase-locked neural activity of a population of neurons which faithfully track dynamic sound features in the auditory brainstem (Bidelman, 2018; Coffey et al., 2019; Marsh et al., 1970; Smith et al., 1975; Tichko and Skoe, 2017; White-Schwoch et al., 2019). Internal arousal state is abstract and inherently subjective rendering it difficult to quantify behaviorally. Some EEG studies investigating arousal revealed that α suppression is induced by emotionally arousing stimuli (Aftanas et al., 2002; Uusberg et al., 2013). As a result, we measured cortical α activity from listeners' running EEG as they executed the speech perception tasks and used it as a neural proxy of arousal (Pivik and Harman, 1995). We hypothesized speech coding along the central auditory system might vary with attention and internal arousal states that actively modulate the interplay of brainstem-cortical signal processing. Furthermore, as α power is predictive of behavioral performance (Gould et al., 2011; Haegens et al., 2011; Kelly et al., 2009; Klatt et al., 2020), we explored correlations between α-yoked FFRs and speech perception measures. Lastly, we also investigated neural decoding of FFRs at different α power as this may provide further evidence revealing the role of α in predicting behavioral performance. Our results provide direct evidence that brainstem speech coding (particularly in noise) is modulated online so that sensory representations are continuously yoked to (i.e., modulated by) the ebb and flow of cortical states to dynamically update perceptual processing.

2. Methods and materials

2.1. Participants

We recruited N=20 young adults (age: 18-35 years, M = 24, SD = 3.4 years; 11 female). All participants exhibited normal hearing thresholds ($\leq$ 25 dB HL; 250 - 8,000 Hz). Since language background and music experience influence FFRs and SIN performance (Mankel and Bidelman, 2018; Parbery-Clark et al., 2009; Zhao and Kuhl, 2018), we required all participants to have < 3 years of formal musical training (M = 0.8 years, SD = 1.2) and speak native English. Participants were predominantly right-handed (M=82.04%, SD = 21.04) with no history of neuropsychiatric disorders. All provided written informed consent in accordance with the Declaration of Helsinki and a protocol approved by the University of Memphis IRB.

2.2. Speech stimuli and task

The speech stimuli and task are described fully in Price and Bidelman (2021) and illustrated in Fig. 1A. Brainstem and cortical EEG were recorded simultaneously (Bidelman et al., 2013; 2019) while subjects performed a continuous monitoring task in which they detected infrequent vowels within a continuous stream of frequent vowels over several minutes. This paradigm was then repeated in noise and during passive listening (requiring no response) to achieve the active/passive and clean/noise blocks (described below). Three synthesized vowel tokens (e.g., /a/, /i/, /u/) were randomly presented back to back over many trials in a continuous stream. These vowels were chosen because they are sustained periodic sounds optimal for evoking FFRs (Bidelman et al., 2019; Skoe and Kraus, 2010). Each individual vowel was 100 ms with a common voice fundamental frequency (F0 = 150 Hz). The first and second formants were 730, 270, 300 Hz (F1) and 1090, 2290, 870 Hz (F2) for /a/, /i/, and /u/, respectively. Notably, the F0 of these stimuli is above the phase-locking limit of cortical neurons and observable FFRs in cortex (Bidelman, 2018; Brugge et al., 2009), ensuring our FFRs would be of brainstem origin (Bidelman, 2018; Coffey et al., 2016). The vowel streams were presented in clean (i.e., no background noise) and noise-degraded conditions. For the noise condition, vowel stimuli were mixed with 8-talker noise babble (Bidelman et al., 2019) at a signal-to-noise ratio (SNR) of +5 dB (speech at 75 dB_A SPL and noise at 70 dB_A SPL) to increase task difficulty.

During active blocks, participants detected infrequent /u/ tokens within the continuous vowel stream via button press. A hit was defined as detection within 5 consecutive tokens following the presentation of a target /u/. For passive blocks, participants watched a captioned (i.e., without sound), self-selected movie and were instructed to ignore any sounds they heard via headphones (i.e., which presented the primary vowel stimuli). This passive task has been shown to maintain arousal without impeding auditory processing (Pettigrew et al., 2004). As demonstrated in Price and Bidelman (2021), because the video was not time locked to the auditory task stimuli, it did not generate a visual ERP and thus only acts to modulate attention to the primary auditory stimuli.

Each block (e.g., clean active, noise active, clean passive, noise passive) was split into 2 runs to allow participants a break halfway through each block. A single run lasted approximately 7.75 min and contained 1000 presentations of each frequent token (/a/, /i/) and 70 trials of the infrequent /u/ token resulting in a total of 2070 trials per run (random order; jittered interstimulus = 95–155 ms, 5 ms steps, uniform distribution; rarefaction polarity) [cf. (Shiga et al., 2015)]. Par-

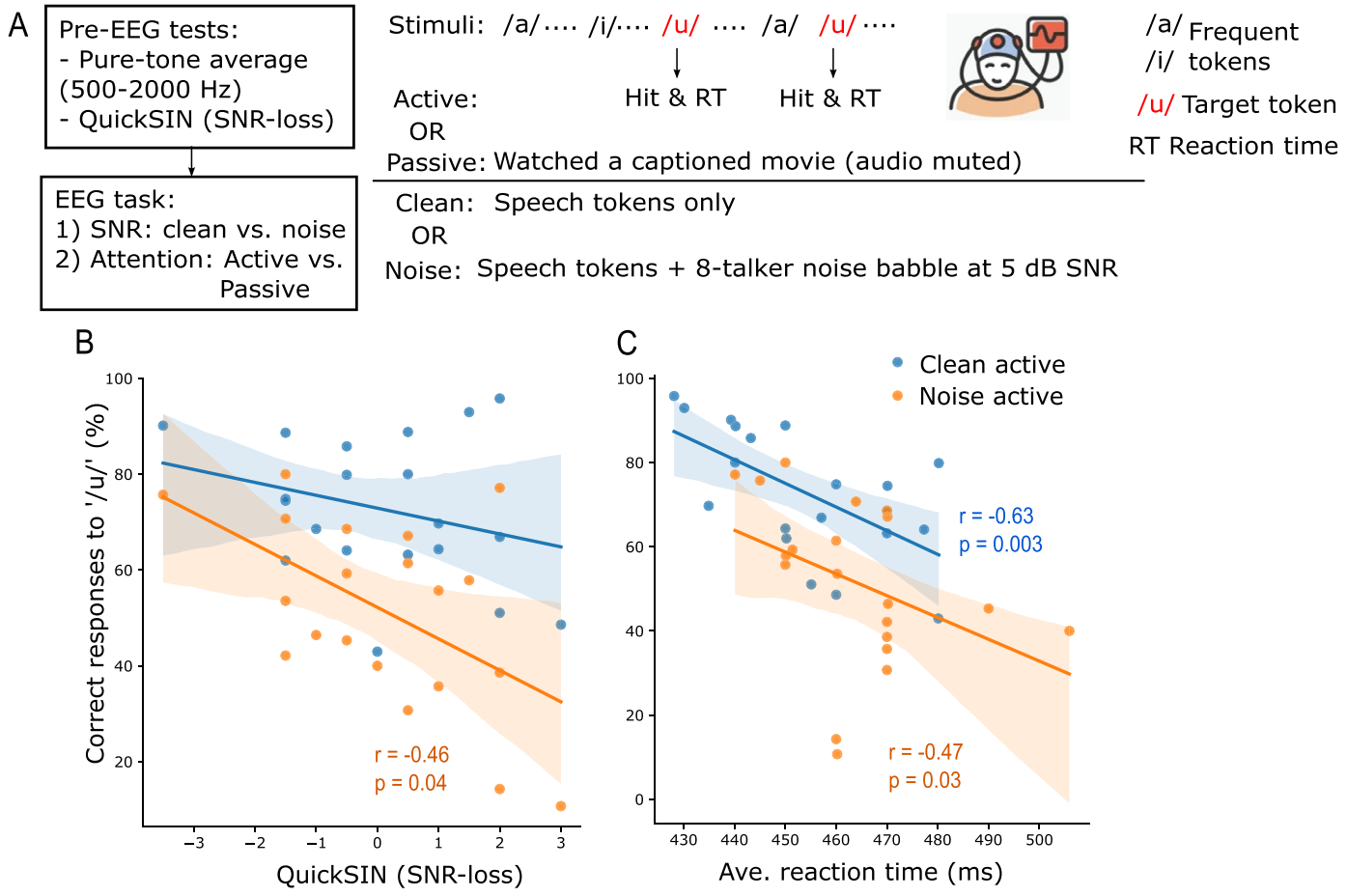

Fig. 1. EEG task performance correlates with normative measures of SIN perception (i.e., QuickSIN scores). (A) Prior to EEG recordings, all participants' pure-tone averages at 500–2000 Hz were obtained and speech-in-noise perception was assessed with QuickSIN. Subsequently, speech-EEGs were recorded under clean or noisy (+5 dB SNR) backgrounds and in active (/u/ detection task) or passive (no task) conditions. (B) QuickSIN scores, in which speech sentences were tested in noise, predict performance during the EEG task for noise-degraded (but not clean) speech. (C) In both clean and noise active sessions, speech detection accuracy predicts reaction times obtained during EEG tasks. r = Spearman's correlation; shaded area = 95% CI of the regression line.

ticipants completed two runs of each attentional block and noise condition (e.g., active clean, active noise, passive clean, passive noise). Block (active/passive) and condition (clean/noisy speech) presentation were counterbalanced across participants to minimize order and fatigue effects (breaks were provided between blocks).Total run time of the EEG recording including breaks was ~70 min (=7.75 min * 2 runs * 4 stimulus blocks + breaks). There were 140 possible /u/ trials that could be logged as hits. On average, participants detected 103.2 ±21.5 (clean) and 72.1 ± 27.5 (noise) targets that were submitted for RT analysis.

Stimulus presentation was controlled by MATLAB (The Mathworks, Inc.; Natick, MA) routed to a TDT RP2 interface (Tucker-Davis Technologies; Alachua, FL) and delivered binaurally through electromagnetically shielded (Campbell et al., 2012) insert earphones (ER-2; Etymotic Research; Elk Grove Village, IL). Phantom bench tests confirmed this shielding fully eliminated stimulus electromagnetic artifact from contaminating EEG recordings (Price and Bidelman, 2021).

2.3. QuickSIN test

We used the Quick Speech-in-Noise (QuickSIN) test to assess listeners' speech perception in noise (Killion et al., 2004). Listeners heard 6 sentences, each with 5 target keywords spoken by a female talker embedded in four-talker babble noise. Target sentences were presented at 70 dB SPL (binaurally) at SNRs decreasing in 5 dB steps from 25 dB (relatively easy) to 0 dB (relatively difficult). SNR-loss scores reflect the difference between a participant's SNR-50 (i.e., SNR required for 50% keyword recall) and the average SNR threshold for normal hearing adults (i.e., 2 dB) (Killion et al., 2004). Higher scores indicate poorer SIN performance. Participants' final SNR-loss was computed from two repetitions of the QuickSIN test (separate sentence lists). Participants' scores in this study ranged from -3.5 to +3 dB of SNR loss (M = 0.1, SD = 1.6), consistent with normal hearing.

2.4. EEG acquisition and preprocessing

EEGs were recorded from 64-channels at 10-10 electrode locations across the scalp (Oostenveld and Praamstra, 2001). Electrodes on the outer canthi and superior/inferior orbit monitored ocular artifacts. Impedances were ≤ 5 kΩ. EEGs were digitized at a high sample rate (5 kHz; DC - 2000 Hz online filters; SynAmps RT amplifiers; Compumedics Neuroscan; Charlotte, NC) to recover both fast (FFR) and slow (cortical) frequency components of the compound potential (Bidelman et al., 2013; Musacchia et al., 2008).

EEG data were processed using Python 3.9.7. For sensor (channel-level) analyses, the data were re-referenced to the mastoid electrodes (M1 and M2). A common average reference was used for subsequent source FFR analysis. Ocular artifacts (saccades and blinks) were automatically reduced within the continuous EEGs using independent component analysis (ICA; method='fastica') (Flexer et al., 2005; Hyvarinen, 1999) based on which ICs best matched the pattern in the electrooculogram (EOG). Responses were then filtered from 130 to 1500 Hz [finite impulse response (FIR) filters; hamming window with 0.02 dB passband

ripple and 53 dB stopband attenuation] to isolate brainstem activity from the EEG (Bidelman et al., 2013; Musacchia et al., 2008).

2.5. Source reconstructed FFRs

We transformed the cleaned sensor data (64-ch) into source space using a virtual source montage (Scherg et al., 2002). The source montage was comprised of a single regional dipole (i.e., current flow in x, y, z planes) positioned in the brainstem midbrain (i.e., inferior colliculus) [for details see Bidelman (2018), Price and Bidelman (2021) and Bidelman and Momtaz (2021)]. Source current waveforms (SWF) from the brainstem dipole were achieved via matrix multiplication of the sensor data (FFR waveform) with the brainstem dipole leadfield (L) matrix (i.e., SWF = L^{-1} x FFR). This applied an optimized spatial filter to all electrodes that calculated their weighted contribution to the scalp-recorded FFRs in order to estimate source activity within the midbrain (Scherg and Ebersole, 1994; Scherg et al., 2002). This model explained >90% of the scalp-recorded FFR (Price and Bidelman, 2021). We used only the z-oriented source activity given the predominantly vertical orientation of current flow in the auditory midbrain pathways relative to the scalp [x- and y-orientations contribute little to the FFR; Bidelman (2018)].

2.6. Cortical EEG activities

Cortical α band activity was also extracted from the EEG and used as a running index of arousal state (high or low) during the speech listening task. α was quantified from the POz channel given the posterior scalp topography of α activity. To isolate α, we filtered recordings at 8–12 Hz (FIR filters). Filtered α activities were epoched with a time window of 195 ms (-50 to 145 ms in which 0 ms corresponded to the onset of an /a/ or /i/ token) to capture approximately 1–2 cycles of α band. Infrequent /u/ tokens were excluded from analysis due to their limited trials. We then measured the root mean square (RMS) amplitude of single trial α activity to quantify cortical arousal level over the duration of the perceptual task. RMS values were then normalized to the median of all RMS values of each run. Next, we visualized the distribution of trial-wise normalized α RMS via a histogram. Trials falling within the 0–25th percentile were categorized as low α power while those falling within the 75–100th percentile were categorized as high α power. This stratification resulted in ~1000 trials each of low- or high-α power per task condition. We similarly measured cortical activity in different frequency bands (e.g., β band; 18–22 Hz) and electrode sites (Fz channel) as negative control analyses. β band is usually dominant at frontocentral scalp areas and is distal from our primary site for α quantification at the posterior (POz). Analyzing POz β and Fz α helped to rule out any unspecific/general effects of high or low EEG activity on FFRs.

2.7. Brainstem FFRs

We then derived FFRs according to the trial-by-trial cortical state, categorizing source FFRs based on whether α amplitude in the same epoch was either high vs. low. Source FFRs were then averaged for each α category, stimulus, condition, and subject. We then analyzed the steady-state portion (10 to 100 ms) of FFR waveforms using the FFT (Blackman window; 11.1 Hz frequency resolution) to capture the spectral composition of the response. F0 amplitude was quantified as the peak spectral maximum within an 11 Hz bin centered around 150 Hz (i.e., F0 of the stimuli). FFR F0 indexes voice pitch coding and has been shown to predict successful SIN perception (Mankel and Bidelman, 2018; Parbery-Clark et al., 2009). To compare F0 amplitudes of low- vs. high-α-indexed FFRs, we calculated an F0 ratio for each condition by dividing F0 amplitudes of high-α-indexed FFRs by those of low-α-indexed FFRs.

$$F0\ ratio = \frac{F0amp[high_{\alpha}]}{F0amp[low_{\alpha}]} \quad (1)$$

F0 ratios > 1 indicate brainstem FFRs were stronger during states of high cortical α power. In contrast, F0 ratios < 1 indicate FFRs were stronger during states of low cortical α power.

2.8. Statistical analysis

We used rmANOVAs to compare brainstem F0 ratios among the four conditions (clean active, noise active, clean passive, noise passive). Multiple pairwise comparisons (with Bonferonni corrections) were performed using the 'pingouin' package in Python. One sample t-tests ('scipy' package in Python) were also used to evaluate whether FFR F0 ratios were significantly different from 1 (and thereby showed α modulation). To compare differences in α RMS values of all subjects across (clean vs. noise or active vs. passive) conditions, we performed post-hoc Conover's test ('scikit_posthocs' package in Python), which is a non-parametric pairwise test, with Bonferroni adjustment. To test for differences in raw F0 amplitudes (log-transformed) across factors of our design, we used a 2 x 2 x 2 (POz α-band power x attention x SNR) mixed model ANOVA ('lme4' package in R). Following Bates et al. (2015), we used a backward selection procedure from the "maximal random effects structure" to include correlated random intercepts and slopes on subject for all terms that still allowed the model to converge and was not overparameterized (i.e., singular fit). The final model was log(FFTamp) = snr*atten*Alpha + (atten|sub). Including any other correlated (or uncorrelated) random effects term(s) on subjects [e.g., (snr|sub), (snr*atten|snr), etc.] resulted in an inestimable or overparameterized model. While the random slopes model did improve the model fit relative to an intercept only model from a statistical standpoint [$\chi^2(2)$=10.3, p=0.0058], the improvement was small (Akaike's information criterion: AIC_{slopes}= 261.1 vs. $AIC_{intercept-only}$ = 267.4). This analysis allowed us to compare raw F0 amplitudes of low- vs. high-α-indexed FFRs, clean vs. noise, and passive vs. active. Initial diagnostics were performed using residual and Q–Q plots to assess heteroscedasticity and normality of data. F0 amplitudes were log-transformed to improve normality and homogeneity of variance assumptions. Effect sizes are reported as η_p^2. We used Spearman's correlations ('scipy' package in Python) to assess pairwise linear relations between neural and behavioral measures.

2.9. Decoding speech tokens from FFRs via machine learning

We used a linear SVM algorithm (Cristianini and Shawe-Taylor, 2000) to determine if the stimulus speech identity (i.e., /a/ vs. /i/ token) could be classified via low- and high-α-indexed FFRs. We focused on the noise active condition for this decoding analysis following procedures described by Xie et al. (2019). The epoched FFRs to /a/ or /i/ tokens were averaged, respectively, across ~500 trials for each of the low- or high-α-indexed FFRs. These average FFR waveforms (10–100 ms steady state portion) were used as input features for SVM. Note this contrasts our F0 analyses in that the entire FFR spectrum was used for token-wise decoding. A total of 450 amplitude-by-time points were input as predictors; vowel types (i.e., /a/ and /i/) served as the ground truth class labels. During one iteration of training and testing, a four-fold cross-validation approach was used to train and evaluate the performance of the linear SVM classifier to obtain a mean decoding accuracy [see Fig. 1 of Xie et al. (2019)]. In this process, subjects were randomly and equally divided into 4 subgroups with 5 unique subjects in each subgroup. Three of the 4 subgroups were selected as the training data while the remaining subgroup was used as the hold-out testing data. This was repeated within each iteration so that each subgroup was held-out as the test data whereas the other 3 subgroups were used to train the SVM classifier. Mean decoding accuracy (for distinguishing vowel tokens from FFRs) was calculated across cross-validated iterations. We performed a total of N=5000 iterations for each low and high α power subset of the data.

To evaluate if the classifier accuracy (mean of N=5000 iterations) was statistically significant, ground truths (i.e., vowel types) were randomly assigned to FFR inputs, and the same training and testing procedures described above were repeated to derive a null distribution of decoding accuracies. We then calculated the p-value to determine the statistical significance of "true" classifier performance using the formula described in Phipson and Smyth (2010):

$$p = \frac{(a+1)}{(n+1)} \tag{2}$$

Where a is the number of decoding accuracies from the null distribution that exceeds the median of the actual distribution of decoding accuracies and n is the total number of decoding accuracies from the null distribution. The same *p*-value calculated using the above equation was also used to compare the prediction performance of the two "true" classifiers (low- vs. high-α-FFRs).

3. Results

3.1. Behavioral speech-in-noise performance

Speech perception during the EEG task was strongly associated with QuickSIN scores (tested prior to EEG) for noise-degraded but not clean speech (Spearman's r = -0.46, $p = 0.04$, Fig. 1B). This confirms the external validity of our laboratory-based task in assessing SIN perception. Both perceptual outcomes were conducted in noisy backgrounds and under active listening conditions, suggesting token-wise vowel detection is a good proxy for sentence-level SIN perception (at least for low-context sentences like the QuickSIN). Behavioral hit responses (percent correct /u/ detections) were also negatively associated with response speeds in the EEG task regardless of SNR (clean active: r = -0.63, $p = 0.003$; noise active: r= -0.47, $p = 0.03$); slower decisions were associated with poorer accuracy (Fig. 1C).

3.2. Categorizing brainstem FFRs according to online cortical arousal state indexed by EEG α power

To determine if brainstem speech-FFRs are yoked to internal cortical arousal state, we measured trial-by-trial changes in EEG α band activity (~10 Hz) to track low and high arousal states. At the same time, trialwise FFR F0 amplitudes were measured to index the magnitude of speech (i.e., voice-pitch) encoding at the brainstem level (Assmann, 1998; Mankel and Bidelman, 2018). FFRs were analyzed at the source level (see Methods) to unmix putative cortical contributions to the FFR (Bidelman, 2018; Coffey et al., 2016). We isolated cortical α band from the posterior POz channel, where waking α activity is maximal at the scalp (Klatt et al., 2020; Moini and Piran, 2020; Nunez, 2016). Unlike α band, β band is more prominent over frontcentral compared to posterior regions of cortex (Kropotov, 2009). To rule out any unspecific or general effects of cortical activity on FFRs, in addition to α band extracted from the POz channel (termed POz α), we also isolated β band from POz (termed POz β) and α from Fz (termed Fz α) to use as negative control analyses.

Continuous POz α responses were epoched with a time window of -50 to 145 ms with respect to vowel onset time (i.e., the same analysis window as FFRs). We then computed the root-mean-square (RMS) power within each epoch. This resulted in approximately 4000 trials of POz α RMS values (Fig. 2A) corresponding to the 4000 frequent speech tokens in the task (i.e., /a/ and /i/). All RMS values were normalized to the RMS median of each run and visualized using a histogram (Fig. 2B). Subsequently, RMS values falling within the range of 0–25th percentile were categorized as low α power while RMS values falling within 75–100th percentile were considered as high α power. This resulted in ~1000 trials in each of the lower and upper percentile ranges. Trials outside these ranges were not analyzed because we wanted to maximize differences between low- and high-α-indexed FFRs.

Speech-evoked source FFRs were then categorized based on their corresponding α power in the same epoch window resulting in a set of low- vs. high-α-indexed FFRs. FFR time waveforms and spectra for a representative subject are shown in Fig. 2C and D for low- and high-α states. Corresponding average α responses are shown in Fig. 2E. The difference in activity level of low vs. high α power is apparent. RMS values of low and high α across all subjects for each of the four conditions is shown in Fig. 2F, with statistically significant differences ($p < 0.01$) observed between some conditions of clean vs. noise or active vs. passive RMS values (non-parametric post-hoc Conover's test with Bonferroni adjustment). The same process of categorizing the level of cortical bands and FFRs into low or high level was repeated on POz β and Fz α as well. These negative controls allowed us to compare the subsequent observations from POZ α to POZ β and Fz α, to ensure the observed changes in speech-evoked FFRs were specifically associated with cortical arousal level (indexed by α power) rather than general fluctuations in the EEG, *per se*.

3.3. Speech FFRs depend on SNR, attention, and cortical α power

Grand averaged FFR spectra for low- and high-α-indexed FFRs across conditions are shown in Fig. 3A–D. In noise, FFR F0 amplitudes during low α power were lower than F0 amplitudes during high α power, especially in the passive condition where low vs. high α F0 responses were significantly different ($t = 2.18$, $p = 0.04$). We computed a normalized (within-subject) measure of α FFR enhancement by dividing F0 amplitude of high-α-indexed FFRs by F0 amplitudes of low-α-indexed FFRs ("F0 ratio"). F0 ratios in noise were higher than in clean backgrounds (rmANOVA, F = 4.84, p = 0.04, $\eta_p^2 = 0.20$) (Fig. 3E). In contrast, responses did not differ in the negative control (POZ β and FZ α) analyses, suggesting FFR modulations were specific to the cortical α band (indexing arousal) and not general properties of the EEG, per se (Fig. A.6). A mixed-model ANOVA on log F0 amplitudes also revealed significant main effects of SNR (clean vs. noise), attention (active vs. passive), and α power (low vs. high) (see Fig. 3F and Table 1). Random effects of the ANOVA analysis can be found in Table A.2

3.4. Brain-behavior relations

We next assessed associations between behavioral performance in the speech detection task collected during EEG recordings (% correct, RTs) and neural measures. The correlations of these behavioral metrics with FFR F0 amplitudes (per condition) were then tested systematically. Correlations for the clean condition were not significant. In contrast, for noise, we found a trend whereby in states of high cortical α, FFR strength in noise decreased for trials with slower RTs (combining both active and passive conditions, Fig. 4A). When focusing on the noise active condition, FFRs during low α states were strongly associated with RT (Spearman's $r = 0.54$, $p = 0.02$). The brain-behavior relation between RTs and FFR amplitudes also differed as a function of cortical α power (Fisher's $z = 3.13$, $p = 0.001$). This opposite direction in correlations was not observed between noise passive FFRs and RTs (recorded in the active task) (Fig. 4C). Repeating the same correlation analysis for POz β and Fz α in the noise active condition did not reveal any significant association between FFRs and RT (Fig. A.7). This suggests that the observed correlation between FFRs and RT was specifically related to α power in POz channel.

3.5. Token decoding from FFRs during low & high α power

The previous analyses showed low-α-indexed FFR amplitudes were correlated with behavior in the noise active condition. We further asked whether FFRs during low α states are actually better at decoding speech representations (i.e., /a/ vs. /i/ tokens) compared to their high α counterparts. We followed the machine learning (ML)-based approach reported by Xie et al. (2019) to analyze FFRs for token decoding. Two

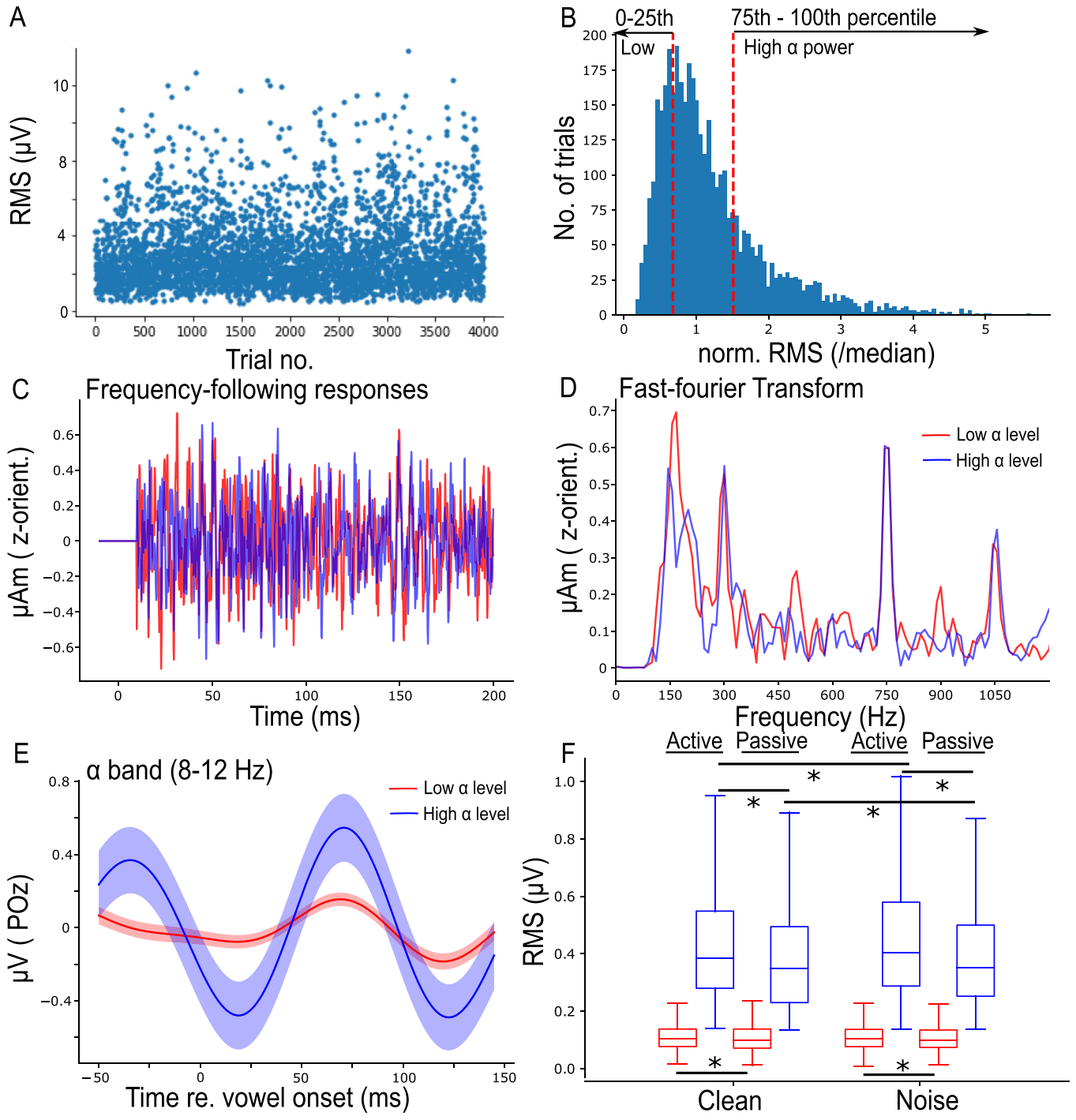

Fig. 2. Examples of data processing (representative subject) to separate FFRs based on cortical arousal state (i.e., αpower). (A) Cortical α band was isolated from the POz channel and epoched at -50-145 ms with respect to vowel stimulus onset time. Root-mean-square (RMS) amplitude was then computed for every epoch (i.e., approximately 4000 single trials) per listener. (B) Histogram of RMS values normalized to the RMS median of each run. Left and right red dashed lines indicate the 25th and 75th percentile, respectively. RMS values falling within the range of 0–25th and 75–100th percentile were categorized as low vs. high α power, respectively, yielding 1000 trials per state. (C) Average source FFRs (onset responses removed) during low (red) vs. high (blue) EEG α power. (D) Frequency spectra of the steady state (10–100 ms) portion of FFR waveforms. (E) Average α waveform of low and high power for the trials extracted from the lower or upper 25th percentiles in (B). (F) Boxplots showing RMS values of low and high α power of all subjects (N=20). Shaded area = ± s.e.m. * p< 0.01 (Conover's test, non-parametric pairwise test, with Bonferroni adjustment)

separate linear support vector machine (SVM) classifiers were trained and tested on FFR waveforms with the goal of decoding the speech stimulus from neural responses. Mean linear SVM classification accuracies using low- or high-α-indexed FFRs are shown as confusion matrices in Fig. 5A. All FFRs yielded classification performance significantly above chance (Fig. 5B), confirming brainstem neural representations closely mirror the spectrotemporal properties of speech. There was more than a 20% increase (78.34 vs 57.77%) in the mean classification accuracy for /a/ tokens when using low-α-FFRs as inputs. The mean classification accuracies for /i/ tokens were at ceiling and exceeded 90% for FFRs

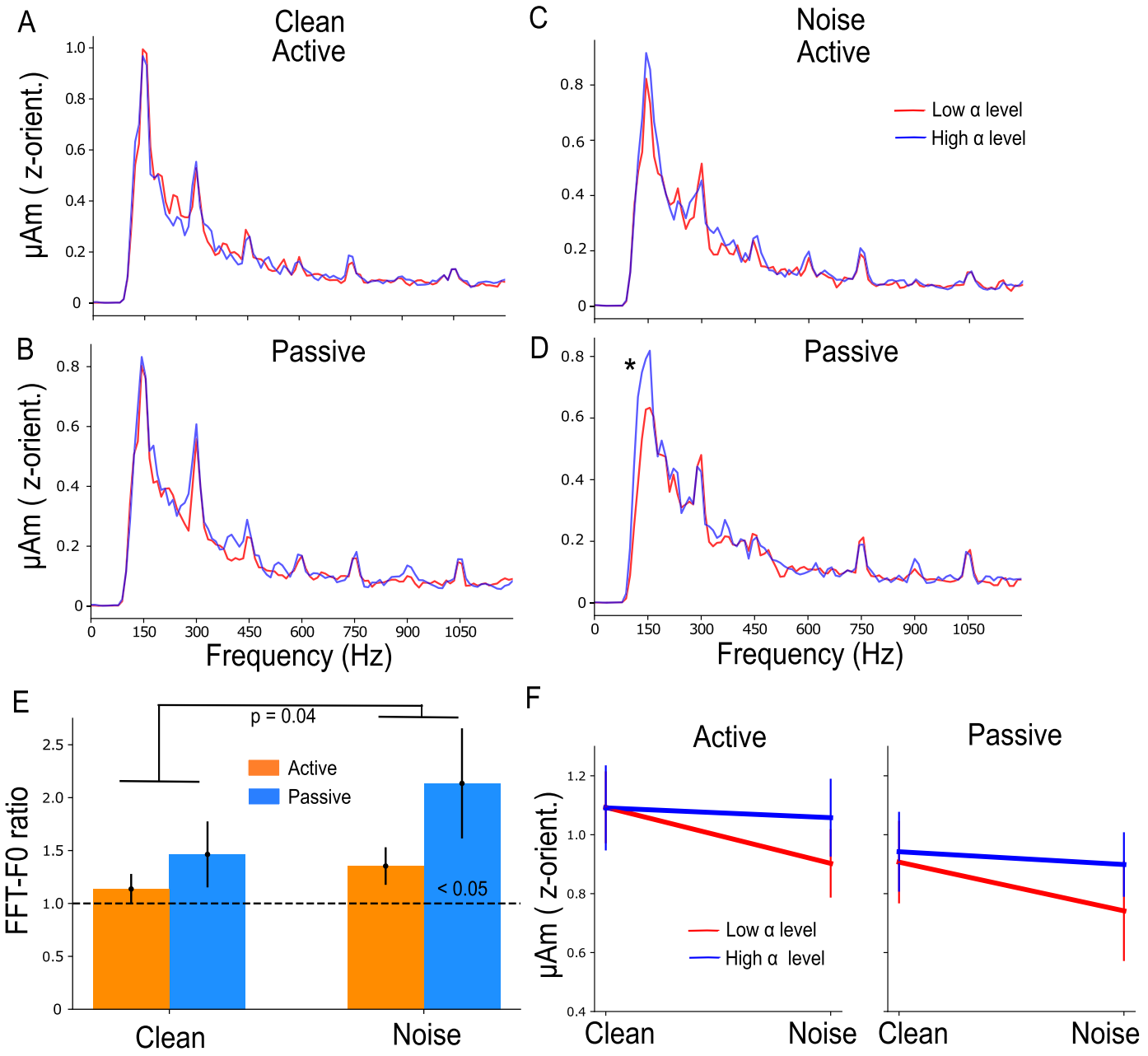

Fig. 3. FFR F0 amplitudes during low α are smaller than during high αin noisy (but not in clean) backgrounds. (A-D) Frequency spectra of speech-evoked FFRs illustrating smaller F0 amplitudes during low α in both active and passive noisy backgrounds (C & D) but not in clean backgrounds (A & B). (E) A noise effect is observed for F0 ratio. Bars marked (<0.05) are significantly larger than 1 (1-sample *t*-test). (F) Raw FFR F0 amplitudes as a function of SNR (clean vs. noise), attention (active vs. passive) and α power (low vs. high). Errorbars= ± s.e.m., *p<0.05.

Table 1
Results of Type III, mixed-model ANOVA on log(FFR F0 amplitudes).

	NumDF	DenDF	*F*-value	*p*-value	η_p^2
SNR	1	140	4.156	0.043 *	0.03
attention	1	140	13.003	<0.001 ***	0.08
α power	1	140	5.027	0.027 *	0.03
SNR x attention	1	140	0.136	0.73	9.71×10^{-4}
SNR x αpower	1	140	2.248	0.136	0.02
attention x αpower	1	140	2.008	0.158	0.01
SNR x attention x αpower	1	140	0.166	0.685	1.18×10^{-3}

* p < 0.05, *** p < 0.001

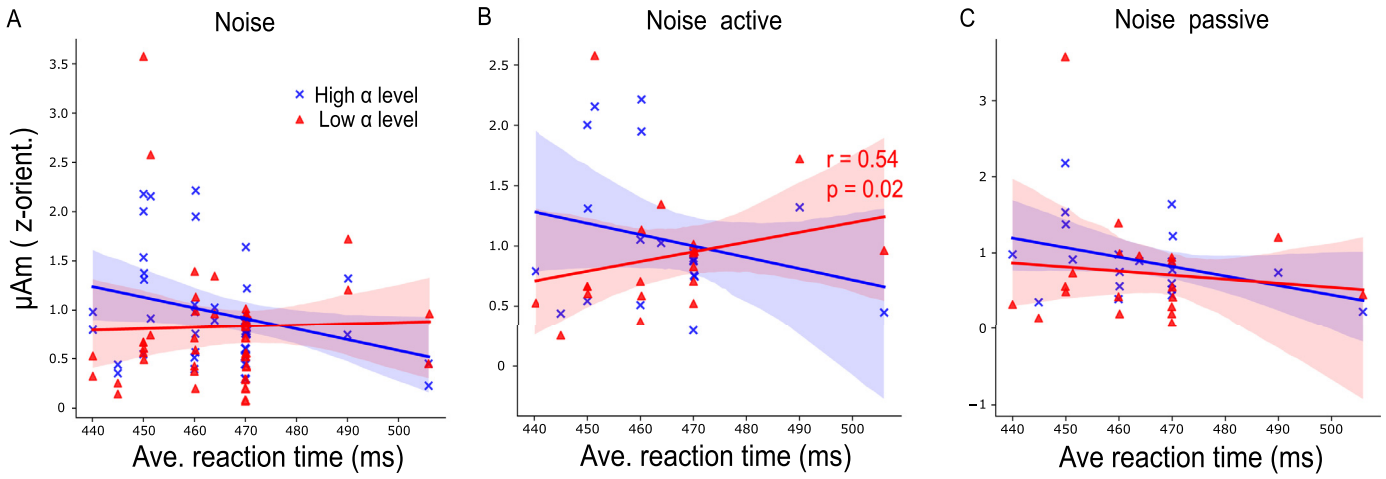

Fig. 4. Brain-behavior associations between FFR and speech perception depend on cortical state. (A) In noise (pooling active and passive conditions), there is a trend of decreased FFR F0 amplitudes during high α power in listeners with slower reaction times. (B-C) When correlations were analyzed separately for noise active (B) and passive (C) conditions, FFRs during low α power correlated with RTs in the active, but not passive, condition. r = Spearman's correlation, shaded area indicates 95% confidence interval of the regression line.

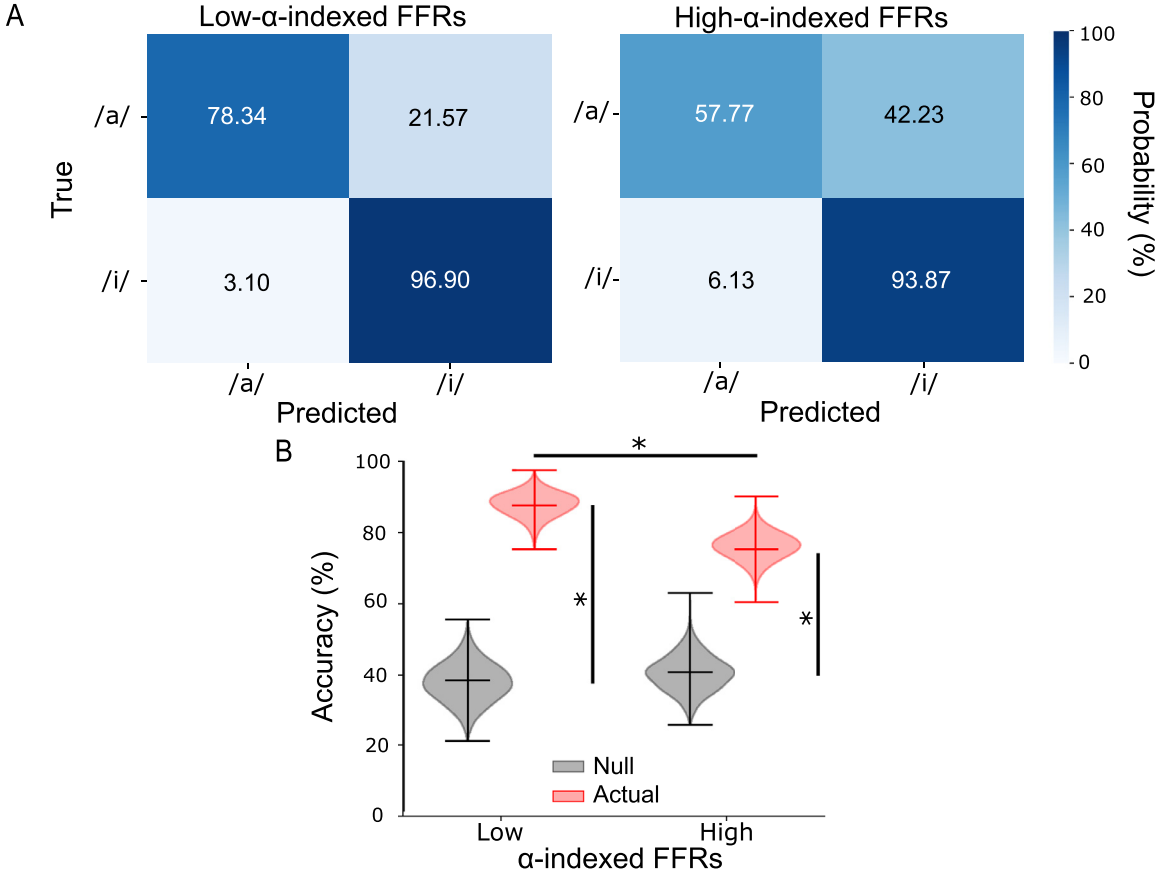

Fig. 5. Average accuracy of linear SVMs in classifying /a/ and /i/ tokens (noise active condition) is better when using low-α-indexed FFRs as inputs. (A) Vowel decoding confusion matrices for low- (left) vs. high (right) α-indexed FFRs (N=5000 bootstrap iterations). (B) Overall distributions of prediction accuracies of SVM classifiers were significantly better using FFRs at low compared to high α power. Distribution of prediction accuracies for FFRs (actual) were also significantly above the null prediction accuracies. Upper/lower ticks=max/min; center tick=medians. * p< 0.001.

at low (96.90%) and high (93.87%) α power. Furthermore, the classifier performance for low α was also significantly better than high α (Fig. 5B). These decoding findings suggest FFR speech representations were of higher fidelity (discriminability) under low compared to high α cortical states. Better vowel decoding may contribute to enhanced token discrimination in noisy backgrounds.

4. Discussion

Our previous study revealed strong attentional enhancements in speech-FFRs at the source level and top-down neural communication from primary auditory cortex to brainstem that aids SIN perception (Price and Bidelman, 2021). Building upon those findings, we show here

the existence of active and dynamic modulation of brainstem speech processing dependent on online changes in listeners' cortical state. We found higher cortical α states positively predict larger speech-evoked FFRs in adverse listening conditions. Furthermore, FFRs during low α power and recorded in the noise active condition predicted behavioral RTs for rapid speech detection and were associated indirectly with other perceptual measures of SIN performance. This positive relationship of FFRs and behavior at low α power were further supported by better token classification using a neural decoding approach. As a whole, these data help reconcile equivocal findings by revealing top-down cortical activity dynamically influences brainstem encoding of speech, which directly imposes constraints on behavior.

Attention effects on brainstem responses have been highly controversial. Some studies reveal attention increases the robustness and temporal precision of speech-evoked FFRs (Forte et al., 2017; Galbraith et al., 2003; Hartmann and Weisz, 2019; Lehmann and Schönwiesner, 2014). However, other studies have observed mixed (Holmes et al., 2018; Saiz-Alía et al., 2019) or no attention effects on FFRs (Galbraith and Kane, 1993; Varghese et al., 2015). Attention effects at the brainstem level could be too subtle to detect in scalp EEG given the inherent mixing of intracranial sources (Price and Bidelman, 2021; Vollmer et al., 2017) and/or require more difficult perceptual tasks (e.g., SIN perception) that challenge speech processing (Galbraith and Arroyo, 1993; Lehmann and Schönwiesner, 2014). In this study, we observed overall enhanced FFRs regardless of the background SNR when subjects actively engaged in challenging speech listening tasks (Fig. 3F, left panel). However, FFR changes due to cortical α were more prominent in noise than clean conditions (Fig. 3E-F). Similarly, even under passive listening conditions, these cortical dependencies in the FFR were stronger under states of low α and for noisy speech (Fig. 3E & right panel of F). These findings suggest that cortical influences like attention (Price and Bidelman, 2021) and internal arousal state (present study) on the FFR are more prominent in adverse listening conditions.

Elevated α oscillations were initially thought to reflect states of wakefulness but without active engagement in tasks (Pfurtscheller et al., 1996). Hence, in this study, we interpret low α power to indicate subjects were in a high arousal state and high α power to index task focus but in a state of wakeful relaxation. We attempted to tease apart arousal and attention by using α power to index ongoing arousal and active trials to index attention. In both active and passive noise conditions, we observed larger speech-evoked FFRs during high α power (i.e., subjects in a relaxed, low arousal state). Our observations do not coincide with the findings of Mai et al. (2019) in which they reported higher phase-locked responses to speech at sub-cortical levels were associated with high arousal state. However, that study recorded EEGs during sleep and not during active tasks, and FFRs were separated according to sleep spindles to index subjects' arousal states rather than online α as done here. These differences may contribute to the discrepancy between studies. Moreover, Makov et al. (2017) reported greater phase-locked FFRs in wakefulness compared to sleep. Although stimuli are processed by the brain during sleep (Issa and Wang, 2008; Nir et al., 2015), neural responses to speech in the auditory subcortex reduce during sleep compared to wakefulness (Portas et al., 2000). Since EEG recordings were conducted in subjects during wakefulness in this study, our results may suggest that subcortical auditory processing in low vs. high arousal states are different in wake vs. sleep states.

In addition, α oscillations play a significant role in functionally inhibiting the processing of task-irrelevant information (Foxe and Snyder, 2011; Jensen and Mazaheri, 2010). Both α power and phase were found to modulate neuronal spike rate (Haegens et al., 2011) and thus can directly affect the efficiency of neural information flow. For example, Wilsch et al. (2015) demonstrated increased α power when participants listened to stimuli presented in noisy backgrounds. Induced α activity is crucial for speech processing in challenging listening conditions as it suppresses irrelevant information (Strauß et al., 2014). Here, we demonstrate that overall α power (both low and high α) is significantly higher ($p < 0.01$, Conover's test) in active than passive speech listening conditions and the mean RMS values of high α power is larger in the noise compared to the clean active condition (Fig. 2F). Another interesting and crucial point regarding α activity is that it fluctuates in amplitude with stimulus and task demands (Klimesch, 2012). Increased α may reflect inhibition to de-prioritize distractors while decreased α might reflect a release from inhibition to prioritize targets (Klatt et al., 2020; Klimesch, 2012). Furthermore, decreased α power has been associated with increased neural firing (Haegens et al., 2011) and improved behavioral performance (Gould et al., 2011; Haegens et al., 2011; Kelly et al., 2009).

In this vein, we observed listeners with smaller amplitudes of low-α-indexed FFRs having shorter behavioral RTs in the noise active condition (Fig. 4B). This highlights a relationship between low α power and behavioral performance. In the vowel identification task in noise, subjects were required to press a button whenever /u/ was presented. Hence, /u/ is a task-relevant target, background noise is task-irrelevant information while /a/ and /i/ can be considered as task-relevant distractors. Although we were unable to measure FFRs to /u/ due to insufficient token counts (only 140 trials in total), we speculate there would be increased neural firing to /u/ (task-relevant target) but lower neural responses to /a/ and /i/ (task-relevant distractors) during low α power. In addition, attention might be maintained during active trials and directed towards the target /u/ stimuli. However, neural responses to non-target stimuli (/a/ and /i/) might become lower especially when participants became more familiarized with the speech detection task. On the other hand, task-irrelevant background noise was inhibited during high α power but not /a/ and /i/ since they are task-relevant stimuli. These two mechanisms provide a possible explanation for our observations in the noise active condition where subjects with faster RTs had larger high-α-indexed FFR amplitudes and smaller low-α-indexed FFR amplitudes (Fig. 4B). Reversing these phenomena resulted in slower RTs when participants were actively identifying stimulus targets in noise. A recent study showed that α does not predict behavioral performance in middle-aged and older listeners (Tune et al., 2021). This discordance could be related to the age differences of participants. Moreover, α activity and behavioral performance were tracked and linked trial-by-trial in that study. However, in the present study, we looked at the relationship of overall behavior performance and α-yoked FFRs.

At this stage, we have evidence showing FFR F0 amplitudes during low and high α power were associated with RT in the noise active condition. Moreover, RT was found to be strongly correlated to the percent of correct responses to /u/ (Fig. 1C) and percent of correct responses to /u/ obtained in noise was strongly predictive of QuickSIN scores (Fig. 1B). Collectively, these results show that FFRs during low α power may have a link with other perceptual performance (e.g., percent of correct responses to /u/ and QuickSIN) tested in noise (Fig. A.8). However, further study is required to investigate if this link with behavior is truly valid.

To further assess the behavioral relevance of α-FFR modulations to SIN listening, we investigated vowel decoding of FFRs at low vs. high α states. Similar ML approaches have been applied to speech-evoked FFRs to decode stimulus classes, e.g., Mandarin lexical tones (Llanos et al., 2017; Xie et al., 2019) and speech tokens (Cheng et al., 2021; Xie et al., 2017). In these studies, decoding performance in correctly classifying FFRs is used as an objective measure of speech discrimination. Moreover, Xie et al. (2017) and Cheng et al. (2021) demonstrated that training and attention can improve FFR classification. The logic behind these ML-based approaches is that attention or training-related plasticity on auditory neural responses produce direct changes in the accuracy with which FFRs are classified. An increase in classification accuracy reflects positive modulation (i.e., enhanced or more robust neural representations to stimuli) while a decrease in accuracy implies negative modulation (i.e., reduced or less robust neural responses). The enhanced classification accuracies for /a/ and /i/ tokens we find under low-α EEG suggests speech representations are of higher fidelity under

high arousal states. This higher fidelity could result from better response SNR or more consistent responses across subjects in the low-α state. This may also account for why brain-behavior relations between FFRs and RTs are mainly observed during low α states (Fig. 4B).

More broadly, difficulties in SIN perception (Duquesnoy, 1983; Fostick et al., 2013; Helfer and Wilber, 1990) and changes in brainstem auditory processing in the presence of noise maskers have been reported in older listeners, even when stimulus intensities are matched for audibility (Lai and Bartlett, 2018). Meanwhile, some previous studies found age has an effect on brain oscillatory activity of α band (Böttger et al., 2002; Klimesch, 1999; Yordanova et al., 1998). Therefore, α oscillations in demanding and challenging listening tasks might be used as an indicator of age-dependent auditory cognitive effort of noise inhibition. Investigating α power and SIN perception in older adults may reveal how aging impacts top-down attentional control to facilitate processing of task-relevant target sounds and inhibit processing of task-irrelevant distractors or maskers. This could partly explain why older listeners have difficulties participating in cocktail party-like listening situations compared with younger listeners (Pichora-Fuller, 2003).

5. Conclusion

Our novel findings reveal that human brainstem FFRs to speech are dynamically modulated by cortical activity (indexed by α power) during speech-in-noise perception. FFRs are thus yoked online to shifts in internal arousal state. Our results also show FFRs during low α (i.e., high arousal state) have more "decodable" speech representations and are more predictive of behavioral SIN performance. Our novel paradigm and single-trial analysis of the FFR help resolve ongoing debates regarding attentional and/or arousal influences on brainstem auditory encoding and corticofugal engagement during active SIN perception in humans. Our data also elucidate possible mechanisms concerning the function of cortical α during SIN recognition whereby sound targets are actively segregated from task-relevant distractors and task-irrelevant background noise to improve perceptual outcomes.

Data and code availability statement

The data supporting the reported findings are available from the corresponding author upon reasonable request.

Ethics statement

All participants provided written informed consent prior to participation in accordance with protocols approved by the University of Memphis IRB.

Declaration of Competing Interest

None.

Credit authorship contribution statement

Jesyin Lai: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Caitlin N. Price:** Formal analysis, Methodology, Writing – review & editing. **Gavin M. Bidelman:** Conceptualization, Data curation, Methodology, Project administration, Writing – original draft, Writing – review & editing.

Acknowledgments

This work was supported by the National Institutes of Health (NIH/NIDCD R01DC016267) (G.M.B.).

Appendix

Table A1
Results of random effects.

Groups	Name	Std. dev.	Corr.
Sub	(Intercept)	0.3368	
	attention passive	0.3505	0.07
Residual		0.4275	

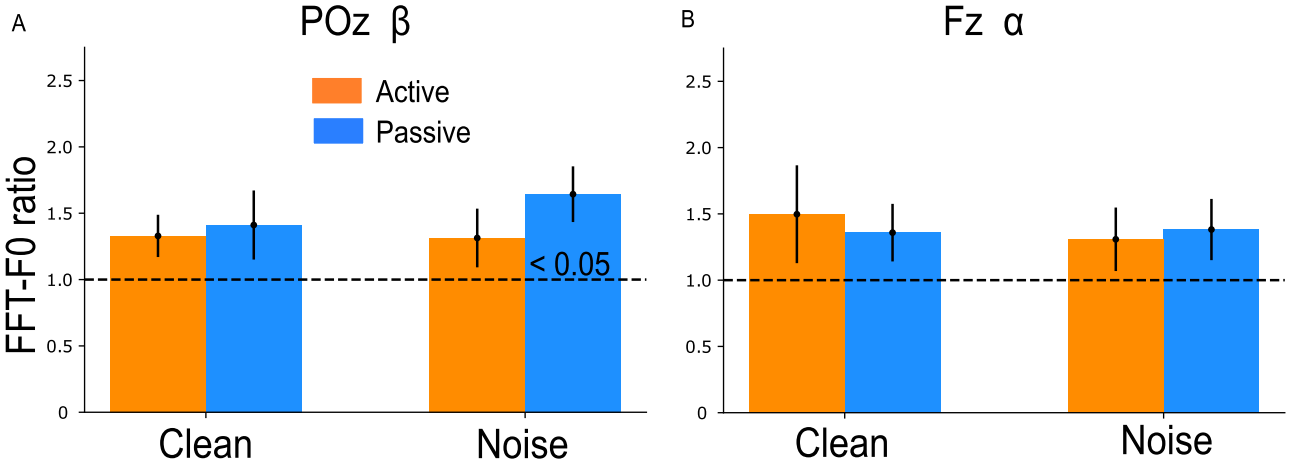

Fig. A1. These control analyses showed no significant dependence of the FFR (as in Fig. 3E but for POz β band and Fz α band). No significant noise effect is observed in F0 ratio when comparing F0 amplitudes of high to low β in POz channel (A) or α in Fz channel (B). <0.05 in bars indicates significantly larger than 1 (1-sample *t*-test).

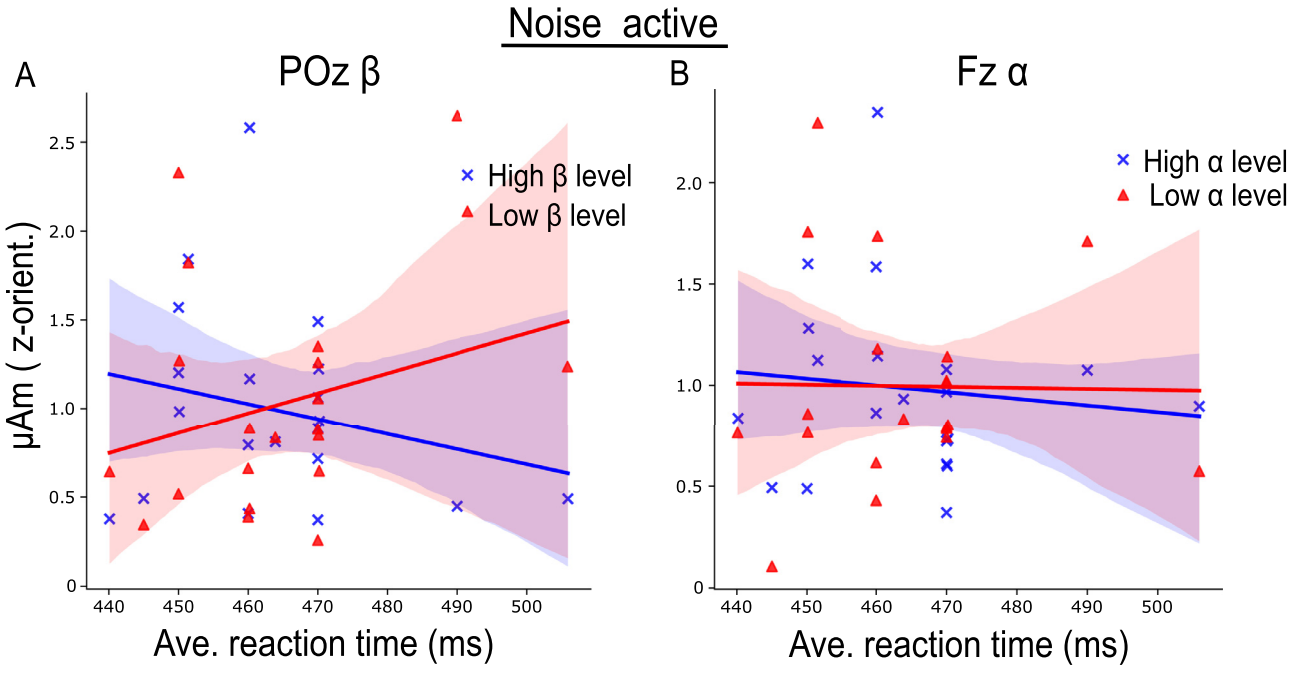

Fig. A2. Repeating the same correlation analysis with a different EEG frequency band (β) and electrode (Fz) produced no significant result. No significant association is observed between F0 amplitudes of FFRs at low (red line) or high (blue line) cortical activity level and reaction time when analyzing β band from POz channel (A) or α band from Fz channel (B). Shaded area indicates 95% confidence intervals of the regression line.

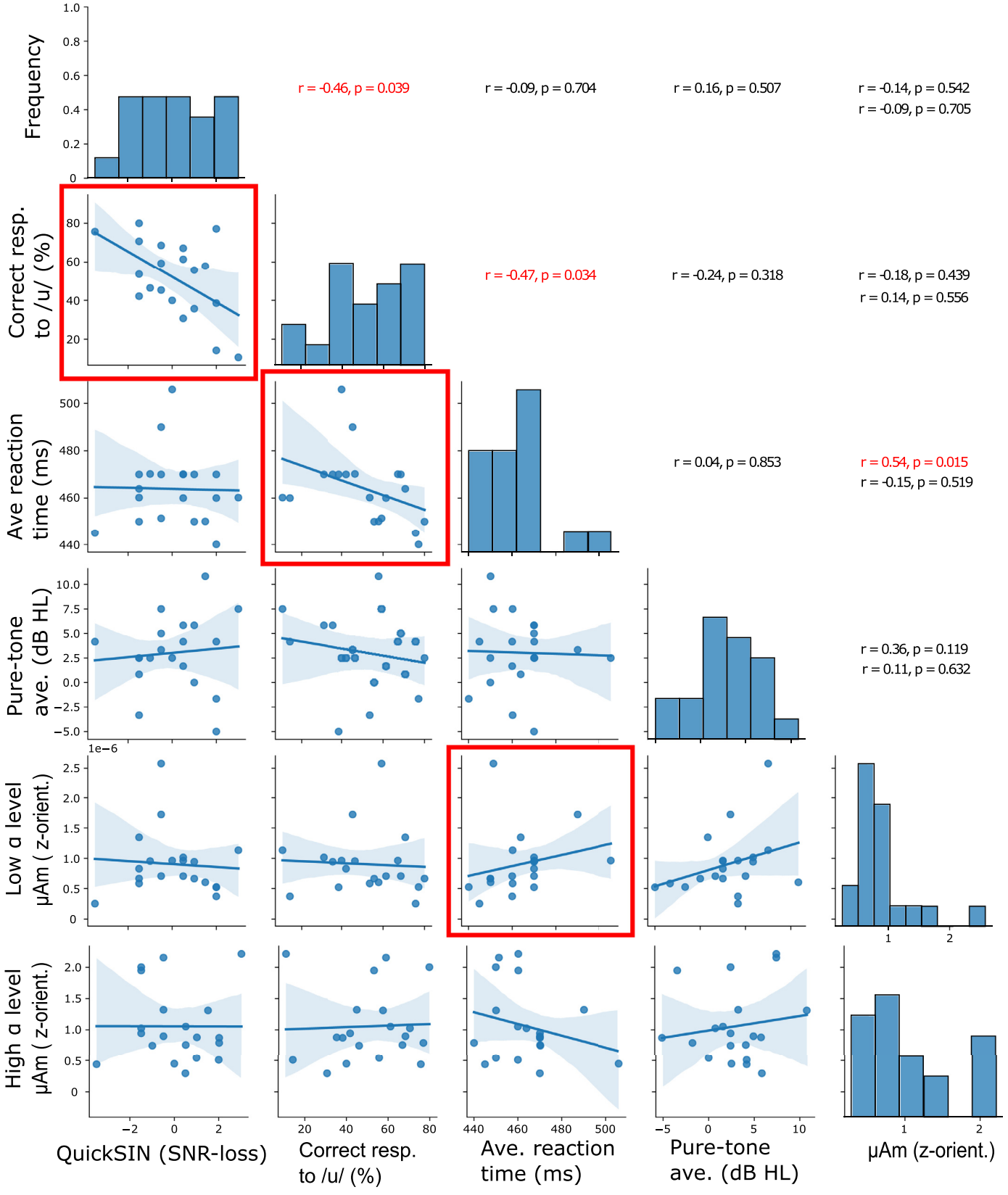

Fig. A3. Summary of all correlations between pre-test perceptual performance vs. EEG behavioral task and EEG behavioral task vs. FFR F0 amplitudes (low or high αpower). Correlations that are statistically significant ($p < 0.05$) are highlighted with red square boxes, and the corresponding *r* and *p*-values are shown in red text. The text in the last column show the *r* and *p*-values for low (upper) and high (lower) α, respectively. The distribution of the behavioral and brainstem response variables are shown along the diagonal. r = Spearman's correlation, and shaded area indicates 95% confidence intervals of the regression line.

References

Aftanas, L.I., Varlamov, A.A., Pavlov, S.V., Makhnev, V.P., Reva, N.V., 2002. Time-dependent cortical asymmetries induced by emotional arousal: EEG analysis of event-related synchronization and desynchronization in individually defined frequency bands. Int. J. Psychophysiol. 44 (1), 67–82. doi:10.1016/S0167-8760(01)00194-5.

Alho, K., Rinne, T., Herron, T.J., Woods, D.L., 2014. Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. Hear. Res. 307, 29–41. doi:10.1016/J.HEARES.2013.08.001.

Alho, K., Vorobyev, V.A., 2007. Brain activity during selective listening to natural speech. Front. Biosci. 12 (8), 3167–3176. doi:10.2741/2304.

Assmann, P.F., 1998. Tracking and glimpsing speech in noise: role of fundamental frequency. J. Acoust. Soc. Am. 100 (4), 2680. doi:10.1121/1.416961.

Atiani, S., Elhilali, M., David, S.V., Fritz, J.B., Shamma, S.A., 2009. Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. Neuron 61 (3), 467–480. doi:10.1016/J.NEURON.2008.12.027.

Bajo, V.M., Moore, D.R., 2005. Descending projections from the auditory cortex to the inferior colliculus in the gerbil, meriones unguiculatus. J. Comp. Neurol. 486 (2), 101–116. doi:10.1002/CNE.20542.

Bates, D., Kliegl, R., Vasishth, S., Baayen, R. H., 2015. Parsimonious mixed models 10.48550/arxiv.1506.04967.

Beyerl, B.D., 1978. Afferent projections to the central nucleus of the inferior colliculus in the rat. Brain Res. 145 (2), 209–223. doi:10.1016/0006-8993(78)90858-2.

Bidelman, G.M., 2015. Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. Hear. Res. 323, 68–80. doi:10.1016/J.HEARES.2015.01.011.

Bidelman, G.M., 2018. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. NeuroImage 175, 56–69. doi:10.1016/J.NEUROIMAGE.2018.03.060.

Bidelman, G. M., Momtaz, S., 2021. Subcortical rather than cortical sources of the frequency-following response (FFR) relate to speech-in-noise perception in normal-hearing listeners 10.1016/j.neulet.2021.135664.

Bidelman, G.M., Moreno, S., Alain, C., 2013. Tracing the emergence of categorical speech perception in the human auditory system. NeuroImage 79, 201–212. doi:10.1016/J.NEUROIMAGE.2013.04.093.

Bidelman, G.M., Price, C.N., Shen, D., Arnott, S.R., Alain, C., 2019. Afferent-efferent connectivity between auditory brainstem and cortex accounts for poorer speech-in-noise comprehension in older adults. Hear. Res. 382, 107795. doi:10.1016/J.HEARES.2019.107795.

Billings, C.J., Gordon, S.Y., McMillan, G.P., Gallun, F.J., Molis, M.R., Konrad-Martin, D., 2020. Noise-induced enhancement of envelope following responses in normal-hearing adults. J. Acoust. Soc. Am. 147 (2), EL201. doi:10.1121/10.0000627.

Binkhamis, G., Léger, A., Bell, S.L., Prendergast, G., O'Driscoll, M., Kluk, K., 2019. Speech auditory brainstem responses: effects of background, stimulus duration, consonantvowel, and number of epochs. Ear Hear. 40 (3), 659. doi:10.1097/AUD.0000000000000648.

Böttger, D., Herrmann, C.S., Von Cramon, D.Y., 2002. Amplitude differences of evoked alpha and gamma oscillations in two different age groups. Int. J. Psychophysiol. 45 (3), 245–251. doi:10.1016/S0167-8760(02)00031-4.

Brugge, J.F., Nourski, K.V., Oya, H., Reale, R.A., Kawasaki, H., Steinschneider, M., Howard, M.A., 2009. Coding of repetitive transients by auditory cortex on Heschl's gyrus. J. Neurophysiol. 102 (4), 2358. doi:10.1152/JN.91346.2008.

Campbell, T., Kerlin, J.R., Bishop, C.W., Miller, L.M., 2012. Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. Ear Hear. 33 (1), 144–150. doi:10.1097/AUD.0B013E3182280353.

Cheng, F.Y., Xu, C., Gold, L., Smith, S., 2021. Rapid enhancement of subcortical neural responses to sine-wave speech. Front. Neurosci. 15, 1772. doi:10.3389/FNINS.2021.747303/BIBTEX.

Coffey, E.B., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zatorre, R.J., Kraus, N., 2019. Evolving perspectives on the sources of the frequency-following response. Nat. Commun. 10 (1), 1–10. doi:10.1038/s41467-019-13003-w.

Coffey, E.B.J., Herholz, S.C., Chepesiuk, A.M.P., Baillet, S., Zatorre, R.J., 2016. Cortical contributions to the auditory frequency-following response revealed by MEG. Nat. Commun. 7, 11070. doi:10.1038/ncomms11070.

Cohen, R.A., 2014. The neuropsychology of attention. Neuropsychol. Attention 1–978. doi:10.1007/978-0-387-72639-7.

Coull, J.T., 1998. Neural correlates of attention and arousal: insights from electrophysiology, functional neuroimaging and psychopharmacology. Prog. Neurobiol. 55 (4), 343–361. doi:10.1016/S0301-0082(98)00011-2.

Cristianini, N., Shawe-Taylor, J., 2000. An introduction to support vector machines : and other kernel-based learning methods, 189.

Davis, R.L., Britt, R.H., 1984. Analysis of the frequency following response in the cat. Hear. Res. 15 (1), 29–37. doi:10.1016/0378-5955(84)90222-3.

Dunlop, C.W., Webster, W.R., Simons, L.A., 1965. Effect of attention on evoked responses in the classical auditory pathway. Nature 206 (4988), 1048–1050. doi:10.1038/2061048b0.

Duquesnoy, A.J., 1983. Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. J. Acoust. Soc. Am. 74 (3), 739. doi:10.1121/1.389859.

Eysenck, M.W., 1982. Attention and arousal. Attent. Arousal doi:10.1007/978-3-642-68390-9.

Flexer, A., Bauer, H., Pripfl, J., Dorffner, G., 2005. Using ICA for removal of ocular artifacts in EEG recorded from blind subjects. Neural Netw. 18 (7), 998–1005. doi:10.1016/J.NEUNET.2005.03.012.

Forte, A.E., Etard, O., Reichenbach, T., 2017. The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. Elife 6. doi:10.7554/ELIFE.27203.

Fostick, L., Ben-Artzi, E., Babkoff, H., 2013. Aging and speech perception: beyond hearing threshold and cognitive ability. J. Basic Clin. Physiol. Pharmacol. 24 (3), 175–183. http://apps.webofknowledge.com/full_record.do?product=UA™search_mode=GeneralSearch™qid=18™SID=3BsaD2wqgJzbR2mhAWS™page=1™doc=1

Foxe, J.J., Snyder, A.C., 2011. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. Front. Psychol. 2 (JUL), 154. doi:10.3389/FPSYG.2011.00154/BIBTEX.

Galbraith, G.C., Arroyo, C., 1993. Selective attention and brainstem frequency-following responses. Biol. Psychol. 37 (1), 3–22. doi:10.1016/0301-0511(93)90024-3.

Galbraith, G.C., Kane, J.M., 1993. Brainstem frequency-following responses and cortical event-related potentials during attention. Percept. Mot. Skills 76 (3 Pt 2), 1231–1241. doi:10.2466/PMS.1993.76.3C.1231.

Galbraith, G.C., Olfman, D.M., Huffman, T.M., 2003. Selective attention affects human brain stem frequency-following response. Neuroreport 14 (5), 735–738. doi:10.1097/00001756-200304150-00015.

Galbraith, G.C., Threadgill, M.R., Hemsley, J., Salour, K., Songdej, N., Ton, J., Cheung, L., 2000. Putative measure of peripheral and brainstem frequency-following in humans. Neurosci. Lett. 292 (2), 123–127. doi:10.1016/S0304-3940(00)01436-1.

Gao, E., Suga, N., 2000. Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: role of the corticofugal system. Proc. Natl. Acad. Sci. 97 (14), 8081–8086. doi:10.1073/PNAS.97.14.8081.

Giard, M.H., Collet, L., Bouchet, P., Pernier, J., 1994. Auditory selective attention in the human cochlea. Brain Res. 633 (1–2), 353–356. doi:10.1016/0006-8993(94)91561-X.

Gould, I.C., Rushworth, M.F., Nobre, A.C., 2011. Indexing the graded allocation of visuospatial attention using anticipatory alpha oscillations. J. Neurophysiol. 105 (3), 1318–1326. doi:10.1152/JN.00653.2010.

Haegens, S., Nácher, V., Luna, R., Romo, R., Jensen, O., 2011. α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. Proc. Natl. Acad. Sci. USA 108 (48), 19377–19382. doi:10.1073/PNAS.1117190108/-/DCSUPPLEMENTAL.

Hartmann, T., Weisz, N., 2019. Auditory cortical generators of the frequency following response are modulated by intermodal attention. NeuroImage 203, 116185. doi:10.1016/J.NEUROIMAGE.2019.116185.

Helfer, K., Wilber, L., 1990. Hearing-loss, aging, and speech-perception in reverberation and noise. J. Speech Hear. Res. 33 (1), 149–155. http://apps.webofknowledge.com.ezproxy.lib.purdue.edu/full_record.do?product=WOS™search_mode=GeneralSearch™qid=23™SID=1FWQvKYVqsHOcsaHt7H™page=1™doc=2

Helfrich, R.F., Fiebelkorn, I.C., Szczepanski, S.M., Lin, J.J., Parvizi, J., Knight, R.T., Kastner, S., 2018. Neural mechanisms of sustained attention are rhythmic. Neuron 99 (4), 854–865. doi:10.1016/J.NEURON.2018.07.032.

Holmes, E., Purcell, D.W., Carlyon, R.P., Gockel, H.E., Johnsrude, I.S., 2018. Attentional modulation of envelope-Following responses at lower (93109 hz) but not higher (217233 hz) modulation rates. JARO – J. Assoc. Res. Otolaryngol. 19 (1), 83–97. doi:10.1007/S10162-017-0641-9/FIGURES/5.

Hyvarinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. 10 (3), 626–634. doi:10.1109/72.761722.

Issa, E.B., Wang, X., 2008. Sensory responses during sleep in primate primary and secondary auditory cortex. J. Neurosci. 28 (53), 14467–14480. doi:10.1523/JNEUROSCI.3086-08.2008.

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/BIBTEX.

Kelly, S.P., Gomez-Ramirez, M., Foxe, J.J., 2009. The strength of anticipatory spatial biasing predicts target discrimination at attended locations: a high-density EEG study. Eur. J. Neurosci. 30 (11), 2224–2234. doi:10.1111/J.1460-9568.2009.06980.X.

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. J. Acoust. Soc. Am. 116 (4), 2395. doi:10.1121/1.1784440.

Klatt, L.I., Getzmann, S., Begau, A., Schneider, D., 2020. A dual mechanism underlying retroactive shifts of auditory spatial attention: dissociating target- and distractor-related modulations of alpha lateralization. Scientific Reports 2020 10:1 10 (1), 1–13. doi:10.1038/s41598-020-70004-2.

Klimesch, W., 1999. EEG Alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Brain Res. Rev. 29 (2–3), 169–195. doi:10.1016/S0165-0173(98)00056-3.

Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. Trends Cognit. Sci. 16 (12), 606–617. doi:10.1016/J.TICS.2012.10.007.

Krone, L., Frase, L., Piosczyk, H., Selhausen, P., Zittel, S., Jahn, F., Kuhn, M., Feige, B., Mainberger, F., Klöppel, S., Riemann, D., Spiegelhalder, K., Baglioni, C., Sterr, A., Nissen, C., 2017. Top-down control of arousal and sleep: fundamentals and clinical implications. Sleep Med. Rev. 31, 17–24. doi:10.1016/J.SMRV.2015.12.005.

Kropotov, J.D., 2009. Beta rhythms. Quant. EEG Event-Related Potentials Neurother. 59–76. doi:10.1016/B978-0-12-374512-5.00003-7.

Kucyi, A., Hove, M.J., Esterman, M., Matthew Hutchison, R., Valera, E.M., 2017. Dynamic brain network correlates of spontaneous fluctuations in attention. Cereb. Cortex 27 (3), 1831–1840. doi:10.1093/CERCOR/BHW029.

Lai, J., Bartlett, E.L., 2018. Masking differentially affects envelope-following responses in young and aged animals. Neuroscience 386, 150–165. doi:10.1016/J.NEUROSCIENCE.2018.06.004.

Lehmann, A., Schönwiesner, M., 2014. Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. PLoS One 9 (1). doi:10.1371/JOURNAL.PONE.0085442.

Llanos, F., Xie, Z., Chandrasekaran, B., 2017. Hidden Markov modeling of frequency-following responses to Mandarin lexical tones. J. Neurosci. Methods 291, 101–112. doi:10.1016/J.JNEUMETH.2017.08.010.

Lukas, J.H., 1980. Human auditory attention: the olivocochlear bundle may function as a peripheral filter. Psychophysiology 17 (5), 444–452. doi:10.1111/J.1469-8986.1980.TB00181.X.

Mai, G., Schoof, T., Howell, P., 2019. Modulation of phase-locked neural responses to speech during different arousal states is age-dependent. NeuroImage 189, 734–744. doi:10.1016/J.NEUROIMAGE.2019.01.049.

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., Golumbic, E.Z., 2017. Sleep disrupts high-Level speech parsing despite significant basic auditory processing. J. Neurosci. 37 (32), 7772–7781. doi:10.1523/JNEUROSCI.0168-17.2017.

Mankel, K., Bidelman, G.M., 2018. Inherent auditory skills rather than formal music training shape the neural encoding of speech. Proc. Natl. Acad. Sci. USA 115 (51), 13129–13134. doi:10.1073/PNAS.1811793115/-/DCSUPPLEMENTAL.

Marsh, J.T., Worden, F.G., Smith, J.C., 1970. Auditory frequency-following response: neural or artifact? Science 169 (3951), 1222–1223. doi:10.1126/SCIENCE.169.3951.1222.

Moini, J., Piran, P., 2020. Cerebral cortex. Funct. Clin. Neuroanat. 177–240. doi:10.1016/B978-0-12-817424-1.00006-9.

Musacchia, G., Strait, D., Kraus, N., 2008. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear. Res. 241 (1–2), 34–42. doi:10.1016/J.HEARES.2008.04.013.

Nir, Y., Vyazovskiy, V.V., Cirelli, C., Banks, M.I., Tononi, G., 2015. Auditory responses and stimulus-specific adaptation in rat auditory cortex are preserved across NREM and REM sleep. Cereb. Cortex 25 (5), 1362–1378. doi:10.1093/CERCOR/BHT328.

Nunez, P. L., 2016. Electroencephalography (EEG). The Curated Reference Collection in Neuroscience and Biobehavioral Psychology. 10.1016/B978-0-12-809324-5.03049-2.

Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP measurements. Clinical Neurophysiology 112, 713–719.

Parbery-Clark, A., Skoe, E., Kraus, N., 2009. Musical experience limits the degradative effects of background noise on the neural processing of sound. J. Neurosci. 29 (45), 14100–14107. doi:10.1523/JNEUROSCI.3256-09.2009.

Petersen, S.E., Posner, M.I., 2012. The attention system of the human brain: 20 years after. Annu. Rev. Neurosci. 35, 73. doi:10.1146/ANNUREV-NEURO-062111-150525.

Pettigrew, C.M., Murdoch, B.E., Ponton, C.W., Kei, J., Chenery, H.J., Alku, P., 2004. Subtitled videos and mismatch negativity (MMN) investigations of spoken word processing. J. Am. Acad. Audiol. 15 (7), 469–485. doi:10.3766/JAAA.15.7.2.

Pfurtscheller, G., Stancák, A., Neuper, C., 1996. Event-related synchronization (ERS) in the alpha band an electrophysiological correlate of cortical idling: a review. Int. J. Psychophysiol. 24 (1–2), 39–46. doi:10.1016/S0167-8760(96)00066-9.

Phipson, B., Smyth, G.K., 2010. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. Stat. Appl. Genet. Mol. Biol. 9 (1). doi:10.2202/1544-6115.1585.

Pichora-Fuller, M., 2003. Cognitive aging and auditory information processing. Int. J. Audiol. 42, 26–32. doi:10.3109/14992020309074641.

Picton, T.W., Hillyard, S.A., 1974. Human auditory evoked potentials. II. Effects of attention. Electroencephalogr. Clin. Neurophysiol. 36 (2), 191–200. doi:10.1016/0013-4694(74)90156-4.

Picton, T.W., Hillyard, S.A., Galambos, R., Schiff, M., 1971. Human auditory attention: a central or peripheral process? Science 173 (3994), 351–353. doi:10.1126/SCIENCE.173.3994.351.

Pivik, R.T., Harman, K., 1995. A reconceptualization of EEG alpha activity as an index of arousal during sleep: all alpha activity is not equal. J. Sleep Res. 4 (3), 131–137. doi:10.1111/J.1365-2869.1995.TB00161.X.

Portas, C.M., Krakow, K., Allen, P., Josephs, O., Armony, J.L., Frith, C.D., 2000. Auditory processing across the sleep-wake cycle: simultaneous EEG and fMRI monitoring in humans. Neuron 28 (3), 991–999. doi:10.1016/S0896-6273(00)00169-0.

Price, C.N., Bidelman, G.M., 2021. Attention reinforces human corticofugal system to aid speech perception in noise. NeuroImage 235, 118014. doi:10.1016/J.NEUROIMAGE.2021.118014.

Price, C.N., Moncrieff, D., 2021. Defining the role of attention in hierarchical auditory processing. Audiol. Res. 11 (1), 112. doi:10.3390/AUDIOLRES11010012.

Rinne, T., Balk, M.H., Koistinen, S., Autti, T., Alho, K., Sams, M., 2008. Auditory selective attention modulates activation of human inferior colliculus. J. Neurophysiol. 100 (6), 3323–3327. doi:10.1152/JN.90607.2008.

Robbins, T.W., Everitt, B.J., 1995. Arousal systems and attention. In: M.S., Gazzaniga (Ed.), The cognitive neurosciences. The MIT Press, pp. 703–720.

Saderi, D., Schwartz, Z.P., Heller, C.R., Pennington, J.R., David, S.V., 2021. Dissociation of task engagement and arousal effects in auditory cortex and midbrain. Elife 10, 1–25. doi:10.7554/ELIFE.60153.

Saiz-Alía, M., Forte, A.E., Reichenbach, T., 2019. Individual differences in the attentional modulation of the human auditory brainstem response to speech inform on speech-in-noise deficits. Sci. Rep. 9 (1), 1–10. doi:10.1038/s41598-019-50773-1.

Scherg, M., Ebersole, J.S., 1994. Brain source imaging of focal and multifocal epileptiform EEG activity. Neurophysiol. Clin. 24 (1), 51–60. doi:10.1016/S0987-7053(05)80405-8.

Scherg, M., Ille, N., Bornfleth, H., Berg, P., 2002. Advanced tools for digital EEG review: virtual source montages, whole-head mapping, correlation, and phase analysis. J. Clin. Neurophysiol. 19 (2), 91–112. doi:10.1097/00004691-200203000-00001.

Shaheen, L.A., Slee, S.J., David, S.V., 2021. Task engagement improves neural discriminability in the auditory midbrain of the marmoset monkey. J. Neurosci. 41 (2), 284–297. doi:10.1523/JNEUROSCI.1112-20.2020.

Shiga, T., Althen, H., Cornella, M., Zarnowiec, K., Yabe, H., Escera, C., 2015. Deviance-related responses along the auditory hierarchy: combined FFR, MLR and MMN evidence. PLoS One 10 (9), e0136794. doi:10.1371/JOURNAL.PONE.0136794.

Skoe, E., Kraus, N., 2010. Auditory brain stem response to complex sounds: atutorial. Ear Hear. 31 (3), 302–324. doi:10.1097/AUD.0B013E3181CDB272.

Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. Electroencephalogr. Clin. Neurophysiol. 39, 465–472. http://www.ncbi.nlm.nih.gov/pubmed/52439

Sohmer, H., Pratt, H., 1977. Identification and separation of acoustic frequency following responses (FFRS) in man. Electroencephalogr. Clin. Neurophysiol. 42 (4), 493–500. doi:10.1016/0013-4694(77)90212-7.

Sohmer, H., Pratt, H., Kinarti, R., 1977. Sources of frequency following responses (FFR) in man. Electroencephalogr. Clin. Neurophysiol. 42, 656–664. http://www.ncbi.nlm.nih.gov/pubmed/67025

Song, J.H., Skoe, E., Banai, K., Kraus, N., 2011. Perception of speech in noise: neural correlates. J. Cognit. Neurosci. 23 (9), 2268. doi:10.1162/JOCN.2010.21556.

Strauß, A., Wöstmann, M., Obleser, J., 2014. Cortical alpha oscillations as a tool for auditory selective inhibition. Front. Hum. Neurosci. 8 (MAY). doi:10.3389/FNHUM.2014.00350.

Suga, N., Ma, X., 2003. Multiparametric corticofugal modulation and plasticity in the auditory system. Nat. Rev. Neurosci. 4 (10), 783–794. doi:10.1038/NRN1222.

Tichko, P., Skoe, E., 2017. Frequency-dependent fine structure in the frequency-following response: the byproduct of multiple generators. Hear. Res. 348, 1–15. doi:10.1016/J.HEARES.2017.01.014.

Tune, S., Alavash, M., Fiedler, L., Obleser, J., 2021. Neural attentional-filter mechanisms of listening success in middle-aged and older individuals. Nat. Commun. 12 (1), 1–14. doi:10.1038/s41467-021-24771-9.

Uusberg, A., Uibo, H., Kreegipuu, K., Allik, J., 2013. EEG alpha and cortical inhibition in affective attention. Int. J. Psychophysiol. 89 (1), 26–36. doi:10.1016/J.IJPSYCHO.2013.04.020.

Varghese, L., Bharadwaj, H.M., Shinn-Cunningham, B.G., 2015. Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. Brain Res. 1626, 146–164. doi:10.1016/J.BRAINRES.2015.06.038.

Vollmer, M., Beitel, R.E., Schreiner, C.E., Leake, P.A., 2017. Passive stimulation and behavioral training differentially transform temporal processing in the inferior colliculus and primary auditory cortex. J. Neurophysiol. 117 (1), 47–64. doi:10.1152/JN.00392.2016.

White-Schwoch, T., Anderson, S., Krizman, J., Nicol, T., Kraus, N., 2019. Case studies in neuroscience: subcortical origins of the frequency-following response. J. Neurophysiol. 122 (2), 844–848. doi:10.1152/JN.00112.2019/ASSET/IMAGES/LARGE/Z9K0081951590003.JPEG.

Wilsch, A., Henry, M.J., Herrmann, B., Maess, B., Obleser, J., 2015. Alpha oscillatory dynamics index temporal expectation benefits in working memory. Cereb. Cortex 25 (7), 1938–1946. doi:10.1093/CERCOR/BHU004.

Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. Institute of Electrical and Electronics Engineers Inc., pp. 5987–5995. doi:10.1109/CVPR.2017.634.

Xie, Z., Reetzke, R., Chandrasekaran, B., 2019. Machine learning approaches to analyze speech-evoked neurophysiological responses. J. Speech Lang. Hear. Res. 62 (3), 587–601. doi:10.1044/2018_JSLHR-S-ASTM-18-0244.

Yordanova, J.Y., Kolev, V.N., Başar, E., 1998. EEG theta and frontal alpha oscillations during auditory processing change with aging. Electroencephalogr. Clin. Neurophysiol. 108 (5), 497–505. doi:10.1016/S0168-5597(98)00028-8.

Zhao, C.T., Kuhl, P.K., 2018. Linguistic effect on speech perception observed at the brainstem. Proc. Natl. Acad. Sci. USA 115 (35), 8716–8721. doi:10.1073/PNAS.1800186115/-/DCSUPPLEMENTAL.