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

Jesyin Lai[1,2], **Caitlin N. Price**[1,2,3], **Gavin M. Bidelman**[1,2,4]

***For correspondence:** gbidel@indiana.edu (GMB)

Present address: [4]2631 East Discovery Parkway, Bloomington, IN, 47408, USA

[1]Institute for Intelligent Systems, University of Memphis; [2]School of Communication Sciences and Disorders, University of Memphis; [3]Department of Audiology and Speech Pathology, University of Arkansas for Medical Sciences; [4]Department of Speech, Language and Hearing Sciences, Indiana University

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 backgrounds aiming to tax perceptual and attentional processing. 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 their α states, low α FFRs in noise were found to be weaker, correlated positively with behavioral response times, more "decodable" via classifiers, and associated indirectly with other signal-in-noise perceptual performance. Our data provide evidence for online corticofugal interplay in humans and establish that brainstem sensory representations are continuously yoked to the ebb and flow of cortical states to dynamically update perceptual processing.

Introduction

Central auditory processing of target sounds is influenced by multiple factors, such as background noise (***Binkhamis et al., 2019***; ***Billings et al., 2020***; ***Song et al., 2011***), task engagement (***Saderi et al., 2021***; ***Shaheen et al., 2021***), attention (***Saiz-Alía et al., 2019***; ***Price and Bidelman, 2021***), 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) (***Beyerl, 1978***; ***Bajo and Moore, 2005***) to provide such 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 brain-

stem activity in humans remains highly controversial (***Galbraith and Kane, 1993***; ***Varghese et al., 2015***; ***Dunlop et al., 1965***; ***Picton et al., 1971***). 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. Using source-reconstructed FFRs, 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 neo-cortical 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 (***Kucyi et al., 2017***; ***Helfrich et al., 2018***). 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 (***Eysenck, 1982***; ***Cohen, 2014***; ***Robbins and Everitt, 1995***), such as sleep, wakefulness, or excitement. When attending to sounds and actively engaging in a task, a subject'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 pathways 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. Brainstem sound 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***; ***Smith et al., 1975***; ***White-Schwoch et al., 2019***; ***Marsh et al., 1970***; ***Coffey et al., 2019***; ***Tichko and Skoe, 2017***). 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 (***Uusberg et al., 2013***; ***Aftanas et al., 2002***). 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 (***Klatt et al., 2020***; ***Haegens et al., 2011***; ***Kelly et al., 2009***; ***Gould et al., 2011***), we explored correlations between α-yoked FFRs and speech perception measures. We also investigated neural decoding of FFRs at different α power as this may provide further evidence revealing the role of α in being predictive of 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 the ebb and flow of cortical states to dynamically update perceptual processing.

Results

We recruited 20 young (age: 18–35 years), normal hearing ($\leq$ 25 dB HL; 250–8000 Hz) adults as participants in this study. Prior to EEG tasks and recordings, participants were assessed with the QuickSIN (for details see Methods and Materials) to quantify their ability to perceive speech-in-

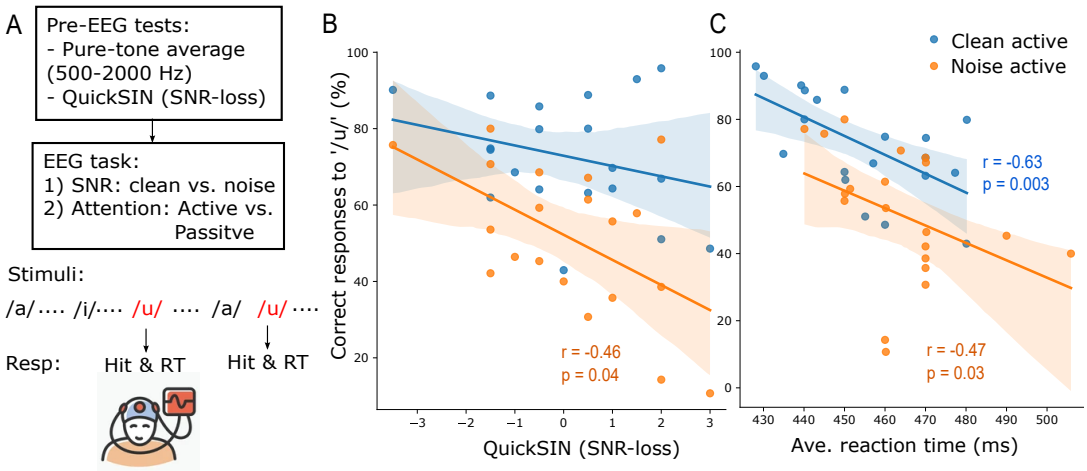

Figure 1. EEG task performance correlates with normative measures of SIN perception (i.e., QuickSIN scores). (A) Prior to EEG recordings, all participants' pure-tone averages at 500-2000 Hz were obtained and speech-in-noise perception was assessed with QuickSIN. Subsequently, speech-EEGs were recorded 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.

noise (***Killion et al., 2004***). Their pure-tone average (PTA) was also calculated as the average hearing threshold (dB HL) at 500, 1000 and 2000 Hz. Subsequently, speech-evoked EEGs were recorded under clean or noisy (+5 dB SNR) backgrounds and in active and passive conditions (for details see Methods and Materials). Vowel sounds (frequent /a/, /i/=distractors; infrequent /u/=targets) were used to evoke auditory brainstem responses. During active sessions, participants were required to press a button (hit) when detecting target /u/ tokens. Hit responses (correct responses to /u/ within 5 tokens) and reaction times (RTs) were recorded (***Figure 1***A).

Behavioral speech-in-noise performance

Speech-in-noise perception during the neuroimaging 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, ***Figure 1***B). This confirms the external validity of our laboratory-based task in assessing clinically-normed 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 speech-in-noise perception. 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 (***Figure 1***C).

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 (***Coffey et al., 2016***; ***Bidelman, 2018***). We isolated cortical α band from the posterior POz channel, where waking α activity is maximal at the scalp (***Moini and Piran, 2020***; ***Nunez, 2016***; ***Klatt et al.,***

2020). 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 be used as negative controls.

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's RMS values (***Figure 2***A) 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 (***Figure 2***B). 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 $\sim$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 ***Figure 2***C and ***Figure 2***D for low- and high-α states. Corresponding average α responses are shown in ***Figure 2***E. 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 ***Figure 2***F, 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.

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 ***Figure 3***A-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, $F = 4.84$, $p = 0.04$, $\eta_p^2 = 0.20$) (***Figure 3***E). However, responses were comparable 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 (***Figure 3–Figure Supplement 1***). 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 ***Figure 3***F and ***Table 1***).

Brain-behavior relations

We next assessed associations between behavioral performance in the speech detection task collected during EEG recordings (percent 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, ***Figure 4***A). 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, however, was not observed between noise passive FFRs and RTs (recorded in the active task) (***Figure 4***C). Repeating the same

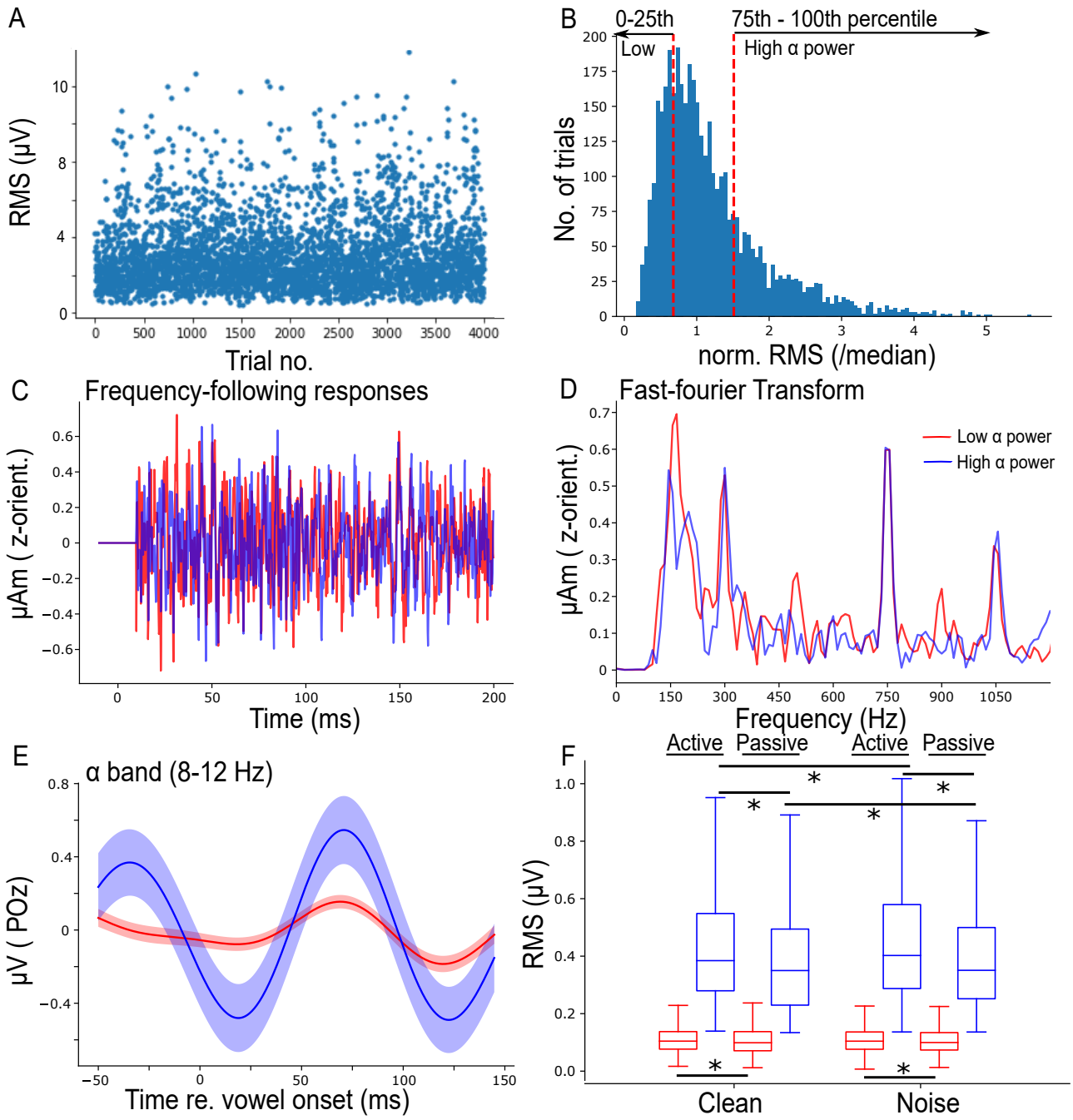

Figure 2. Examples of data processing (representative subject) to separate FFRs based on cortical arousal state (i.e., α power). (A) Cortical α band was isolated from the POz channel and epoched at -50-145 ms with respect to vowel stimulus onset time. Root-mean-square (RMS) value was computed for every epoch (i.e., approximately 4000 single trials) per listener. (B) Histogram of RMS values normalized to the RMS median of each run. The first red dashed line (left) indicates the 25th percentile and the second (right) indicates the 75th percentile. 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 (with removed onset responses) 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 = $\pm$ s.e.m. * $p < 0.01$ (Conover's test, non-parametric pairwise test, with Bonferroni adjustment)

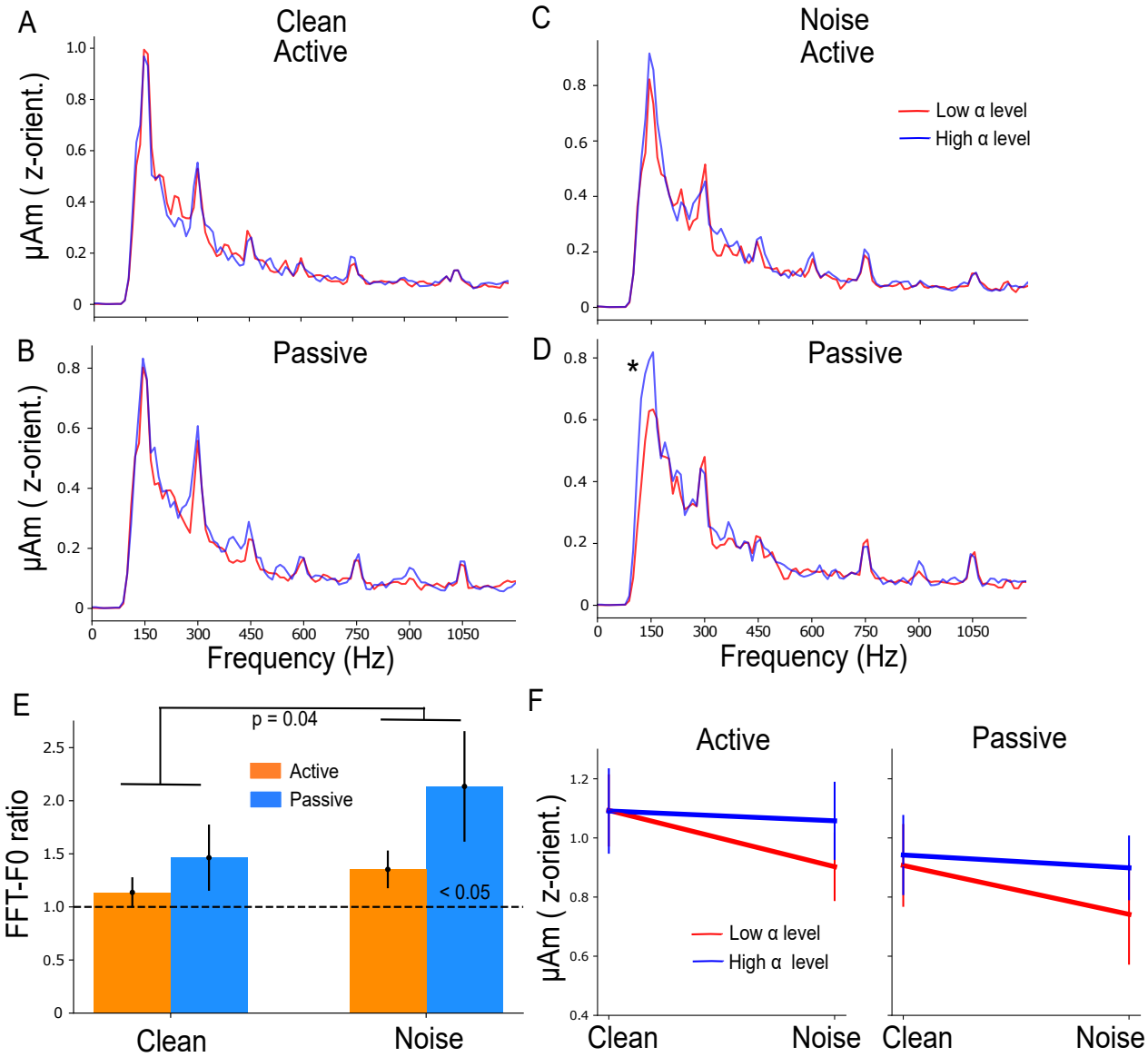

Figure 3. FFR F0 amplitudes during low α are lower than during high α in noisy (but not in clean) backgrounds. (A-D) Frequency spectra of speech-evoked FFRs illustrating lower F0 amplitudes during low α in both active and passive noisy backgrounds (C & D) but not in clean backgrounds (A & B). (E) A noise 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= $\pm$ s.e.m., *p<0.05

Figure 3–Figure supplement 1. As in Figure 3E but for POz β band and Fz α band. These control analyses showed no significant dependence of the FFR.

Table 1. Results of Type III, mixed-model ANOVA on ln(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 α power	1	140	2.248	0.136	0.02
attention x α power	1	140	2.008	0.158	0.01
SNR x attention x α power	1	140	0.166	0.685	1.18×10^{-3}

* $p < 0.05$, *** $p < 0.001$

Table 1–source data 1. FFR F0 amplitudes measured from all subjects and stimulus conditions.

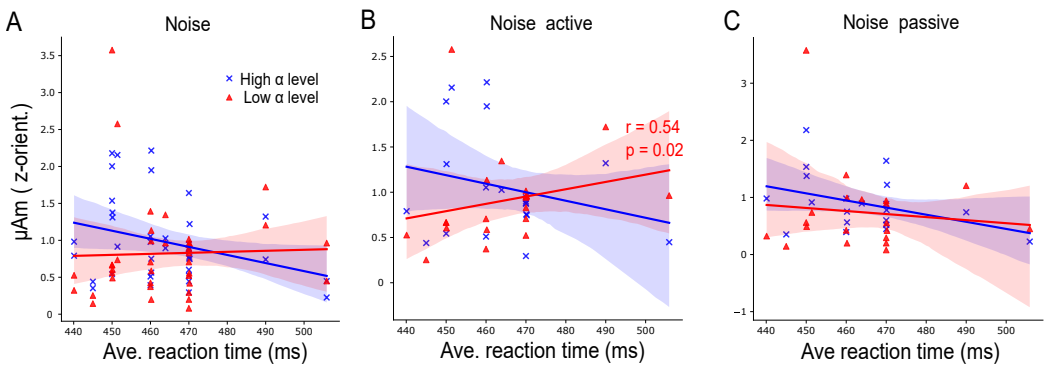

Figure 4. Brain-behavior associations between FFR and speech perception depend on cortical states. (A) In noise (pooling active and passive conditions), there is a trend of decreased FFR F0 amplitudes during high α power in listeners with slower reaction times. (B-C) When correlations were analyzed separately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

Figure 4–Figure supplement 1. Repeating the same correlation analysis with a different EEG frequency band (β) and electrode (Fz) produced no significant result.

Figure 4–Figure supplement 2. Summary of all correlations between pre-test perceptual performance vs. EEG behavioral task and EEG behavioral task vs. FFR F0 amplitudes (low or high α power).

correlation analysis for POz β and Fz α in the noise active condition did not reveal any significant association between FFRs and RT (***Figure 4–Figure Supplement 1***). This suggests that the observed correlation between FFRs and RT was specifically related to α power in POz channel.

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 percent of correct response to /u/ (***Figure 1***C) and percent of correct response to /u/ obtained in noise conditions was strongly predictive of QuickSIN score (***Figure 1***B). Collectively, these results show FFRs during low α power have an indirect relationship with other perceptual performance (e.g., percent of correct response to /u/ and QuickSIN) tested in noise (***Figure 4–Figure Supplement 2***). In another words, cortical α modulation of FFRs in noisy backgrounds is indirectly associated with speech detection performance and QuickSIN via behavioral RT.

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

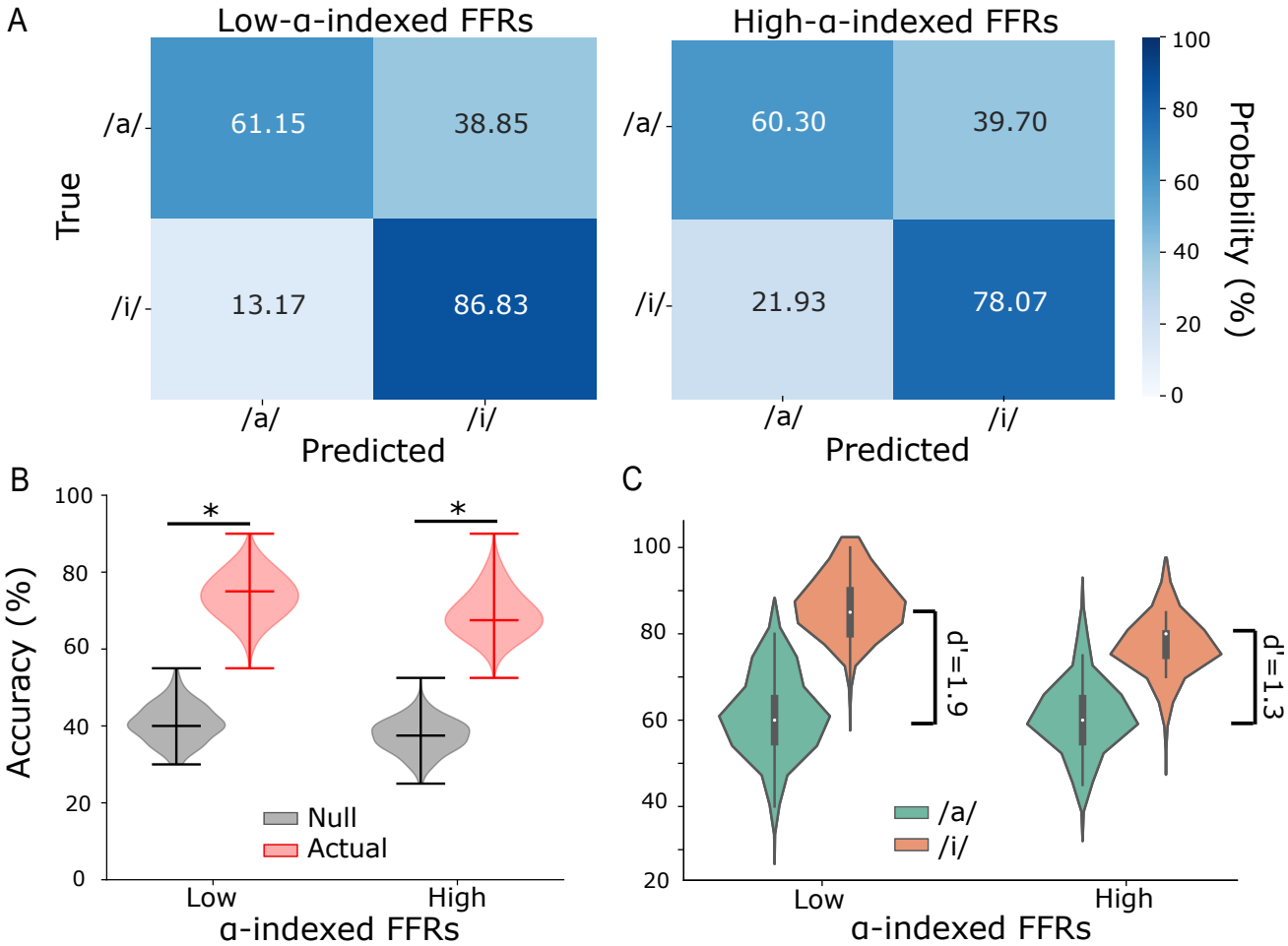

Figure 5. Average accuracy of linear SVMs in classifying /i/ from /a/ tokens (noise active condition) is better when using low-α-indexed FFRs as inputs. (A) Vowel decoding confusion matrices for low- (left) vs. high (right) α-indexed FFRs (N=5000 bootstrap iterations). (B) Distribution of prediction accuracies for FFRs (actual) were significantly above the null prediction accuracies. Upper/lower ticks=max/min; center tick=medians. (C) Sensitivity (d') in discriminating /a/ & /i/ tokens was higher for FFRs at low compared to high α power. More robust representation of /i/ tokens during low α power leads to a larger difference between /a/ & /i/ classification accuracies that may enhance token discrimination in noisy backgrounds. * p< 0.001

followed the machine learning (ML)-based approach reported by ***Xie et al.*** (***2019***) to analyze FFRs for token decoding. Two 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 ***Figure 5***A. All FFRs yielded classification performance significantly above chance ***Figure 5***B, confirming brainstem neural representations closely mirror the spectrotemporal properties of speech. While mean classification accuracies for /a/ token were similar (61.16 % vs 60.3 %) in both α conditions, /i/ tokens were classified better using FFRs at low (86.83 %) compared to high (78.07 %) α power. We further calculated d' as an objective measure to compare token discriminability at low vs. high α power ***Figure 5***C. The d' of low-α-indexed FFRs was 0.6 higher (d' = 1.9 vs. 1.3) than that of high-α-indexed FFRs. 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.

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 the listener's cortical state. We found higher cortical α states positively associate with 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 advocated by better token discrimination 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 (***Galbraith et al., 2003***; ***Forte et al., 2017***; ***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 (***Vollmer et al., 2017***; ***Price and Bidelman, 2021***) and/or require more difficult perceptual tasks (e.g., SIN perception) that challenge speech processing (***Lehmann and Schönwiesner, 2014***; ***Galbraith and Arroyo, 1993***). In this study, we observed overall enhanced FFRs regardless of the background SNR when subjects actively engaged in challenging speech listening tasks (***Figure 3***F, left panel). However, FFR changes due to cortical α were more prominent in noise than clean conditions (***Figure 3***E-F). Similarly, even under passive listening conditions, these cortical dependencies in the FFR were stronger under states of low α and for noisy speech (***Figure 3***E & 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 the 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. 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 (***Jensen and Mazaheri, 2010***; ***Foxe and Snyder, 2011***). 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 irrele-

vant information (***Strauß et al., 2014***). Here, we demonstrate that α power 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 (***Figure 2***F). 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. Indeed, α oscillations have been interpreted to have two mechanistic roles underlying the functions of selective attention (***Klimesch, 2012***; ***Klatt et al., 2020***). ***Klatt et al.*** (***2020***) were able to tease apart these two mechanisms by showing a shifting of attention towards lateralized targets resulted in contralateral decreases in α power whereas shifting attention away from lateralized distractors resulted in contralateral increases in α power. Furthermore, decreased α power has been associated with increased neural firing (***Haegens et al., 2011***) and improved behavioral performance (***Haegens et al., 2011***; ***Kelly et al., 2009***; ***Gould et al., 2011***).

In this vein, we observed subjects with smaller amplitudes of low-α-indexed FFRs having shorter behavioral RTs in the noise active condition (***Figure 4***B). 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. 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 (***Figure 4***B). Reversing these phenomena resulted in slower RTs when subjects were actively identifying stimulus targets in noise.

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 (***Xie et al., 2017***; ***Cheng et al., 2021***). 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 larger difference in classification accuracies and higher sensitivity between /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 (***Figure 4***B).

More broadly, difficulties in SIN perception (***Duquesnoy, 1983***; ***Helfer and Wilber, 1990***; ***Fostick et al., 2013***) 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 (***Yordanova et al., 1998***; ***Klimesch, 1999***; ***Böttger et al., 2002***). 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***).

In conclusion, our 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 directly 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.

Methods and Materials

Participants

We recruited twenty young adults (age: 18–35 years, M = 24, SD = 3.4 years; 11 female). All participants exhibited normal hearing thresholds ($\leq$ 25 dB HL; 250–8000 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.

Speech stimuli and task

The speech stimuli and task are described fully in (***Price and Bidelman, 2021***). Brainstem and cortical EEG were recorded simultaneously (***Bidelman et al., 2013***, ***2019***) during SIN-listening tasks to investigate the attentional effects and cortical influences on brainstem responses to speech. Three synthesized vowel tokens (e.g., /a/, /i/, /u/) were presented during the recording of EEGs. These vowels were chosen because they are sustained periodic sounds optimal for evoking FFRs (***Bidelman et al., 2019***; ***Skoe and Kraus, 2010***). Each 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 vowels 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 (***Killion et al., 2004***) at a signal-to-noise ratio (SNR) of +5 dB (speech at 75 dB_A SPL and noise at 70 dB_A SPL).

During active blocks, participants detected /u/ tokens, which were infrequent, via button press. A “hit” was defined as detection within 5 token presentations of /u/. For passive blocks, participants watched a captioned movie and were instructed to ignore any sounds they heard. This passive “task” has been shown to maintain arousal without impeding auditory processing (***Pettigrew et al., 2004***).

Each attentional block and noise condition (e.g., clean active, noise active, clean passive, noise passive) was divided into 2 runs allowing for breaks as needed. A single run lasted approximately 7.75 minutes and contained 1000 trials 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***)). Participants 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) pre-

sentation were counterbalanced across participants to minimize order and fatigue effects (breaks were provided between blocks).

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***).

QuickSIN test

The Quick Speech-in-Noise (QuickSIN) test assessed listeners' speech perception in noise (***Killion et al., 2004***). Listeners heard lists of 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' scores in this study ranged from -3.5 to 3 dB of SNR loss (M = 0.1, SD = 1.6), consistent with normal hearing.

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 $\leq$ 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 activities) frequency components of the EEG (***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 corrected in the continuous EEGs using independent component analysis (ICA; method='fastica') (***Flexer et al., 2005***; ***Hyvarinen, 1999***). In this process, ICs best matching the pattern in the electrooculogram (EOG) were automatically identified and excluded from the data. 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 further emphasize brainstem activity (***Bidelman et al., 2013***; ***Musacchia et al., 2008***).

Source FFRs

We transformed the cleaned sensor data (64 ch) into source space using a virtual source montage. The source montage 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***); ***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 mibdrain (***Scherg and Ebersole, 1994***; ***Scherg et al., 2002***). This model explains >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***)].

Cortical 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. 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 us to rule out any unspecific or general effects of high or low cortical activity on FFRs.

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 (10–100 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 a F0 ratio for each condition by dividing F0 amplitudes of high-α-indexed FFRs by those of low-α-indexed FFRs.

$$F0\ ratio = \frac{F0amp[high_{\alpha}]}{F0amp[low_{\alpha}]} \tag{1}$$

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.

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 (subjects = random factor) ANOVA ('lme4' package in Rstudio). This 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.

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 140 trials of the smallest and largest α RMS values for each of the low- or high-α-indexed FFRs. These average FFR waveforms (10 to 100 ms steady state portion) were used as input features for SVM. Note this contrasts the 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/ & /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 [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.

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***):

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

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.

To compare decoding accuracies for discriminating /a/ and /i/ tokens at low vs. high α power FFRs, we used d' (sensitivity) as a measure for token discrimination. d' is advantageous because it is independent of sample size and it was calculated as below:

$$d' = \frac{|\mu_a - \mu_i|}{\sigma_{total}} \tag{3}$$

where μ is the grand mean classification accuracy for /a/ or /i/ tokens, and σ_{total} is the square root of variance derived from all classification accuracies of /a/ and /i/ tokens for low- and high-α-indexed FFRs.

Acknowledgments

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

References

Aftanas LI, Varlamov AA, Pavlov SV, Makhnev VP, Reva NV. Time-dependent cortical asymmetries induced by emotional arousal: EEG analysis of event-related synchronization and desynchronization in individually defined frequency bands. International journal of psychophysiology : official journal of the International Organization of Psychophysiology. 2002; 44(1):67–82. https://pubmed.ncbi.nlm.nih.gov/11852158/, doi: 10.1016/S0167-8760(01)00194-5.

Alho K, Rinne T, Herron TJ, Woods DL. Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. Hearing research. 2014 1; 307:29–41. https://pubmed.ncbi.nlm.nih.gov/23938208/, doi: 10.1016/J.HEARES.2013.08.001.

Alho K, Vorobyev VA. Brain activity during selective listening to natural speech. Frontiers in bioscience : a journal and virtual library. 2007 5; 12(8):3167–3176. https://pubmed.ncbi.nlm.nih.gov/17485291/, doi: 10.2741/2304.

Assmann PF. Tracking and glimpsing speech in noise: Role of fundamental frequency. The Journal of the Acoustical Society of America. 1998 6; 100(4):2680. https://asa.scitation.org/doi/abs/10.1121/1.416961, doi: 10.1121/1.416961.

Atiani S, Elhilali M, David SV, Fritz JB, Shamma SA. Task Difficulty and Performance Induce Diverse Adaptive Patterns in Gain and Shape of Primary Auditory Cortical Receptive Fields. Neuron. 2009 2; 61(3):467–480. /pmc/articles/PMC3882691//pmc/articles/PMC3882691/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3882691/, doi: 10.1016/J.NEURON.2008.12.027.

Bajo VM, Moore DR. Descending projections from the auditory cortex to the inferior colliculus in the gerbil, Meriones unguiculatus. Journal of Comparative Neurology. 2005 5; 486(2):101–116. https://onlinelibrary.wiley.com/doi/full/10.1002/cne.20542https://onlinelibrary.wiley.com/doi/abs/10.1002/cne.20542https://onlinelibrary.wiley.com/doi/10.1002/cne.20542, doi: 10.1002/CNE.20542.

Beyerl BD. Afferent projections to the central nucleus of the inferior colliculus in the rat. Brain Research. 1978 4; 145(2):209–223. doi: 10.1016/0006-8993(78)90858-2.

Bidelman GM. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. NeuroImage. 2018 7; 175:56–69. doi: 10.1016/J.NEUROIMAGE.2018.03.060.

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

Bidelman GM, Moreno S, Alain C. Tracing the emergence of categorical speech perception in the human auditory system. NeuroImage. 2013 10; 79:201–212. doi: 10.1016/J.NEUROIMAGE.2013.04.093.

Bidelman GM, Price CN, Shen D, Arnott SR, Alain C. Afferent-efferent connectivity between auditory brainstem and cortex accounts for poorer speech-in-noise comprehension in older adults. Hearing Research. 2019 10; 382:107795. doi: 10.1016/J.HEARES.2019.107795.

Billings CJ, Gordon SY, McMillan GP, Gallun FJ, Molis MR, Konrad-Martin D. Noise-induced enhancement of envelope following responses in normal-hearing adults. The Journal of the Acoustical Society of America. 2020 2; 147(2):EL201. https://asa.scitation.org/doi/abs/10.1121/10.0000627, doi: 10.1121/10.0000627.

Binkhamis G, Léger A, Bell SL, Prendergast G, O'Driscoll M, Kluk K. Speech Auditory Brainstem Responses: Effects of Background, Stimulus Duration, Consonant–Vowel, and Number of Epochs. Ear and Hearing. 2019 5; 40(3):659. /pmc/articles/PMC6493675//pmc/articles/PMC6493675/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6493675/, doi: 10.1097/AUD.0000000000000648.

Böttger D, Herrmann CS, Von Cramon DY. Amplitude differences of evoked alpha and gamma oscillations in two different age groups. International journal of psychophysiology : official journal of the International Organization of Psychophysiology. 2002 9; 45(3):245–251. https://pubmed.ncbi.nlm.nih.gov/12208531/, doi: 10.1016/S0167-8760(02)00031-4.

Brugge JF, Nourski KV, Oya H, Reale RA, Kawasaki H, Steinschneider M, Howard MA. Coding of Repetitive Transients by Auditory Cortex on Heschl's Gyrus. Journal of Neurophysiology. 2009 10; 102(4):2358. /pmc/articles/PMC2775384//pmc/articles/PMC2775384/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2775384/, doi: 10.1152/JN.91346.2008.

Campbell T, Kerlin JR, Bishop CW, Miller LM. Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. Ear and hearing. 2012 1; 33(1):144–150. https://pubmed.ncbi.nlm.nih.gov/21760513/, doi: 10.1097/AUD.0B013E3182280353.

Cheng FY, Xu C, Gold L, Smith S. Rapid Enhancement of Subcortical Neural Responses to Sine-Wave Speech. Frontiers in Neuroscience. 2021 12; 15:1772. doi: 10.3389/FNINS.2021.747303/BIBTEX.

Coffey EBJ, Herholz SC, Chepesiuk AMP, Baillet S, Zatorre RJ. Cortical contributions to the auditory frequency-following response revealed by MEG. Nature Communications. 2016 3; 7:11070. http://www.ncbi.nlm.nih.gov/pubmed/27009409http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4820836http://www.nature.com/doifinder/10.1038/ncomms11070, doi: 10.1038/ncomms11070.

Coffey EBJ, Nicol T, White-Schwoch T, Chandrasekaran B, Krizman J, Skoe E, Zatorre RJ, Kraus N. Evolving perspectives on the sources of the frequency-following response. Nature Communications 2019 10:1. 2019 11; 10(1):1–10. https://www.nature.com/articles/s41467-019-13003-w, doi: 10.1038/s41467-019-13003-w.

Cohen RA. The Neuropsychology of Attention. The Neuropsychology of Attention. 2014 1; p. 1–978. doi: 10.1007/978-0-387-72639-7.

Coull JT. Neural correlates of attention and arousal: insights from electrophysiology, functional neuroimaging and psychopharmacology. Progress in neurobiology. 1998 7; 55(4):343–361. https://pubmed.ncbi.nlm.nih.gov/9654384/, doi: 10.1016/S0301-0082(98)00011-2.

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

Dunlop CW, Webster WR, Simons LA. Effect of Attention on Evoked Responses in the Classical Auditory Pathway. Nature 1965 206:4988. 1965; 206(4988):1048–1050. https://www.nature.com/articles/2061048b0, doi: 10.1038/2061048b0.

Duquesnoy AJ. Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. The Journal of the Acoustical Society of America. 1983; 74(3):739. http://scitation.aip.org/content/asa/journal/jasa/74/3/10.1121/1.389859, doi: 10.1121/1.389859.

Eysenck MW. Attention and Arousal. Attention and Arousal. 1982; doi: 10.1007/978-3-642-68390-9.

Flexer A, Bauer H, Pripfl J, Dorffner G. Using ICA for removal of ocular artifacts in EEG recorded from blind subjects. Neural Networks. 2005 9; 18(7):998–1005. doi: 10.1016/J.NEUNET.2005.03.012.

Forte AE, Etard O, Reichenbach T. The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. eLife. 2017 10; 6. doi: 10.7554/ELIFE.27203.

Fostick L, Ben-Artzi E, Babkoff H. Aging and speech perception: beyond hearing threshold and cognitive ability. Journal of Basic and Clinical Physiology and Pharmacology. 2013; 24(3):175–83. http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=18&SID=3BsaD2wqgJzbR2mhAWS&page=1&doc=1.

Foxe JJ, Snyder AC. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. Frontiers in Psychology. 2011; 2(JUL):154. doi: 10.3389/FPSYG.2011.00154/BIBTEX.

Galbraith GC, Kane JM. Brainstem frequency-following responses and cortical event-related potentials during attention. Perceptual and motor skills. 1993; 76(3 Pt 2):1231–1241. https://pubmed.ncbi.nlm.nih.gov/8337070/, doi: 10.2466/PMS.1993.76.3C.1231.

Galbraith GC, Arroyo C. Selective attention and brainstem frequency-following responses. Biological psychology. 1993; 37(1):3–22. https://pubmed.ncbi.nlm.nih.gov/8110920/, doi: 10.1016/0301-0511(93)90024-3.

Galbraith GC, Olfman DM, Huffman TM. Selective attention affects human brain stem frequency-following response. Neuroreport. 2003 4; 14(5):735–738. https://pubmed.ncbi.nlm.nih.gov/12692473/, doi: 10.1097/00001756-200304150-00015.

Gao E, Suga N. Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: Role of the corticofugal system. Proceedings of the National Academy of Sciences. 2000 7; 97(14):8081–8086. https://www.pnas.org/content/97/14/8081https://www.pnas.org/content/97/14/8081.abstract, doi: 10.1073/PNAS.97.14.8081.

Giard MH, Collet L, Bouchet P, Pernier J. Auditory selective attention in the human cochlea. Brain research. 1994 1; 633(1-2):353–356. https://pubmed.ncbi.nlm.nih.gov/8137171/, doi: 10.1016/0006-8993(94)91561-X.

Gould IC, Rushworth MF, Nobre AC. Indexing the graded allocation of visuospatial attention using anticipatory alpha oscillations. Journal of neurophysiology. 2011 3; 105(3):1318–1326. https://pubmed.ncbi.nlm.nih.gov/21228304/, doi: 10.1152/JN.00653.2010.

Haegens S, Nácher V, Luna R, Romo R, Jensen O. α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. Proceedings of the National Academy of Sciences of the United States of America. 2011 11; 108(48):19377–19382. https://www.pnas.org/content/108/48/19377https://www.pnas.org/content/108/48/19377.abstract, doi: 10.1073/PNAS.1117190108/-/DCSUPPLEMENTAL.

Hartmann T, Weisz N. Auditory cortical generators of the Frequency Following Response are modulated by intermodal attention. NeuroImage. 2019 12; 203:116185. doi: 10.1016/J.NEUROIMAGE.2019.116185.

Helfer K, Wilber L. Hearing-loss, aging, and speech-perception in reverberation and noise. Journal of Speech and Hearing Research. 1990; 33(1):149–55. 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 RF, Fiebelkorn IC, Szczepanski SM, Lin JJ, Parvizi J, Knight RT, Kastner S. Neural Mechanisms of Sustained Attention Are Rhythmic. Neuron. 2018 8; 99(4):854–865. doi: 10.1016/J.NEURON.2018.07.032.

Holmes E, Purcell DW, Carlyon RP, Gockel HE, Johnsrude IS. Attentional Modulation of Envelope-Following Responses at Lower (93–109 Hz) but Not Higher (217–233 Hz) Modulation Rates. JARO - Journal of the Association for Research in Otolaryngology. 2018 2; 19(1):83–97. https://link.springer.com/article/10.1007/s10162-017-0641-9, doi: 10.1007/S10162-017-0641-9/FIGURES/5.

Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks. 1999 5; 10(3):626–34. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=761722, doi: 10.1109/72.761722.

Issa EB, Wang X. Sensory Responses during Sleep in Primate Primary and Secondary Auditory Cortex. Journal of Neuroscience. 2008 12; 28(53):14467–14480. https://www.jneurosci.org/content/28/53/14467https://www.jneurosci.org/content/28/53/14467.abstract, doi: 10.1523/JNEUROSCI.3086-08.2008.

Jensen O, Mazaheri A. Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. Frontiers in Human Neuroscience. 2010; 4:186. doi: 10.3389/FNHUM.2010.00186/BIBTEX.

Kelly SP, Gomez-Ramirez M, Foxe JJ. The strength of anticipatory spatial biasing predicts target discrimination at attended locations: a high-density EEG study. The European journal of neuroscience. 2009 12; 30(11):2224–2234. https://pubmed.ncbi.nlm.nih.gov/19930401/, doi: 10.1111/J.1460-9568.2009.06980.X.

Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S. 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. 2004 10; 116(4):2395. https://asa.scitation.org/doi/abs/10.1121/1.1784440, doi: 10.1121/1.1784440.

Klatt LI, Getzmann S, Begau A, Schneider D. A dual mechanism underlying retroactive shifts of auditory spatial attention: dissociating target- and distractor-related modulations of alpha lateralization. Scientific Reports 2020 10:1. 2020 8; 10(1):1–13. https://www.nature.com/articles/s41598-020-70004-2, doi: 10.1038/s41598-020-70004-2.

Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain research Brain research reviews. 1999 4; 29(2-3):169–195. https://pubmed.ncbi.nlm.nih.gov/10209231/, doi: 10.1016/S0165-0173(98)00056-3.

Klimesch W. Alpha-band oscillations, attention, and controlled access to stored information. Trends in Cognitive Sciences. 2012 12; 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. Top-down control of arousal and sleep: Fundamentals and clinical implications. Sleep medicine reviews. 2017 2; 31:17–24. https://pubmed.ncbi.nlm.nih.gov/26883160/, doi: 10.1016/J.SMRV.2015.12.005.

Kropotov JD. Beta Rhythms. Quantitative EEG, Event-Related Potentials and Neurotherapy. 2009; p. 59–76. doi: 10.1016/B978-0-12-374512-5.00003-7.

Kucyi A, Hove MJ, Esterman M, Matthew Hutchison R, Valera EM. Dynamic Brain Network Correlates of Spontaneous Fluctuations in Attention. Cerebral Cortex. 2017 3; 27(3):1831–1840. https://academic.oup.com/cercor/article/27/3/1831/3056303, doi: 10.1093/CERCOR/BHW029.

Lai J, Bartlett EL. Masking Differentially Affects Envelope-following Responses in Young and Aged Animals. Neuroscience. 2018 8; 386:150–165. https://pubmed.ncbi.nlm.nih.gov/29953908/, doi: 10.1016/J.NEUROSCIENCE.2018.06.004.

Lehmann A, Schönwiesner M. Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. PloS one. 2014 1; 9(1). https://pubmed.ncbi.nlm.nih.gov/24454869/, doi: 10.1371/JOURNAL.PONE.0085442.

Llanos F, Xie Z, Chandrasekaran B. Hidden Markov modeling of frequency-following responses to Mandarin lexical tones. Journal of neuroscience methods. 2017 11; 291:101–112. https://pubmed.ncbi.nlm.nih.gov/28807860/, doi: 10.1016/J.JNEUMETH.2017.08.010.

Lukas JH. Human auditory attention: the olivocochlear bundle may function as a peripheral filter. Psychophysiology. 1980; 17(5):444–452. https://pubmed.ncbi.nlm.nih.gov/7465714/, doi: 10.1111/J.1469-8986.1980.TB00181.X.

Mai G, Schoof T, Howell P. Modulation of phase-locked neural responses to speech during different arousal states is age-dependent. NeuroImage. 2019 4; 189:734–744. https://pubmed.ncbi.nlm.nih.gov/30703520/, doi: 10.1016/J.NEUROIMAGE.2019.01.049.

Makov S, Sharon O, Ding N, Ben-Shachar M, Nir Y, Golumbic EZ. Sleep Disrupts High-Level Speech Parsing Despite Significant Basic Auditory Processing. Journal of Neuroscience. 2017 8; 37(32):7772–7781. https://www.jneurosci.org/content/37/32/7772https://www.jneurosci.org/content/37/32/7772.abstract, doi: 10.1523/JNEUROSCI.0168-17.2017.

Mankel K, Bidelman GM. Inherent auditory skills rather than formal music training shape the neural encoding of speech. Proceedings of the National Academy of Sciences of the United States of America. 2018 12; 115(51):13129–13134. https://www.pnas.org/content/115/51/13129https://www.pnas.org/content/115/51/13129.abstract, doi: 10.1073/PNAS.1811793115/-/DCSUPPLEMENTAL.

Marsh JT, Worden FG, Smith JC. Auditory frequency-following response: neural or artifact? Science (New York, NY). 1970; 169(3951):1222–1223. https://pubmed.ncbi.nlm.nih.gov/5450700/, doi: 10.1126/SCIENCE.169.3951.1222.

Moini J, Piran P. Cerebral cortex. Functional and Clinical Neuroanatomy. 2020 1; p. 177–240. doi: 10.1016/B978-0-12-817424-1.00006-9.

Musacchia G, Strait D, Kraus N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hearing research. 2008 7; 241(1-2):34–42. https://pubmed.ncbi.nlm.nih.gov/18562137/, doi: 10.1016/J.HEARES.2008.04.013.

Nir Y, Vyazovskiy VV, Cirelli C, Banks MI, Tononi G. Auditory responses and stimulus-specific adaptation in rat auditory cortex are preserved across NREM and REM sleep. Cerebral cortex (New York, NY : 1991). 2015 5; 25(5):1362–1378. https://pubmed.ncbi.nlm.nih.gov/24323498/, doi: 10.1093/CERCOR/BHT328.

Nunez PL. Electroencephalography (EEG). The Curated Reference Collection in Neuroscience and Biobehavioral Psychology. 2016 1; doi: 10.1016/B978-0-12-809324-5.03049-2.

Parbery-Clark A, Skoe E, Kraus N. Musical experience limits the degradative effects of background noise on the neural processing of sound. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2009 11; 29(45):14100–14107. https://pubmed.ncbi.nlm.nih.gov/19906958/, doi: 10.1523/JNEUROSCI.3256-09.2009.

Petersen SE, Posner MI. The Attention System of the Human Brain: 20 Years After. Annual review of neuroscience. 2012 7; 35:73. /pmc/articles/PMC3413263//pmc/articles/PMC3413263/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3413263/, doi: 10.1146/ANNUREV-NEURO-062111-150525.

Pettigrew CM, Murdoch BE, Ponton CW, Kei J, Chenery HJ, Alku P. Subtitled videos and mismatch negativity (MMN) investigations of spoken word processing. Journal of the American Academy of Audiology. 2004; 15(7):469–485. https://pubmed.ncbi.nlm.nih.gov/15484597/, doi: 10.3766/JAAA.15.7.2.

Pfurtscheller G, Stancák A, Neuper C. Event-related synchronization (ERS) in the alpha band — an electrophysiological correlate of cortical idling: A review. International Journal of Psychophysiology. 1996 11; 24(1-2):39–46. doi: 10.1016/S0167-8760(96)00066-9.

Phipson B, Smyth GK. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. Statistical applications in genetics and molecular biology. 2010; 9(1). https://pubmed.ncbi.nlm.nih.gov/21044043/, doi: 10.2202/1544-6115.1585.

Pichora-Fuller MK. Cognitive aging and auditory information processing - PubMed. Int J Audiol. 2003; 42:26–32. https://pubmed.ncbi.nlm.nih.gov/12918626/, doi: 10.3109/14992020309074641.

Picton TW, Hillyard SA. Human auditory evoked potentials. II. Effects of attention. Electroencephalography and clinical neurophysiology. 1974; 36(2):191–200. https://pubmed.ncbi.nlm.nih.gov/4129631/, doi: 10.1016/0013-4694(74)90156-4.

Picton TW, Hillyard SA, Galambos R, Schiff M. Human auditory attention: a central or peripheral process? Science (New York, NY). 1971; 173(3994):351–353. https://pubmed.ncbi.nlm.nih.gov/5559908/, doi: 10.1126/SCIENCE.173.3994.351.

Pivik RT, Harman K. A reconceptualization of EEG alpha activity as an index of arousal during sleep: all alpha activity is not equal. Journal of sleep research. 1995; 4(3):131–137. https://pubmed.ncbi.nlm.nih.gov/10607151/, doi: 10.1111/J.1365-2869.1995.TB00161.X.

Portas CM, Krakow K, Allen P, Josephs O, Armony JL, Frith CD. Auditory processing across the sleep-wake cycle: simultaneous EEG and fMRI monitoring in humans. Neuron. 2000; 28(3):991–999. https://pubmed.ncbi.nlm.nih.gov/11163282/, doi: 10.1016/S0896-6273(00)00169-0.

Price CN, Bidelman GM. Attention reinforces human corticofugal system to aid speech perception in noise. NeuroImage. 2021 7; 235:118014. doi: 10.1016/J.NEUROIMAGE.2021.118014.

Price CN, Moncrieff D. Defining the Role of Attention in Hierarchical Auditory Processing. Audiology Research. 2021 3; 11(1):112. /pmc/articles/PMC8006147//pmc/articles/PMC8006147/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8006147/, doi: 10.3390/AUDIOLRES11010012.

Rinne T, Balk MH, Koistinen S, Autti T, Alho K, Sams M. Auditory selective attention modulates activation of human inferior colliculus. Journal of neurophysiology. 2008 12; 100(6):3323–3327. https://pubmed.ncbi.nlm.nih.gov/18922948/, doi: 10.1152/JN.90607.2008.

Robbins TW, Everitt BJ. Arousal systems and attention. In: *The cognitive neurosciences*, in m. s. gazzaniga ed. The MIT Press.; 1995.p. 703–720. https://psycnet.apa.org/record/1994-98810-044.

Saderi D, Schwartz ZP, Heller CR, Pennington JR, David SV. Dissociation of task engagement and arousal effects in auditory cortex and midbrain. eLife. 2021 2; 10:1–25. doi: 10.7554/ELIFE.60153.

Saiz-Alía M, Forte AE, Reichenbach T. Individual differences in the attentional modulation of the human auditory brainstem response to speech inform on speech-in-noise deficits. Scientific Reports 2019 9:1. 2019 10; 9(1):1–10. https://www.nature.com/articles/s41598-019-50773-1, doi: 10.1038/s41598-019-50773-1.

Scherg M, Ebersole JS. Brain source imaging of focal and multifocal epileptiform EEG activity. Neurophysiologie clinique = Clinical neurophysiology. 1994; 24(1):51–60. https://pubmed.ncbi.nlm.nih.gov/8121338/, doi: 10.1016/S0987-7053(05)80405-8.

Scherg M, Ille N, Bornfleth H, Berg P. Advanced tools for digital EEG review: virtual source montages, whole-head mapping, correlation, and phase analysis. Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society. 2002; 19(2):91–112. https://pubmed.ncbi.nlm.nih.gov/11997721/, doi: 10.1097/00004691-200203000-00001.

Shaheen LA, Slee SJ, David SV. Task Engagement Improves Neural Discriminability in the Auditory Midbrain of the Marmoset Monkey. Journal of Neuroscience. 2021 1; 41(2):284–297. https://www.jneurosci.org/content/41/2/284https://www.jneurosci.org/content/41/2/284.abstract, doi: 10.1523/JNEUROSCI.1112-20.2020.

Shiga T, Althen H, Cornella M, Zarnowiec K, Yabe H, Escera C. Deviance-Related Responses along the Auditory Hierarchy: Combined FFR, MLR and MMN Evidence. PLOS ONE. 2015 9; 10(9):e0136794. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0136794, doi: 10.1371/JOURNAL.PONE.0136794.

Skoe E, Kraus N. Auditory brain stem response to complex sounds: A tutorial. Ear and Hearing. 2010 6; 31(3):302–324. https://journals.lww.com/ear-hearing/Fulltext/2010/06000/Auditory_Brain_Stem_Response_to_Complex_Sounds__A.2.aspx, doi: 10.1097/AUD.0B013E3181CDB272.

Smith JC, Marsh JT, Brown WS. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. Electroencephalogr Clin Neurophysiol. 1975; 39:465–72. http://www.ncbi.nlm.nih.gov/pubmed/52439.

Song JH, Skoe E, Banai K, Kraus N. Perception of Speech in Noise: Neural Correlates. Journal of Cognitive Neuroscience. 2011 9; 23(9):2268. /pmc/articles/PMC3253852//pmc/articles/PMC3253852/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3253852/, doi: 10.1162/JOCN.2010.21556.

Strauß A, Wöstmann M, Obleser J. Cortical alpha oscillations as a tool for auditory selective inhibition. Frontiers in human neuroscience. 2014 5; 8(MAY). https://pubmed.ncbi.nlm.nih.gov/24904385/, doi: 10.3389/FNHUM.2014.00350.

Suga N, Ma X. Multiparametric corticofugal modulation and plasticity in the auditory system. Nature reviews Neuroscience. 2003; 4(10):783–794. https://pubmed.ncbi.nlm.nih.gov/14523378/, doi: 10.1038/NRN1222.

Tichko P, Skoe E. Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators. Hearing research. 2017 5; 348:1–15. https://pubmed.ncbi.nlm.nih.gov/28137699/, doi: 10.1016/J.HEARES.2017.01.014.

Uusberg A, Uibo H, Kreegipuu K, Allik J. EEG alpha and cortical inhibition in affective attention. International Journal of Psychophysiology. 2013 7; 89(1):26–36. doi: 10.1016/J.IJPSYCHO.2013.04.020.

Varghese L, Bharadwaj HM, Shinn-Cunningham BG. Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. Brain research. 2015 11; 1626:146–164. https://pubmed.ncbi.nlm.nih.gov/26187756/, doi: 10.1016/J.BRAINRES.2015.06.038.

Vollmer M, Beitel RE, Schreiner CE, Leake PA. Passive stimulation and behavioral training differentially transform temporal processing in the inferior colliculus and primary auditory cortex. Journal of neurophysiology. 2017 1; 117(1):47–64. https://pubmed.ncbi.nlm.nih.gov/27733594/, doi: 10.1152/JN.00392.2016.

White-Schwoch T, Anderson S, Krizman J, Nicol T, Kraus N. Case studies in neuroscience: subcortical origins of the frequency-following response. Journal of Neurophysiology. 2019 8; 122(2):844–848. https://journals.physiology.org/doi/abs/10.1152/jn.00112.2019, doi: 10.1152/JN.00112.2019/ASSET/IMAGES/LARGE/Z9K0081951590003.JPEG.

Wilsch A, Henry MJ, Herrmann B, Maess B, Obleser J. Alpha Oscillatory Dynamics Index Temporal Expectation Benefits in Working Memory. Cerebral cortex (New York, NY : 1991). 2015 7; 25(7):1938–1946. https://pubmed.ncbi.nlm.nih.gov/24488943/, doi: 10.1093/CERCOR/BHU004.

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

Xie Z, Reetzke R, Chandrasekaran B. Machine Learning Approaches to Analyze Speech-Evoked Neurophysiological Responses. Journal of speech, language, and hearing research : JSLHR. 2019 3; 62(3):587–601. https://pubmed.ncbi.nlm.nih.gov/30950746/, doi: 10.1044/2018{_}JSLHR-S-ASTM-18-0244.

Yordanova JY, Kolev VN, Başar E. EEG theta and frontal alpha oscillations during auditory processing change with aging. Electroencephalography and clinical neurophysiology. 1998 9; 108(5):497–505. https://pubmed.ncbi.nlm.nih.gov/9780020/, doi: 10.1016/S0168-5597(98)00028-8.

Zhao CT, Kuhl PK. Linguistic effect on speech perception observed at the brainstem. Proceedings of the National Academy of Sciences of the United States of America. 2018 8; 115(35):8716–8721. https://www.pnas.org/content/115/35/8716https://www.pnas.org/content/115/35/8716.abstract, doi: 10.1073/PNAS.1800186115/-/DCSUPPLEMENTAL.

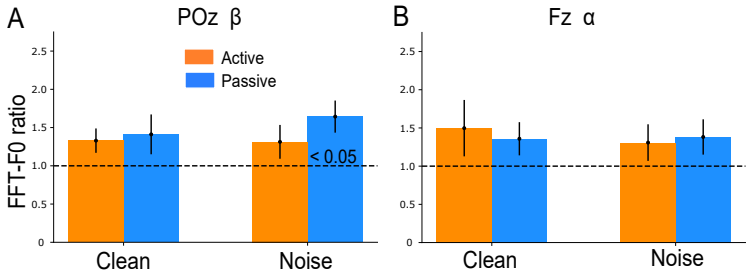

Figure 3–Figure supplement 1. No significant noise effect is observed in F0 ratio when comparing F0 amplitudes of high to low β in POz channel (A) or α in Fz channel (B). <0.05 in bars indicates significantly larger than 1 (1-sample t-test).

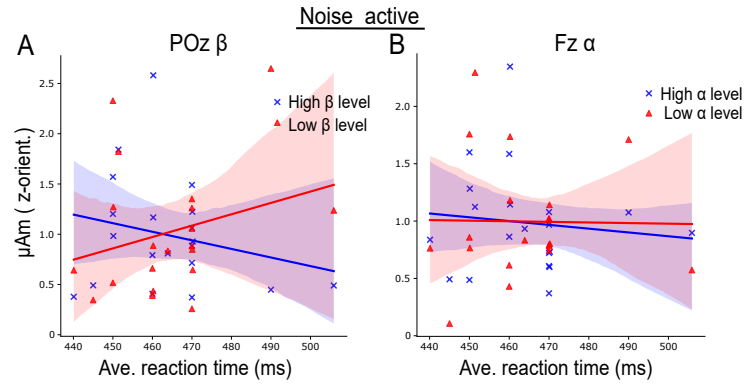

Figure 4–Figure supplement 1. No significant association is observed between F0 amplitudes of FFRs at low (red line) or high (blue line) cortical activity level and reaction time when analyzing β band from POz channel (A) or α band from Fz channel (B). Shaded area indicates 95 % confidence intervals of the regression line.

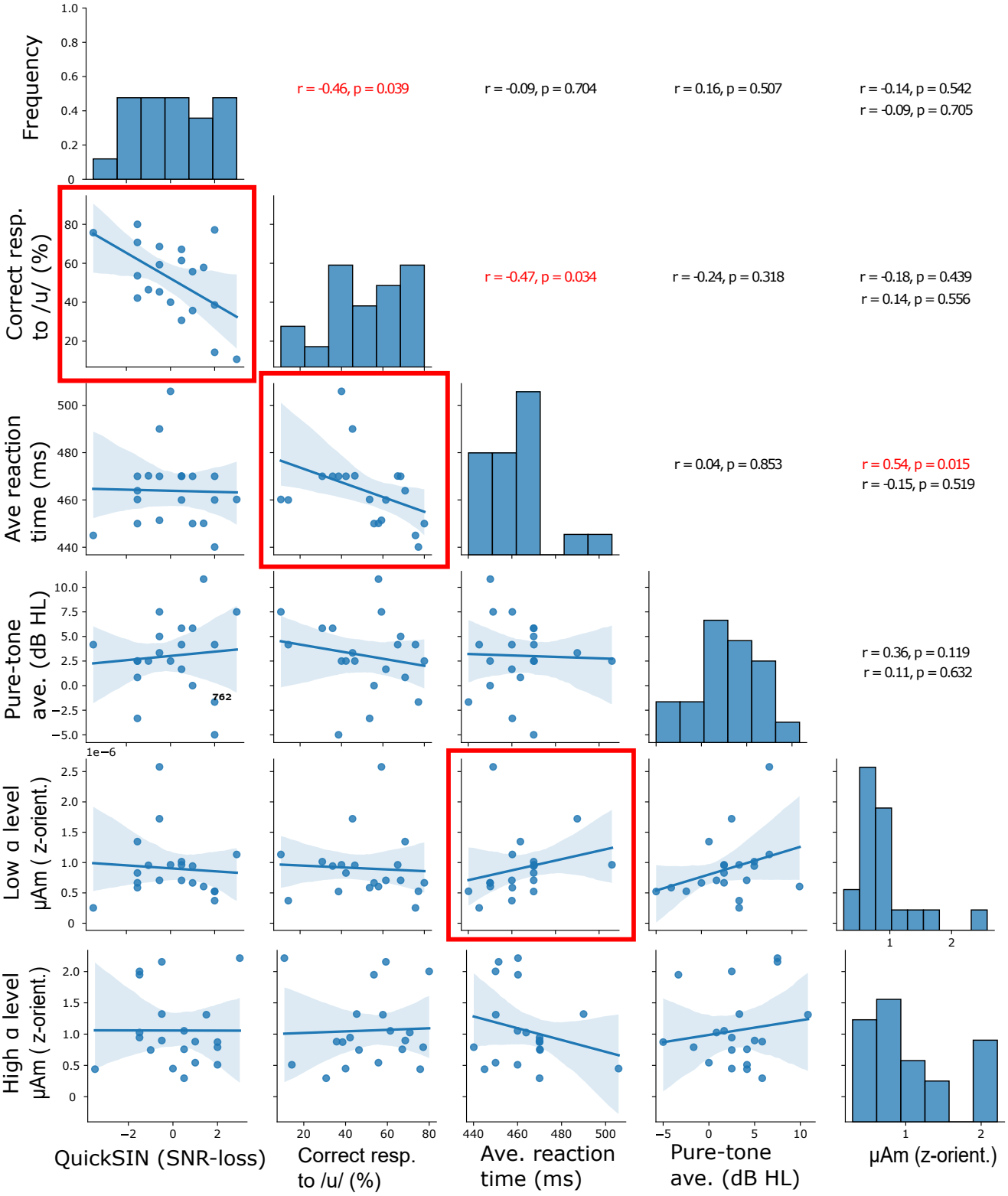

Figure 4–Figure supplement 2. Indirect relationships between FFR F0 amplitudes at low/high α power and behavioral auditory performance are observed in the noise active condition. Correlations that are statistically significant ($p < 0.05$) are highlighted with red square boxes, and the corresponding r and p-values are shown in red texts. 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.