NeuroImage 235 (2021) 118014

Contents lists available at ScienceDirect

NeuroImage

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

Attention reinforces human corticofugal system to aid speech perception in noise

Caitlin N. Price[a,b,*], Gavin M. Bidelman[a,b,c]

[a] Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
[b] School of Communication Sciences and Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA
[c] Department of Anatomy and Neurobiology, University of Tennessee Health Sciences Center, Memphis, TN, USA

ARTICLE INFO

Keywords:
EEG
Functional connectivity
FFR
Corticofugal tuning
Hierarchical speech processing

ABSTRACT

Perceiving speech-in-noise (SIN) demands precise neural coding between brainstem and cortical levels of the hearing system. Attentional processes can then select and prioritize task-relevant cues over competing background noise for successful speech perception. In animal models, brainstem-cortical interplay is achieved via descending corticofugal projections from cortex that shape midbrain responses to behaviorally-relevant sounds. Attentional engagement of corticofugal feedback may assist SIN understanding but has never been confirmed and remains highly controversial in humans. To resolve these issues, we recorded source-level, anatomically constrained brainstem frequency-following responses (FFRs) and cortical event-related potentials (ERPs) to speech via high-density EEG while listeners performed rapid SIN identification tasks. We varied attention with active vs. passive listening scenarios whereas task difficulty was manipulated with additive noise interference. Active listening (but not arousal-control tasks) exaggerated both ERPs and FFRs, confirming attentional gain extends to lower subcortical levels of speech processing. We used functional connectivity to measure the directed strength of coupling between levels and characterize "bottom-up" vs. "top-down" (corticofugal) signaling within the auditory brainstem-cortical pathway. While attention strengthened connectivity bidirectionally, corticofugal transmission disengaged under passive (but not active) SIN listening. Our findings (i) show attention enhances the brain's transcription of speech even prior to cortex and (ii) establish a direct role of the human corticofugal feedback system as an aid to cocktail party speech perception.

1. Introduction

Complex speech-in-noise (SIN) perception requires the listener to isolate a talker's voice from competing background noise. Extraction of cues necessary for SIN perception relies on accurate and efficient bottom-up perceptual and top-down cognitive processes. Bottom-up encoding serves to provide rich neural representations of acoustic features while higher-order cognitive processes enable the prioritization and integration of inputs for comprehension. A breakdown at any point within this integrated system could lead to SIN deficits (Humes, 1996; Moore, 2015).

Attention is a top-down cognitive process that alerts and orients listeners to focus concentration on environmental stimuli (Petersen and Posner, 2012). 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. Given that top-down mechanisms fine-tune auditory neural coding (Atiani et al., 2009; Gao and Suga, 2000; Suga and Ma, 2003), attention is thought to influence all stages of auditory processing from the inner ear to cortex (Galbraith et al., 2003; Giard et al., 1994; Hernandez-Peon, 1966; Lukas, 1980; Picton and Hillyard, 1974; Rinne et al., 2008).

Attentional modulation of auditory cortical activity is well documented (Alho et al., 2014; Alho and Vorobyev, 2007; Picton and Hillyard, 1974), but whether similar enhancements exist at earlier stages prior to cortex (e.g., brainstem) remains contentiously debated. Human brainstem responses to nonspeech stimuli are largely invariant to attentional state (Hirschhorn and Michie, 1990; Picton and Hillyard, 1974; Picton et al., 1971; Woods and Hillyard, 1978). Newer contradictory findings have emerged from more recent studies on frequency-following responses (FFRs)—microphonic-like potentials generated predominantly from brainstem (i.e., subcortical structures including cochlear nucleus, medial geniculate body of the thalamus, and inferior colliculus) which index neural phase-locking to dynamic sound features (Bidelman, 2018; Coffey et al., 2019; Gardi et al., 1979;

* Corresponding author at: School of Communication Sciences and Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA.
E-mail addresses: cenelms@memphis.edu (C.N. Price), gmbdlman@memphis.edu (G.M. Bidelman).

https://doi.org/10.1016/j.neuroimage.2021.118014
Received 24 December 2020; Received in revised form 9 March 2021; Accepted 25 March 2021
Available online 29 March 2021

Marsh et al., 1970; Smith et al., 1975; Tichko and Skoe, 2017; White-Schwoch et al., 2019). Some electrophysiological studies suggest attention enhances the robustness and temporal precision of speech FFRs (Forte et al., 2017; Galbraith et al., 2003; Hartmann and Weisz, 2019; Lehmann and Schonwiesner, 2014). Still, others demonstrate mixed (Holmes et al., 2018; Saiz-Alia et al., 2019) or even null attention-related FFR effects (Galbraith and Kane, 1993; Varghese et al., 2015). Brainstem responses are reliably recorded during sleep and sedation (Skoe and Kraus, 2010a) which bolsters long-held assumptions that subcortical processing is largely pre-attentive and automatic (Tzounopoulos and Kraus, 2009). Attention effects at the brainstem level might be too subtle to detect in scalp EEG (Varghese et al., 2015) or require highly specialized/unnatural perceptual tasks that overly challenge speech processing (Galbraith and Arroyo, 1993; Lehmann and Schonwiesner, 2014). Equivocal findings have even led to assertations that “efferent-mediated inhibitory mechanisms have [no] role to play in selective attention” and do not “suppress irrelevant information at early stages of the auditory system” (Hirschhorn and Michie, 1990, p. 507).

Architecturally, the auditory neuroaxis contains both afferent (ear-to-brain) and efferent (brain-to-ear) projections. Germane to the present experiments, descending cortico-collicular (i.e., corticofugal)[1] fibers from primary auditory cortex (PAC) with targets in brainstem (BS; i.e., inferior colliculus) recalibrate sound processing of midbrain neurons by fine-tuning their receptive fields in response to behaviorally relevant stimuli (Suga, 2008; Suga et al., 2000). Corticofugal efferents also drive learning-induced plasticity in animals (Bajo et al., 2010; Suga et al., 2000). Given the midbrain (upper BS) is the primary source of human scalp FFRs (Bidelman, 2018; Bidelman and Momtaz, 2021; Marsh et al., 1970; Saiz-Alia and Reichenbach, 2020; White-Schwoch et al., 2019), these efferents may account for experience-dependent plasticity observed in seminal human FFR studies (Chandrasekaran and Kraus, 2010; Kraus and White-Schwoch, 2015; Musacchia et al., 2007; Wong et al., 2007). Despite experience-dependent changes observed in human FFRs and ample evidence for online subcortical modulation in animals (Bajo et al., 2010; Slee and David, 2015; Suga et al., 2000; Vollmer et al., 2017), there have been no direct measurements of corticofugal system function in humans. Theoretically, efferent control of brainstem activity should occur for behaviorally relevant stimuli (Suga, 2008), in states of goal-directed attention (Slee and David, 2015; Vollmer et al., 2017), and strengthen in more taxing listening conditions (e.g., SIN tasks) (Ahissar and Hochstein, 2004). Attention-based neuronal feedback provided by the descending hearing system could help refine early sensory speech representations and assist the brain in navigating complex listening environments.

Here, we aimed to elucidate two highly disputed topics regarding our understanding of the human auditory system. First, we aimed to determine the degree to which attention actively reinforces early neural representations for speech prior to neocortex. To this end, we recorded high-density EEGs as listeners performed rapid SIN tasks varying in attentional demand. To overcome challenges of prior work, we used anatomically constrained source imaging to disentangle speech-evoked responses generated from brainstem (FFR) and cortex (ERP). This approach allowed us to jointly index attentional gain in speech processing at both stages of the hearing pathway and evaluate hierarchical processing in auditory attention. Secondly, we aimed to quantify human corticofugal function in SIN perception. We used functional connectivity to characterize “bottom-up” vs. “top-down” (corticofugal) signaling within the brainstem-cortical pathway. Under the premise that the corticofugal system shapes brainstem signal processing only for perceptually taxing and behaviorally-relevant sounds (i.e., "ego-centric selection," Ahissar and Hochstein, 2004; Suga et al., 2000), we predicted stronger PAC→BS connectivity for noise-degraded relative to clear conditions. Furthermore, if speech perception conforms to early “sensory gating” and “bottleneck” models of attentional locus (Broadbent, 1971; Hirschhorn and Michie, 1990; Picton et al., 1971; Woods and Hillyard, 1978), we hypothesized attention would enhance brainstem phase-locking to speech as measured via source-resolved FFRs (with similar or stronger gains in cortical ERPs). Our results provide the first direct evidence that attention reinforces brainstem auditory processing in humans via corticofugal signaling and confirm this feedback is critical for cocktail party speech perception.

[1] EEG does not afford the level of granularity necessary to assess the relative roles of sub-nuclei or associated interconnecting neural fibers comprising the auditory afferent/efferent pathways. With source and functional connectivity analyses, we are able to evaluate and interpret activity within and between gross anatomical levels (midbrain and auditory cortex), so our analyses should be taken as broad measures of neural encoding and transmission rather than reflecting activity from and between unitary nuclei within the pathway.

2. Materials and methods

2.1. Participants

Twenty young adults (age: 18–35 years, $M = 24$, $SD = 3.4$ years; 11 female) participated in the study. An a priori power analysis (t-test, 2-tailed, $\alpha = 0.05$, power = 95%) revealed this sample was sufficient to detect similar sized effects ($d = 0.84$-1.0) as in previous FFR/ERP SIN studies (Bidelman et al., 2019; Faul et al., 2007). All participants exhibited normal hearing thresholds (≤ 25 dB HL; 250–8000 Hz). Because language background and music experience influence FFRs/ERPs and SIN performance (Mankel and Bidelman, 2018; Parbery-Clark et al., 2009; Zhao and Kuhl, 2018), we required participants to have < 3 years of formal musical training ($M = 0.8$ years, $SD = 1.2$) and be native English speakers. They were monolingual except for 2 participants who rated their second language (e.g., Spanish, ASL) proficiency as functional-to-good. Participants were predominantly right-handed ($M = 82.04\%$, $SD = 21.04$) (Oldfield, 1971) with no history of neuropsychiatric disorders. All provided written informed consent prior in accordance with the Declaration of Helsinki for experiments involving humans and joint EEG and MRI protocols approved by the University of Memphis IRB.

2.2. Speech stimuli & task

We recorded FFRs and ERPs simultaneously (Bidelman et al., 2013, 2019) during SIN-listening tasks designed to evaluate attentional effects on the brain’s hierarchical encoding of speech. Three synthesized vowel tokens (e.g., /a/, /i/, /u/) were presented during the recording of EEGs. We chose vowels as sustained periodic sounds optimally evoke FFRs and ERPs (Bidelman et al., 2019; Skoe and Kraus, 2010a). 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). Fig. S1 confirms FFRs to these high-frequency stimuli were absent from PAC and occurred only in the BS source. Tokens were matched in average root mean square (r.m.s.) amplitude. We presented the vowels in clean (i.e., no background noise) and noise-degraded conditions. For the noise condition, vowel stimuli were mixed with 8 talker noise babble (cf. 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).

Attention was varied via active vs. passive listening blocks. During active blocks, participants detected infrequent tokens (/u/) via button press. We defined a “hit” as detection within 5 token presentations of a target. 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). We included a control block ($n = 8$ of the 20 participants) of visual-only stimulation to rule out potential visual confounds inherent to the passive condition. In these runs, the headphone

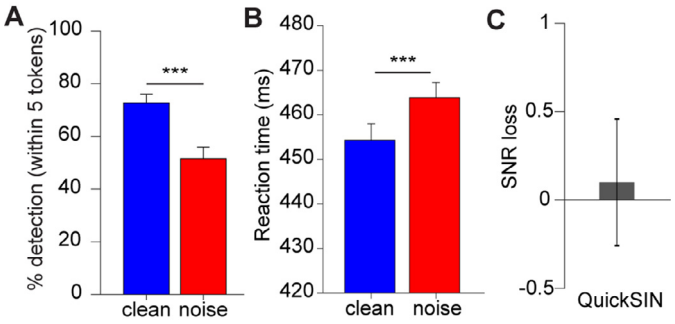

Fig. 1. Behavioral target speech detection is hindered by noise. (**A**) Behavioral accuracy and (**B**) reaction times for detecting infrequent /u/ tokens in clean and noise-degraded conditions. Noise hinders speech perception by reducing perceptual accuracy and slowing decision speeds. **(C)** Average QuickSIN scores across participants. errorbars = ± s.e.m., $^{***}p < 0.001$.

transducers were unplugged from the wall, and participants watched the captioned movie (i.e., as in passive block but muted audio). The success of these controls was confirmed by the absence of any time-locked response in the control condition indicating no visual artifacts in the recordings (see Fig. 2A, B).

Each attentional block and noise condition (e.g., active clean, active noise, passive clean, passive noise) 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). The control block consisted of a single run of clean and noise conditions (e.g., control clean, control noise). Block (active/passive/control) and condition (clean/noisy speech) presentation was counterbalanced across participants to minimize order and fatigue effects. During breaks, participants were encouraged to stretch and rest. No additional cognitive tasks were performed during breaks. 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 EEG recordings (Fig. S2).

2.3. QuickSIN test

The Quick Speech-in-Noise (QuickSIN) test assessed listeners' speech reception thresholds 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. We presented target sentences at 70 dB SPL (binaurally) at SNRs decreasing in 5 dB steps from 25 dB (very easy) to 0 dB (very 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. We averaged scores from two lists per listener (Fig. 1C).

2.4. EEG acquisition and preprocessing

We recorded EEGs from 64-channels at 10-10 electrode locations across the scalp (Oostenveld and Praamstra, 2001). During EEG acquisition, the ground electrode was placed ~1 cm anterior to Fz, and the reference electrode was located just posterior to Cz. 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 (ERP) frequency components of the compound speech-evoked potential (Bidelman et al., 2013; Musacchia et al., 2008).

We processed EEG data in Curry 7 (Compumedics Neuroscan) and BESA Research v7.0 (BESA, GmbH). Ocular artifacts (saccades and blinks) were corrected in continuous EEGs using principal component analysis (PCA) (Picton et al., 2000). Cleaned EEGs were epoched (-10–200 ms), pre-stimulus baselined, and ensemble averaged to obtain compound speech-evoked potentials (Bidelman et al., 2013). We bandpass filtered full-band responses from 130 to 1500 Hz and 1 to 30 Hz (zero-phase Butterworth filters; slope = 48 dB/octave) to isolate FFRs and ERPs, respectively (Bidelman et al., 2013; Musacchia et al., 2008). Unless otherwise noted (cf. scalp FFR analysis), data were re-referenced to the common average of all electrodes. We excluded infrequent /u/ tokens from the analyses due to their limited number of trials and to avoid mismatch negativities in the data. We co-registered the functional EEG data to each individual's MRI anatomy to perform source analysis (detailed in Sections 2.5 and 2.6).

2.5. MRI scans and EEG co-registration

3D T1-weighted anatomical volumes were obtained on a Siemens 1.5T Symphony TIM scanner (tfl3d1 GR/IR sequence; TR=2000 ms, TE=3.26 ms, inversion time=900 ms, phase encoding steps=341, flip angle=8°, FOV=256 × 256 acquisition matrix, 1.0 mm axial slices). Scanning was conducted at the Semmes Murphey Neurology Clinic (Memphis, TN). Scans were segmented in BESA MRI 2.0. Following inhomogeneity correction (Scherg et al., 2002), images were automatically partitioned into scalp, skull, CSF, and brain compartments (Chan and Vese, 2001) and the cortical surface was reconstructed to allow optional inflation of the brain volume (Fischl et al., 1999). MRI volumes were rendered in both ACPC and Talairach and Tournoux (1988) spaces using 3D spline interpolation.

Following MRI segmentation, electrode locations were warped to the scalp surface (anchored to the nasion and preauricular fiducials) to co-register sensor locations to each individual's anatomy. Electrode positions were mapped with a quad sensor Polhemus Fastrak digitizer (Polhemus, Colchester, VT). We then generated a 4-layer finite element head model (FEM) based on the MRI segmentation (Wolters et al., 2007) to construct each individual's leadfield (forward volume conductor). We used relative conductivities (S/m) of 0.33, 0.0042, 1.79, 0.33 (defaults for adult brain; BESA MRI 2.0) for the scalp, skull, cerebrospinal fluid, and brain tissue compartments, respectively (Baumann et al., 1997). The FEM leadfield described the magnitude each source signal contributed at each sensor (Scherg, 1990) and is less prone to spatial errors than other head models (e.g., concentric spherical conductor) (Fuchs et al., 2002). Collectively, this approach allowed us to source localize each participant's cortical ERPs and brainstem FFRs with high precision, constrained to their individual brain anatomy. MRIs were not available for 4 participants. In these cases, we used a 4-shell spherical volume conductor head model and an adult template brain (Berg and Scherg, 1994).

2.6. Source analysis

We transformed listeners' scalp potentials to source space using BESA which allowed for connectivity analysis between BS and PAC. We used a virtual source montage (Fig. 2E; adapted from Bidelman, 2018) comprised of regional dipoles (i.e., current flow in x, y, z planes) positioned in the brainstem midbrain (i.e., inferior colliculus) and bilateral PAC based on template brain atlases (Bidelman, 2018).[2] This applied an op-

[2] The dipole model used by Bidelman (2018) was adapted by extracting the BS and PAC dipoles from the original 6 source solution and applied to our data.

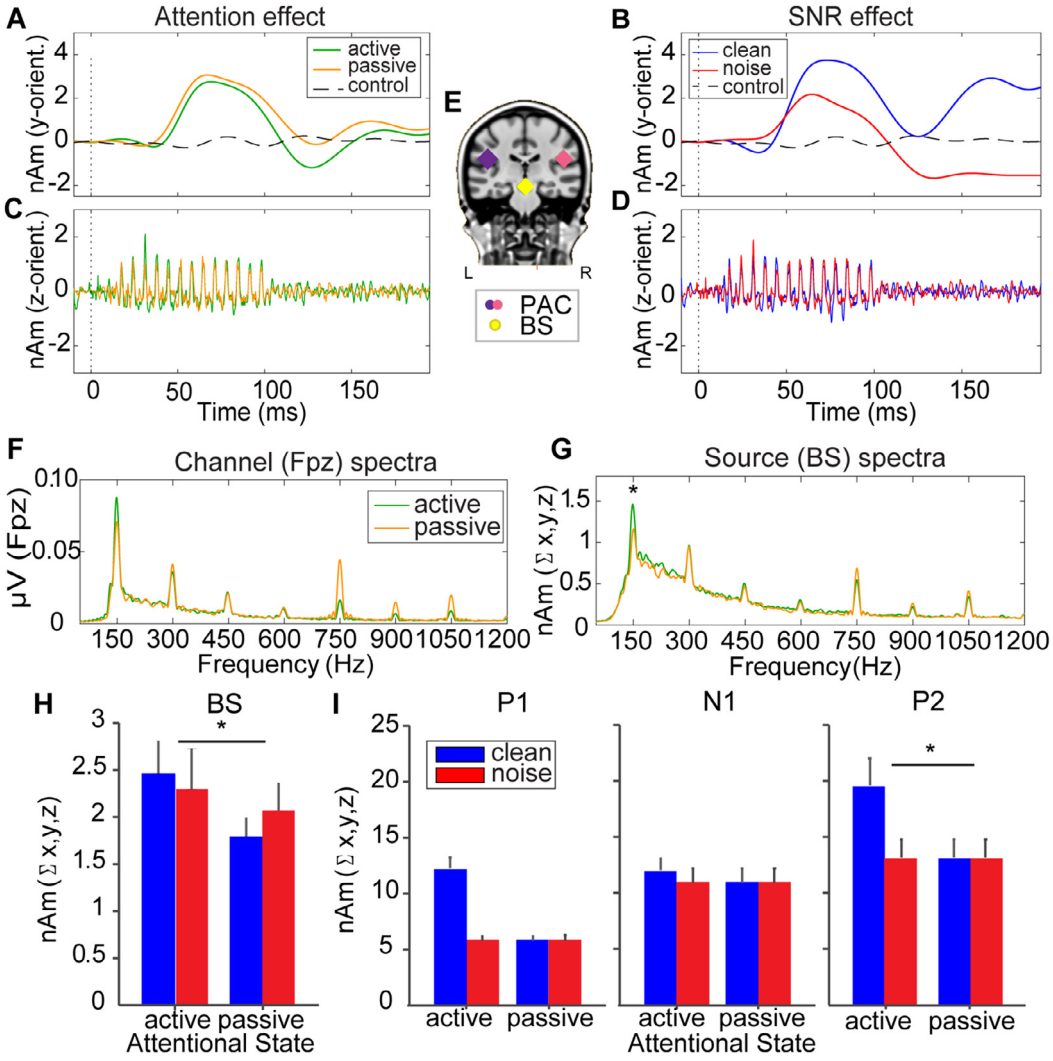

Fig. 2. Attention enhances brainstem (FFR) and cortical (ERP) encoding of noise-degraded speech. (**A-B**) Cortical ERPs and (**C-D**) brainstem FFRs extracted from dipoles located in midbrain and bilateral PAC (**E**, inset brain). For clarity, only one dipole orientation is shown per ROI (collapsing tokens). Noise exerts larger effects on cortical vs. brainstem speech coding. Attention enhances both ERPs (P2; 150 ms) and FFRs. (**F-G**) FFR spectra (collapsing tokens/SNRs) illustrating attentional enhancements of speech-FFRs in source (but not channel-level) data. (**H**) Source FFR-F0 amplitudes are stronger during active listening but are resistant to noise. (**I**) Source ERP magnitudes. Attention increases only P2 magnitudes. errorbars = ± s.e.m, $^{*}p < 0.05$. BS, brainstem; PAC, primary auditory cortex.

timized spatial filter to all electrodes that calculated their weighted contribution to the scalp-recorded FFRs/ERPs in order to estimate activity within each source location within the head (Scherg and Ebersole, 1994; Scherg et al., 2002).[3] For each cortical dipole source, we combined activity from the three orientations—computed as the L2-norm of x, y, z time waveforms—to yield an unbiased measure of the aggregate response within each ROI. For source FFRs, following Coffey et al, (2016). We summed the three orientations in the frequency domain since FFRs are most readily analyzed spectrally (Coffey et al., 2016, 2017). Across stimulus conditions, average goodness of fit (GoF) for our 3-dipole model was 90.6% [residual variance (RV) = 9.4 ± 1.6%], indicating excellent fit to the scalp data.

2.7. FFR and ERP quantification

2.7.1. Brainstem FFRs

We analyzed the steady-state portion (10–100 ms) of FFR waveforms using a Fast Fourier transform (FFT; Gaussian window; 11 Hz frequency resolution) which captured the spectral composition of the response. F0 amplitude was quantified as the maximum FFT amplitude within a 10 Hz bin centered around 150 Hz (i.e., F0 of the stimuli). FFR F0 indexes voice pitch coding and predicts successful SIN perception

That original model was fit to template brains so anatomical locations reflect normalized (i.e., atlas) locations. Identical dipole locations were utilized in this study so no "fitting" was performed, *per se*. The main difference is that we improved the approach by (i) using each individual's MRI (rather than a template brain) to construct their own FEM leadfield and thereby improve source reconstruction accuracy and (ii) converting each single dipole to a regional source to consider activity in all three (x, y, z) dipole orientations from each ROI. This approach provided more accurate source estimates because it exploited each individual's brain anatomy and considered all possible activity from each source (i.e., current flow in all directions).

[3] The process is essentially an inverse spatial filter which computes the source waveforms (currents) at each dipole location as SWF = L^{-1} x ERP, where ERP is the scalp voltage data [64 (electrodes) x 1051 (samples)] and L is the leadfield matrix [64 (weights) x 9 (3 dipoles per x, y, z orientation)] describing the weights of how each source location contributes to each scalp electrode. L is defined by the dipole model (anatomical locations + orientations and conductivities of each brain compartment layer).

(Mankel and Bidelman, 2018; Parbery-Clark et al., 2009). We analyzed FFRs at both the electrode (Fpz; linked mastoid reference; Bidelman, 2015; Bidelman and Momtaz, 2021; Krishnan, 2002) and source level to compare our findings to previous literature investigating attentional effects on FFRs.

2.7.2. Cortical ERPs

We quantified source-level ERP wave (i.e., P1, N1, P2) amplitude and latency using automated peak analysis. Latency windows were determined following visual inspection of grand average traces. P1 was identified as the maximum positive deflection occurring within 40–80 ms; N1 as the greatest negative deflection between 90 and 145 ms; and P2 as the maximum positive deflection within 145–175 ms (Hall, 1992).

2.8. Functional connectivity

We measured functional connectivity between PAC and BS source waveforms using phase transfer entropy (PTE), a measure of nonlinear, directed (causal) signal dependency (Bidelman et al., 2018, 2019; Price et al., 2019).[4] For details, see Lobier et al., 2014 and Bidelman et al., 2019. PTE was calculated using a Matlab function distributed in the Brainstorm package (Tadel et al., 2011).[5] PTE was computed in both directions to quantify differences in the strength of bottom-up afferent (BS→PAC) vs. top-down efferent (PAC→BS) connectivity within the auditory brainstem-cortical pathway. This allowed us to evaluate how attending to auditory signals and listening in more adverse conditions influenced bidirectional signaling including critical corticofugal (PAC→BS) function.

2.9. Statistical analyses

We performed 2 × 2 × 2 (vowel x attention x SNR) mixed model (subjects=random factor) ANOVAs (GLIMMIX, SAS® 9.4, SAS Institute; Cary, NC). Tukey-Kramer adjustments corrected multiple comparisons. Initial diagnostics were performed using residual and Q-Q plots to assess heteroscedasticity and normality of data. Unless otherwise noted, data were log-transformed to improve normality and homogeneity of variance assumptions. Effect sizes are reported as η_p^2. Paired samples *t*-tests (two-tailed) were used to compare behavioral performance between conditions. Generalized linear models (GLM) evaluated brain-behavior relationships. Neural measures from each ROI (FFR_{BS}: F0 amplitude; ERP_{PAC}: P2 magnitude), connectivity (afferent, efferent), behavioral [QuickSIN, pure-tone average (PTA) hearing thresholds], and stimulus factors (SNR) were included as predictors of (i) target detection accuracy and (ii) reaction time (RT) for the active SIN detection task. Two models were analyzed, one for each outcome variable (i.e., accuracy, RT). Variance inflation factors (VIFs) ranged from 1.01 to 1.32 for accuracy and 1.0–1.6 for RT indicating negligible multicollinearity among the predictors (Lüdecke, 2020).

[4] PTE results in scalars (given in unit bits) describing the strength of connectivity between each pair of ROIs in the A→B and B→A directions (i.e., 2 x 2 matrix). Pertinent to the present study, the bi-directionality of the metric enabled us to evaluate strength of neural signaling in both afferent (BS→PAC) and efferent (PAC→BS) directions. Similar to a correlation, PTE of 0 indicates completely random activity and negligible connectivity between given ROIs. However, unlike correlations, PTE cannot be negative and has no upper bound. The asymmetry of the metric (i.e., A→B ≠ B→A) allows testing of directed signal dependencies. Larger PTE values indicate stronger connectivity between ROIs.

[5] Available at https://github.com/brainstormtools/brainstorm3/blob/master/external/fraschini/PhaseTE_MF.m

3. Results

3.1. Behavioral data

Noise expectedly reduced listeners' target speech detection accuracy (t_{19} = 7.32, p < 0.001) and slowed response speeds (t_{19} = -4.23, p < 0.001), confirming poorer SNRs were detrimental to speech perception (Fig. 1A, B). False alarm rates were low for both clean and noise conditions (clean: 11.9 ± 1.1%, noise: 10.2 ± 2.8%). Moreover, pooling SNRs, both accuracy (t_{19} = 0.54, p = 0.59) and RTs (t_{19} = -1.87, p = 0.08) showed no difference when comparing performance between the first and second halves of the active trials. This suggests that listeners maintained vigilance throughout the experiment.

3.2. Electrophysiological data

Source-level FFRs and ERPs reflecting brainstem and cortical activity contrast the effects of attentional state (Fig. 2A, C) and noise (Fig. 2B, D) on the neural encoding of speech during the perceptual task.

3.2.1. Brainstem FFRs

To determine if attention modulates brainstem SIN processing, we measured FFR F0 amplitudes, a neural proxy of voice pitch encoding that serves as an important cue for tracking speech in background noise (Assmann, 1996; Mankel and Bidelman, 2018). We analyzed FFRs at both electrode and source levels to replicate previous literature (cf. Galbraith and Kane, 1993; Galbraith et al., 2003; Holmes et al., 2018; Lehmann and Schonwiesner, 2014; Varghese et al., 2015). Noise weakened scalp FFRs (Fig. S3; $F_{1, 136}$ = 8.15, p = 0.01, η_p^2 = 0.06) as in previous studies. Critically, we found no evidence for attentional effects in scalp (electrode-level) data (Fig. 2F; $F_{1, 136}$ = 1.68, p = 0.20), corroborating prior null attention effects on FFRs (Picton and Hillyard, 1974; Picton et al., 1971; Woods and Hillyard, 1978). Contrastively, *source-level* FFRs were enhanced during active listening (Fig 2G, H; $F_{1, 136}$ = 5.39, p = 0.02, η_p^2 = 0.04) and were invariant to noise (Fig. S3; $F_{1, 136}$ = 2.22, p = 0.14). These findings reveal attention modulates brainstem speech encoding but only when viewed at the source level.

3.2.2. Cortical ERPs

For cortical source ERPs (i.e., activity from PAC), pooling hemispheres, P1 and P2 magnitudes decreased with noise (P1: $F_{1, 136}$ = 77.33, p < 0.0001, η_p^2 = 0.36; P2: $F_{1, 136}$ = 6.21, p = 0.01, η_p^2 = 0.04), whereas N1 differences were negligible ($F_{1, 136}$ = 0.04, p = 0.84) (Fig. 2I). Critically, attention enhanced P2 ($F_{1, 136}$ = 4.97, p = 0.03, η_p^2 = 0.04), confirming its role in active listening (Naatanen, 1975; Picton and Hillyard, 1974) and a biomarker of SIN perception (Bidelman et al., 2019). Although attentional enhancements appear most evident in the clean condition, we found no significant attention x SNR interaction ($F_{1, 136}$ = 0.28, p = 0.60). Thus, we cannot conclude that attentional enhancement differed across noise conditions. No other significant effects or interactions were found.

3.3. Brainstem-cortical functional connectivity

We used phase-transfer entropy (PTE) to quantify directional functional connectivity and evaluate effects of attention and task demands on bottom-up (i.e., afferent) and top-down (i.e., efferent) neural signaling within the brainstem-cortical pathway (see Section 2.8). This analysis was conducted on untransformed data since initial diagnostics for parametric tests were satisfied. However, results were qualitatively similar with and without transformation. All PTE values differed from zero (i.e., were greater than random noise) indicating significant connectivity strength between ROIs (all p < 0.0001). Connectivity was stronger during active vs. passive listening overall (Fig. 3A, C; attention main effect: $F_{1, 295}$ = 7.92, p = 0.01, η_p^2 = 0.03). More critically, we found a direction

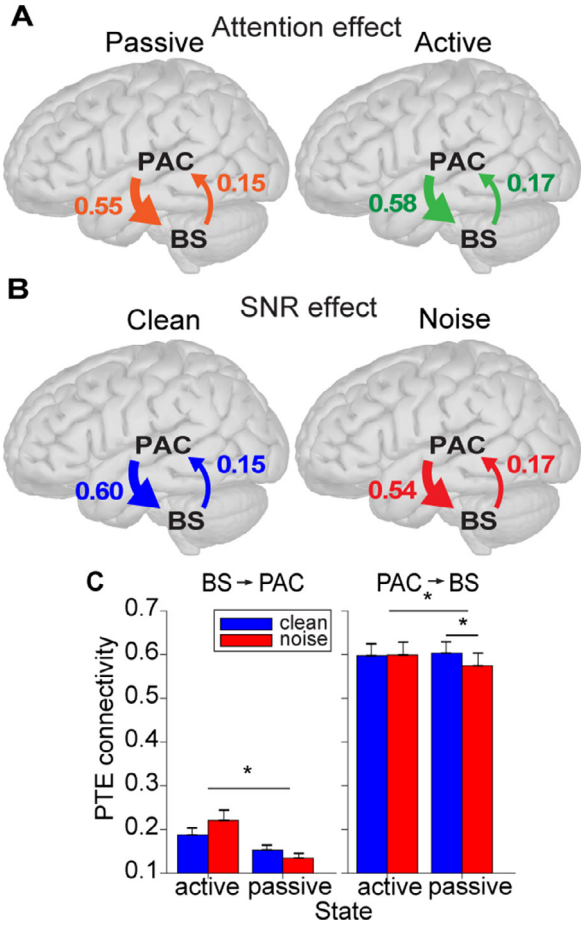

Fig. 3. Attention and noise modulate afferent and efferent connectivity between brainstem and cortex during speech perception. Functional connectivity measured via phase-transfer entropy between BS and PAC source waveforms. Main effect of (A) attention and (B) SNR on connectivity strength. (C) Attention increases bidirectional communication between BS and PAC. Corticofugal efferent connectivity decreases in noise, but only during passive listening. Attention maintains efferent signaling in more challenging listening conditions. errorbars = ± s.e.m, $^{*}p < 0.05$.

x SNR interaction on connectivity strength (Fig. 3B, C; $F_{1,\ 295} = 11.59$, $p < 0.001$, $\eta_p^2 = 0.04$). Efferent (PAC→BS) connectivity was more robust for clean speech and weakened in noise ($t_{295} = 3.32$, $p = 0.01$), whereas afferent (BS→PAC) connectivity was invariant to noise ($t_{295} = -1.50$, $p = 0.44$). Additional planned contrasts revealed noise decreased efferent connectivity during passive ($t_{292} = 3.12$, $p = 0.04$) but not active SIN perception ($t_{292} = 1.57$, $p = 0.77$). These findings reveal corticofugal transmission remained stable during active SIN perception but disengaged while passively coding otherwise identical speech stimuli. No significant difference was noted between active and passive noise conditions for efferent connectivity ($t_{292} = 2.51$, $p = 0.20$).

3.4. Brain-behavior correlations

Having established attentional gain effects within and between BS and PAC, we next used GLM regression to evaluate relations between neural, behavioral, and stimulus factors that drive behavioral SIN processing. For accuracy, the overall multivariate model was highly significant ($F_{7,32} = 11.15$, $p < 0.0001$; $r^2 = 0.86$, $\eta_p^2 = 0.71$). Evaluating individual terms (see Supplementary Table S1) revealed significant predictors in QuickSIN scores (t_{32}= -2.37, $p = 0.02$) and SNR ($t_{32} = -6.53$, $p < 0.0001$). For reaction time, the overall multivariate model was highly significant ($F_{7,32} = 5.04$, $p = 0.001$; $r^2 = 0.89$, $\eta_p^2 = 0.52$). Evaluating individual terms (Table S2) revealed significant predictors in afferent ($t_{32} = 2.34$, $p = 0.03$) and efferent connectivity ($t_{32} = -2.11$, $p = 0.04$) and SNR ($t_{32} = 3.43$, $p = 0.002$). These results confirm the behavioral relevance of BS-PAC connectivity in predicting the speed of listeners' behavioral decisions during our SIN perception task.

4. Discussion

By simultaneously recording speech-evoked brainstem and cortical EEGs and manipulating attentional engagement and task difficulty during active SIN perception, our findings show (i) attention actively modulates SIN processing across the auditory neuroaxis, exerting influences on speech representations as early as the midbrain; and (ii) attention maintains top-down neural communication from PAC to BS in adverse listening conditions. Our neuroimaging results establish a direct role of the human corticofugal feedback system as an aid to cocktail party speech perception.

At the cortical level, we found source ERPs were diminished in noise, confirming acoustic degradation weakens neural representations of speech in PAC (Bidelman et al., 2019; Du et al., 2014). Contrasting cortex, brainstem FFR-F0 was surprisingly resistant to noise. The resilience of FFR voice pitch encoding in low-level noise is attributable to a reinforcement of neural activity as midbrain neurons phase-lock to both F0 and upper harmonics of vowel stimuli (Bidelman and Krishnan, 2010; Bidelman and Momtaz, 2021; Parbery-Clark et al., 2009; Smith et al., 1978). Collectively, our concurrent subcortical-cortical recordings expose a differential pattern in how noise challenges the brain's speech processing with stronger changes in cortex relative to brainstem levels.

Our findings corroborate previous neuroimaging studies suggesting both FFRs and the cortical P2 are strong predictors of perceptual accuracy for SIN in both younger (Bidelman et al., 2018; Bidelman and Momtaz, 2021; Coffey et al., 2017) and older (Bidelman et al., 2019) adults. That P2 is more strongly related to SIN perception than earlier ERP waves (e.g., P1, N1) is consistent with prior neuroimaging studies linking P2 to auditory perceptual object formation (Bidelman et al., 2013) and P1/N1 to exogenous acoustic properties (Bidelman et al., 2018; Parbery-Clark et al., 2011). A potential explanation for our cross-level differences is that cortical ERPs reflect aggregate coding of multiple and integrated acoustic features (e.g., pitch, timbre, etc.), whereas FFR-F0 reflects primary voice pitch encoding (Bidelman and Krishnan, 2010; Skoe and Kraus, 2010a). Collectively, our electrophysiological data coupled with prior studies provide converging evidence that noise differentially influences speech coding at brainstem and cortical levels.

We found pervasive enhancements in neural coding with active listening at both subcortical and cortical levels. Our ERP data replicate well-known attentional effects in auditory cortex and its links to complex speech processing (Bidelman et al., 2019; Parbery-Clark et al., 2011). ERP amplitudes typically show attention-driven enhancement of the sensory-perceptual N1 response with associated reductions in P2 (Naatanen, 1975). Even larger attentional gains are observed in late cortical activity indexing post-perceptual processing and response selection (Picton and Hillyard, 1974; Picton et al., 1971). Our ERP data reveal larger P2 magnitudes but invariant N1 across attentional states. These findings are consistent with notions that N1 reflects changes in arousal and acoustic feature coding (Bidelman et al., 2013; Coull, 1998; Naatanen and Picton, 1987), whereas P2 indexes active perceptual processes including stimulus classification, speech identity, and auditory object formation (Bidelman et al., 2013). The lack of attentional effects on N1 might also be attributable to heavier neural adaptation given the rapid delivery of our speech stimuli. Indeed, PAC attention effects are particularly susceptible to stimulus presentation rate in the timeframe of N1 (~100 ms) (Neelon et al., 2006).

Critically, our data expose strong attentional enhancements in brainstem speech-FFRs when viewed at the *source* level. Attention effects on FFRs have been highly controversial (Dunlop et al., 1965; Galbraith and Kane, 1993; Picton et al., 1971; Varghese et al., 2015); brainstem re-

sponses recorded during speech perception tasks typically fail to vary with listening state (Varghese et al., 2015), despite concomitant changes in cortical ERPs. Our data help reconcile equivocal findings by revealing top-down influences in source but not channel-level (i.e., scalp electrode) FFR data. Previous failures to consistently observe attentional changes in FFRs might rest in the overwhelming analysis of scalp-level data which blur activity of multiple generators underlying the FFR including BS and PAC (Bidelman, 2018; Coffey et al., 2016), but also cochlear sources (Bidelman, 2018) that may be too peripheral for the purview of attention. Moreover, the high voice pitch (F0=150 Hz) of our stimuli rules out cortical contributions to our FFRs, which were recorded above the phase-locking limit of PAC neurons (<100 Hz) (Fig. S1; Bidelman, 2018). Consequently, our source analysis reveals that "purer" FFRs localized to brainstem (cf. Hartmann and Weisz, 2019) are highly sensitive to attention-dependent reshaping. Our results provide convincing evidence that attention enhances speech coding online as early as the midbrain and suggest subcortical structures may function as an early filtering mechanism as posited by early attention theories (Broadbent, 1971; Treisman, 1960).

Beyond local enhancements, our data further show attention strengthens neural signaling in both feedforward (afferent) and feedback (efferent) directions within the primary auditory pathways. In the visual system, attention increases inter-regional connectivity involved in sensory processing (Buchel and Friston, 1997) with attentional selection being driven by interactions between feedforward and feedback mechanisms (Khorsand et al., 2015). In audition, we found top-down feedback is stronger than its feedforward counterpart, surprisingly, regardless of attentional state or task difficulty (low-level noise). Yet, attention enhanced BS-PAC neural communication bidirectionally. Higher engagement of the efferent system regardless of attention may explain why some FFR studies have observed response enhancements even in passive speech listening tasks (Chandrasekaran et al., 2009; Skoe and Kraus, 2010b).

Animal studies show corticofugal efferents tune subcortical auditory signal processing during short-term auditory learning suggesting cortically-guided feedback shapes earlier sensory coding (Bajo et al., 2010; Suga, 2008). Analogous corticofugal modulation is speculated to account for signal enhancements observed in human FFRs among listeners with long-term experience and training (Chandrasekaran et al., 2009; Galbraith et al., 2003; Lukas, 1980; Skoe and Kraus, 2010b; Tzounopoulos and Kraus, 2009; Wong et al., 2007). Furthermore, predictive coding models suggest that enhanced top-down modulation optimizes early neural encoding thus reducing the strain on feedforward signaling (Carbajal and Malmierca, 2018). For instance, more challenging conditions result in the recruitment of additional top-down resources. As top-down influences fine-tune early neural representations, the afferent signaling strength between subcortical and cortical levels required to transmit an enhanced signal with equivalent fidelity would lessen. This implies that afferent connectivity may weaken in more challenging listening conditions (cf. Carter and Bidelman, 2021). To date, evidence for direct corticofugal involvement in human hearing has been unverified. Theoretically, increased top-down contributions are expected in more challenging scenarios (e.g., during learning, degraded listening environments, increased attentional demands) to sharpen earlier sensory processing and facilitate transmission of faithful neural representations of the acoustic input. Using direct measures of brainstem-cortical connectivity, we find attention reinforces not only brainstem speech coding but also corticofugal signaling in more difficult (noisy) conditions as evidenced by weakened efferent connectivity for passive listening in noise yet sustained connectivity for active listening. We also find that afferent connectivity, although increased with attention, is invariant to noise. This suggests that afferent signaling might serve to provide the most robust, clean signal to subsequent levels of processing regardless of SNR. Together, our findings suggest attention maintains top-down (and bottom-up) neural signaling in noise to tune and enhance early speech encoding which is associated with improved behavioral performance.

While our data show clear attentional modulation of brainstem speech processing, they cannot adjudicate the domain generality of these effects. It remains possible that similar attentional benefits in the cortico-collicular system exist for non-speech sounds, as suggested in animal data (Suga et al., 2000; Suga and Ma, 2003). Furthermore, varying the required engagement and thus difficulty of the task may reveal additional attentional changes in hierarchical speech processing. Load theories of attention suggest that attention is only truly engaged and necessary for selection when tasks are sufficiently difficult (Lavie, 1995; Woldorff et al., 1987). This may explain contradictory findings of studies investigating attentional effects on subcortical auditory encoding (Dittmann-Balcar et al., 1999; Galbraith and Kane, 1993; Galbraith et al., 2003). Additionally, our analyses were restricted to early portions of the auditory-speech network (i.e., BS↔PAC) that reflect high-fidelity sensory coding but do not encompass the later neural networks implicated in semantic, lexical, and other post-perceptual processing necessary for spoken word recognition (e.g., frontoparietal cortices). Interhemispheric connections contribute to auditory attention (Bamiou et al., 2007; Petersen and Posner, 2012), and activation within secondary auditory and prefrontal regions (i.e., posterior superior temporal gyrus, inferior frontal gyrus, prefrontal cortex) strengthens during demanding speech perception tasks (Hickok and Poeppel, 2000). Importantly, we do not deny the very necessary contributions of higher-order brain structures and related attentional networks that engage in challenging listening conditions to decode speech (Alain et al., 2018; Du et al., 2014). Rather, our data argue these processes commence much earlier in the speech hierarchy with attention tuning the representation and formation of speech percepts prior to cortex. To further assess these brainstem-cortical functions during online speech processing, neural decoding analyses at both subcortical and cortical levels could be used to evaluate how cognitive processes and early auditory encoding contribute to successful speech perception in noise (e.g., Al-Fahad et al., 2020; Etard et al., 2019).

5. Conclusions

In sum, our results emphasize the complex interaction between attention and hierarchical auditory processing during SIN perception. Our findings resolve ongoing debates regarding attentional influences on early brainstem encoding and corticofugal engagement during active listening in humans. We show attention enhances neural encoding in cortex but also in brainstem, a surprisingly early stage of auditory processing. Furthermore, using functional connectivity, we demonstrate attentional enhancement in two-way communication between subcortical and cortical levels with stronger communication occurring in the corticofugal direction. Thus, our results suggest attention serves as a mechanism to overcome detrimental noise effects and maintain efficient top-down signaling in challenging listening conditions. Overall, our findings provide novel measurements of corticofugal function in human hearing and establish its involvement as an attention-dependent gain control in speech perception.

Data and code availability statement

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

Author contributions

Caitlin N. Price: Conceptualization, methodology, investigation, formal analysis, visualization, writing – original draft and review & editing. **Gavin M. Bidelman:** Conceptualization, methodology, formal analysis, writing – original draft and review & editing, supervision.

Ethics statement

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

Declaration of competing interest

None.

Acknowledgments

This work was supported by the National Institutes of Health (NIH/NIDCD R01DC016267) (G.M.B.) and the UofM Institute for Intelligent Systems Dissertation grant (C.N.P.). The authors would also like to thank Claude Alain, Jane Brown, and Kelsey Mankel who provided comments during the preparation of this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118014.

References

Ahissar, M., Hochstein, S., 2004. The reverse hierarchy theory of visual perceptual learning. Trends Cognit. Sci. 8, 457–464.
Al-Fahad, R., Yeasin, M., Bidelman, G.M., 2020. Decoding of single-trial EEG reveals unique states of functional brain connectivity that drive rapid speech categorization decisions. J. Neural Eng. 17, 016045.
Alain, C., Du, Y., Bernstein, L.J., Barten, T., Banai, K., 2018. Listening under difficult conditions: An activation likelihood estimation meta-analysis. Hum. Brain Mapp. 39 (7), 2695–2709.
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.
Alho, K., Vorobyev, V.A., 2007. Brain activity during selective listening to natural speech. Front. Biosci. 12, 3167–3176.
Assmann, P.F., 1996. Tracking and glimpsing speech in noise: role of fundamental frequency. J. Acoust. Soc. Am. 100, 2680.
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, 467–480.
Bajo, V.M., Nodal, F.R., Moore, D.R., King, A.J., 2010. The descending corticocollicular pathway mediates learning-induced auditory plasticity. Nat. Neurosci. 13, 253–260.
Bamiou, D.E., Sisodiya, S., Musiek, F.E., Luxon, L.M., 2007. The role of the interhemispheric pathway in hearing. Brain Res. Rev. 56, 170–182.
Baumann, S.B., Wozny, D.R., Kelly, S.K., Meno, F.M., 1997. The electrical conductivity of human cerebrospinal fluid at body temperature. IEEE Trans. Biomed. Eng. 44, 220–223.
Berg, P., Scherg, M., 1994. A fast method for forward computation of multiple-shell spherical head models. Electroencephalogr. Clin. Neurophysiol. 90, 58–64.
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.
Bidelman, G.M., 2018. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. Neuroimage 175, 56–69.
Bidelman, G.M., Davis, M.K., Pridgen, M.H., 2018. Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. Hear. Res. 367, 149–160.
Bidelman, G.M., Krishnan, A., 2010. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. Brain Res. 1355, 112–125.
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. Neurosci. Lett. 746, 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.
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.
Broadbent, D.E., 1971. Decision and Stress. Academic P., London, New York.
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, 2358–2374.
Buchel, C., Friston, K.J., 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb. Cortex 7, 768–778.
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, 144–150.
Carbajal, G.V., Malmierca, M.S., 2018. The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. Trends Hear. 22, 2331216518784822.
Carter, J.A., Bidelman, G.M., 2021. Auditory cortex is susceptible to lexical influence as revealed by informational vs. energetic masking of speech categorization. Brain Res. 1759, 147385.
Chan, T.F., Vese, L.A., 2001. Active contours without edges. IEEE Trans. Image Process. 10, 266–277.
Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., Kraus, N., 2009. Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. Neuron 64, 311–319.
Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. Psychophysiology 47, 236–246.
Coffey, E.B., Herholz, S.C., Chepesiuk, A.M., Baillet, S., Zatorre, R.J., 2016. Cortical contributions to the auditory frequency-following response revealed by MEG. Nat. Commun. 7, 11070.
Coffey, E.B.J., Chepesiuk, A.M.P., Herholz, S.C., Baillet, S., Zatorre, R.J., 2017. Neural correlates of early sound encoding and their relationship to speech-in-noise perception. Front. Neurosci. 11, 479.
Coffey, E.B.J., 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, 5036.
Coull, J.T., 1998. Neural correlates of attention and arousal: insights from electrophysiology, functional neuroimaging and psychopharmacology. Prog. Neurobiol. 55, 343–361.
Dittmann-Balcar, A., Thienel, R., Schall, U., 1999. Attention-dependent allocation of auditory processing resources as measured by mismatch negativity. Neuroreport 10, 3749–3753.
Du, Y., Buchsbaum, B.R., Grady, C.L., Alain, C., 2014. Noise differentially impacts phoneme representations in the auditory and speech motor systems. PNAS 111, 1–6.
Dunlop, C.W., Webster, W.R., Simons, L.A., 1965. Effect of attention on evoked responses in the classical auditory pathway. Nature 206, 1048–1050.
Etard, O., Kegler, M., Braiman, C., Forte, A.E., Reichenbach, T., 2019. Decoding of selective attention to continuous speech from the human auditory brainstem response. Neuroimage 200, 1–11.
Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175–191.
Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207.
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.
Fuchs, M., Kastner, J., Wagner, M., Hawes, S., Ebersole, J.S., 2002. A standardized boundary element method volume conductor model. Clin. Neurophysiol. 113, 702–712.
Galbraith, G.C., Arroyo, C., 1993. Selective attention and brainstem frequency-following responses. Biol. Psychol. 37, 3–22.
Galbraith, G.C., Kane, J.M., 1993. Brainstem frequency-following responses and cortical event-related potentials during attention. Percept. Mot. Skills 76, 1231–1241.
Galbraith, G.C., Olfman, D.M., Huffman, T.M., 2003. Selective attention affects human brain stem frequency-following response. Neuroreport 14, 735–738.
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. U.S.A. 97, 8081–8086.
Gardi, J., Merzenich, M.M., McKean, C., 1979. Origins of the scalp-recorded frequency–following response in the cat. Audiology 18, 353–381.
Giard, M.H., Collet, L., Bouchet, P., Pernier, J., 1994. Auditory selective attention in the human cochlea. Brain Res. 633, 353–356.
Hall, J.W., 1992. Handbook of auditory evoked responses. Allyn and Bacon, Needham Heights.
Hartmann, T., Weisz, N., 2019. Auditory cortical generators of the frequency following response are modulated by intermodal attention. Neuroimage 203, 116185.
Hernandez-Peon, R., 1966. Physiological mechanisms in attention. In: Russell, R.W. (Ed.), Frontiers in Physiological Psychology. Academic, New York.
Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cognit. Sci. 4, 131–138.
Hirschhorn, T.N., Michie, P.T., 1990. Brainstem auditory evoked potentials (BAEPS) and selective attention revisited. Psychophysiology 27, 495–512.
Holmes, E., Purcell, D.W., Carlyon, R.P., Gockel, H.E., Johnsrude, I.S., 2018. Attentional modulation of envelope-following responses at lower (93–109 Hz) but not higher (217–233 Hz) modulation rates. J. Assoc. Res. Otolaryngol. 19, 83–97.
Humes, L.E., 1996. Speech understanding in the elderly. J. Am. Acad. Audiol. 7, 161–167.
Khorsand, P., Moore, T., Soltani, A., 2015. Combined contributions of feedforward and feedback inputs to bottom-up attention. Front. Psychol. 6, 155.
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, 2395–2405.
Kraus, N., White-Schwoch, T., 2015. Unraveling the biology of auditory learning: a cognitive-sensorimotor-reward framework. Trends Cognit. Sci. 19, 642–654.
Krishnan, A., 2002. Human frequency-following responses: representation of steady-state synthetic vowels. Hear. Res. 166, 192–201.
Lavie, N., 1995. Perceptual load as a necessary condition for selective attention. J. Exp. Psychol. Hum. Percept. Perform. 21, 451–468.
Lehmann, A., Schonwiesner, M., 2014. Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. PLoS One 9, e85442.

Lobier, M., Siebenhuhner, F., Palva, S., Palva, J.M., 2014. Phase transfer entropy: a novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. Neuroimage 85 (Pt 2), 853–872.
Lüdecke, D., 2020. sjPlot: Data visualization for statistics in social science. R Package Version 2.8.6, https://CRAN.R-project.org/package=sjPlot.
Lukas, J.H., 1980. Human auditory attention: the olivocochlear bundle may function as a peripheral filter. Psychophysiology 17, 444–452.
Mankel, K., Bidelman, G.M., 2018. Inherent auditory skills rather than formal music training shape the neural encoding of speech. PNAS 115, 13129–13134.
Marsh, J.T., Worden, F.G., Smith, J.C., 1970. Auditory frequency-following response: neural or artifact? Science 169, 1222–1223.
Moore, D.R., 2015. Sources of pathology underlying listening disorders in children. Int. J. Psychophysiol. 95, 125–134.
Musacchia, G., Sams, M., Skoe, E., Kraus, N., 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. PNAS 104, 15894–15898.
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, 34–42.
Naatanen, R., 1975. Selective attention and evoked potentials in humans–a critical review. Biol. Psychol. 2, 237–307.
Naatanen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24, 375–425.
Neelon, M.F., Williams, J., Garell, P.C., 2006. The effects of attentional load on auditory ERPs recorded from human cortex. Brain Res. 1118, 94–105.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP measurements. Clin. Neurophysiol. 112, 713–719.
Parbery-Clark, A., Marmel, F., Bair, J., Kraus, N., 2011. What subcortical-cortical relationships tell us about processing speech in noise. Eur. J. Neurosci. 33, 549–557.
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, 14100–14107.
Petersen, S.E., Posner, M.I., 2012. The attention system of the human brain: 20 years after. Annu. Rev. Neurosci. 35, 73–89.
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, 469–485.
Picton, T.W., Hillyard, S.A., 1974. Human auditory evoked potentials. II. Effects of attention. Electroencephalogr. Clin. Neurophysiol. 36, 191–199.
Picton, T.W., Hillyard, S.A., Galambos, R., Schiff, M., 1971. Human auditory attention: a central or peripheral process? Science 173, 351–353.
Picton, T.W., van Roon, P., Armilio, M.L., Berg, P., Ille, N., Scherg, M., 2000. The correction of ocular artifacts: a topographic perspective. Clin. Neurophysiol. 111, 53–65.
Price, C.N., Alain, C., Bidelman, G.M., 2019. Auditory-frontal channeling in alpha and beta bands is altered by age-related hearing loss and relates to speech perception in noise. Neuroscience 423, 18–28.
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, 3323–3327.
Saiz-Alia, 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, 14131.
Saiz-Alia, M., Reichenbach, T., 2020. Computational modeling of the auditory brainstem response to continuous speech. J. Neural Eng. 17, 036035.
Scherg, M., 1990. Fundamentals of dipole source potential analysis. In: Grandori, F., Hoke, M., Romani, G.L. (Eds.), Auditory Evoked Magnetic Fields and Electric Potentials. Karger, Basel, pp. 40–69 Advances in Audiology.
Scherg, M., Ebersole, J.S., 1994. Brain source imaging of focal and multifocal epileptiform EEG activity. Neurophysiol. Clin. 24, 51–60.
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, 91–112.
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, e0136794.
Skoe, E., Kraus, N., 2010a. Auditory brain stem response to complex sounds: a tutorial. Ear Hear. 31.
Skoe, E., Kraus, N., 2010b. Hearing it again and again: on-line subcortical plasticity in humans. PLoS One 5, e13645.
Slee, S.J., David, S.V., 2015. Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. J. Neurosci. 35, 13090–13102.
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.
Smith, J.C., Marsh, J.T., Greenberg, S., Brown, W.S., 1978. Human auditory frequency–following responses to missing fundamental. Science 201, 639–641.
Suga, N., 2008. Role of corticofugal feedback in hearing. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. 194, 169–183.
Suga, N., Gao, E., Zhang, Y., Ma, X., Olsen, J.F., 2000. The corticofugal system for hearing: recent progress. Proc. Natl. Acad. Sci. USA 97, 11807–11814.
Suga, N., Ma, X., 2003. Multiparametric corticofugal modulation and plasticity in the auditory system. Nat. Rev. Neurosci. 4, 783–794.
Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. 2011, 879716.
Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: an Approach to Cerebral Imaging. Georg Thieme Stuttgart; New York.
Tichko, P., Skoe, E., 2017. Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators. Hear. Res. 348, 1–15.
Treisman, A.M., 1960. Contextual cues in selective listening. Q. J. Exp. Psychol. 242–248.
Tzounopoulos, T., Kraus, N., 2009. Learning to encode timing: mechanisms of plasticity in the auditory brainstem. Neuron 62, 463–469.
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.
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, 47–64.
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, 844–848.
Woldorff, M., Hansen, J.C., Hillyard, S.A., 1987. Evidence for effects of selective attention in the mid-latency range of the human auditory event-related potential. Electroencephalogr. Clin. Neurophysiol. 40, 146–154 Suppl.
Wolters, C.H., Anwander, A., Berti, G., Hartmann, U., 2007. Geometry-adapted hexahedral meshes improve accuracy of finite-element-method-based EEG source analysis. IEEE Trans. Biomed. Eng. 54, 1446–1453.
Wong, P.C., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat. Neurosci. 10, 420–422.
Woods, D.L., Hillyard, S.A., 1978. Attention at the cocktail party: brainstem evoked responses reveal no peripheral gating. In: Otto, D. (Ed.), Multidisciplinary Perspectives in Event-Related Brain Potential Research. GPO, Washington, D.C.
Zhao, T.C., Kuhl, P.K., 2018. Linguistic effect on speech perception observed at the brainstem. PNAS 115, 8716–8721.

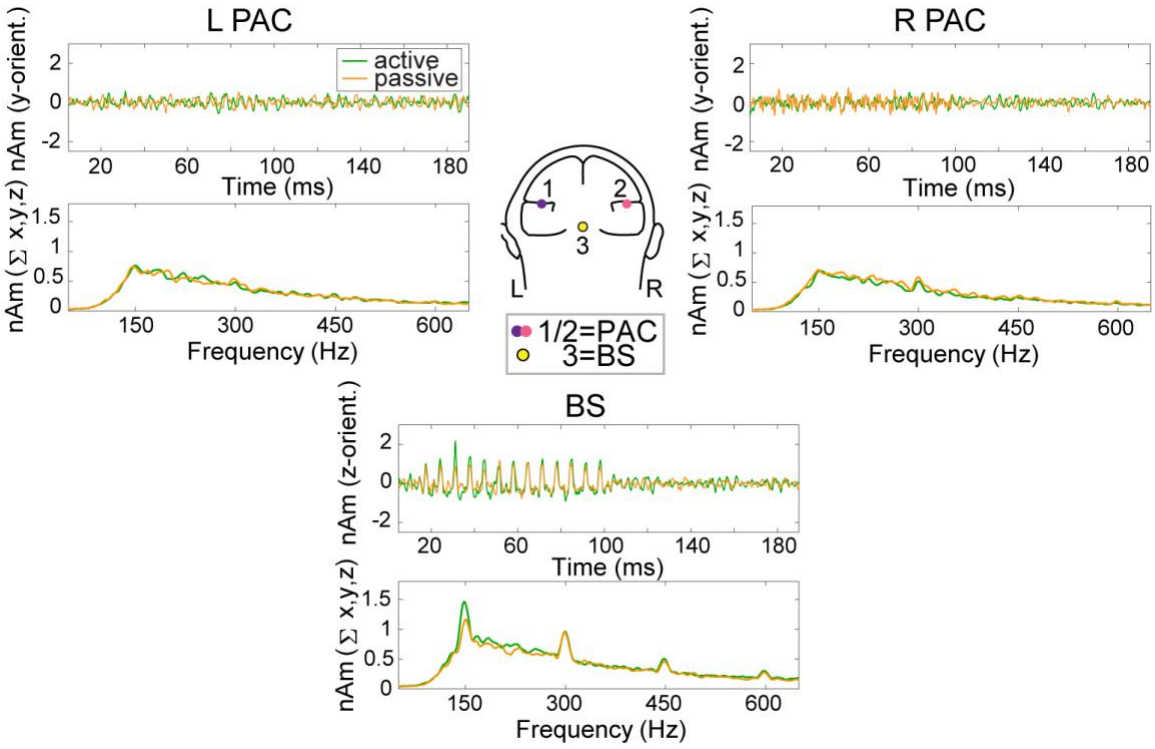

Figure S1. Phase-locked FFRs to vowel stimuli are only observed subcortically. FFRs are plotted in the time and frequency domain for each dipole (collapsing tokens/SNR) illustrating robust phase-locking in only the BS source.

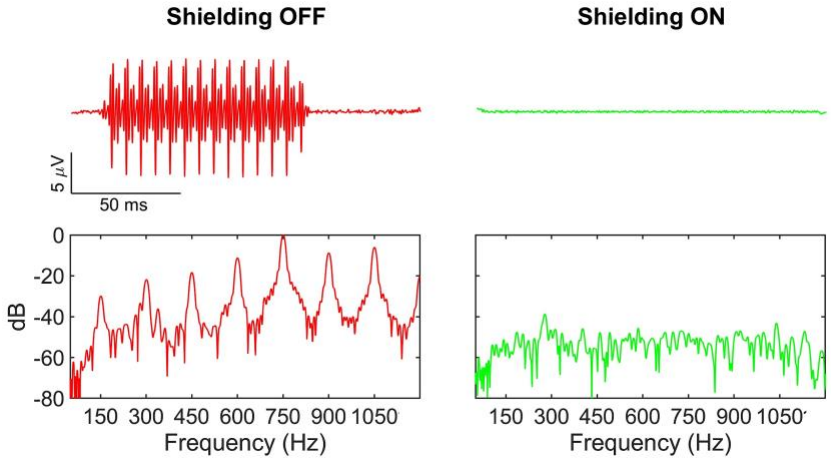

Figure S2. Control runs confirm absence of stimulus artifact in EEG recordings. Shielding was accomplished by encasing the headphone transducer boxes and cabling with Scotch® Electrical Shielding Tape 24 (3M Company; St. Paul, MN) and grounding its termination at the soundbooth wall (Campbell et al., 2012). We used a simulated phantom head consisting of a triangular resistive circuit of three 4.7 kΩ resistors to model the impedance (~ 5 kΩ) of a human load (see Fig. 3.3; Luck, 2005; p. 115). Recordings (N≈200 trials; vowel /a/, single rarefaction polarity) were then made with the headphones placed near the phantom with and without the shielding grounded. Top=time waveforms, bottom=spectra. Without (i.e., bypassing) shielding (*left*), substantial stimulus electromagnetic artifact contaminates recordings, masquerading as "FFR." Reinstating shielding (*right*) effectively eradicates pickup of stimulus artifact and ensures clean biological responses during FFR recording.

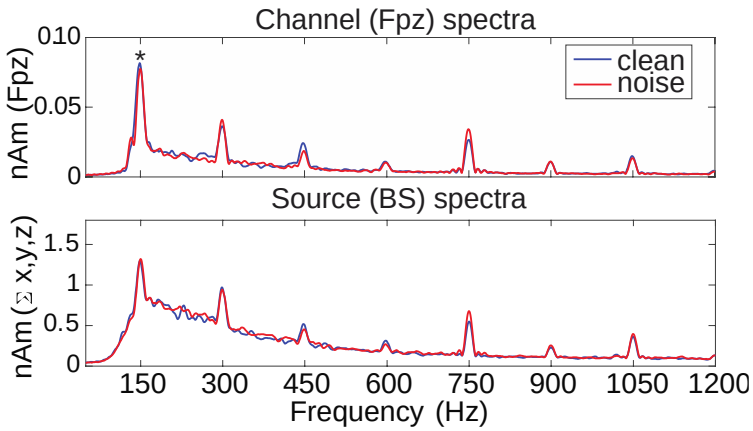

Figure S3. Noise weakens FFR-F0 encoding at the channel-level. FFR spectra (collapsing tokens/attention) illustrating detrimental noise effects on channel-level FFRs.

Table S1. GLME model fit parameters for prediction of target detection accuracy. Coefficients and significance tests for individual predictor variables including neural and hearing measures and SNR. Level of significance denoted *p< 0.05, ***p< 0.001

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	62.85	11.14	5.64	32	3.09e-06	40.15	85.55
FFR_{F0}	2.70	2.47	1.09	32	0.28	-2.34	7.73
ERP_{P2}	-0.008	0.20	0.04	32	0.97	-0.39	0.41
aff	-39.44	23.30	-1.69	32	0.10	-86.91	8.03
eff	23.16	13.90	1.67	32	0.11	-5.14	51.47
QSIN*	-4.57	1.93	-2.37	32	0.02	-8.50	-0.64
PTA	-0.18	0.83	-0.21	32	0.83	-1.88	1.52
SNR***	-19.47	2.98	-6.53	32	2.40e-07	-25.54	-13.39

Table S2. GLME model fit parameters for prediction of reaction time. Coefficients and significance tests for individual predictor variables including neural and hearing measures and SNR. Level of significance denoted *p< 0.05, ***p< 0.001

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	462.99	9.04	51.23	32	2.77e-32	444.58	481.40
FFR_{F0}	-2.12	1.99	-1.06	32	0.29	-6.18	1.94
ERP_{P2}	-0.03	0.15	-0.23	32	0.82	-0.35	0.28
aff*	43.08	18.45	2.34	32	0.03	5.50	80.67
eff*	-22.70	10.73	-2.11	32	0.04	-44.56	-0.84
QSIN	-0.94	1.91	-0.49	32	0.62	-4.82	2.94
PTA	0.28	0.82	0.35	32	0.73	-1.39	1.96
SNR***	7.71	2.25	3.43	32	0.002	3.13	12.29

References

Campbell, T., Kerlin, J.R., Bishop, C.W., Miller, L.M., 2012. Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. Ear and Hearing 33, 144-150.

Luck, S., 2005. An Introduction to the Event-Related Potential Technique. MIT Press., Cambridge, MA, USA.