Title: Subcortical sources drive the relation between frequency-following responses (FFRs) and speech-in-noise perception

Authors: Gavin M. **Bidelman**[1,2,3] & Sara **Momtaz**[1]

Affiliations:

[1]School of Communication Sciences & Disorders, University of Memphis, Memphis, TN

[2]Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA

[3]University of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, TN

Address for editorial correspondence:
Gavin M. Bidelman, PhD
School of Communication Sciences & Disorders
University of Memphis
4055 North Park Loop
Memphis, TN, 38152
TEL: (901) 678-5826
FAX: (901) 678-2472
EMAIL: gmbdlman@memphis.edu

ABSTRACT

The frequency-following response (FFR) is a scalp-recorded potential reflecting a mixture of phase-locked neural activity generated from several nuclei along the auditory pathway. FFRs have been widely used as a neural barometer of complex listening skills especially performance on speech-in noise (SIN) tasks: across listeners with various hearing profiles and ages, more robust speech-evoked FFRs are associated with improved SIN perception. Applying individually optimized source reconstruction to speech-FFRs recorded via EEG (FFR_{EEG}), we assessed the relative contributions of subcortical [auditory nerve (AN), brainstem (BS)] and cortical [bilateral primary auditory cortex, PAC] generators to the scalp response with the aim of identifying which source(s) drives the brain-behavior relation between FFRs and perceptual SIN skills. We found FFR strength declined precipitously from AN to PAC, consistent with the roll-off of phase-locking at progressively higher stages of the auditory neuroaxis. FFRs at the speech fundamental frequency (F0) were resistant to moderate noise interference across all sources, but FFRs were largest in BS relative to all other sources (BS > AN >> PAC). Cortical PAC FFRs were only weakly observed above the noise floor in a restricted bandwidth around the low pitch of speech stimuli (F0≈100 Hz). Brain-behavior regressions revealed (i) AN and BS FFRs were sufficient to describe listeners' QuickSIN scores and (ii) contrary to neuromagnetic (MEG) FFRs, neither left nor right PAC FFR_{EEG} predicted SIN performance. Our findings suggest subcortical sources not only dominate the electrical FFR but also the link between speech-FFRs and SIN processing as observed in previous EEG studies.

Keywords: Auditory event-related potentials (ERPs); auditory brainstem response (ABR); cocktail party scenario; cortical FFR; noise-degraded speech perception

New & Noteworthy: The frequency-following responses (FFRs) is an auditory neurophonic brain response that correlates with complex listening skills including speech-in-noise perception. FFRs origin from phase-locked activity across the auditory pathway including sub- and neo-cortex. By recording speech-FFRs via EEG, we show that brainstem sources dominate FFRs measured at the scalp and are most predictive of behavior. Findings underscore the brainstem-centric view of the FFR and its relation to nascent auditory perceptual processing.

1. INTRODUCTION

In real-world listening environments, acoustic interference hinders the successful extraction of critical speech information. Speech-in-noise (SIN) understanding is often exacerbated in cases of hearing impairment (Chmiel & Jerger, 1996; Popelka et al., 2016), and figure-ground deficits are among the hallmarks of central hearing disorders (Medwetsky, 2002), putative "hidden hearing loss" (Kujawa & Liberman, 2015), and listening deficits in older adults (Gordon-Salant et al., 2011; Schneider et al., 2002). However, SIN comprehension varies substantially even among normal-hearing individuals (Bidelman et al., 2018; Bidelman & Dexter, 2015; Bidelman & Howell, 2016; Killion & Niquette, 2000; Middelweerd et al., 1990; Song et al., 2011). This has led to recent interest in the neurophysiological basis of SIN processing which has focused on the role of central auditory brain mechanisms in complex listening skills and how robust neural encoding supports successful listening abilities (for review, see Bidelman, 2017).

Insight into the neuroimaging markers of speech perception is challenging due to the various underlying mechanisms, including the perceptual aspects of hearing (i.e., sensitivity to spatial and spectrotemporal cues) as well as cognitive processes (e.g., working memory). In this regard, the human frequency-following response (FFR) has provided considerable insight into these former processes, that is, how well auditory neural-perceptual coding governs adverse listening skills. FFRs are neurophonic potentials generated by a mixture of subcortical and cortical structures of the auditory system that phase-lock to spatiotemporal features of periodic acoustic stimuli including speech (Krishnan, 2007). Links between speech-FFRs (recorded with EEG) and SIN processing have been reported in a wide range of studies over the past decade (e.g., Bidelman & Krishnan, 2010; Parbery-Clark et al., 2011; Parbery-Clark et al., 2009; Saiz-Alía et al., 2019; Song et al., 2011; Yellamsetty & Bidelman, 2019). For instance, Song et al. (2011) showed the magnitude of the speech fundamental frequency (F0) encoded in the FFR positively predicted SIN performance on the QuickSIN (Killion et al., 2004): "Top SIN" performers on the task had more robust FFRs than "Bottom SIN" performers, who had both weaker neural representation of the speech F0 and poorer perceptual scores. Complementary findings were reported by Anderson et al. (2010) who showed poorer (i.e., lower median) SIN listeners experienced greater (~0.5-1 ms) noise-related shifts in the timing of their speech FFR from quiet to noise than top performing listeners. These collective findings suggest a strong relationship between the quality of neural representation of complex sounds, as reflected via FFRs, and perceptual SIN skills, whereby faster and more robust speech encoding is associated with better behavioral outcomes. While there is convincing evidence to implicate brainstem as the active driver of FFR in these studies—and

therefore the brain-behavior (i.e., FFR-SIN) relation—scalp recorded data alone cannot fully adjudicate this interpretation.

Though traditionally viewed as a brainstem response (Sohmer et al., 1977), it has long been recognized there are multiple sources of FFRs stemming from throughout the hearing pathway. These include the cochlear microphonic (Sohmer & Pratt, 1977), auditory nerve (Bidelman, 2015, 2018), brainstem (midbrain) (Bidelman, 2015, 2018; Sohmer et al., 1977), and under some circumstances, primary auditory cortex (Bidelman, 2018; Coffey et al., 2016). EEG studies have interpreted FFR correlates of SIN perception based on how well (*brainstem*) FFRs encode important acoustic properties of speech (e.g., voice pitch and timbre cues) (Bidelman & Krishnan, 2010; Parbery-Clark et al., 2012; Parbery-Clark et al., 2011; Parbery-Clark et al., 2009; Song et al., 2011). Still, acknowledging FFRs likely contain contributions from cortex for low-frequency (~100 Hz) stimuli (Bidelman, 2018; Coffey et al., 2016), it is possible the link between this neurophonic and SIN behaviors is at least partially driven by higher-level “cortical FFRs” (cf. Coffey et al., 2016; Coffey et al., 2017). Indeed, based on MEG-recorded FFRs, Coffey et al. suggested higher-level cognitive tasks such as SIN perception were dominated by FFRs generated in right auditory cortex (Coffey et al., 2017; see also Hartmann & Weisz, 2019). This finding is in stark contrast to the brainstem-centric view of the FFR and its relation to auditory perception (Chandrasekaran et al., 2012; Parbery-Clark et al., 2009; Saiz-Alía et al., 2019; Song et al., 2011; Zhang & Gong, 2019)—now suggested even in recent MEG studies (Kulasingham et al., 2019; Zhao & Kuhl, 2018).

We suspect ambiguity of cortical vs. subcortical structures in accounting for perceptual correlates between FFR and behavior might be driven by limitations of different neuroimaging modalities. Whereas subcortical auditory structures are among the most significant sources of the scalp FFR_{EEG} (Bidelman, 2015, 2018), MEG is more sensitive to superficial neuronal activity (deep brainstem sources become invisible; Cohen & Cuffin, 1983; Hillebrand & Barnes, 2002; Kulasingham et al., 2019). Given the differential pickup of deeper brainstem (EEG) vs. shallower cortical (MEG) contributions to the FFR, it is reasonable to assume that the imaging modality could bias associations between the FFR and behavior. Because FFRs are recorded more commonly with EEG, we were interested to reevaluate the neural origins of the FFR_{EEG}-SIN association, with particular attention to *which* subcortical and/or cortical structure(s) drives this brain-behavior relation.

In light of emerging controversies on sources of the FFR and their relation to complex auditory behaviors (Bidelman, 2018; Coffey et al., 2019; Holmes & Herrmann, 2017), we aimed to

unravel the link between speech-FFRs and SIN processing using a more comprehensive, systems-level neuroimaging approach. Using high-density EEG, we recorded multichannel FFRs to noise-degraded speech tokens. Source analysis of the responses allowed us to parse region-specific activity to the electrical FFR and evaluate the relative contribution of each nuclei to SIN perception. Our findings show phase-locked activity peripheral to cortex dominates the EEG-based FFR as well as its link to perceptual speech-in-noise listening abilities.

2. METHODS & MATERIALS

2.1 Participants

Data reported herein reflect a new and comprehensive source analysis of FFRs recorded in *n*=12 normal-hearing adults (age: 24.7 ± 2.7 years) reported in Bidelman et al. (2018). All participants were native English speakers with a similar level of education (having completed at least an undergraduate degree) and <3 years formal music training (1.3 ± 1.8 years) that occurred at least five years before the study. Hearing thresholds were normal (≤ 25 dB HL) bilaterally at octave frequencies between 250-8000 Hz in all participants. None reported a history of neuropsychiatric disorders. All gave written informed consent for the study protocol approved by the University of Memphis Institutional Review Board.

2.2 Stimuli

FFRs were elicited by the 300 ms male speech token /ama/ (for details, see Bidelman et al., 2018). The F0 pitch fell gradually over the token's duration (F0= 120 Hz to 88 Hz). Importantly, the lower F0 of the stimulus (~100 Hz) was expected to elicit phase-locked FFRs of both brainstem and cortical origin (Bidelman, 2018; Coffey et al., 2016). Vowel formant frequencies (F1-F3) were 830, 1200, and 2760 Hz, respectively. In addition to this "clean" stimulus (SNR = ∞ dB), speech was presented in noise-degraded conditions. Degraded speech was created by adding four-talker babble noise (Killion et al., 2004) to the clean token at signal-to-noise ratios (SNRs) of +10 and +5 dB. Babble was presented continuously to avoid it being time-locked to the stimulus (e.g., Alain et al., 2012; Bidelman & Howell, 2016). Listeners heard 2000 trials of each speech token (passive listening) for each noise condition, presented binaurally with rarefaction polarity (jittered interstimulus: 400-700 ms) at 81 dB SPL through ER-30 insert earphones (Etymotic Research). Extended acoustic tubing (20 ft) prevented electromagnetic stimulus artifact from contaminating the biological FFR response.

2.3 EEG

EEG recording procedures and preprocessing are described fully in Bidelman et al. (2018). Briefly, EEGs were recorded from 64 electrodes at 10-10 scalp locations (Oostenveld & Praamstra, 2001) and were digitized at a sampling rate of 5000 Hz (SynAmps RT amplifiers; Compumedics Neuroscan) using an online passband of DC-2500 Hz. Data were re-referenced to a common average reference (CAR) for analyses. Following ocular artifact correction (Picton et al., 2000), cleaned EEGs were epoched (-200-550 ms), baseline-corrected to the pre-stimulus period, and averaged in the time domain to obtain FFRs for each stimulus condition per participant and electrode. FFRs were then bandpass filtered (80-1500 Hz; zero-phase; -24 dB/oct) for subsequent source analysis and quantification.

2.3.1 Source waveform derivations. Scalp FFR potentials (sensor-level recordings) were transformed to source space using a virtual source montage implemented in BESA® Research v7 (BESA, GmbH) (Scherg et al., 2019; Scherg et al., 2002). This digital re-montaging applies a spatial filter to all electrodes (defined by the foci of our dipole configuration) to transform electrode recordings to a reduced set of source signals reflecting the neuronal current (in units nAm) as seen *within* each anatomical region of interest. We adopted the dipole configuration described in Bidelman (2018) to assess the relative contribution of brainstem and cortical FFR sources to SIN processing. Full details of this model, its derivation, and fit accuracy are reported elsewhere (Bidelman, 2018). The model consisted of 5 dipoles seeded in bilateral auditory cortex (PAC; source #1–2), the brainstem (midbrain) (BS; source #3), and bilateral auditory nerve (AN; sources #4–5) (**Fig. 2**, inset). This allowed us to reduce each listener's electrode recordings (64-channels) to 5 unmixed source waveforms describing their scalp FFR data. For each participant, dipole locations were held fixed and were used as a spatial filter to derive FFR source waveforms (Bidelman, 2018). Critically, we fit individual dipole orientations to each participant's data (anatomical locations remained fixed) to maximize the explained variance of the model at the individual subject level. The model provided a robust fit to the grand averaged (clean) scalp data in the 0-350 ms response interval (goodness of fit: 77%), confirming that FFRs could be adequately described by a restricted number of sources along the primary auditory neuroaxis.

2.3.2 FFR source waveform analysis. FFR source waveforms were analyzed in the spectral domain via BESA. From each waveform (per source, SNR, and listener), we first computed Fast Fourier Transforms (FFTs) in the post-stimulus response interval (0-350 ms, $\cos^2$ windowed; see **Fig. 1**, shading). We then measured the peak amplitude in the F0 frequency bin. We focused our analysis on F0 because (i) this component was present across all sources of the

FFR (see **Fig. 2**) and thus could be measured at each anatomical level and (ii) FFR-F0 has been explicitly related to SIN processing in previous EEG studies (Bidelman, 2016; Bidelman et al., 2018; Bidelman & Krishnan, 2010; Prevost et al., 2013; Russo et al., 2004; Smalt et al., 2012). F0 amplitude was measured manually as the largest peak in the frequency range of 88-120 Hz (i.e., F0 of the stimulus), relative to the noise floor (computed via the FFT of the pre-stimulus interval, -200-0ms). Subtracting post-stimulus from pre-stimulus interval amplitudes thus retained only evoked neural responses (i.e., FFRs) above common EEG noise. This allowed us to assess the degree to which subcortical and cortical FFR sources showed veridical phase-locked activity (i.e., above noise-floor). Two measurements were made for each spectral amplitude and averaged for reliability. Test-retest agreement was excellent (r = 0.99).

2.4 Behavioral speech-in-noise task

We measured listeners' speech reception thresholds in noise using the QuickSIN test (Killion et al., 2004). Participants heard two lists of six sentences embedded in four-talker babble noise. Each contained five key words. Sentences were presented at 70 dB SPL via insert earphones (bilaterally) using pre-recorded signal-to-noise ratios (SNRs) that decreased in 5 dB steps from 25 dB (very easy) to 0 dB (very difficult). They scored one point for each key word correctly repeated. "SNR loss" (computed in dB) was determined as the SNR required to correctly identify 50% of the key words (Killion et al., 2004). Each ear was tested separately (two lists each) and subsequently averaged to obtain a singular measure of SIN perception for each participant[1].

2.5 Statistical analysis

We used two-way, mixed-model ANOVAs (source x SNR; subjects=random factor; SAS® 9.4, GLIMMIX) to assess changes in FFR F0 amplitude across anatomical levels and noise conditions. Amplitudes were SQRT-transformed to satisfy normality and homogeneity of variance assumptions necessary for parametric ANOVAs. We used generalized linear mixed effects (GLME) regression to evaluate links between behavioral QuickSIN scores and the F0 component of source-localized FFRs [pooled across the noise (+10, +5 dB SNR) conditions]. Subjects were included as a random factor in the regression to model random intercepts per listener [e.g., QuickSIN ~ PAC_{left} + PAC_{right} + BS + AN_{left} + AN_{right} + (1| subject)]. This multivariate model allowed us to assess the relative contribution of subcortical and cortical FFR sources to SIN perception. GLME regression was achieved using the 'fitglme' function in MATLAB 2019b.

[1] One individual's QuickSIN data was lost due to an error in data logging (EEG data were unaffected). This value was imputed using mean of the remaining participants' QuickSIN scores.

3. RESULTS

FFRs appeared as phase-locked potentials that mirrored the acoustic periodicity of speech (**Fig. 1**). Their voltage distribution on the scalp showed maximal amplitude at frontocentral electrode sites (e.g., Fpz) and polarity inversion at the mastoids. These observations are consistent with FFR sources in the deep midbrain that point obliquely in an anterior orientation to the vertex (parallel to the brainstem) (Bidelman, 2015, 2018). The volume-conducted nature of these sensor-space (i.e., electrode) recordings did not allow us to separate the underlying sources that contribute to the FFR and SIN processing. Consequently, subsequent analyses were conducted in source space.

[Insert Figure 1 near here]

3.1 Source FFR waveforms and spectra

Grand averaged source FFR waveforms and spectra parsing region-specific activity within the three source levels (AN, BS, and PAC) are shown in **Figure 2**. FFRs showed strong phase-locking at the speech F0 frequency (~100 Hz) in all source foci. However, only subcortical sources (AN, BS) showed additional response energy at higher harmonics of the speech signal (H2 and H3), consistent with the higher phase-locking limits of more peripheral auditory nuclei (Joris et al., 2004), including those which generate the FFR (Bidelman, 2018; Herdman et al., 2002). In contrast, cortical PAC sources were largely band-limited to F0 (~100 Hz) and failed to show reliable FFRs above the noise floor at higher frequencies. Raw waveforms similarly show variation in the strength of the response across the neuroaxis. FFRs were maximal at the level of BS, were robust but to a lesser extent in AN, and reduced dramatically at the level of PAC. These results confirm a mixture of generators produces the speech FFR, including brainstem and cortical levels of the auditory system (Bidelman, 2018; Coffey et al., 2016). More critically, they further confirm a dominance of midbrain and lower structures to the electrical FFR (e.g., Bidelman, 2015; Bidelman, 2018; Kiren et al., 1994; Smith et al., 1975; Sohmer et al., 1977; White-Schwoch et al., 2019; Zhang & Gong, 2019).

[Insert Figure 2 near here]

3.2 FFR F0 amplitudes as a function of source level and noise

An ANOVA conducted on F0 amplitudes revealed significant differences in FFR strength among various source levels ($F_{4,126}$= 10.42, p <0.0001) but not between SNRs ($F_{2,126}$= 0.47, p = 0.6286) (SNR x level interaction: $F_{8,126}$= 0.38, p=0.9313) (**Fig. 3**). Tukey-Kramer adjusted multiple

comparisons revealed the amplitude of F0 was weaker in PAC vs. BS (p<0.0001), AN vs. BS (p=0.0002), and PAC vs. AN (p=0.0260). There was no difference between PAC_{right} vs. PAC_{left} (p=1.000). Thus, our data did not reveal evidence of a functional asymmetry (i.e., right hemisphere bias) for cortical FFRs as observed in MEG studies (Coffey et al., 2016). Moreover, we found a brainstem-dominant gradient in F0 amplitudes across levels (BS > AN >> PAC), further suggesting BS is the largest contributor to the FFR_{EEG} response (Bidelman, 2018).

[Insert Figure 3 near here]

3.3 Brain-behavior relations

Following previous channel-based FFR studies, we first evaluated associations between scalp-level recordings (Fpz electrode) and QuickSIN scores. We selected this channel as FFRs are strongest at frontocentral scalp locations (e.g., Fig. 1) (Bidelman, 2015, 2018). GLME regression confirmed FFR F0 measured at the scalp was negatively associated with QuickSIN scores; stronger FFRs at Fpz predicted better SIN perception (i.e., lower QuickSIN scores) ($F_{1,10}$=4.75, p=0.05, adj-R^2=0.69). These channel-level results replicate the well-known link between FFRs and SIN processing observed in scalp-recorded data (e.g., Bidelman & Krishnan, 2010; Parbery-Clark et al., 2011; Parbery-Clark et al., 2009; Song et al., 2011; Yellamsetty & Bidelman, 2019).

We next aimed to adjudicate which FFR sources drive this brain-behavior relation and predict SIN performance. To this end, we used multivariate GLME regression to tease apart relations between source-specific speech FFRs (AN, BS, PAC) and QuickSIN scores. This analysis revealed the aggregate of FFR sources predicted SIN performance [$F_{5,4}$=7.95, p=0.0333; null hypothesis coefficients=0], accounting for 87% of the variance in QuickSIN scores (adj-R^2=0.87) (**Fig. 4**). Scrutiny of individual model terms revealed significant neural predictors in BS (p=0.0071) and bilateral AN (AN_{left}: p=0.0062; AN_{right}: p=0.011). Contrary to MEG FFRs (Coffey et al., 2017), neither left nor right PAC FFR_{EEG} predicted QuickSIN performance (PAC_{left}: p=0.13; PAC_{right}: p=0.695). Collectively, these findings suggest subcortical sources not only dominate the FFR_{EEG} but also the link between speech-FFRs and SIN processing as observed in previous EEG studies (e.g., Bidelman & Krishnan, 2010; Parbery-Clark et al., 2011; Parbery-Clark et al., 2009; Saiz-Alía et al., 2019; Song et al., 2011; Yellamsetty & Bidelman, 2019).

[Insert Figure 4 near here]

4. DISCUSSION

By measuring source-level FFRs via EEG, our results confirm a mixture of auditory neural generators contribute to the aggregate FFR and these contributions vary systematically across anatomical level of the hearing pathway. Whereas subcortical sources (AN, BS) showed robust phase-locked FFRs, PAC showed comparably weak responses. In contrast to auditory nerve and midbrain FFR, auditory cortical FFRs were absent >120 Hz. Thus, PAC contributions were restricted to the speech F0. In contrast, we find subcortical FFRs are not only larger at F0 than their cortical counterparts but persist at higher spectral frequencies to maintain coding of the harmonics and timbral elements of speech (see also Bidelman, 2018). Our findings further imply that previously observed relations between speech-FFRs and SIN perception (e.g., Bidelman et al., 2018; Bidelman & Krishnan, 2010; Parbery-Clark et al., 2011; Parbery-Clark et al., 2009; Saiz-Alía et al., 2019; Song et al., 2011) are mediated by subcortical sources and the quality of neural representation at subcortical stages of auditory processing.

4.1 Neural phase locking diminishes along the ascending auditory system

Our neuroimaging data establish that the relative contribution of sources to the scalp FFR_{EEG} varies critically across levels of the auditory neuroaxis. In agreement with prior source imaging work (Bidelman, 2018), we found a gradient in contributions to the response, with the BS (midbrain) forming the largest component of the FFR_{EEG} (i.e., BS > AN >> PAC). Our findings converge with human and animal studies suggesting subcortex (e.g., inferior colliculus and lower structures) as the primary source(s) of the FFR within the 20-200 Hz frequency bandwidth (Bidelman, 2018; Kiren et al., 1994; Smith et al., 1975; Sohmer & Pratt, 1977; Sohmer et al., 1977). Our data lead us to reiterate the critical point that structures producing FFRs critically vary in both a level- (i.e., stage of brain) and stimulus- (i.e., frequency) dependent manner.

Diminishing FFR strength from brainstem to cortex is broadly consistent with the known phase-locking tolerance of nuclei along the ascending auditory pathway (Joris et al., 2004; Wallace et al., 2007; Wallace et al., 2000; Wallace et al., 2002). Physiologically, reduced phase-locking could be due to increased synaptic jitter, organizational complexities of the pathway including (re)convergence, increasing time constants of neuronal membranes, and/or different ionic channels that cause a deterioration in temporal processing in more central neurons (Trussell, 1999). Collectively, these processes act as a low pass filter resulting in each anatomical division having a reduced upper limit of temporal precision. Indeed, phase-locking in auditory nerve persists up to >4-5 kHz (Joris et al., 2004), whereas the upper limit in midbrain neurons is ~1100-1200 Hz (Bidelman & Powers, 2018; Liu et al., 2006). In stark contrast, locking in auditory cortical neurons among several species (monkeys, guinea pigs) is generally

not observed above ~100-120 Hz (Behroozmand et al., 2016; Brugge et al., 2009; Joris et al., 2004; Wallace et al., 2000)—even lower by some accounts (< 50 Hz; Lu et al., 2001). Our EEG data therefore agree with prior animal studies by showing limited phase-locked FFR activity in cortex up to a nominal frequency of 120 Hz. Nevertheless, given the reduced nature of these responses we are left to reaffirm that cortical contributions to the speech-FFR, when weakly present, are largely band-limited to the F0 component of low-pitch stimuli (e.g., male speech) (Bidelman, 2018).

4.2 Neural encoding of speech F0 is robust to noise

While our data show clear level-dependent changes in the strength of FFR coding, we found noise had little impact on F0 responses. The lack of noise effects on pitch (F0) coding is consistent with prior FFR studies (Bidelman, 2016; Bidelman et al., 2018; Bidelman & Krishnan, 2010; Laroche et al., 2013; Prevost et al., 2013; Russo et al., 2004; Smalt et al., 2012). Such resilience (and even increment) of F0 responses in the presence of noise is hypothesized to be due to stochastic resonance (Henry, 1999), the recruitment of low-frequency “tails” of high-frequency cochlear neurons due to high-intensity stimulus levels, and/or across channel integration of F0-related harmonics or intermodulation distortion products (Smalt et al., 2012). All of these putative mechanisms could lead to more robust neural synchronization and therefore redundancy of pitch-relevant cues carried in FFRs even with additive noise (for discussion, see Bidelman, 2017).

4.3 Subcortical sources drive the link between speech-FFRs and SIN perception

Our data reveal SIN processing is governed by a coordinated orchestration of phase-locked activity to the spectrotemporal details of speech generated from different levels of the auditory system. However, our multivariate regression showed that subcortical sources (AN, BS) better predict listeners’ QuickSIN scores. In stark contrast, “cortical FFRs” failed to predict perception. Moreover, we find PAC responses are eradicated for the most meaningful frequencies in speech (i.e., energy above F0) which carry timbral information on talker identity, including important formant cues (cf. Banai et al., 2009; Bidelman & Krishnan, 2010; Kraus & Nicol, 2005). This is not to say that cortex is not involved in SIN perception. On the contrary, there is overwhelming ERP evidence that auditory and non-auditory brain regions in both cerebral hemispheres are heavily involved in SIN perception (Bidelman et al., 2020; Bidelman & Howell, 2016; Bidelman & Yellamsetty, 2017; Billings et al., 2013; Du et al., 2014; Parbery-Clark et al., 2011). Rather, our data reveal little-to-no involvement of cortex when it comes to the relation between sustained, stimulus phase-locked *FFRs* and SIN processing.

Results from MEG studies have suggested associations between FFR periodicity and SIN perception might be driven by right PAC (Coffey et al., 2016). Our findings challenge notions that cortical FFRs mediate SIN processing (Coffey et al., 2016; Coffey et al., 2017) and fail to show a right hemisphere bias in cortical FFRs as measured via MEG (Coffey et al., 2016; Hartmann & Weisz, 2019; Ross et al., 2007). Rather, we find brainstem foci (auditory nerve, midbrain) to be the dominant predictors of SIN listening skills as implied (but never confirmed) in previous FFR studies examining scalp recordings (e.g., Anderson et al., 2010; Bidelman et al., 2018; Bidelman & Krishnan, 2010; Parbery-Clark et al., 2011; Parbery-Clark et al., 2009; Saiz-Alía et al., 2019; Song et al., 2011). In fact, despite being weakly observable, cortical FFRs in the present study did not predict QuickSIN scores.

Presumably, discrepancies between neuroimaging studies reflect differences in the sensitivity of MEG vs. EEG for detecting directed neuronal currents and deep vs. superficial sources. Given the bias of MEG to more superficial and tangential brain activity (deeper brainstem sources become magnetically invisible), we have previously argued this may over-exaggerate and falsely “color” FFRs with neocortical contributions that filter out important subcortical sources (Bidelman, 2018). This may account for the differential brain-behavior relations between FFRs and SIN perception observed in this (EEG) vs. previous (MEG) FFR studies. Indeed, even recent MEG studies have suggested a major brainstem locus when describing evoked periodic neural activity characteristic of the FFR (Kulasingham et al., 2019; Zhao & Kuhl, 2018). Based on EEG and computational modeling, we have suggested that brainstem provide a 3x larger contribution to scalp FFRs (at F0) than cortical sources (Bidelman, 2018). Similar conclusions were recently drawn by Ross et al. (2020) using MEG.

Collectively, our findings corroborate recent work demonstrating functional connectivity between brainstem and auditory cortex (i.e., afferent signal flow) is important in differentiating “good” vs. “poor” perceivers in SIN listening tasks (Bidelman et al., 2018). These previous studies, along with current data, suggest the degree to which speech representations are effectively passed from early to later stages of the neural hierarchy are highly predictive of complex listening skills. We infer that the burden of SIN perception—with regards to the FFR—is mainly on subcortical parts of the auditory neuroaxis and the degree to which their neural representations capture and maintain the dynamic, time-frequency cues of speech.

Acknowledgments

We thank Dr. Caitlin Price for comments on earlier version of the manuscript. This work was supported by grants from the American Hearing Research Foundation (AHRF), American Academy of Audiology Foundation (AAAF), and the National Institute on Deafness and Other

Communication Disorders of the National Institutes of Health (NIDCD R01DC016267) awarded to G.M.B.

Figure legends

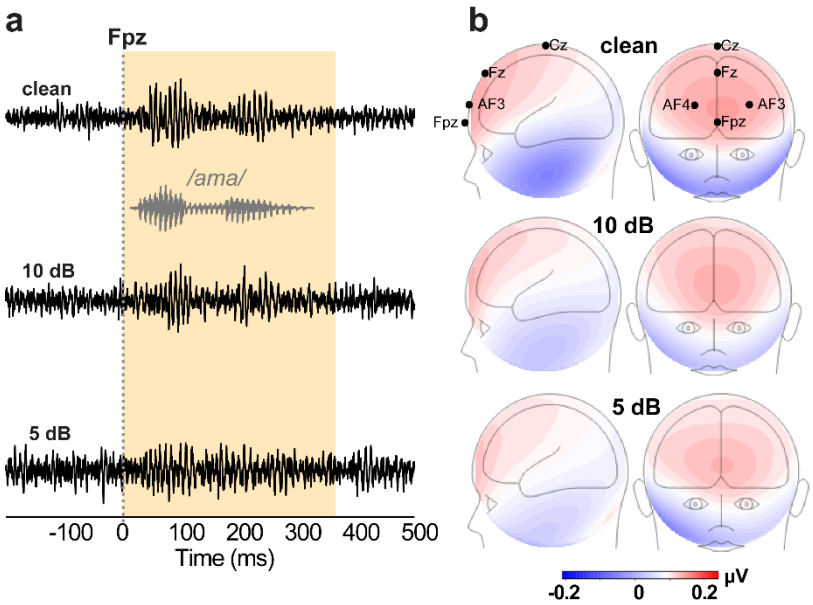

Figure 1: FFR (sensor-level) waveforms and scalp topographies as a function of SNR. (a) Electrode recordings at channel Fpz across noise levels. Gray trace= stimulus waveform. FFRs appear as phase-locked potentials that mirror the acoustic periodicity of speech. (**b**) FFR topographies. Maps reflect the voltage distribution on the scalp, averaged across the periodic "wavelets" of the FFR (i.e., 25 most prominent positive peaks; see Bidelman, 2015) in the response time window (0-350 ms; yellow shading). Maximal FFR amplitude near frontocentral sites (e.g., Fpz) and polarity inversion at the mastoids are consistent with deep midbrain sources that point obliquely in an anterior orientation to the vertex (parallel to the brainstem) (Bidelman, 2015, 2018). Red/blue shading = positive/negative voltage.

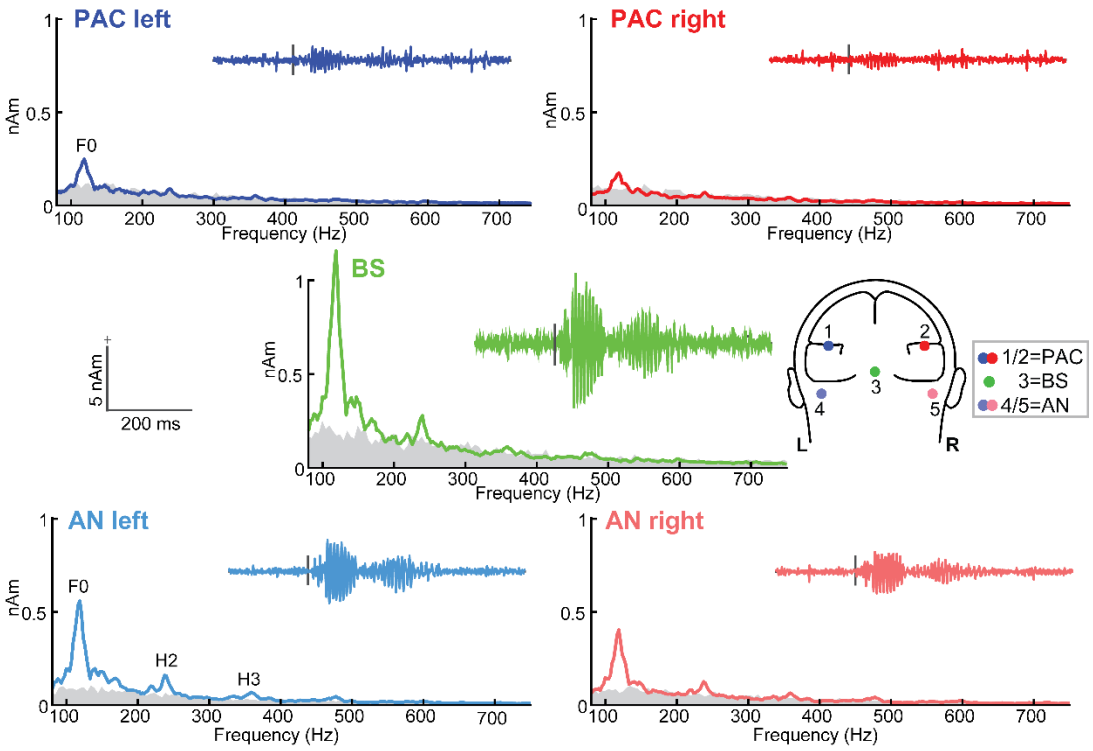

Figure 2: Source-level FFR waveforms and spectra along the ascending auditory neuroaxis. Grand average response spectra and waveforms (*insets*) at each dipole source of the FFR (head model; Bidelman, 2018). Only FFRs to clean speech are shown for clarity. Gray shading = spectral noise floor measured in the pre-stimulus interval. FFRs show strong phase-locking at the speech F0 frequency (~100 Hz) in both subcortical and cortical sources. However, subcortical sources (AN, BS) show additional response energy at higher harmonics of speech (H2 and H3). Cortical (PAC) sources do not show reliable FFRs at higher harmonics (i.e., > 100 Hz). This is also confirmed by raw source waveforms, which show robust responses in AN and BS but weak FFRs in PAC. AN, auditory nerve; BS, brainstem midbrain; PAC, primary auditory cortex.

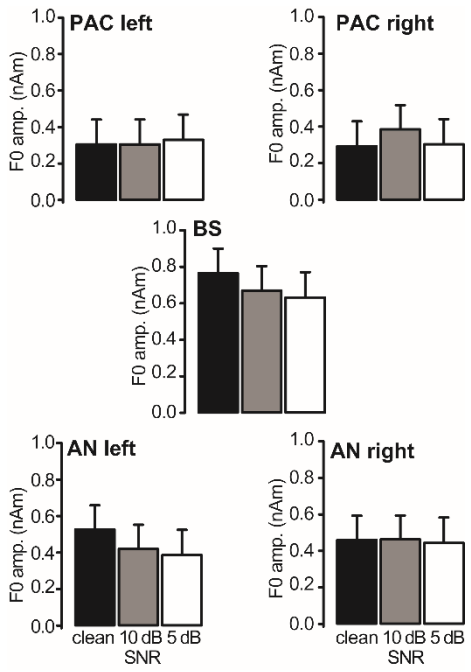

Figure 3: FFR F0 amplitudes as a function of source and SNR. F0 amplitude (re. noise floor) varies across source level and is stronger in subcortical (AN, BS) compared to cortical (PAC) FFR generators. Among all sources, F0 amplitude is largest at the BS level. F0 did not vary across SNRs. Error bars = ±1 s.e.m.

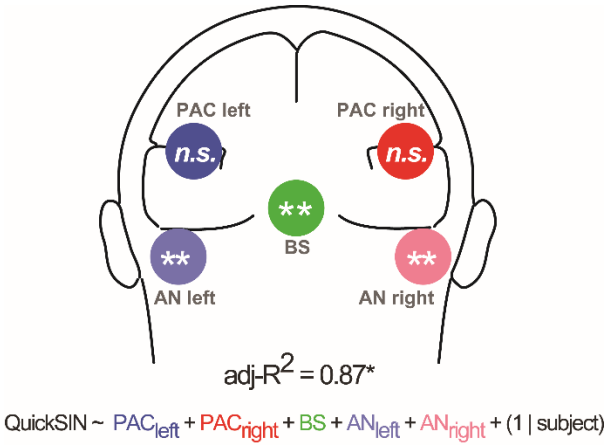

Figure 4: Brainstem speech coding is sufficient to explain SIN listening skills. Subcortical and cortical sources of the FFR collectively account for 87% of the variance (adj-R^2) in behavioral QuickSIN scores. Statistical flags mark significant predictors in the GLME model. Subcortical sources (left/right AN, midbrain BS) are strongly associated with SIN performance, whereas cortical FFRs (left/right PAC) fail to predict behavior. *$p < 0.05$; **$p \leq 0.01$

References

Alain, C., McDonald, K., & Van Roon, P. (2012). Effects of age and background noise on processing a mistuned harmonic in an otherwise periodic complex sound. *Hearing Research, 283*, 126-135.

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. *Journal of Neuroscience, 30*(14), 4922-4926.

Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Reading and subcortical auditory function. *Cerebral Cortex, 19*(11), 2699-2707.

Behroozmand, R., Oya, H., Nourski, K. V., Kawasaki, H., Larson, C. R., Brugge, J. F., Howard, M. A., 3rd, & Greenlee, J. D. W. (2016). Neural correlates of vocal production and motor control in human Heschl's gyrus. *Journal of Neuroscience, 36*(7), 2302-2315.

Bidelman, G., & Powers, L. (2018). Response properties of the human frequency-following response (FFR) to speech and non-speech sounds: level dependence, adaptation and phase-locking limits. *International Journal of Audiology, 57*(9), 665–672.

Bidelman, G. M. (2015). Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. *Hearing Research, 323*, 68-80.

Bidelman, G. M. (2016). Relative contribution of envelope and fine structure to the subcortical encoding of noise-degraded speech. *Journal of the Acoustical Society of America, 140*(4), EL358-363.

Bidelman, G. M. (2017). Communicating in challenging environments: Noise and reverberation. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *Springer Handbook of Auditory Research: The frequency-following response: A window into human communication*. Springer Nature.

Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage, 175*, 56–69.

Bidelman, G. M., Bush, L. C., & Boudreaux, A. M. (2020). Effects of noise on the behavioral and neural categorization of speech. *Frontiers in Neuroscience*.

Bidelman, G. M., Davis, M. K., & Pridgen, M. H. (2018). Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hearing Research, 367*, 149-160.

Bidelman, G. M., & Dexter, L. (2015). Bilinguals at the "cocktail party": Dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners' speech-in-noise recognition deficits. *Brain and Language, 143*, 32-41.

Bidelman, G. M., & Howell, M. (2016). Functional changes in inter- and intra-hemispheric auditory cortical processing underlying degraded speech perception. *Neuroimage, 124*, 581-590.

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research, 1355*, 112-125.

Bidelman, G. M., & Yellamsetty, A. (2017). Noise and pitch interact during the cortical segregation of concurrent speech. *Hearing Research, 351*, 34-44.

Billings, C. J., McMillan, G. P., Penman, T. M., & Gille, S. M. (2013). Predicting perception in noise using cortical auditory evoked potentials. *Journal of the Association for Research in Otolaryngology, 14*(6), 891-903.

Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., Kawasaki, H., Steinschneider, M., & Howard, M. A., 3rd. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *Journal of Neurophysiology, 102*(4), 2358-2374.

Chandrasekaran, B., Kraus, N., & Wong, P. C. (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology, 107*(5), 1325-1336.

Chmiel, R., & Jerger, J. (1996). Hearing aid use, central auditory disorder, and hearing handicap in elderly persons. *Journal of the American Academy of Audiology, 7*(3), 190-202.

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. *Nature Communications, 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 [Original Research]. *Frontiers in Neuroscience, 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. *Nature Communications, 10*(1), 5036.

Cohen, D., & Cuffin, B. N. (1983). Demonstration of useful differences between magnetoencephalogram and electroencephalogram. *Electroencephalogr Clin Neurophysiol, 56*(1), 38-51.

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences of the United States of America, 111*(19), 1-6.

Gordon-Salant, S., Fitzgibbons, P. J., & Yeni-Komshian, G. H. (2011). Auditory temporal processing and aging: Implications for speech understanding of older people. *Audiology Research, 1*(e4), 9-15.

Hartmann, T., & Weisz, N. (2019). Auditory cortical generators of the Frequency Following Response are modulated by intermodal attention. *Neuroimage, 203*, 116185.

Henry, K. R. (1999). Noise improves transfer of near-threshold, phase-locked activity of the cochlear nerve: evidence for stochastic resonance? *Journal of Comparative Physiology A, 184*(6), 577-584.

Herdman, A. T., Lins, O., van Roon, P., Stapells, D. R., Scherg, M., & Picton, T. (2002). Intracerebral sources of human auditory steady-state responses. *Brain Topography, 15*(2), 69-86.

Hillebrand, A., & Barnes, G. R. (2002). A Quantitative Assessment of the Sensitivity of Whole-Head MEG to Activity in the Adult Human Cortex. *Neuroimage, 16*(3, Part A), 638-650.

Holmes, E., & Herrmann, B. (2017). Revisiting the contribution of auditory cortex to frequency-following responses. *Journal of Neuroscience, 37*(21), 5218-5220.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews, 84*, 541-577.

Killion, M., & Niquette, P. (2000). What can the pure-tone audiogram tell us about a patient's snr loss? *The Hearing Journal, 53*(3), 46-53.

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. *Journal of the Acoustical Society of America, 116*(4 Pt 1), 2395-2405.

Kiren, T., Aoyagi, M., Furuse, H., & Koike, Y. (1994). An experimental study on the generator of amplitude-modulation following response. *Acta Otolaryng. Suppl. (Stockh.), 511*, 28-33.

Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends in Neurosciences, 28*(4), 176-181.

Krishnan, A. (2007). Human frequency following response. In R. F. Burkard, M. Don, & J. J. Eggermont (Eds.), *Auditory evoked potentials: Basic principles and clinical application* (pp. 313-335). Lippincott Williams & Wilkins.

Kujawa, S. G., & Liberman, M. C. (2015). Synaptopathy in the noise-exposed and aging cochlea: Primary neural degeneration in acquired sensorineural hearing loss. *Hearing Research, 330, Part B*, 191-199.

Kulasingham, J. P., Brodbeck, C., Presacco, A., Kuchinsky, S. E., Anderson, S., & Simon, J. Z. (2019). High frequency cortical processing of continuous speech in younger and older listeners. *bioRxiv, doi: https://doi.org/10.1101/2019.12.19.883314*, .

Laroche, M., Dajani, H. R., Prevost, F., & Marcoux, A. M. (2013). Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise. *Ear and Hearing, 34*(1), 63-74.

Liu, L. F., Palmer, A. R., & Wallace, M. N. (2006). Phase-locked responses to pure tones in the inferior colliculus. *Journal of Neurophysiology, 95*(3), 1926-1935.

Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates [10.1038/nn737]. *Nat Neurosci, 4*(11), 1131-1138.

Medwetsky, L. (2002). Central auditory processing testing: A battery approach. In J. Katz (Ed.), *Handbook of Clinical Audiology* (5th ed., pp. 510-524). Lippincott Williams & Wilkins.

Middelweerd, M. J., Festen, J. M., & Plomp, R. (1990). Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology, 29*(1), 1-7.

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

Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. *Frontiers in aging neuroscience, 4*, 30.

Parbery-Clark, A., Marmel, F., Bair, J., & Kraus, N. (2011). What subcortical-cortical relationships tell us about processing speech in noise [Research Support, Non-U.S. Gov't

Research Support, U.S. Gov't, Non-P.H.S.]. *European Journal of Neuroscience, 33*(3), 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. *Journal of Neuroscience, 29*(45), 14100-14107.

Picton, T. W., van Roon, P., Armilio, M. L., Berg, P., Ille, N., & Scherg, M. (2000). The correction of ocular artifacts: A topographic perspective. *Clinical Neurophysiology, 111*(1), 53-65.

Popelka, G. R., Moore, B. C. J., Popper, A. N., & Fay, R. R. (2016). *Hearing Aids*. Springer Science+Business Media, LLC.

Prevost, F., Laroche, M., Marcoux, A. M., & Dajani, H. R. (2013). Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. *Clinical Neurophysiology, 124*(1), 52-60.

Ross, B., Tremblay, K. L., & Alain, C. (2020). Simultaneous EEG and MEG recordings reveal vocal pitch elicited cortical gamma oscillations in young and older adults. *Neuroimage, 204*, 116253.

Ross, D., Choi, J., & Purves, D. (2007). Musical intervals in speech. *Proceedings of the National Academy of Sciences of the United States of America, 104*(23), 9852-9857.

Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology, 115*(9), 2021-2030.

Saiz-Alía, M., Forte, A. E., & Reichenbach, T. (2019). Individual differences in the attentional modulation of the human auditory brainstem response to speech inform on speech-in-noise deficits. *Scientific Reports, 9*(1), 14131.

Scherg, M., Berg, P., Nakasato, N., & Beniczky, S. (2019). Taking the eeg back into the brain: The power of multiple discrete sources [Original Research]. *Frontiers in Neurology, 10*(855).

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. *Journal of Clinical Neurophysiology, 19*(2), 91-112.

Schneider, B. A., Daneman, M., & Pichora-Fuller, M. K. (2002). Listening in aging adults: From discourse comprehension to psychoacoustics. *Canadian Journal of Experimental Psychology, 56*(3), 139-152.

Smalt, C. J., Krishnan, A., Bidelman, G. M., Ananthakrishnan, S., & Gandour, J. T. (2012). Distortion products and their influence on representation of pitch-relevant information in the human brainstem for unresolved harmonic complex tones. *Hearing Research, 292*(1-2), 26-34.

Smith, J. C., Marsh, J. T., & Brown, W. S. (1975). Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources. *Electroencephalography and Clinical Neurophysiology, 39*(5), 465-472.

Sohmer, H., & Pratt, H. (1977). Identification and separation of acoustic frequency following responses (FFRs) in man. *Electroencephalography and Clinical Neurophysiology, 42*(4), 493-500.

Sohmer, H., Pratt, H., & Kinarti, R. (1977). Sources of frequency-following responses (FFR) in man. *Electroencephalography and Clinical Neurophysiology, 42*, 656-664.

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011). Perception of speech in noise: Neural correlates. *Journal of Cognitive Neuroscience, 23*(9), 2268-2279.

Trussell, L. O. (1999). Synaptic mechanisms for coding timing in auditory neurons. *Annual review of physiology, 61*(1), 477-496.

Wallace, M. N., Anderson, L. A., & Palmer, A. R. (2007). Phase-locked responses to pure tones in the auditory thalamus. *Journal of Neurophysiology, 98*(4), 1941-1952.

Wallace, M. N., Rutkowski, R. G., Shackleton, T. M., & Palmer, A. R. (2000). Phase-locked responses to pure tones in guinea pig auditory cortex. *Neuroreport, 11*(18), 3989-3993.

Wallace, M. N., Shackleton, T. M., & Palmer, A. R. (2002). Phase-locked responses to pure tones in the primary auditory cortex. *Hearing Research, 172*(1-2), 160-171.

White-Schwoch, T., Anderson, S., Krizman, J., Nicol, T., & Kraus, N. (2019). Case studies in neuroscience: subcortical origins of the frequency-following response. *Journal of Neurophysiology, 122*(2), 844-848.

Yellamsetty, A., & Bidelman, G. M. (2019). Brainstem correlates of concurrent speech identification in adverse listening conditions. *Brain Research, 1714*, 182-192.

Zhang, X., & Gong, Q. (2019). Frequency-following responses to complex tones at different frequencies reflect different source configurations [Original Research]. *Frontiers in Neuroscience, 13*(130).

Zhao, T. C., & Kuhl, P. K. (2018). Linguistic effect on speech perception observed at the brainstem. *Proceedings of the National Academy of Sciences of the United States of America, 115*(35), 8716-8721.