Subcortical sources dominate the neuroelectric auditory frequency-following response to speech

Gavin M. Bidelmana,b,c,*

a School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA
b Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
c University of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, TN, USA

ARTICLE INFO

Keywords:
Auditory evoked potentials (AEPs)
Auditory brainstem response (ABR) to speech
Classical low resolution electromagnetic tomography analysis recursively applied (CLARA)
Intracerebral dipole generators
Time-frequency analysis

ABSTRACT

Frequency-following responses (FFRs) are neurophonic potentials that provide a window into the encoding of complex sounds (e.g., speech/music), auditory disorders, and neuroplasticity. While the neural origins of the FFR remain debated, renewed controversy has reemerged after demonstration that FFRs recorded via magnetoencephalography (MEG) are dominated by cortical rather than brainstem structures as previously assumed. Here, we recorded high-density (64 ch) FFRs via EEG and applied state-of-the art source imaging techniques to multi-channel data (discrete dipole modeling, distributed imaging, independent component analysis, computational simulations). Our data confirm a mixture of generators localized to bilateral auditory nerve (AN), brainstem inferior colliculus (BS), and bilateral primary auditory cortex (PAC). However, frequency-specific scrutiny of source waveforms showed the relative contribution of these nuclei to the aggregate FFR varied across stimulus frequencies. Whereas AN and BS sources produced robust FFRs up to ~700 Hz, PAC showed weak phase-locking with little FFR energy above the speech fundamental (100 Hz). Notably, CLARA imaging further showed PAC activation was eradicated for FFRs >150 Hz, above which only subcortical sources remained active. Our results show (i) the site of FFR generation varies critically with stimulus frequency; and (ii) opposite the pattern observed in MEG, subcortical structures make the largest contribution to electrically recorded FFRs (AN/C21 BS > PAC). We infer that cortical dominance observed in previous neuromagnetic data is likely due to the bias of MEG to superficial brain tissue, underestimating subcortical structures that drive most of the speech-FFR. Cleanly separating subcortical from cortical FFRs can be achieved by ensuring stimulus frequencies are >150–200 Hz, above the phase-locking limit of cortical neurons.

Introduction

The auditory frequency-following response (FFR) is a neurophonic potential recorded at the scalp that reflects sustained, phase-locked neural ensemble activity of the auditory system. FFRs reflect the neural encoding of dynamic, spectrotemporal features of periodic acoustic stimuli and consequently, provide a “neural fingerprint” of sound within the human electroencephalogram (EEG). The remarkable fidelity of FFRs is evident in behavioral experiments in which the neural responses are replayed as audio signals. These studies demonstrate that FFRs evoked by speech sounds (neural potentials) are highly intelligible to the point they can be reliably identified by external observers (Bidelman, 2018; Grafton et al., 1995; Weiss and Bidelman, 2015). Given their remarkable spectrotemporal detail, FFRs have provided important insight into auditory processing including individual differences in speech listening skills (Anderson et al., 2011; Bidelman, 2017b; Bidelman and Alain, 2015; Song et al., 2011), neuroplasticity of learning and language experience (Carcagno and Plack, 2011; Chandrasekaran et al., 2012; Kraus and Chandrasekaran, 2010; Krishnan and Gandour, 2009; Krieman et al., 2012), the neurobiology of music (Bidelman, 2013; Bidelman and Krishnan, 2009, 2011; Bidelman et al., 2011b; Bone et al., 2014; Cousineau et al., 2015), auditory aging (Anderson et al., 2013; Bidelman et al., 2014a; Parthasarathy and Bartlett, 2012), and abnormal encoding of complex sounds in clinical populations (Bellier et al., 2015b; Bidelman et al., 2017; Billiet and Bellis, 2011; Chandrasekaran et al., 2009; Cunningham et al., 2001; Kraus et al., 2017; Rocha-Muniz et al., 2012; Song et al., 2008; White-Schwoch et al., 2015).

Despite an abundance of FFR studies, surprisingly little is understood
about its basic characteristics, first among them, its anatomical origin(s) (i.e., source generators). Converging evidence from single-unit recordings in animal models, human scalp M/EEG, and lesion studies in several species suggest that an array of sources contribute to FFR generation (Bidelman, 2015b; Chandrasekaran and Kraus, 2010; Coffey et al., 2016; Smith et al., 1975; Sohmer et al., 1977). Many of these studies have proposed the inferior colliculus (IC) of the brainstem as the FFR’s primary neural generator. The notion of a midbrain origin to the FFR is supported by the fact that the short latency of the response (~5–10 ms) aligns with first spike latencies in the IC (Langner and Schreiner, 1988)—earlier than possible from cortical generators (L雕gos-Chauvel et al., 1994), FFRs contain phase-locked activity (e.g., >1000 Hz) well beyond the upper limit of phase-locking of cortical neurons (i.e., ~100 Hz) (Aiken and Picton, 2008; Akhoun et al., 2008; Wallace et al., 2000), there is a high correspondence between far-field and near-field intracranial FFRs recorded directly from the IC (Smith et al., 1975), cryogenic cooling of the IC results in disappearance of FFRs within the colliculi and at the scalp (Smith et al., 1975), and the response is eradiated with focal lesions to the IC (Sohmer and Pratt, 1977).

Nevertheless, the dominant sources of the FFR are likely to vary across species. In cat, the auditory nerve (AN), cochlear nucleus (CN) and other olivary nuclei may contribute ~50% of the amplitude to the scalp-recorded FFR, with weaker contributions from IC (Gardi et al., 1979).1 One study has also documented FFRs recorded from medial geniculate body (MGB) (Weinberger et al., 1970). However, Smith et al. (1975) noted that the small amplitude of these responses may have been a far-field radiation of FFRs picked up from lower nuclei and concluded it improbable that the scalp-FFR originates from a locus rostral to IC. Nevertheless, it is clear from these early invasive and cross-species studies that a mixture of brainstem sources are involved in the generation of FFRs (Bidelman, 2015b; Chandrasekaran and Kraus, 2010; Stillman et al., 1978; Tichko and Skoe, 2017).

Localizing scalp potentials in humans is challenged by the fact that far-field potentials are volume conducted to the scalp making it difficult to circumscribe evoked responses to any one anatomical site. In a recent study (Bidelman, 2015b), we recorded high-density (64 channel) speech-evoked FFRs which allowed us the unique opportunity to map the FFR’s topography and triangulate the response using source modeling techniques commonly applied to the auditory cortical event-related potentials (ERPs) (Picton et al., 1999). Dipole mapping and 3-channel Lisnajous analyses (Pratt et al., 1984, 1987) were used to localize the most likely FFR generator and the orientation of its voltage trajectory in 3D space. We found that FFRs were described by two proximal dipole sources within the midbrain (i.e., brainstem IC) each having an oblique, fronto-centrally oriented voltage gradient that oscillated parallel to the brainstem. Our findings were consistent with a midbrain (IC) origin to the FFR noted in previous studies (Bidelman, 2015b; Smith et al., 1975; Sohmer and Pratt, 1977; Zhang and Gong, 2017). However, we also demonstrated that the strength of the response varied dramatically across the scalp; electrodes near the mastoids were able to record FFRs up to ~1100 Hz whereas locations over cerebral sites showed FFRs only up to ~100 Hz, corresponding to the fundamental frequency (F0) of our speech stimulus (see Fig. 6 of Bidelman, 2015b). Higher frequency FFRs near the mastoids is consistent with the notion these channels predominantly record more peripheral following responses, i.e., cochlear microphonic or neural FFRs emitted from auditory nerve (AN) (Chimento and Schreiner, 1990; Marsh et al., 1970).

Despite ample evidence for a brainstem origin (Bidelman, 2015b; Gardi et al., 1979; Smith et al., 1975; Sohmer et al., 1977), renewed controversy surrounding its source(s) emerged after demonstration that FFRs to the voice pitch (F0 = 100 Hz) of speech were observable in MEG responses recorded near auditory cortex (Coffey et al., 2016). Moreover, distributed source analysis suggested that auditory cortex accounted for the highest relative percentage of the neuromagnetic FFR signal, with brainstem nuclei (e.g., CN, IC, MGB) accounting for surprisingly little (~10%) of the response variance (see Fig. S6 of Coffey et al., 2016). Since the FFR is typically interpreted as reflecting a subcortical (brainstem) origin (Chandrakaran and Kraus, 2010; Krishnan and Gandour, 2009; Skoe and Kraus, 2010; Tsounopoulos and Kraus, 2009; Wong et al., 2007), a cortical contribution would qualify theoretical understanding of the response and the locus of experience-dependent plasticity often interpreted in the context of human FFR studies (Bidelman, 2013; Chandrasekaran et al., 2009; Kraus and Chandrasekaran, 2010; Krishnan and Gandour, 2009). However, an important property of sustained potentials like the FFR and auditory steady-state response (ASSR) is that the

---

1 The cochlear microphonic (CM) receptor potential also contributes to the FFR. CM contribution is typically minimized by using alternating polarity stimuli which eradicates the phase-varying (CM) but maintains the phase-invariant neural FFR (Aiken and Picton, 2008; Chimento and Schreiner, 1990; Marsh et al., 1970).
relative contribution of brainstem and cortical sources varies systematically with stimulus frequency (Bidelman, 2015b). Higher frequencies (>80 Hz) evoke peripheral (AN) and brainstem generators whereas low frequencies (<100 Hz) recruit mainly cortical neural ensembles (Herdman et al., 2002; Kuwada et al., 2002). Indeed, MEG-FFRs are only observable at the speech F0 (100 Hz) (Coffey et al., 2016), indicating that cortical contributions, if present, are restricted to the lowest frequencies of the speech spectrum. Thus, a critical but overlooked point in previous studies is that frequency of the phase-locked activity must be considered to properly interpret the generation sites of the FFR. Problematically, MEG is largely insensitive to deep sources (Baillet, 2017; Baillet et al., 2001; Cohen and Cuffin, 1983; Hillebrand and Barnes, 2002).
Consequently, the natural bias of MEG to superficial brain tissue makes it unclear if the “cortical dominance” in FFRs suggested by Coffey et al. (2016) is idiosyncratic to the MEG modality, which tends to inherently underemphasize deep source contributions (i.e., brainstem) known to play an important role in FFR generation (e.g., Bidelman, 2015b; Chi-mento and Schreiner, 1990; Gardi et al., 1979; Smith et al., 1975; Sohmer et al., 1977; Tichko and Skoe, 2017; Zhang and Gong, 2017). Given equivocal findings, the aim of the present study was to better elucidate the various neural generators contributing to the FFR. To this end, we recorded multichannel speech-FFRs from normal-hearing listeners and applied state-of-the-art source imaging techniques to the neural recordings (discrete dipole modeling, distributed imaging, independent component analysis, computational simulations). With the exception of our recent work (Bidelman, 2015b), we are not aware of any EEG studies that have undertaken such a comprehensive source modeling of the FFR as most reports are limited to a single-channel montage. Under similar channel counts, the spatial resolution of EEG source estimation is comparable to MEG (~8–59 mm) (Cohen and Cuf-fin, 1991; Cohen et al., 1990; Hedrich et al., 2017). Yet, a major advantage is that activity of deep sources is more readily recorded with EEG than MEG, allowing us the opportunity to more veridically image both deep (brainstem) and putative superficial (cortical) sources of the FFR (cf. Coffey et al., 2016). We reasoned that cortical contributions to the FFR (if present in EEG recordings) would diminish at higher frequencies and cease altogether above the phase-locking limit of cortical neurons. To test this hypothesis, we performed time-frequency analysis on source-level FFRs extracted from various subcortical and cortical regions of interest (ROIs) to evaluate the frequency-dependence of different FFR sites along the auditory neuroaxis. Our findings confirm a mixture of FFR generators localized to bilateral auditory nerve (AN), upper brainstem (BS) IC, and bilateral primary auditory cortex (PAC). However, we show that the site of FFR generation varies critically with stimulus frequency with subcortical structures making the largest contribution to the electrically recorded FFR (i.e., AN ≥ BS > PAC), opposite the pattern observed in MEG (Coffey et al., 2016). While weak cortical FFRs are apparent for low-frequency stimulation (~100 Hz), we find only subcortical FFR sources (AN, BS) are active for frequencies spanning the majority of the speech bandwidth (i.e., all spectral cues above a voice pitch of F0 = 100 Hz).

**Methods & materials**

**Participants**

We recorded multichannel FFRs from n = 18 young adults (μ ± SD age: 25.0 ± 2.7 years; 4 males, 14 females). All had obtained a similar level of formal education (at least an undergraduate degree) and were monolingual speakers of American English. All participants were required to have minimal (<3 years) formal musical training (μ ± SD: 1.8 ± 1.9 years) as musicianship is known to enhance brainstem and cortical auditory evoked potentials (e.g., Bidelman and Alain, 2015; Bidelman et al., 2014b; Musacchia et al., 2007). Pure-tone audiograms confirmed normal hearing thresholds (i.e., ≤25 dB HL) at octave frequencies between 250 and 8000 Hz. Participants reported no history of neuropsychiatric disorders. Each gave informed written consent in compliance with a protocol approved by the University of Memphis Institutional Review Board and were compensated monetarily for their time.

**Stimulus presentation**

We used a 300 ms (10 ms cos² ramping)/vCv/speech token/ama/to evoke brainstem FFRs (see Fig. 1 of Bidelman and Howell, 2016). The 50 ms nasal (/m/) was flanked by each vowel phoneme (/a/), both 125 ms in duration. Vowel formant frequencies (F1-F3) were 830, 1200, and 2760 Hz. Intensity was fixed across time. The token’s pitch fell gradually over its duration from an F0 of 120 Hz–88 Hz (slope = −0.11 Hz/ms). This low F0 (~100 Hz) was expected to generate FFRs with some proportion of cortical involvement (Coffey

---

**Fig. 6.** Band-specific spectral time courses reveal the spectro-spatio-temporal dynamics of speech coding along the auditory neuroaxis. Band time courses extracted from ITPL maps (see Fig. 5) index the phase-locked source activity contributing the FFR across different speech frequencies. A running t-test (Gathrie and Buchwald, 1991) shows contiguous segments of at least 15 ms duration where spectral power differs from the control Occ band amplitude (■ segments; p < 0.05). Top gray trace, stimulus waveform. (a) Encoding at the stimulus F0 (88–120 Hz). All sources (except the Occ control) show robust responses at the F0 in the 100–200 ms interval during the duration of the speech token. (b) As in (a) but for H5 (440–600 Hz). Only subcortical structures (AN, BS) show reliable encoding of the vowel at higher frequencies. Shading = ± 1 s. e. m.
et al., 2016) in addition to lower subcortical sources (e.g., AN and brainstem; Bidelman, 2015b). Listeners heard 2000 trials (passive listening) presented with fixed, rarefaction polarity. The interstimulus interval (ISI) was jittered randomly between 400 and 700 ms (20-ms steps, rectangular distribution) to avoid rhythmic entrainment of the EEG (Bidelman et al., 2014a; Luck, 2005).

Stimulus presentation was controlled by MATLAB® 2013b (The MathWorks) routed to a TDT RP2 interface (Tucker-Davis Technologies). Tokens were delivered binaurally at 81 dB SPL through ER-30 insert earphones (Etymotic Research). ER-30 earphones feature extended acoustic tubing (20 ft) that allowed us to place their transducers well outside the testing booth. Spatially separating the transcender from participants avoided the possibility of electromagnetic stimulus artifact contaminating neural responses (Aiken and Picton, 2008; Campbell et al., 2012). The lowpass frequency response of the headphone apparatus was corrected with a dual channel 15 band graphical equalizer (dbx EQ Model 215s; Harman) to achieve a relatively flat frequency response through 4 kHz (Bidelman, 2018; Bidelman and Howell, 2016). All response latencies were corrected for the acoustic delay of the headphone.

Electrophysiological recordings and data analysis

Electrophysiological recording procedures were identical to our previous FFR studies (e.g., Bidelman, 2015b). Subjects were instructed to relax and refrain from extraneous body movement (to minimize myogenic artifacts), ignore the sounds they hear (to divert attention to the auditory stimuli), and were allowed to watch a muted subtitled movie to maintain a calm yet wakeful state. Neuroelectric activity was recorded from 64 sintered Ag/AgCl electrodes at standard 10-10 scalp locations (Oostenveld and Praamstra, 2001). EEGs were digitized at a sampling rate of 5000 Hz (SynAmps RT amplifiers; Compumedics Neuroscan) using an online passband of DC-2500 Hz. Electrodes placed on the outer canthi of the eyes and the superior and inferior orbit were used to monitor ocular activity. During online acquisition, all electrodes were referenced to an additional sensor placed ~1 cm posterior to Cz. Data were re-referenced off-line to a common average reference for subsequent analyses. Contact impedances were maintained ≤5 kΩ throughout the duration of the experiment.

Subsequent data processing was performed in Curry 7 (Compumedics Neuroscan) and BESA® Research v6.1 (BESA, GmbH). Evoked potentials were highpass filtered (80–2500 Hz) to isolate the sustained and periodic FFR (Bidelman et al., 2013; Musacchia et al., 2008). Note that this highpass setting also eradicated ocular artifacts (e.g., blinks, saccades) which are restricted to <5–10 Hz (Lins et al., 1993; Picton et al., 2000). Trials whose voltage amplitude/gradient exceed 120/75 μV were automatically rejected prior to averaging. Participants’ EEGs (sensor recordings) were then transformed to source space using an empirically-derived virtual source montage (detailed below), epoched (±200–550 ms), baseline corrected, and subject to (i) conventional time-domain averaging to derive source FFR waveforms, (ii) single-trial time-frequency analysis to examine spectral properties of FFRs emitted from different structures along the auditory neuroaxis, (iii) independent component analysis (ICA), and (iv) distributed source imaging.

FFR dipole analysis and source montage derivation

Dipole modeling. Neuroelectric signals recorded at scalp electrodes reflect a mixture of overlapping potentials originating from different brain sources. Consequently, discrete source analysis was used to noninvasively estimate the location and orientation of the most likely neuronal dipoles underlying the human FFR to speech (e.g., Bidelman, 2015b; Bidelman and Grall, 2014). Dipole locations (e.g., Fig. 3a) reflect the mean “center of gravity” of neural activity that best explains (in a variance minimization sense) the sensor data recorded at the scalp (e.g., Fig. 3a). Fitting was carried out in BESA® (v6.1). We used a four-shell spherical volume conductor head model (Berg and Scherg, 1994; Sarvas, 1987) with BESA default settings, i.e., relative conductivities (1/Ωm) of 0.33, 0.33, 0.0042, and 1 for the head, scalp, skull, and cerebrospinal fluid, respectively, and sizes of 85 mm (radius), 6 mm (thickness), 7 mm (thickness), and 1 mm (thickness) (Herdman et al., 2002; Picton et al., 1999). Importantly, spherical head models yield comparable accuracy when localizing both deep (brainstem) and superficial (cortical) sources (Cuffin et al., 2001).

Localization and orientation fitting was first performed on the group averaged scalp-recorded waveforms (common average referenced potentials) and used to guide individualized fits. FFRs contain many closely spaced periodic peaks, reflecting phase-locked activity to each pitch period of the stimulus. Following our previous source modeling of FFRs (Bidelman, 2015b), we assume that each of these peaks is generated by the same underlying neural structure(s). To this end, we extracted the 20 most prominent peaks of the grand averaged FFR to create a template “wavelet” (see “peaks Fig. 1; Bidelman, 2015b), which we then fit with a six dipole configuration (see Fig. 3b-c). The model included two symmetric sources in bilateral auditory cortex (PAC; source #1–2), upper brainstem near IC (BS; source #3), and approximate locations of bilateral AN (AN; sources #4–5). Importantly, our head model included coverage external to cortex which avoided the possibility that deep, subcortical and/or peripheral sources (e.g., AN) known to elicit FFRs (Bidelman, 2015b) might be misattributed to more superficial brain areas (cf. Coffey et al., 2016). While simplistic, this set of a priori sources was selected based on previous work identifying FFR contributions from AN (Bidelman, 2015b; Chimento and Schreiner, 1990; Wever and Bray, 1930), brainstem inferior colliculus (Smith et al., 1975; Sohmer et al., 1977), and cortical PAC (Coffey et al., 2016). It also allowed specific hypothesis testing on these ROIs. Sparse source models have also been shown to better resolve subcortical sources in M/EEG than denser configurations (Krishnaswamy et al., 2017), further justifying a restricted dipole count. In addition, we seeded a control dipole in posterior occipital cortex (Occ; source #6). This ROI was not expected to show any auditory response and thus, served as a control to evaluate cross-talk between dipoles (i.e., activity of one source leaking into another). To be acceptably robust, we required the model’s goodness of fit (GoF) to explain ≥80% of the data variance, consistent with previous source analysis criterion applied to auditory evoked potentials (Alain et al., 2009; Bidelman, 2015b; Butler and Trainor, 2012; Picton et al., 1999). GoF was computed automatically in BESA as 1-RV, were RV is the normalized sum of squares of the residual activity (–recorded – modeled data) not accounted for by the source waveforms (Picton et al., 1999).

Source montages. To extract source responses within each ROI of our dipole model, we transformed each listener’s raw continuous EEG into source-level recordings using a virtual source montage (Scherg et al., 2002). This digital re-montaging approach applies a spatial filter to all electrodes of the EEG (defined by the foci of our dipole configuration). Relative weights were then optimized to image activity within each ROI while largely suppressing overlapping activity stemming from other concurrently active brain regions (for details, see Scherg and Ebersole, 1994; Scherg et al., 2002). This allowed us to reduce each listener’s EEG (64-channels) to 6 source channels, each of which estimated activity generated within a single ROI of our dipole model. In each participant, the model was held fixed and was used as a spatial filter to derive source waveforms for each individual’s FFR (Zendel and Alain, 2014). The resulting source waveforms reflect the neuronal current (measured in nA) as seen within each anatomical ROI of Fig. 3b. Comparing source waveforms across ROIs allowed us to assess the relative contribution of different generators that might explain the FFR (e.g., AN, BS, PAC).

3 Dipole fitting was conducted on the average of FFR peak “wavelets” (Fig. 3c). This effectively normalized the time-varying nature of the FFR.
**Time-frequency analysis of source FFRs**

Phase-locking in the auditory system varies with anatomical level. Synchronization becomes progressively weaker at higher levels of the pathway: phase-locking in AN fibers persists up to frequencies of ~5000 Hz but cortical units are limited to ~100 Hz (for review, see Joris et al., 2004). Similarly, the dominant generator(s) of the scalp FFR are likely to vary depending on the frequency of the stimulus (Bidelman, 2015b).

With these points in mind, we analyzed FFRs spectrographically (Hoechstetter et al., 2004) to index the joint time-frequency information contained in the neural responses and evaluate frequency-dependent contributions to the FFR across each anatomical site. From artifact-free, single trial epochs we computed intertrial phase-locking (ITPL) (Delorme and Makeig, 2004) on the source montaged data. ITPLs reflect phase-locked neuronal activity consistent across stimulus trials at each time and frequency and are akin to the power spectrum of the evoked potential. Here, ITPL was necessary because it explicitly removed induced, high-frequency neuronal activity (e.g., cortical gamma oscillations) that could masquerade as a “cortical FFR” and confound the interpretation of our data (for discussions, see Bidelman et al., 2015a; Shahin et al., 2009). ITPL values range from 0 (random trial-to-trial phase) to 1 (identical phase across trials). Higher values in the ITPL spectrogram indicate neural responses that are more temporally aligned between trials and thus reflect stronger phase-locking. ITPL spectrograms were computed between 80 and 1000 Hz using a frequency/time resolution of 10 Hz/5 ms. These analyses allowed us to compare how the spectral details of speech are captured in peripheral (AN), brainstem (IC), and cortical (PAC) FFR sources.

**Independent component analysis (ICA) of FFRs**

ICA decomposes the EEG by modeling it as a collection of statistically independent sub-signals via linear spatial filtering (e.g., Bashivan et al., 2014; Jung et al., 2001; Makeig et al., 1996; Makeig et al., 2004). ICA was used to further describe the FFR’s generators and provide converging evidence to our more comprehensive dipole approach. We applied ICA on the raw, grand averaged electrode recordings using the extended infomax ICA algorithm (Lee et al., 1999). In this implementation, the dimensionality of data is optionally reduced by principal component analysis (PCA) prior to ICA calculation. PCA components explaining less than 1% variance are ignored. The resulting ICA decomposition provided a set of orthogonal signals, where components loading lower in number explain a larger percentage of variation in the EEG. ICA topographies were then localized via dipole fitting. One spatial component, corresponding to each ICA topography was included in this source solution. While we have previously used CLARA to image electrophysiological speech processing at high resolution (e.g., Alain et al., 2017), Functional CLARA activation maps were overlaid onto the BESA adult MRI template for visualization with respect to the brain anatomy (Richards et al., 2016). While a template brain (as opposed to individual subject MRIs) was expected to reduce the absolute precision of localization by ~5 mm (Acar and Makeig, 2013), this error was uniform across individuals and was still much smaller than the distance between PAC and brainstem. We computed CLARA maps from FFR data highpass filtered at (i) 80 Hz and (ii) 180 Hz to contrast activity that did (80 Hz filter) and did not include (180 Hz filter) the low 100 Hz response at the speech F0 (i.e., voice pitch). We reasoned that “PAC FFRs” should be silent above 100–150 Hz, the upper limit of phase-locking for cortical neurons (Joris et al., 2004; Wallace et al., 2000) and frequencies at which near-field FFRs from human PAC have been rarely observed (Brugge et al., 2009). In the present study, CLARA distributed imaging provided a qualitative view of speech activations underlying the FFR and helped corroborate the more quantitative time-frequency analysis of our discrete dipole modeling.

**Relative latencies of subcortical and cortical sources**

We measured the latency of each FFR source to verify anatomical plausibility of our dipole model and estimate the propagation delay between successive auditory stages. The time lag between pairs of nuclei was estimated by cross-correlating their source waveforms (Bidelman et al., 2014a; Coffey et al., 2016; Galbraith and Brown, 1990). This provided a running correlation as a function of lag between each source pair. The lag within a search window (1–9 ms) producing the maximum cross-correlation was taken as the latency shift between sources. This constraint was necessary given the periodic nature of the FFR and the ~100 Hz F0 of our stimulus (10 ms period) which yielded a periodic cross-correlation function. Values near the lower bound (1 ms) indicate poor separation of sources. The cumulative delay between consecutive FFR sources was used to ensure latencies fell within a plausible range of transmission time between auditory structures (cf. Coffey et al., 2016).

**Comparison of forward projection of source signals with scalp data and relative contributions of different FFR sources**

To assess the relative contribution of peripheral, brainstem, and cortical sources to the FFR we compared the topographic distribution of each ROI with that of the scalp-recorded sensor data (Coffey et al., 2016). This was accomplished by projecting each ROI source waveform through the forward model (i.e., leadfield), describing the magnitude each source signal contributes at each electrode sensor (Scherg, 1999). The leadfield for a particular source is essentially a spatial filter describing its outward projection to the scalp. The simulated scalp projection of a source is estimated as $$\hat{d}(t) = L \cdot s(t)$$, where $$\hat{d}(t)$$ is the simulated signal at the electrodes (i.e., sensor topography of that source), L is the leadfield matrix of the source model, and $$s(t)$$ is its source waveform (Scherg, 1999), $$\hat{d}(t) = L \cdot s(t)$$ where $$\hat{d}(t)$$ is the simulated signal at the electrodes (i.e., sensor topography of that source), L is the leadfield matrix of the source model, and $$s(t)$$ is its source waveform (Scherg, 1999). $$\hat{d}(t) = L \cdot s(t)$$, where $$\hat{d}(t)$$ is the simulated signal at the electrodes (i.e., sensor topography of that source), L is the leadfield matrix of the source model, and $$s(t)$$ is its source waveform (Scherg, 1999). $$\hat{d}(t) = L \cdot s(t)$$, where $$\hat{d}(t)$$ is the simulated signal at the electrodes (i.e., sensor topography of that source), L is the leadfield matrix of the source model, and $$s(t)$$ is its source waveform (Scherg, 1999).
1990). Following Coffey et al. (2016), we calculated the RMS of the simulated data across all electrode channels and expressed the relative contribution of each ROI source as a percentage of the total signal simulated from each of the six dipoles (per subject). For each participant, we also measured the correlation between each ROI scalp map (simulated projections) and that of their raw FFR (recorded data). These analyses allowed us to assess the relative contribution and similarity of each source ROI topography to the distribution of the empirical FFR signals actually recorded across all electrode locations.

Results

Speech-evoked FFRs appeared as a sustained neurophonic potential, reflecting phase-locked activity to the spectrotemporal features of speech (Fig. 1). As an initial, data-driven approach to unravel the sources of the FFR, we conducted ICA decomposition on the scalp-recorded data (Fig. 2). Of the first 10 IC topographies accounting for ~80% of the data variance, three components explaining 29.1, 12.3 and 2.6% of the variance were localized spatially to the caudal brainstem (midway between IC and AN; likely CN), primary auditory cortex (PAC), and a lateral area proximal to AN. Topographies show current source density (CSD) maps (units = μV/cm²), a reference free measure computed as the second spatial derivative of the field potential at each electrode (Kayser and Tenke, 2015). Sources can be loosely inferred from CSD maps at locations where current flow shows polarity reversal (e.g., hot to cold colors representing a current source-sink). As an example, IC2A shows at least two active dipoles on each side of the head, a vertically oriented source in the middle of the head just above the ear and another anterior-posterior oriented source just behind the ear (see also, raw voltage topography of sensor-level FFRs, Fig. 3a). A mirror image of the configuration shown here is equally likely given that ICA does not typically account for the bilateral symmetry of auditory evoked responses (i.e., FFRs generated in both left/right AN; Dien et al., 1997; Picton et al., 1999).

To corroborate these qualitative observations, we fit a discrete dipole source model to the scalp FFR topography (Fig. 3a) consisting of five sources positioned in peripheral (left/right AN), brainstem (IC), and cortical sources (left/right PAC) (Fig. 3b). A sixth dipole in occipital lobe (Occ) functioned as a control location. While this dipole model is no doubt a simplistic picture of the FFR and does not preclude other possible sources, our model accurately described the mean FFR “wavelet” (i.e., * points, Fig. 1) with a residual variance of 3.7%, corresponding to a GoF of 96.3% (Fig. 3c). Importantly, the model also accurately described the broader time course of the full FFR periodicity, accounting for 79% of the variance in voltage distribution at the scalp (Fig. 3d). These results confirm our discrete source model accurately described the speech FFR.

We transformed each listener's raw FFRs into source-level recordings using a virtual source montage (Scherg et al., 2002) in order to extract source activity within each dipole ROI. This enabled us to view FFR waveforms generated within the AN, IC, and PAC (see Fig. 5). We confirmed biological plausibility of the source responses by verifying that the propagation delay between consecutive structures of the auditory pathway aligned with known transmission times between nuclei (Coffey et al., 2016) (Fig. 4). Response latencies were estimated via cross-correlation analysis between pairs of source waveforms. We found that the propagation delay between AN and IC was 4.49 ± 1.89 ms, consistent with the central transmission time from the cochlea to midbrain estimated via the click-ABR I-V interval (Don and Eggermont, 1978; Edwards et al., 1983; Hall, 1992; Harte et al., 2009; Hashimoto et al., 1981). The cumulative delay between IC and PAC was 8.66 ± 3.56 ms, corresponding to an inter-region transmission time of ~4.2 ms (±8.6–4.49). Lastly, the cumulative lag between AN and PAC was 13.57 ± 4.85 ms, consistent with the earliest response onset observed in human depth electrode recordings from primary auditory cortex (~13 ms) (Légeois-Chauvel et al., 1994) and first spike-latencies in animal models (~11–18 ms; Hell, 1997; Wallace et al., 2002). Few participants showed maximum correlations at the lower bound of the search window (1 ms), indicating good spatiotemporal separation among the source signals of our model.

Source waveforms and time-frequency analysis of single-trial FFRs are shown for each ROI in Fig. 5. Spectrographic maps show intertrial phase-locking (ITPL) (Delorme and Makeig, 2004) at each time and frequency per source. Phase-consistent activity (i.e., FFR) appears in hotter colors. Notable in each spectral map are the two segments of harmonic energy, corresponding to responses to the first and second vowel portions of the /ama/ speech stimulus. AN and BS sources showed strong neural phase-locking up to about the sixth (H6 = 624 Hz) or seventh (H7 = 728 Hz) harmonic. In contrast, bilateral auditory cortex (PAC) (and Occ control) showed weaker phase-locking in both the time- and frequency-domains with little energy above the stimulus F0 (~100 Hz). A partial response is observable in the Occ source at ~100 Hz. Because this control location is not expected to show an auditory response, this could reflect leakage between sources, i.e., pickup of distal auditory responses at the Occ location. We rule out this possibility in subsequent quantitative analyses.

To directly compare the frequency dependence and strength of neural phase-locking between anatomical sites, we extracted the band time courses from each neural spectrogram (e.g., Bidelman, 2017a). We then used a running t-test (unpaired, unpoled variance) to identify contiguous segments in each band where the spectral power differed from both (i) the baseline interval and (ii) the amplitude in the Occ control band. We required that significant segments persist for >15 ms to further minimize false positive rates (Bidelman and Chung, 2015; Guthrie and Buchwald, 1991).

For the F0–H1 band (spanning 88–120 Hz), we found that right AN, BS, and right PAC showed statistically reliable phase-locking during the vowel durations of the stimulus that was both larger than the baseline and Occ control source amplitudes (Fig. 6a). This suggests a putative cortical contribution to the FFR at low-frequencies (~100 Hz) (e.g., Coffey et al., 2016). The fact that the Occ source strength did not differ from baseline (within any harmonic band), confirms good spatial separation between sources and minimal cross-talk (i.e., spatial bleed) from distal FFR sources. In stark contrast to amplitudes at H1, PAC showed no reliable FFRs for higher harmonics of speech (H2-H5) whereas BS and AN FFR phase-locking remained robust (e.g., H5 band, Fig. 6b). Stronger subcortical representation (i.e., AN/BS > PAC) to the FFR was apparent beginning as low as the second harmonic (176–240 Hz).

Direct comparisons of phase-locking strength (pooling mirrored sources across left/right hemispheres) revealed FFRs were modulated by stimulus frequency and anatomical level (Fig. 7). Amplitudes became progressively weaker in all sources with increasing frequency, consistent with the roll-off of neural phase-locking at higher anatomical levels. However, subcortical sources showed better phase-locking than cortical generators (i.e., AN/BS > PAC) regardless of the spectral band. A mixed-model ANOVA (SAS® 9.4) with fixed effects of source (4 levels: PAC, AN, BS, Occ) and harmonic # (5 levels: H1-H5) (subjects = random factor) conducted on sqrt-transformed ITPLs (used to normalize heteroscedastic variance) confirmed main effects of source (F(4,232) = 20.27, p < 0.0001, n² = 0.16) and harmonic (F(4,232) = 8.75, p < 0.0001, n² = 0.10) on phase-locking strength (with no interaction). Pooled across harmonics, Tukey-corrected comparisons revealed stronger phase locking in AN (p = 0.0020) and BS (p = 0.0008) relative to PAC. Both subcortical sources showed similar phase locking (i.e., BS = AN; p = 0.99) that was, as

4 Some listeners showed lower than expected lags (e.g., 6 ms for AN–PAC, Fig. 4a). This artifact can occur when using cross-correlation for lag estimation. As explained by Coffey et al. (2016), the transmission time to PAC (13 ms) (Légeois-Chauvel et al., 1994) is longer than the average period of our stimulus F0 (104 Hz ξ 9.6 ms), which results in multimodal cross-correlation distributions. Selecting peak values between 1 and 9 ms helped minimize these effects but given the periodic nature of a cross-correlation function, smaller than expected lags can sometimes still occur. Similar issues were observed in Coffey et al. (2016).
expected, more robust than the Occ control (p < 0.0001). Mean phase-locking to speech in PAC did not differ from the control source (i.e., PAC = Occ; p = 0.074). Similarly, whereas PAC showed a systematic decline in phase-locking between H1 and H5 [linear effect: p = 0.0087], BS remained stable in phase-locking to stimuli up to H5 (440–600 Hz) [linear effect: p = 0.072]\(^5\). Collectively, these findings suggest (i) the FFR is dominated by activity from subcortical sources (i.e., AN, BS) which also provide richer encoding of speech through a more expansive frequency bandwidth, (ii) the relative weighting of FFR in terms of spectral detail varies from peripheral to central auditory structures (i.e., AN > BS > PAC), and (iii) contrary to MEG findings (cf. Coffey et al., 2016), cortex plays a weak (even negligible) role in generating the electrical FFR.

Functional CLARA images of the FFR to speech projected onto T1-weighted MRI anatomies are shown in Fig. 8. When FFR data were high-pass filtered at 80 Hz to include responses to the low F0-pitch of speech, functional maps confirmed activation in midbrain inferior colliculus (IC) and areas adjacent to bilateral superior temporal gyrus (PAC). Contrastively, when FFRs were filtered at 180 Hz (removing the F0) we found that cortical activity was eradicated at these higher frequencies and activation was circumscribed mainly to brainstem IC. These distributed source images corroborate our discrete dipole and spectral results (cf. Fig. 6) by confirming some (albeit weak) involvement of cortical structures to the FFR (Coffey et al., 2016). Critically however, cortical contributions are apparent only for low stimulus frequencies (~100 Hz); we observe only subcortical FFR sources are active for frequencies spanning the majority of the speech bandwidth (>150 Hz).

Having demonstrated that an assembly of brainstem and cortical sources participate in the generation of the speech-FFR, we next aimed to establish the relative contribution of these different sources to the scalp-recorded response. To this end, we forward projected individual source activities to the scalp through the dipole model’s leadfield. This allowed us to visualize the voltage topography produced by each generator and assess similarity between simulated and empirical (recorded) FFR data.

Source-specific projections are shown in Fig. 9a along with the scalp topography of the raw FFR (i.e., electrode recorded data). Of the four distinct sources examined, it is apparent that the frontocentral topography of the scalp FFR is best described by a combination of AN and IC (subcortical) source activity. This topography has been described in previous FFR/ASSR studies (Bidelman et al., 2015a; Bidelman, 2015b; Krishnaswamy et al., 2017; Vanvooren et al., 2014) and is consistent with deep, obliquely oriented source(s) of the brainstem. Indeed, the proportion of subcortical vs. cortical sources contributing to the FFR depended critically on anatomical level \(F_{3,51} = 48.63, p < 0.0001, \eta^2 = 0.74\) (Fig. 9b). AN, IC, and PAC explained 23%, 39%, and 23% of the recorded FFR’s scalp topography, respectively. Tukey contrasts revealed that brainstem IC explained a larger portion of FFR recordings than either PAC (p < 0.0001) or AN (p < 0.0001). We also found that the contribution of PAC differed from the control Occ region (p = 0.001), again indicating cortex makes a statistically reliable contribution to the FFR when the response bandwidth includes low-frequency (100 Hz) information of the speech F0. However, subcortical sources (AN, IC) accounted for the majority (62%) of the overall FFR compared to the combination of other generators (\(t_{27} = 7.32, p < 0.00001\)). The correlation between each source projection’s topography and that of the recorded FFR data was similarly modulated by anatomical level \(F_{3,51} = 6.60, p = 0.0007, \eta^2 = 0.31\) (Fig. 9c). A linear contrast showed that the correspondence with the FFR topography became progressively weaker with ascending source along the auditory pathway (i.e., AN > IC > PAC > Occ; p < 0.0001).

To corroborate our empirical findings, we conducted model simulations to further understand the relative weightings of different source generators of the FFR. We first created a synthetic source waveform that captured the most important spectrottemporal characteristics of the recorded FFR (Fig. 10, left). This template response was constructed using 10 harmonics of 104 Hz (i.e., the mean F0 of our stimulus) with amplitudes decaying according to 1/f. The lowpass nature of this template waveform mimicked the natural spectral roll-off of the FFR, which has weak energy above ~1200 Hz (see Fig. 5; Bidelman, 2015b; Chandrasekaran and Kraus, 2010). We then used the BESA\(^6\) Dipole Simulator (http://www.besa.de/downloads/besa-simulator/) to simulate multichannel FFRs as if they were recorded at the scalp. Using the simulator, we placed the template FFR into each of the six dipoles of our discrete source model (Fig. 3b). For all simulations, the time course of the control Occ was turned off since it made a negligible contribution to FFR (see Fig. 9). We assumed that each dipole generated an FFR with the same spectral characteristics of the template that varied only in amplitude, latency, and phase. We time shifted the AN, BS, and PAC waveforms based on the latencies of measured in the empirical data (see Fig. 4; AN = 0 ms, IC = 4.5 ms, PAC: 13 ms).

We then simulated multichannel FFRs in an iterative manner, changing the amplitude scaling of different sources on each step. For each simulation, we varied each source’s amplitude and recomputed the scalp topography of the resulting (simulated) data. Thus, there were three open parameters during a single model fit (scaling factors for AN, BS, and PAC; left and right were treated identically). Amplitude scaling was varied between ~1 and 1 nAm (10 steps). This resulted in a total parameter space of 1000 amplitude settings, corresponding to all three-way combinations of possible AN, BS, and AN scaling factors (i.e., 10 x 10 x 10).

\(^5\) In contrast to the strength of neural phase-locking (Fig. 7), pitch tracking accuracy (i.e., cross-correlation between stimulus and FFR F0 track; Bidelman et al., 2011a; Krishnan et al., 2005) did not differ across sources [Pearson’s r: AN: 0.27 ± 0.21; BS: 0.23 ± 0.18; AN: 0.29 ± 0.22; \(F_{2,54} = 0.49, p = 0.62\)]. However, this might be expected given the shallow F0 of our speech stimulus—which was well within the phase-locking limits of brainstem and cortical neurons (Joris et al., 2004)—and the fact that pitch tracking accuracy is less sensitive than pitch strength to stimulus and subject factors (Bidelman et al., 2011a; Krishnan et al., 2005).
Fig. 8. CLARA source imaging of speech-FFRs. Functional source activation images \((n = 10\) fit iterations; voxel size \(= 5\) mm (Talairach); regularization \(= 0.01\% SVD\)) for the speech FFR. PAC = primary auditory cortex; IC = brainstem inferior colliculus. Functional maps are projected onto the BESA template brain. For FFRs highpass filtered at 80 Hz (a), maps show activation to speech in both auditory brainstem and cortex. (b) In contrast, when FFRs are filtered at 180 Hz (removing the F0), only brainstem IC sources remain active (PAC is eradicated).

Fig. 9. Relative weighting of sources to the FFRs generation reveals the strongest contributor is peripheral (AN, IC) sources. (a) (top) Scalp topography of the FFR. (bottom) Sensor space projection of each source FFR waveform based on the 6-dipole model shown in Fig. 3b. Maps show the scalp distribution of the FFR (full bandwidth response) at a representative peak of the FFR (i.e., see *Figs. 1b and 3c). Topography of the FFR is similar to AN and IC projections. (b) Relative percentage of the FFR signal explained by each sensor-space simulation. The strongest contributor to the EEG-FFR is brainstem IC. (c) Correlations between individual source topographies (i.e., panel a) and the scalp distribution of the recorded FFR for each listener. Peripheral sources (AN, IC) show the highest correlation with empirically recorded FFRs. error bars \(= \pm 1\) s.e.m.

Fig. 10. FFR model simulations. (left) Synthetic template source waveform and spectrum used to simulate FFRs. (middle) Results of the optimal solution minimizing the MSE between the scalp topographies of the actual (see Fig. 3a) and simulated data. Scalp maps show the individual topographies of each dipole source projection. Waveforms show the response at each individual site where individual amplitudes are weighted according to the optimal parameters determined via minimization fitting. Source waveforms are time shifted based on response latencies measured from empirical data (see Fig. 4; AN = 0 ms, IC = 4.5 ms, PAC: 13 ms). The Occ control source was set to zero for all simulations. (right) Simulated FFR topography based on the optimally determined weighting of sources: \(\text{FFR} \approx -0.11\text{AN} + 0.33\text{BS} - 0.11\text{PAC}\). Brainstem contribution to the FFR is 3x that of other sources. Note the similarity between simulated scalp map and actual data shown in Fig. 3a. The vertical line in the source time courses shows the latency at which the scalp map was computed.
For a given simulation (i.e., amplitude triplet), we compared the simulated FFR scalp map to that of the actual measured FFR data (see Fig. 3a) via mean-squared error (MSE). Computing MSE for all 1000 iterations allowed us to determine which combination of AN-BS-PAC scaling minimized the error between simulated and recorded FFRs.

Fig. 10 shows the results of the optimal solution that minimized MSE between the scalp topographies of the model and actual FFR recordings. The optimal scaling factors were FFR ~ -0.11AN + 0.33BS – 0.11PAC. Notably, the BS source contributed nearly 3x the energy to the scalp FFR relative to other sites. The negative scaling of the AN and PAC sources indicates that the phases of these source responses were of opposite polarity (i.e., 180°) relative to BS. The resulting scalp topography of the simulated data (Fig. 10, right) closely resembles the empirical FFR data (see Fig. 3a). While our model and parameter space are non-exhaustive and make simplifying assumptions, simulations corroborate our empirical results (e.g., Fig. 9) by revealing the primary source generator of the electric-FFR is dominantly of subcortical origin. Collectively, both our empirical findings (inverse source estimation) and modeling (forward estimation) reveal under various converging measures, that peripheral (AN) and brainstem (IC) structures describe the largest proportion of the electrically-recorded FFR.

Discussion

By recording multichannel FFRs to speech via EEG, our data confirm a mixture of generators localized to bilateral AN, brainstem IC, and bilateral PAC. These findings corroborate MEG-recordings which have shown that portions of the speech-FFR are generated at both subcortical and cortical levels of the auditory system (cf. Coffey et al., 2017; Coffey et al., 2016). However, frequency-specific scrutiny of source waveforms revealed that the relative contribution of these nuclei to the aggregate FFR varied systematically with stimulus frequency. Whereas AN and BS sources produced robust FFRs up to about the sixth or seventh harmonic (~600–700 Hz), PAC showed markedly weaker phase-locking and was unable to encode stimulus energy for frequencies above the speech fundamental (F0=100 Hz). Moreover, we found that subcortical structures made the largest contribution to the electrically-recorded FFR: AN and BS dominate the response (i.e., AN > BS > PAC) for a majority of the speech bandwidth above the voice pitch—a finding confirmed via computational simulations of the FFR (Fig. 10). By our estimates, when speech contains low-pitch energy near ~100 Hz, more than 60% of the scalp FFR is dominated by phase-locked activity emitted from brainstem, 100% for speech signals >150 Hz (Figs. 8–9). Our findings directly contrast previous reports which have suggested that auditory cortex accounts for the highest relative percentage of the neuromagnetic FFR signal (see Fig. S6 of Coffey et al., 2016). Taken together, our findings challenge the generalizability of MEG-based descriptions of the FFR (Coffey et al., 2017; Coffey et al., 2016) and reveal important frequency-specific modulations to its source interpretation.

There are several features of the Coffey et al. data that warrant scrutiny and might account for differences in MEG (Coffey et al., 2017; Coffey et al., 2016) vs. EEG-based FFRs (present study; Bidelman, 2015b). First, magnetic FFRs contain remarkably little energy above a speech fundamental of 100 Hz (see Fig. 1 of Coffey et al., 2016). This contrasts the rich spectrum observed in electrical recordings in which EEG-FFRs show strong phase-locked energy well up to ~1000 Hz, covering 10+ harmonics of the F0 and higher formant information (e.g., Fig. 5; Bidelman, 2015d; Parbery-Clark et al., 2009; Skoe and Kraus, 2010). Higher spectral frequencies (i.e., formants) are arguably more important to perception than F0-pitch alone since they carry information on the timbral identity of speech and thus, determine “what” is being said (Assmann and Summerfield, 1989, 1990). Moreover, in agreement with our EEG data (Figs. 6–8), MEG-FFRs localize to auditory cortex (PAC) but only for stimulus/response frequencies < 100 Hz. Yet, this finding might be expected given the known phase-locking limits of PAC neurons and near-field FFR recordings from cortex which are similarly band limited to 100 Hz (Brugge et al., 2009; Joris et al., 2004; Wallace et al., 2000) or even lower by some accounts (<50 Hz; Lu et al., 2001). In stark contrast, the bandwidth observed in electrical FFRs extends up to 1000 Hz (Fig. 5; Bidelman, 2015d; Parbery-Clark et al., 2009; Skoe and Kraus, 2010), more consistent with the phase-locking capabilities of brainstem IC neurons (i.e., ~1100 Hz; Liu et al., 2006).

Secondly, previous source analysis of neuromagnetic FFRs suggest that auditory cortex accounts for the highest relative percentage of the FFR signal, with brainstem nuclei (e.g., CN, IC, MGB) describing surprisingly little (~10%) of the response variance (see Fig. S6 of Coffey et al., 2016). However, these findings are at odds with the aforementioned animal work and human EEG studies which, through converging techniques, suggest FFRs are dominated by subcortical and peripheral sources (e.g., AN; Bidelman, 2015b; Gardi et al., 1979; Smith et al., 1975). Our data here demonstrate that the FFR is dominated by early, subcortical sources as peripheral as the auditory nerve (Figs. 6–7). Our findings are thus in accordance with previous EEG studies which have similarly reported that brainstem structures and AN dominate the electrically-recorded scalp FFR (see Fig. 4 of Bidelman, 2015b).

In their analysis, Coffey et al. (2016) also noted similarities (correlations) between the topographies of the obligatory cortical ERPs (P1-N1-P2) and the FFR (their Fig. 3), concluding their “striking resemblance [supported] the existence of a cortical contribution to the FFR (Coffey et al., 2016, p. 3).” Known properties of MEG could have driven these results. In particular, radial sources are largely invisible to MEG which is also far less sensitive to deep sources than EEG (Baillet, 2017; Baillet et al., 2001; Cohen and Cuffin, 1983; Hillebrand and Barnes, 2002). A dipole positioned at the center of a spherical volume conductor (i.e., the head) is always radial to the surface and elicits zero magnetic external field. Activity from deep sources in the brain (e.g., brainstem: IC, CN, etc.) will be highly suppressed in MEG recordings, contributing little to no measurable response. Consequently, the natural bias of MEG to superficial brain tissue makes it unclear if the cortical dominance in magnetic FFRs (Coffey et al., 2016) is idiosyncratic to the MEG modality, which inherently underemphasizes brainstem contributions. Moreover, assertions based on topographic comparisons with the cortical ERPs are potentially flawed given the different filtering that was applied to FFR (80–2000 Hz) vs. ERP (2–40 Hz) signals (Coffey et al., 2016, p. 9). While FFRs and cortical ERPs are spatially distributed similarly across the scalp (e.g., Bellier et al., 2015a; Bidelman, 2015b), each class of response resides in a distinct bandwidth of the EEG spectrum [ERPs: < 40 Hz; FFR: F0 (stimulus dependent)–1100 Hz]. As shown in this study, inference on the FFR’s sources must critically consider the frequency of the phase-locked activity, since the relative contribution of brainstem and cortical sources to sustained potentials is known to vary with stimulus frequency (Bidelman, 2015b); higher frequencies evoke peripheral (AN) and brainstem generators whereas low frequencies recruit mainly cortical neural ensembles (Herdman et al., 2002; Kuwada et al., 2002).

In this vein, our discrete and distributed source imaging results corroborate our spectral results by confirming some (albeit weak) involvement of cortical structures in generating the FFR (Coffey et al., 2016). However, our data temper and extend recent studies by demonstrating different sources to the sustained speech FFR vary critical depending on the stimulus frequency and that a cortical contribution to the FFR is eradicated for speech frequencies above ~100–150 Hz. For the majority of meaningful frequencies in speech (i.e., all higher harmonics above F0, formant frequencies) we find no evidence of cortical involvement (Fig. 9). Cortical FFRs need to be observable but only under the restrictive circumstance when speech contains a low frequencies (e.g., F0 pitch=100 Hz), as might be the case with a male speaker. However, we note that the absence of cortical FFRs above ~100 Hz does not preclude that fact that cortex may represent frequency (pitch) via a population rate codes (Michel et al., 2013), rather than a temporal-based scheme as reflected in FFRs. Our results also align with notions that cortical neurons might reflect the temporal encoding of speech envelope (i.e., F0) (Abrams et al., 2017) but not higher frequency, temporal fine structure
(i.e., harmonics) that is robustly coded by subcortical ensembles (White-Schwoch et al., 2016). Nevertheless, our findings are in close agreement with the known phase-locking limits of PAC neurons (Joris et al., 2004; Wallace et al., 2000) and nearfield FFRs recorded in human PAC (Brugge et al., 2009) which similarly fail to show phase-locked responses in either single-unit or ensemble neural responses above ~150 Hz. Similarly, multi-unit recordings in PAC of non-human mammals (akin to the ensemble FFR) reveal phase-locked activity at 100 Hz but not 250 Hz (Abrams et al., 2017; Steinschneider et al., 1980).

Our results also corroborate recent modeling studies which suggest that <100 Hz FFRs are best described by long latency (13 ms and 26 ms) generators (presumably of cortical origin; see Fig. 4a) (Tichko and Skoe, 2017), with higher frequencies of the FFR (> > 100 Hz) drawing contributions from earlier, subcortical origins. Of practical implication, our data imply that to achieve clean separation of “subcortical”- from “cortical-FFRs”, investigators should ensure stimulus frequencies fall above >150 Hz. This range is above the known phase-locking limit of cortical neurons but where brainstem correlates of auditory-perceptual phenomena and plasticity are still readily apparent (Bidelman et al., 2011b; Bidelman and Krishnan, 2009); Bidelman et al., 2017, p.3615; Bidelman et al., 2014b; Clinard et al., 2010).

The stark frequency-dependence we find in speech FFRs is consistent with the broader literature on the auditory steady-state response (ASSR), a sustained potential evoked by modulated tones. Indeed, source modeling of the ASSR shows high modulation rates (>80 Hz) yield dominant contributions from midbrain sources (Herdmann et al., 2002; Kuwada et al., 2002). Furthermore, high-frequency ASSRs are still observed despite bilateral aspiration of the auditory cortices which rules out the possibility of cortical generator(s) (Kiren et al., 1994). Moreover, sleep (Aoyagi et al., 1993) and general anesthesia (e.g., propofol; Pflourde et al., 2008) attenuate the 40 Hz (cortical) ASSR but have little effect on the higher 80 Hz (brainstem) component. Insomuch as the FFR and ASSR represent similar neurophonic phenomena (i.e., auditory phase-locked activity), it is perhaps predictable that we find sources of the speech-FFR vary in a frequency-dependent manner (Figs. 5–8).

It is worth noting that both this and previous studies (Bidelman, 2015b; Coffey et al., 2016) localizing the FFR used passive listening tasks (i.e., participants viewing silent films). Thus, an interesting yet open question is the extent to which brainstem and cortical contributions to the speech FFR change with attentional state or task demands, e.g., as might be involved in learning (Chandrasekaran et al., 2012; Xie et al., 2017a) or taxing behavioral tasks (e.g., speech-in-noise processing; Bidelman, 2017b; Parbery-Clark et al., 2011). Additionally, a limitation of our study is that our stimulus design considered a single speech stimulus. While this token contained broadband information that allowed us to assess phase-locking to a wide bandwidth of speech frequencies (100–1000 Hz), future studies that parametrically vary F0 in speech and non-speech stimuli (Kuwada et al., 2002) would be interesting to assess the weighting of different FFR generators and replicate the results here. To this point, our simulations using tone complexes are promising because they provide converging evidence of a brainstem locus to the FFR using non-speech stimuli. Our approach also did not evaluate possible contributions of other lower brainstem nuclei (e.g., CN) shown to generate FFRs in animal studies (Gardi et al., 1979). Given the close proximity of nuclei in the caudal brainstem, it is possible that if present, some CN activity was misattributed to our AN source(s). This is a known limitation of all M/EEG source localization studies on tightly spaced generators (Coffey et al., 2016). Intracranial and/or single-unit recordings along the entirety of the auditory neuroaxis are necessary to unambiguously determine the sources of the FFR.

Cortical and brainstem auditory responses are differentially susceptible to drug manipulations (Selinger et al., 2016) and auditory learning (Elmer et al., 2017). Moreover, studies have shown that cortical and brainstem speech processing are differentially affected by experience-dependent plasticity (Bidelman and Alain, 2015; Bidelman et al., 2014b), age-related hearing loss (Bidelman et al., 2014a), and neurodegenerative disorders such as mild cognitive impairment (Bidelman et al., 2017). Collectively, while studies point to functional distinctions, it is likely that a coordinated interplay between brainstem and cortex is necessary to drive certain aspects of sound learning (Cans and Sanes, 2017; Xie et al., 2017b; Yi et al., 2016), hone listening skills in complex environments (Parbery-Clark et al., 2011), and enhance auditory brain function following short- or long-term auditory experience(s) (Bajc et al., 2010; Bidelman et al., 2014b; Musacchia et al., 2007; Suga et al., 2000). On this point, we have recently reported that the functional coupling between brainstem and cortical auditory speech processing is disrupted during the aging process (Bidelman et al., 2014a), but is strengthened (offset) in older adults with musical training (Bidelman and Alain, 2015). Addressing similar questions of subcortical-cortical reciprocity in future studies may require the use of more imaginative stimulus paradigms (e.g., Bidelman, 2015d) and novel recording/analysis approaches (e.g., Bidelman et al., 2013; Forte et al., 2017; Söhmmer and Fienmesser, 1970; Tietze, 1979) that allow for the concurrent monitoring of brainstem and cortical neural activity.

Conclusions

Results of the present study reveal that the methodology used to record FFRs is not unbiased and acts to “color” the response, akin to how different microphones might color the recording of sounds they detect. Clearly, it is important to keep the methodology in mind when interpreting far-field responses called the “FFR” and interpreting its underlying generators in a given task or stimulus paradigm. MEG seems to yield FFRs with (overemphasized) neocortical contributions (Coffey et al., 2016) that “filter out” important subcortical contributions. In contrast, EEG-based FFRs better detect deep brainstem FFRs (auditory nerve, midbrain) that reflect subcortical auditory processing.

Acknowledgements

The author thanks Megan Howell and Mary Katherine Davis for assistance in data collection and Bonnie Brown for comments on earlier versions of the manuscript. This work was supported by grants from the American Hearing Research Foundation (AHRF), American Academy of Audiology (AAA) Foundation, and University of Memphis Research Investment Fund (UMRIF) awarded to G.M.B.

References


---

6 As noted by Abrams et al. (2017), the propensity of cortical multi-unit (ensemble) activity to follow periodicities as fast as 100 Hz contrasts single-unit data, in which neurons show considerably poorer phase-locking for frequencies above even 50 Hz (Lu et al., 2001). Conceivably, ensemble activity like the FFR might be able to phase lock at higher rates than single units if activity across frequencies is optimally combined in a “volleying” scheme (Weyer and Bray, 1937). However, despite some improvement in cortical synchronization rates, available evidence indicates that even ensemble activity is still restricted to 100–150 Hz (Brugge et al., 2009; Abrams et al., 2017).


