Causal relationship between the right auditory cortex and speech-evoked frequency-following response: Evidence from combined tDCS and EEG

Guangting Mai\textsuperscript{1,\*}, Peter Howell\textsuperscript{1,\*}

\textsuperscript{1}Department of Experimental Psychology, University College London, London, UK, WC1H 0AP

*Correspondence: guangting.mai.15@ucl.ac.uk (G.M.); p.howell@ucl.ac.uk (P.H.)

Abbreviated title: Causal relation between auditory cortex and FFR

Conflict of interest: The authors declare no competing financial interests

Author contributions: G.M. designed research, performed research, and analysed the data; P.H. provided laboratory resources and supervised the study; G.M. drafted the manuscript; both authors contributed to writing, editing and approving the paper.

Acknowledgments: We thank Dr Naheem Bashir for instructions on tDCS experiment setups and Mr Andrew Clark for technical support for EEG experiments. We also thank the Dominic Barker Trust who paid for the tDCS equipment. G.M. was supported by a UCL Graduate Research Scholarship for Cross-disciplinary Training.
Abstract

Speech-evoked frequency-following response (FFR) reflects the neural encoding of speech periodic information in the human auditory systems. FFR is of fundamental importance for pitch and speech perception and serves as clinical biomarkers for various auditory and language disorders. While it is suggested that the main neural source of FFR is in the auditory brainstem, recent studies have shown a cortical contribution to FFR predominantly in the right hemisphere. However, it is still unclear whether auditory cortex and FFR are causally related. The aim of this study was to establish this causal relationship using a combination of transcranial direct current stimulation (tDCS) and scalp-recorded electroencephalography (EEG). We applied tDCS over the left and right auditory cortices in right-handed normal-hearing participants and examined the after-effects of tDCS on FFR using EEG during monaural listening to a repeatedly-presented speech syllable. Our results showed that: (1) before tDCS was applied, participants had greater FFR magnitude when they listened to speech from the left than the right ear, illustrating right-lateralized hemispheric asymmetry for FFR; (2) anodal and cathodal tDCS applied over the right, but not left, auditory cortex significantly changed FFR magnitudes compared to the sham stimulation; specifically, such after-effects occurred only when participants listened to speech from the left ear, emphasizing the right auditory cortical contributions along the contralateral pathway. The current finding thus provides the first causal evidence that validates the relationship between the right auditory cortex and speech-evoked FFR and should significantly extend our understanding of speech encoding in the brain.

Significance Statement

Speech-evoked frequency-following response (FFR) is a neural activity that reflects the brain’s encoding of speech periodic features. The FFR has great fundamental and clinical importance for auditory processing. Whilst convention maintains that FFR derives mainly from the brainstem, it has been argued recently that there are additional contributions to FFR from the auditory cortex. Using a combination of tDCS, that altered neural excitability of auditory cortices, and EEG recording, the present study provided the first evidence to validate a causal relationship between the right auditory cortex and speech-evoked FFR. The finding supports the right-asymmetric auditory cortical contributions to processing of speech periodicity and...
advances our understanding of how speech signals are encoded and analysed along the central auditory pathways.

Introduction

Speech-evoked frequency-following response (FFR) is a phase-locked neural activity that reflects early processing of periodic features of input speech signals in the human brain (Picton and Aiken, 2008; Coffey et al., 2019).

The FFR is closely related to fundamental auditory processes. For instance, it plays an important role in pitch perception. FFR reflects the neural fidelity of linguistic pitch and is stronger in tonal language than non-tonal language speakers (Krishnan et al., 2004, 2005, 2009). It has greater strength in musicians who have better pitch discrimination ability than people without musical training (Musacchia et al., 2007; Wong et al., 2007; Strait et al., 2009; Bidelman et al., 2011). Furthermore, FFR is important for speech-in-noise perception. Greater FFR magnitudes are associated with better speech recognition ability in noisy environments (Song et al., 2011; Parbery-Clark et al., 2011). FFR also reflects neural plasticity related to fundamental cognitive and physiological processes such as auditory learning (Skoe et al., 2014), changes in arousal (Mai et al., 2019) and attention (Lehmann and Schönwiesner, 2014; Hartmann and Weisz, 2019).

Clinically, FFR is proposed as a biomarker for various auditory and language disorders. FFR declines with age (Anderson et al., 2012; Presacco et al., 2016) and can predict word recognition ability during speech-in-noise perception in older adults (Anderson et al., 2011; Fujihira and Shiraishi, 2015; Mai et al., 2018). This indicates that degradations to FFR could potentially explain the increased speech-in-noise difficulty experienced during aging. FFRs are also associated with hearing deficits such as cochlear synaptopathy (Encina-Llamas et al., 2019) and auditory processing disorders (Schochat et al., 2017). Furthermore, FFR is a potential marker for detecting functional impairments in learning and cognitive disorders in children, such as learning difficulties in literacy (Cunningham et al., 2001; Banai et al., 2007; White-Schwoch et al., 2015), dyslexia (Hornickel et al., 2013) and autism (Russo et al., 2008).

It is argued that the fundamental and clinical importance of FFR is linked to the neural fidelity of speech in the inferior colliculus at the brainstem, which has been proposed as the
main neural origin of FFR (Chandrasekaran and Kraus, 2010; Bidelman, 2015, 2018). Recent studies, however, have shown an additional source of FFR in the right auditory cortex associated with musical experience, pitch discrimination ability (Coffey et al., 2016), speech-in-noise perception (Coffey et al., 2017a) and intermodal attention (Hartmann and Weisz, 2019). FFR strength is associated with right-lateralized hemodynamic activity in the auditory cortex (Coffey et al., 2017b), consistent with the relative specialization of right auditory cortex for pitch and tonal processing (Zatorre and Berlin, 2001; Patterson et al., 2002; Hyde et al., 2008; Albouy et al., 2013; Cha et al., 2016).

Despite findings that show the potential cortical contribution to FFRs, it is unclear whether the relationship between auditory cortex and FFR is causal. The aim of the present study was to determine such relationship. Here, transcranial direct current stimulation (tDCS) was applied to alter neural excitability in the left and right auditory cortices. We examined the after-effects of tDCS on speech-evoked FFR using electroencephalography (EEG). tDCS is a non-invasive neuro-stimulation that modulates cortical excitability (Jacobson et al., 2012). By applying direct currents over the scalp, tDCS leads to neural excitation or inhibition in proximal parts of the cortex that last for up to 90 minutes post-stimulation (Nitsche and Paulus, 2001). Previous studies showed that applying tDCS over the right, compared to the left, auditory cortex can significantly change pitch discrimination performances, supporting the causal role of the right auditory cortex for pitch perception (Mathys et al., 2010; Matsushita et al., 2015). However, such causality has not been established for neurophysiological signatures like FFR. The present study tested the hypothesis that tDCS over the right auditory cortex should change the FFR strength during monaural listening to speech syllables. We further predicted that such after-effects should occur particularly along the contralateral auditory pathway where participants listen to speech from the left ear.

**Materials and Methods**

**Participants**

Ninety participants (18-40 years old; 45 females) were recruited and completed the entire experiment. Two other participants dropped out during the tDCS phase because they felt uncomfortable with the skin sensation when stimulation was applied. All participants had normal hearing (pure-tone audiometric thresholds <25 dB HL within frequency range of 0.25–
6 kHz for both ears) tested using a MAICO MA41 Audiometer (MAICO Diagnostics, Germany). Participants were non-tonal language speakers, had no long-term musical training and reported no history of neurological or speech/language disorders. They had not participated in any brain stimulation experiments in the two weeks prior to the present experiment.

All participants were right-handed (Handedness Index (HI) > 40; Oldfield, 1971). They were assigned at random to one of five groups, each of which received different types of tDCS (detailed in Experimental design). HI did not differ significantly between the five groups (all \( p > 0.4 \), uncorrected), indicating that the degree of handedness was well-matched across stimulation types. The absence of HI differences across groups is important because it has been argued that handedness influences functional hemispheric specialization (Carey et al., 2014; Willems et al., 2014). Hence matching the HI across groups ensured that any effects of tDCS were not confounded by handedness.

**Syllable stimulus for the FFR recording**

A 120-ms-long syllable /i/ spoken by a male with a static fundamental frequency (F\(_0\)) at 136 Hz was used for the FFR recordings. The waveform and spectrum of the syllable are shown as Figure 1. The syllable has three formants (F1, F2 and F3 at ~280, 2400 and 3100 Hz, respectively). It has a stable amplitude profile across the syllable period except for the 5-ms rising and falling cosine ramps applied at the onset and offset to avoid transients.

**Figure 1. The syllable stimulus for FFR recordings.** (A) Temporal waveform of the syllable /i/. (B) Spectrum of the syllable (0–4000 Hz) showing the formant locations (F1, F2 and F3). (C) The same spectrum as (B) that shows the first four harmonics with F\(_0\) at 136 Hz. N.B., the spectrum was obtained via Fast Fourier Transform (FFT) after zero-padding the temporal waveform to 1 second.
Experimental design

The experimental procedure is summarized in Figure 2. FFRs were recorded pre- and post-tDCS during monaural listening to the syllable stimulus to test for any after-effects of tDCS.

FFR recording

EEG were recorded over participants’ scalps using an ActiveTwo system (Biosemi ActiView, The Netherlands) with sampling rate of 16,384 Hz whilst they listened to the repeatedly-presented syllable /i/ (see Syllable stimulus for the FFR recording) both pre- and post-tDCS. The recording site was at the vertex (Cz) localized using a standard Biosemi cap, which is the conventional site used for obtaining FFRs (Skoe and Kraus, 2010). Bilateral earlobes served as the reference and ground electrodes were CMS and DRL at the parieto-occipital sites. Electrode impedance was kept below 35 mV. The syllable stimulus was presented at ~4 times per second with inter-stimulus interval (ISI) fixed at 120 ms. The stimulus was played monaurally via electrically-shielded inserted earphone (ER-3 insert earphone, Intelligent Hearing Systems, Miami, FL) at 85 dB SL (excluding ISIs) in each ear (e.g. left-ear listening followed by right-ear listening or vice versa with order of ear presentation counterbalanced across participants). Monaural listening ensured that after-effects of ipsilateral and contralateral tDCS (relative to the listening ears) could be tested separately (see Statistical analyses). For each ear, there were 1,500 sweeps for the positive and 1,500 sweeps for the negative polarity presented in an intermixed order (i.e., 3,000 sweeps in total).

Participants were seated comfortably in an armchair in an electromagnetically- and sound-shielded booth. They listened passively to the stimulus sequence whilst keeping their eyes on a fixation cross in the centre of a computer screen. The 3,000 syllable sweeps in each ear were broken into six 2-minute-long blocks (500 sweeps each) with ~40 second breaks between blocks. Participants were required to keep awake and refrain from body and head movements whilst they were listening to the sounds. The FFR recording lasted for ~30 minutes for both pre- and post-tDCS. The post-tDCS recording was completed within 45 minutes post-tDCS for all participants to ensure that any after-effects of tDCS on FFRs were sustained (Nitsche and Paulus, 2001).

tDCS
tDCS was applied over the scalp using a battery-driven direct current stimulator (Magstim HDCStim, UK) with a pair of rubber-surface electrodes (5×5 cm) contained in saline-soaked cotton pads. Participants were assigned at random to one of the five groups (18 participants (9 females) per group; single-blinded). The five groups received the following different types of tDCS: (1) anodal stimulation on the left auditory cortex (AC) (Left-Anod); (2) cathodal stimulation on the left AC (Left-Cathod); (3) anodal stimulation on the right AC (Right-Anod); (4) cathodal stimulation on the right AC (Right-Cathod); and (5) Sham, with electrode configurations randomly chosen from (1)–(4) for each participant (in this group, the active electrode was put on the left AC for half of the participants and on the right AC for the other half). Centre position of the active electrode was on T7/T8 (according to the 10/20 EEG system) for the left/right AC. The reference electrode was placed on the forehead above the eyebrow contralateral to the active electrode (see Matsushita et al., 2015; also see Figure 2).

For groups (1)–(4), tDCS was applied at 1 mA for 25 minutes with the currents ramping up/down for 15 seconds at the stimulation onset/offset. Sham applied tDCS only for 30 seconds in total (15 seconds ramping up and down respectively) at the onset of stimulation. This created the usual sensations associated with tDCS in Sham but without actual stimulation during the remainder of the run. All experimental sessions were conducted during the day time (mornings or early afternoons) and all participants had enough sleep (at least 6 hrs) the night before (based on self-report prior to the experiment) to ensure adequate cortical plasticity triggered by tDCS (Salehinejad et al., 2019).

During tDCS, participants completed a pitch discrimination task while they listened to sound stimuli over a loudspeaker 1 metre in front of them in the same sound-shielded booth used for the FFR recordings. Three short complex tones (400 ms) were presented on each trial at a calibrated level of 75 dB SL at the 1 metre position. The task was an ‘ABX’ task. In each trial, two tones ‘A’ and ‘B’ with different fundamental frequencies (F₀) were played consecutively followed by a third tone ‘X’ randomly selected from ‘A’ or ‘B’. Participants had to identify whether ‘X’ was the same as ‘A’ or ‘B’. They gave their best guess when they were unsure of the answer. The process followed a ‘2-down, 1-up’ adaptive procedure, in which the F₀ difference between ‘A’ and ‘B’ decreased by $\sqrt{2}$ times following two consecutive correct trials and increased by $\sqrt{2}$ times following an incorrect trial. No feedback about response accuracy was provided. Half-minute breaks were taken every 4 minutes. This task was included during tDCS because tDCS preferentially modulates neural networks that are currently active (Reato et al., 2010; Ranieri et al., 2012; Bikson and Rahman, 2013).
Concurrent tDCS and the pitch discrimination task could therefore maintain auditory cortical activity during neuro-stimulation, hence maximizing the effect of tDCS on neural excitability.

Figure 2. Illustrations for the experiment design. Participants first listened to a repeated syllable /i/ monaurally while FFR was recorded over scalp-EEGs at Cz. tDCS was then applied over the auditory cortex (AC) along with a pitch discrimination task. The same syllable listening task as in the first step was finally performed following tDCS to detect any after-effects of neuro-stimulation.

EEG Signal processing

All EEG signal processing was conducted via Matlab R2017a (The Mathworks).

Pre-processing

As mentioned, the FFR was captured from Cz. The EEG signals were first re-referenced to the bilateral earlobes and bandpass-filtered between 90 and 4000 Hz using a 2nd-order zero-phase Butterworth filter. The filtered signals were then segmented for each sweep (-50 to 150 ms relative to the syllable onset). Each segment was baseline-corrected by subtracting the average of the pre-stimulus (-50–0 ms) period. Segments that exceeded ±25 mV were rejected to minimize movement artefacts. The resultant rejection rates were < 2.5% averaged across participants for all cases (pre- and post-tDCS for the five stimulation groups for both left and right ear conditions).

FFR magnitudes
FFRs with the positive and negative polarities (FFRPos and FFRNeg) were first obtained by temporally averaging the pre-processed signals across sweeps with the respective polarities. FFRs for envelopes of F0 and its harmonics (i.e., periodicity; FFREnv) and temporal fine structures (TFS; FFRTFS) were obtained by adding and subtracting FFRPos and FFRNeg, respectively (Aiken and Picton, 2008). The addition and subtraction minimized the responses to TFS in FFREnv and to envelopes in FFRTFS, so that purer FFRs to envelopes and TFS were obtained separately (Aiken and Picton, 2008). Spectral magnitudes of FFREnv and FFRTFS were then calculated.

**Figure 3. A representative sample of FFR.** Sample waveforms (top panels) and the corresponding spectrograms (lower panels) of FFR Env (left) and FFR TFS (right) were obtained from a single participant in the left ear listening condition before tDCS was applied. The first two harmonics of F0 (F0 and 2F0) dominate the power of FFR Env as indicated in the FFR Env spectrogram (lower left). The three formants (F1, F2 and F3) in FFR TFS are shown and indicated in the FFR TFS spectrogram (lower right); F1 occurs at H2 for this vowel (the 2nd harmonic).

For FFREnv, FFREnv,F0 and FFREnv,2F0 (FFREnv at F0 and its 2nd harmonic, 2F0) that dominate the power of FFREnv (see Figure 3 left panel) were focused on. Whereas FFREnv,F0 and FFREnv,2F0 reflect neural phase-locking to the stimulus envelope periodicity in the central auditory systems, higher harmonics (≥ 3) of FFREnv may reflect distortion products resulting from non-linearities in response to acoustic stimuli on the basilar membrane (Smalt et al.,...
Whilst it is expected that FFR$_{ENV, F0}$ plays the major role in phase-locking to speech periodicity, FFR$_{ENV, 2F0}$ also makes contributions (e.g., Aiken and Picton, 2008) because of the non-sinusoidal characteristics of speech periodicity (Holmberg et al., 1988; also see discussions in Smalt et al., 2012). The procedure for measuring the magnitudes of FFR$_{ENV, F0}$ and FFR$_{ENV, 2F0}$ was as follows: a set of 120 ms (same length as the stimulus syllable) sliding windows (1 ms per step), each with a 5-ms rising/falling cosine ramp at the onset/offset, was applied to the FFR$_{ENV}$ waveform. As FFR$_{ENV}$ occurs at the auditory brainstem (Chandrasekaran and Kraus, 2010; Bidelman, 2015, 2018) and/or primary auditory cortex (Coffey et al., 2016), the neural transmission delays were set at 5–20 ms. Onsets of the windows were therefore set at 6–21 ms (allowing for an additional ~1 ms sound transmission through the plastic tube of the earphone to the cochlea) after the syllable onset. The windowed FFR$_{ENV}$ waveform in each step was then zero-padded to 1 second to allow for a frequency resolution of 1 Hz and the log-transformed FFT-powers (10*log$_{10}[$power$]$) centred at $F_0$ and $2F_0$ were measured (averaged across 136 ± 2 Hz and 272 ± 2 Hz, respectively). Finally, the FFR$_{ENV, F0}$ and FFR$_{ENV, 2F0}$ magnitudes were taken as the powers at the optimal neural delays (i.e., when powers are maximal across all steps for $F_0$ and $2F_0$, respectively).

For FFR$_{TFS}$, FFR$_{TFS, H2}$ and FFR$_{TFS, F2F3}$ (FFR$_{TFS}$ at the 2$^{nd}$ harmonic that represents $F_1$ for this vowel, and at $F_2$ and $F_3$, respectively; see Figure 4.3 right panel) were focused on. FFR$_{TFS, H2}$ reflects FFR to TFS at the resolved-harmonic region while FFR$_{TFS, F2F3}$ reflects FFR to TFS at the unresolved-harmonic region. The same procedure used when measuring magnitudes of FFR$_{ENV, F0}$ and FFR$_{ENV, 2F0}$ was followed, except that: (1) the procedure was applied on FFR$_{TFS}$ at H2 (for FFR$_{TFS, H2}$) and at H16–H27 (the 16$^{th}$ to 27$^{th}$ harmonics corresponding to the range of $F_2$ and $F_3$ for FFR$_{TFS, F2F3}$; the final magnitude was taken as the mean magnitude across all harmonics in this range); (2) the neural delays during analyses were set at 1–6 ms (0–5 ms delays allowing an additional 1 ms sound transmission through the plastic tube of the earphone) as FFR$_{TFS}$ arises at earlier stages of auditory processing in the periphery (Aiken and Picton, 2008).

Because of the different neural origins of FFR$_{ENV}$ (brainstem/auditory cortex) and FFR$_{TFS}$ (periphery), the present study thus allows us to confirm whether tDCS applied to auditory cortex affects FFR that arise at different levels of the auditory systems.

**Statistical analyses**
Before testing the after-effects of tDCS, analyses were first conducted to check whether baseline (pre-tDCS) characteristics were matched across stimulation. ANOVAs were conducted using the baseline magnitudes and optimal neural delays of FFRs as dependent variables, Stimulation (Left-Anod, Left-Cathod, Right-Anod, Right-Cathod and Sham) and Ear (left vs. right) as independent variables. Post-hoc analyses were conducted following significant interactions or main effects.

After-effects of tDCS (differences in FFR magnitudes between post- and pre-tDCS) were tested using linear mixed-effect regressions. These were conducted using after-effects as dependent variables, Stimulation and Ear as fixed-effect factors and Participant as the random-effect factor. Post-hoc analyses were conducted following significant interactions or main effects.

Furthermore, regardless of whether interaction effects occurred between Stimulation and Ear, planned comparisons for the after-effects were conducted between different stimulation types in the left and right ear conditions, respectively. This was because collapsing the left and right ears would smear the distinctions between any after-effects along the contralateral pathway (ears with tDCS on the opposite side) and those along the ipsilateral pathway (ears with tDCS on the same side), which was one of the aspects addressed in the present study. As multiple comparisons were conducted for each ear (5 stimulation types leading to 10 comparisons), the critical $\alpha$ value for detecting significance was adjusted at 0.005. It was predicted that, compared to Sham, significantly greater after-effects of tDCS over the right auditory cortex (Right-Anod and Right-Cathod), but not the left auditory cortex (Left-Anod or Left-Cathod), should be found, consistent with the current hypothesis that the right auditory cortex makes specific contributions to FFR.

FFR magnitudes were magnitudes of $\text{FFR}_{\text{ENV}}$ ($\text{FFR}_{\text{ENV,F0}}$ and $\text{FFR}_{\text{ENV,2F0}}$) and $\text{FFR}_{\text{TFS}}$ ($\text{FFR}_{\text{TFS,H2}}$ and $\text{FFR}_{\text{TFS,F2F3}}$) (see EEG signal processing). For $\text{FFR}_{\text{ENV}}$, the present study combined the magnitudes of $\text{FFR}_{\text{ENV,F0}}$ and $\text{FFR}_{\text{ENV,2F0}}$, rather than use them as separate dependent variables. The reason was that, it was observed that the summed $\text{FFR}_{\text{ENV,F0}}$ and $\text{FFR}_{\text{ENV,2F0}}$ magnitude yielded greater effect sizes during planned comparisons where statistical significance ($p < 0.05$, uncorrected) was detected using $\text{FFR}_{\text{ENV,F0}}$ or $\text{FFR}_{\text{ENV,2F0}}$ magnitude alone: Cohen’s $d = 0.752$ and 1.001 for $\text{FFR}_{\text{ENV,F0}}$ and for the summed $\text{FFR}_{\text{ENV,F0}}$ and $\text{FFR}_{\text{ENV,2F0}}$ magnitude, respectively, when Right-Anod was compared with Sham in the left ear listening condition; Cohen’s $d = 0.934$ and 1.140 for $\text{FFR}_{\text{ENV,F0}}$ and for combined $\text{FFR}_{\text{ENV,F0}}$ and $\text{FFR}_{\text{ENV,2F0}}$. 
and FFR_{ENV, 2F0} magnitude, respectively, when Right-Cathod was compared with Sham in the left ear listening condition (see Results for further details).

### Results

#### Baseline characteristics

Table 1 and 2 shows the baseline magnitudes and neural delays for FFR_{ENV}, FFR_{TFS,H2} and FFR_{TFS,F2F3} in both the left and right ear conditions. ANOVAs were conducted for baseline magnitudes and optimal neural delays of FFR_{ENV}, FFR_{TFS,H2} and FFR_{TFS,F2F3}.

For FFR_{ENV}, a significant main effect of Ear was found for the magnitude (F(1, 85) = 12.318, p < 0.001; greater magnitude in the left than in the right ear condition) but not for the neural delay (F(1, 85) = 0.055, p = 0.815); no main effects of Stimulation (magnitude: F(4, 85) = 0.932, p = 0.450; neural delay: F(4, 85) = 0.799, p = 0.529) or [Stimulation × Ear] interactions were found (magnitude: F(4, 85) = 0.541, p = 0.706; neural delay: F(4, 85) = 0.046, p = 0.996). Furthermore, no significant differences were found between any stimulation type in either ear condition (magnitude: all p > 0.07; neural delay: all p > 0.1). Figure 4 illustrates the comparison of baseline magnitudes for FFR_{ENV} between the left and right ear conditions after collapsing across stimulation types (due to the significant main effect of Ear but no main effect of Stimulation).

For FFR_{TFS,H2}, there were no significant main effects of Stimulation (magnitude: F(4, 85) = 0.692, p = 0.600; neural delay: F(4, 85) = 1.421, p = 0.234) or Ear (magnitude: F(1, 85) = 3.483, p = 0.065; neural delay: F(1, 85) = 1.842, p = 0.178), or [Stimulation × Ear] interactions (magnitude: F(4, 85) = 0.744, p = 0.565; neural delay: F(4, 85) = 0.587, p = 0.673). No significant differences were found between any stimulation type in either ear condition (magnitude: all p > 0.1; neural delay: all p > 0.05).

For FFR_{TFS,F2F3}, significant main effects of Stimulation (F(4, 85) = 40.872, p < 0.001) and Ear (F(1, 85) = 4.225, p = 0.002; greater in the right than the left ear condition) were found for the magnitude, but not for the neural delay (Stimulation: F(4, 85) = 1.504; p = 0.208; Ear: F(1, 85) = 0.324, p = 0.571). A significant [Stimulation × Ear] interaction was found for the neural delay (F(4, 85) = 2.549, p = 0.045), but not for the magnitude (F(4, 85) = 1.763, p =
Post-hoc analyses found significant differences in magnitudes between several stimulation types (collapsing the left and right ears: Left-Anod vs. Right-Anod, \( t(34) = -2.110, p = 0.042 \); Left-Anod vs. Sham, \( t(34) = -2.713, p = 0.010 \); Left-Cathod vs. Right-Anod, \( t(34) = -2.796, p = 0.008 \); Left-Cathod vs. Right-Cathod, \( t(34) = -2.566, p = 0.015 \); Left-Cathod vs. Sham, \( t(34) = -3.498, p = 0.001 \)). Significant differences were found between stimulation types for the neural delay in both the left ear (Left-Anod vs. Right-Cathod, \( t(34) = -2.703, p = 0.011 \)) and the right ear condition (Right-Anod vs. Right-Cathod, \( t(34) = 2.279, p = 0.029 \); Left-Anod vs. Right-Anod, \( t(34) = -2.240, p = 0.032 \); Right-Anod vs. Sham, \( t(34) = 2.629, p = 0.013 \)). All \( p \)-values here are reported without correction.

The results thus indicate that the baseline characteristics of \( \text{FFR}_{\text{ENV}} \) and \( \text{FFR}_{\text{TFS,H2}} \), but not \( \text{FFR}_{\text{TFS,F2F3}} \), were well matched across stimulation types. As such, although after-effects were tested for all three FFR signatures, \( \text{FFR}_{\text{ENV}} \) and \( \text{FFR}_{\text{TFS,H2}} \) are focused on. In addition, the main effects of Ear for \( \text{FFR}_{\text{ENV}} \) and \( \text{FFR}_{\text{TFS,F2F3}} \) magnitudes may reflect the laterality of speech encoding at the subcortical (Chandrasekaran and Kraus, 2010; Bidelman, 2015, 2018) and/or cortical levels (Coffey et al., 2016, 2017b), which will be discussed further (see Discussion).

**Table 1.** Baseline magnitudes (standard deviations shown in the brackets; in \( \text{dB} \)) for \( \text{FFR}_{\text{ENV}} \), \( \text{FFR}_{\text{TFS,H2}} \) and \( \text{FFR}_{\text{TFS,F2F3}} \) across stimulation types in the left and right ear conditions.

<table>
<thead>
<tr>
<th>FFRs</th>
<th>Ear</th>
<th>Left-Anod</th>
<th>Left-Cathod</th>
<th>Right-Anod</th>
<th>Right-Cathod</th>
<th>Sham</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{FFR}_{\text{ENV}} )</td>
<td>Left</td>
<td>76.28 (5.27)</td>
<td>78.84 (7.03)</td>
<td>76.05 (6.96)</td>
<td>76.24 (8.56)</td>
<td>73.45 (10.05)</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>75.05 (5.79)</td>
<td>75.42 (4.72)</td>
<td>72.82 (6.62)</td>
<td>74.24 (7.91)</td>
<td>72.17 (10.92)</td>
</tr>
<tr>
<td>( \text{FFR}_{\text{TFS,H2}} )</td>
<td>Left</td>
<td>30.35 (5.70)</td>
<td>30.70 (7.71)</td>
<td>32.52 (7.12)</td>
<td>31.68 (6.24)</td>
<td>33.37 (6.86)</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>32.71 (3.88)</td>
<td>30.63 (6.98)</td>
<td>32.59 (7.97)</td>
<td>33.40 (7.51)</td>
<td>34.36 (5.66)</td>
</tr>
<tr>
<td>( \text{FFR}_{\text{TFS,F2F3}} )</td>
<td>Left</td>
<td>15.31 (7.26)</td>
<td>13.21 (7.24)</td>
<td>19.45 (7.16)</td>
<td>20.14 (6.58)</td>
<td>20.46 (5.75)</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>17.13 (7.07)</td>
<td>16.28 (6.58)</td>
<td>22.97 (7.57)</td>
<td>20.90 (7.82)</td>
<td>23.58 (6.09)</td>
</tr>
</tbody>
</table>

**Table 2.** Baseline neural delays (standard deviations shown in the brackets; in \( \text{ms} \)) for \( \text{FFR}_{\text{ENV}} \), \( \text{FFR}_{\text{TFS,H2}} \) and \( \text{FFR}_{\text{TFS,F2F3}} \) across stimulation types in the left and right ear conditions.
<table>
<thead>
<tr>
<th>FFRs</th>
<th>Ear</th>
<th>Left-Anod</th>
<th>Left-Cathod</th>
<th>Right-Anod</th>
<th>Right-Cathod</th>
<th>Sham</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFR_{ENV}</td>
<td>Left</td>
<td>8.75 (2.45)</td>
<td>9.42 (2.44)</td>
<td>9.67 (2.70)</td>
<td>8.56 (2.81)</td>
<td>8.81 (2.71)</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>8.78 (2.02)</td>
<td>9.47 (3.49)</td>
<td>9.50 (3.25)</td>
<td>8.58 (1.69)</td>
<td>9.08 (2.33)</td>
</tr>
<tr>
<td>FFR_{TFS,H2}</td>
<td>Left</td>
<td>3.50 (2.28)</td>
<td>4.50 (1.82)</td>
<td>3.50 (1.82)</td>
<td>3.67 (2.06)</td>
<td>4.28 (1.60)</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>3.61 (2.30)</td>
<td>4.94 (1.59)</td>
<td>4.44 (1.95)</td>
<td>4.11 (2.00)</td>
<td>4.06 (1.92)</td>
</tr>
<tr>
<td>FFR_{TFS,F2F3}</td>
<td>Left</td>
<td>2.90 (0.36)</td>
<td>3.04 (0.27)</td>
<td>3.03 (0.44)</td>
<td>3.20 (0.31)</td>
<td>3.05 (0.53)</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>2.93 (0.48)</td>
<td>3.03 (0.48)</td>
<td>3.28 (0.47)</td>
<td>2.97 (0.34)</td>
<td>2.87 (0.48)</td>
</tr>
</tbody>
</table>

Figure 4. Comparison of baseline magnitude for FFR_{ENV} between the left and the right ear conditions. The comparison was conducted by collapsing the stimulation types following the ANOVA results which showed a significant main effect of Ear, but no significant main effect of Stimulation or [Stimulation × Ear] interaction for the baseline FFR_{ENV} magnitude. The left and the right ear conditions are indicated as blue and orange, respectively. (A) Waveforms of FFR_{ENV} averaged across stimulation types. (B)(C) FFT-power spectra averaged across stimulation types, obtained using the individual optimal neural delays for (B) FFR_{ENV,F0} (showing 110–160 Hz peaking at F₀ of 136 Hz) and (C) FFR_{ENV,2F0} (showing 250–300 Hz peaking at 2F₀ 272 Hz) (shaded areas in the spectra cover the ranges of ±1 standard errors (SEs)). (D) FFR_{ENV} magnitude (summed magnitude of FFR_{ENV,F0} and FFR_{ENV,2F0}). Significant greater FFR_{ENV} magnitude was found in the left than in the right ear condition (***, **p < 0.001, uncorrected). Error bars indicate the SEs.

After-effects on FFR_{ENV}

FFR_{ENV} magnitude refers to the summed FFR_{ENV,F0} and FFR_{ENV,2F0} magnitudes (see Statistical analyses). Figure 5 shows the waveforms and FFT-power spectra for FFR_{ENV} across...
participants. Linear mixed-effect regression showed a significant main effect of Stimulation ($F(4, 85) = 2.549, p = 0.045$). No main effect of Ear ($F(1, 85) = 0.784, p = 0.378$) or [Stimulation × Ear] interaction ($F(4, 85) = 1.309, p = 0.273$) was found. Post-hoc independent-sample t-tests were thus conducted between different stimulation types following the main effect of Stimulation (collapsing the left and right ear due to the lack of [Stimulation × Ear] interaction). After-effects of tDCS over the right AC were significantly lower than that of Sham (Right-Anod vs. Sham, $t(34) = -2.569, p = 0.015$ (uncorrected), Cohen’s $d = 0.856$; Right-Cathod vs. Sham, $t(34) = -2.219, p = 0.033$ (uncorrected), Cohen’s $d = 0.740$) (Figure 6).

**Figure 5.** Waveforms and power spectra for $\text{FFR}_\text{ENV}$ averaged across participants. (A) and (B) show the waveforms and FFT-power spectra in the left and right ear condition, respectively. Pre- and post-tDCS were indicated as black and red, respectively (shaded areas in the spectra cover the ranges of ±1 SEs from the means). From left to right are different...
stimulation types (Left-Anod, Left-Cathod, Right-Anod, Right-Cathod and Sham). Upper panels: waveforms of FFR\textsubscript{ENV}; mid and lower panels: power spectra obtained using the individual optimal neural delays for FFR\textsubscript{ENV,F0} (mid; showing 110–160 Hz peaking at F\textsubscript{0} of 136 Hz) and FFR\textsubscript{ENV,2F0} (lower; showing 250–300 Hz peaking at 2F\textsubscript{0} of 272 Hz).

**Figure 6.** After-effects of tDCS on FFR\textsubscript{ENV} magnitudes comparing across stimulation types after collapsing the left and right ears. Collapsing the left and right ears was conducted following the ANOVA results which showed a significant main effect of Stimulation but no significant main effect of Ear or [Stimulation × Ear] interaction. Red circles indicate individual data for the corresponding stimulation types (Left-Anod, Left-Cathod, Right-Anod, Right-Cathod and Sham). Post-hoc paired comparisons showed significant differences between tDCS over the right AC (Right-Anod and Right-Cathod) and Sham (*p < 0.05, uncorrected). Error bars indicate the SEs.

Planned comparisons between different stimulation types were subsequently conducted for the left and right ear listening conditions to determine whether tDCS has effects along the contralateral or ipsilateral pathway. The critical α value for detecting significance was adjusted to 0.005 (there were 10 pairs of comparisons in each ear condition). The results are illustrated in **Figure 7** (upper panels). In the left ear condition, significant differences were found between tDCS over the right AC and Sham (Right-Anod vs. Sham, t(34) = -3.024, p < 0.005, Cohen’s \(d\)
= 1.001; Right-Cathod vs. Sham, t(34) = -3.420, p = 0.002, Cohen’s d = 1.140). No significant effects were found for any other comparison (all p > 0.2). In the right ear condition, no significant effects were found for any pair of comparison (all p > 0.2). All p-values shown here are reported without correction.

**Figure 7. After-effects of tDCS on FFR magnitudes.** Upper, mid and lower panels indicate the after-effects on magnitudes of FFR$_{ENV}$, FFR$_{TFS_H2}$ and FFR$_{TFS_F2F3}$, respectively. Planned comparisons were conducted between different stimulation types in both the left and right ear conditions, with the critical $\alpha$ value set at 0.005 according to multiple comparisons. Significant differences were found between tDCS over the right auditory cortex (Right-Anod and Right-Cathod) and Sham in the left ear condition for FFR$_{ENV}$. (**p < 0.005, uncorrected; i.e., p <
0.05 after correction according to multiple comparisons) Red circles indicate individual data for the corresponding stimulation types. Error bars indicate the SEs.

**After-effects on FFR\textsubscript{TFS}**

Equivalent analyses to those conducted for the FFR\textsubscript{ENV} magnitude were conducted for the magnitudes of FFR\textsubscript{TFS\_H2} and FFR\textsubscript{TFS\_F2F3}. The linear mixed-effect regressions did not show significant main effects of Stimulation (FFR\textsubscript{TFS\_H2}: F(4, 85) = 0.528, \( p = 0.715 \); FFR\textsubscript{TFS\_F2F3}: F(4, 85) = 0.613, \( p = 0.655 \)) or Ear (FFR\textsubscript{TFS\_H2}: F(1, 85) = 0.496, \( p = 0.467 \); FFR\textsubscript{TFS\_F2F3}: F(1, 85) = 0.213, \( p = 0.646 \)), or significant [Stimulation \times Ear] interactions (FFR\textsubscript{TFS\_H2}: F(4, 85) = 0.530, \( p = 0.714 \); FFR\textsubscript{TFS\_F2F3}: F(4, 85) = 1.189, \( p = 0.322 \)).

Planned comparisons did not find significant after-effects between different stimulation types in the left or right ear condition (FFR\textsubscript{TFS\_H2}: all \( p > 0.6 \) in the left ear condition and all \( p > 0.1 \) in the right ear condition; FFR\textsubscript{TFS\_F2F3}: all \( p > 0.09 \) in the left ear condition and all \( p > 0.1 \) in the right ear condition; see Figure 7, mid and lower panels). All \( p \)-values are reported without correction.

**Discussion**

The current study used a combined tDCS and EEG approach to test for a causal contribution of auditory cortex to speech-evoked FFR in healthy right-handed participants. The left and right auditory cortices were neuro-stimulated in different groups of participants and the after-effects of tDCS on the FFR were examined during monaural listening to a repeated speech syllable. The results showed that tDCS, both anodal and cathodal, over the right auditory cortex, generated significantly greater after-effects on FFR\textsubscript{ENV} magnitude compared to sham stimulation. Specifically, such effects were present only in the left ear listening condition, indicating that the changes in processing of speech periodicity information occur along the contralateral pathway (i.e., from the left ear to the right auditory cortex). The results thus agree with previous studies that have shown a close relation between the right auditory cortex and FFR (Coffey et al., 2016, 2017a, 2017b; Hartmann and Weisz, 2019) and provide the first evidence for a causal relationship.
Laterality for FFR\textsubscript{ENV} at the baseline

Ear laterality for the baseline FFR\textsubscript{ENV} and FFR\textsubscript{TFS,F2F3} magnitudes were found (see \textit{Baseline characteristics}). The discussion here focuses on FFR\textsubscript{ENV} alone due to the significant main effect of Stimulation for the baseline FFR\textsubscript{TFS,F2F3} magnitude (which means that the baseline was not well matched across stimulation types) and lack of a significant after-effect of tDCS on FFR\textsubscript{TFS,F2F3} magnitude.

The present study found that baseline FFR\textsubscript{ENV} had significantly greater magnitude in the left than in the right ear condition, supporting the lateralization of speech periodicity encoding along the contralateral auditory pathway from the left ear to the right auditory cortex. This echoes the previous findings that showed the right-hemispheric lateralization of the classic 40 Hz auditory steady-state response (ASSR) (Ross et al., 2005; Luke et al., 2017). Right lateralization was also found in ASSR at 80 Hz (Vanvooren et al., 2014). ASSRs are phase-locked responses to amplitude-modulated tones/noise and both ASSRs and FFR\textsubscript{ENV} are envelope-following responses (Dimitrijevic et al., 2004). Whilst the 40 Hz ASSR has its main generator at the cortical level (Herdman et al., 2002; Ross et al., 2002, 2005), prominent activities occur at the brainstem level for the 80 Hz ASSR (Herdman et al., 2002) and speech-evoked FFR\textsubscript{ENV} (Chandrasekaran and Kraus, 2010; Bidelman, 2015, 2018; this can also be seen in the present study where average optimal neural delays were between 5 and 10 ms, see Table 2). Recent studies, however, have shown that FFR has additional sources in the auditory cortex (Coffey et al., 2016, 2017a; Hartmann and Weisz, 2019). It is thus not clear whether the observed laterality of FFR\textsubscript{ENV} in the present study occurs at the subcortical or cortical level or, more equivocally, whether auditory cortex contributes to this laterality. As such, the current combined tDCS and EEG approach showed how altering neural excitability of auditory cortex in the left or right hemisphere can lead to changes in FFR which therefore provides confirmatory evidence for a causal cortical contribution.

Causal role of the right auditory cortex for FFR\textsubscript{ENV}

After-effects found for FFR\textsubscript{ENV} but not FFR\textsubscript{TFS} indicate that tDCS had impacts on the responses at the subcortical and/or cortical levels above the auditory periphery. The findings thus argue for a causal role of the right auditory cortex in processing speech periodicity information along the contralateral pathway in the central auditory systems.
The present study thus advances our understanding of the relationship between FFR and pitch processing in the right auditory cortex. Previous studies have shown that FFR is closely related to pitch perception. FFR strength can be enhanced by both short-term perceptual training of pitch discrimination (Carcagno and Plack, 2011) as well as long-term musical experience (Musacchia et al., 2007; Wong et al., 2007; Strait et al., 2009; Bidelman et al., 2011). Furthermore, FFR has been used as an index of neural fidelity of linguistic pitch and the fidelity is greater in tonal language than in non-tonal language speakers (Krishnan et al., 2004, 2005, 2009). Despite this, however, rather than reflecting the result of pitch extraction, FFR has been suggested to reflect subcortical responses to monaural temporal information (e.g., periodicity cues) that are important for extracting pitch of complex sounds (i.e., ‘pitch-bearing’ information; Gockel et al., 2011). On the other hand, the process of pitch extraction itself takes place in the auditory cortex (Penagos et al., 2004; Bendor and Wang, 2005; Puschmann et al., 2010) with a right hemispheric specialization (Zatorre and Berlin, 2001; Patterson et al., 2002; Hyde et al., 2008; Mathys et al., 2010; Albouy et al., 2013). In this respect, the current after-effects of tDCS may reflect a top-down corticofugal modulation process in which the right auditory cortex affects the processing of pitch-bearing information that occurs at the subcortical level. Alternatively, although EEG mainly captures FFR signals originating from the brainstem (Bidelman, 2015, 2018), cortical sources have been found dominated in the right hemisphere (Coffey et al., 2016; 2017a). It therefore cannot be excluded that tDCS may affect the FFR magnitude directly at the cortical level. It is noteworthy that the current finding could not disentangle whether the effects emerge at the subcortical or cortical level, or both.

Also, stronger evidence would be provided for the specific contributions of the right auditory cortex to FFR if significant differences in after-effects were further found between tDCS over the right and the left auditory cortex. However, the present results did not show such differences. A possible explanation is that tDCS not only alters excitability of regions in which electrodes are located but can yield widespread changes across the brain (see a review: Filmer et al., 2014). This could be due to the diffuse nature of the tDCS where currents do not only flow between electrodes, but also spread widely through various other regions (Faria et al., 2011; Bai et al., 2014; Unal and Bikson, 2018). tDCS also changes functional connectivity (Sehm et al., 2012; Kunze et al., 2016) by which interactions of auditory cortices between the two hemispheres may be further activated. Therefore, tDCS over the left auditory cortex could also cause some changes in the right side that yield similar (but smaller) after-effects as direct stimulation over the right auditory cortex.
Neurophysiological consequences of tDCS

An intriguing finding of the present study is that anodal and cathodal tDCS over the right auditory cortex resulted in the same direction of changes, both causing decreases in FFR<sub>ENV</sub> magnitude compared to sham. Conventionally, anodal and cathodal stimulations reflect depolarization and hyperpolarization of neurons, respectively, which should lead to opposite directions of after-effects (Jacobson et al., 2012). However, it is not unusual that tDCS has polarity-independent effects due to the underlying complexity of its neurophysiological consequences. For example, several studies have shown that anodal and cathodal tDCS have the same effects on excitability of motor cortex (Antal et al., 2007), motor learning (de Xivry et al., 2011), cerebellar functions for working memory (Ferrucci et al., 2008) and visuomotor learning (Shah et al., 2013). The first possible explanation would be the non-linear effects of tDCS depending on the current density. It has been shown that cathodal tDCS with an electrode size of 35 cm<sup>2</sup> can lead to inhibition in the motor cortex at 1 mA but excitation at 2 mA (Batsikadze et al., 2013). The present study used a current intensity at 1 mA but with smaller electrode size (25 cm<sup>2</sup>; hence greater current density). It could be that this current density through the auditory cortex would lead to non-linear effects as resulted in the motor cortex. Second, it is possible that similar changes in concentrations of relevant neurotransmitters are caused by anodal and cathodal tDCS. It was found that with 1 mA currents, anodal tDCS causes decreases in GABA concentration that lead to cortical excitation; cathodal tDCS also causes decreases in GABA, but with greater concurrent decreases in glutamate that lead to cortical inhibition (Stagg et al., 2009). It is possible that GABA concentrations, which decrease following both anodal and cathodal tDCS, play an important role for changes in FFR<sub>ENV</sub> magnitude.

Conclusion

The current results validate the previous findings that the right auditory cortex makes significant contributions to speech-evoked FFR (Coffey et al., 2016, 2017a, 2017b; Hartmann and Weisz, 2019) by establishing a causal relationship between the two. To our knowledge, this is the first evidence for this causality and it could be essential due to the fundamental and clinical importance of the FFR. Thus, these findings should advance our understanding of how speech periodicity and pitch information are processed along the central auditory pathways in the human brain. Future research is needed to further clarify where exactly this causality emerges, i.e., to disentangle whether the effects are realized through top-down corticofugal
modulations on the subcortical level, or modulations directly in the cortex. Moreover, it will be worthwhile to further investigate how changes in concentrations of neurotransmitters by neuro-stimulation relate to this causality, which can help us better understand the underlying mechanisms of the cortical contributions to FFR.

References


