Oscillatory Entrainment of the Frequency-following Response in Auditory Cortical and Subcortical Structures

Emily B. J. Coffey,^{1,2,3,4} Isabelle Arseneau-Bruneau,^{2,3,4,5} Xiaochen Zhang,^{2,6} Sylvain Baillet,^{2,4} and Robert J. Zatorre^{2,3,4,5}

¹Department of Psychology, Concordia University, Montreal, Quebec H4B 1R6, Canada, ²Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada, ³Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, Quebec H3C 3J7, Canada, ⁴Centre for Research on Brain, Language and Music (CRBLM), Montreal, Quebec H3G 2A8, Canada, ⁵Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), McGill University, Montreal, Quebec H3A 1E3, Canada, and ⁶Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, People's Republic of China

There is much debate about the existence and function of neural oscillatory mechanisms in the auditory system. The frequency-following response (FFR) is an index of neural periodicity encoding that can provide a vehicle to study entrainment in frequency ranges relevant to speech and music processing. Criteria for entrainment include the presence of poststimulus oscillations and phase alignment between stimulus and endogenous activity. To test the hypothesis of entrainment, in experiment 1 we collected FFR data for a repeated syllable using magnetoencephalography (MEG) and electroencephalography in 20 male and female human adults. We observed significant oscillatory activity after stimulus offset in auditory cortex and subcortical auditory nuclei, consistent with entrainment. In these structures, the FFR fundamental frequency converged from a lower value over 100 ms to the stimulus frequency, consistent with phase alignment, and diverged to a lower value after offset, consistent with relaxation to a preferred frequency. In experiment 2, we tested how transitions between stimulus frequencies affected the MEG FFR to a train of tone pairs in 30 people. We found that the FFR was affected by the frequency of the preceding tone for up to 40 ms at subcortical levels, and even longer durations at cortical levels. Our results suggest that oscillatory entrainment may be an integral part of periodic sound representation throughout the auditory neuraxis. The functional role of this mechanism is unknown, but it could serve as a fine-scale temporal predictor for frequency information, enhancing stability and reducing susceptibility to degradation that could be useful in real-life noisy environments.

Key words: auditory cortex; brainstem; entrainment; frequency-following response; magnetoencephalography; oscillations

Significance Statement

Neural oscillations are proposed to be a ubiquitous aspect of neural function, but their contribution to auditory encoding is not clear, particularly at higher frequencies associated with pitch encoding. In a magnetoencephalography experiment, we found converging evidence that the frequency-following response has an oscillatory component according to established criteria: poststimulus resonance, progressive entrainment of the neural frequency to the stimulus frequency, and relaxation toward the original state on stimulus offset. In a second experiment, we found that the frequency and amplitude of the frequency-following response to tones are affected by preceding stimuli. These findings support the contribution of intrinsic oscillations to the encoding of sound, and raise new questions about their functional roles, possibly including stabilization and low-level predictive coding.

Introduction

From brainstem to cortex, the auditory system responds in a synchronized manner to acoustical energy containing regular

repeating elements. In low-frequency ranges (<10 Hz), it may function to parse the temporal structure of speech and music (Nozaradan et al., 2011; Arnal and Giraud, 2012; Giraud and

Received Aug. 20, 2020; revised Feb. 23, 2021; accepted Feb. 24, 2021.

Author contributions: E.B.J.C. and R.J.Z. designed research; E.B.J.C., I.A.-B., and X.Z. performed research; S.B. contributed unpublished reagents/analytic tools; E.B.J.C. and X.Z. analyzed data; E.B.J.C. and R.J.Z. wrote the paper.

This work was supported by a Foundation Grant from the Canadian Institutes for Health Research and a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to R.J.Z. E.B.J.C. was supported by a Banting Postdoctoral Fellowship. R.J.Z. is a fellow of the Canadian Institute for Advanced Research. We acknowledge the financial support of Health Canada, through the Canada Brain Research Fund in partnership with the Montreal Neurological Institute, and of the Healthy Brains for Healthy Lives initiative

of McGill University. We would like to thank Annie Qin for assisting with piloting; the team at the McConnell Brain Imaging Center (Montreal Neurological Institute) for supporting and making accommodations for the study, particularly Elizabeth Bock; and the Brainstorm team.

The authors declare no competing financial interests.

Correspondence should be addressed to Emily B. J. Coffey at emily.coffey@concordia.ca.

<https://doi.org/10.1523/JNEUROSCI.2313-20.2021>

Copyright © 2021 the authors

Poeppel, 2012), to integrate information (Schroeder et al., 2008), or to enable selective attention (Schroeder and Lakatos, 2009; Zion Golumbic et al., 2013; Lakatos et al., 2013). In higher frequency ranges associated with pitch information (80–400 Hz), frequency-following responses (FFRs) are observed (Skoe and Kraus, 2010; Kraus et al., 2017; Coffey et al., 2019; Krizman and Kraus, 2019).

Although rhythmic oscillations appear to be ubiquitous in the auditory system (Gourévitch et al., 2020; Neymotin et al., 2020), it is unclear whether such responses represent merely the summation of evoked responses to periodic sound elements (Bidelman, 2015) or arise from truly oscillatory intrinsic properties of the auditory system (Obleser and Kayser, 2019). The distinction is important to understanding their function (Haegens and Zion Golumbic, 2018; Gourévitch et al., 2020) and fits into a bigger picture concerning the roles of intrinsic oscillations throughout the brain, in which they appear to play a role coordinating neuronal and network activity to support complex cognitive processes (Buzsáki and Draguhn, 2004; Thut et al., 2012).

The main aim of the present study is to clarify whether pitch encoding via the FFR involves a “true” entrainment mechanism, in the sense of involving an intrinsic oscillatory mechanism, as appears to be the case in lower frequency ranges (Lakatos et al., 2013) as well as in other sensory modalities (e.g., alpha band entrainment in vision; Mathewson et al., 2012; Spaak et al., 2014). The FFR provides an excellent vehicle to address this question because its frequency content closely resembles that of the evoking stimulus, its strength correlates with behavioral and perceptual variables, and it is used widely in noninvasive human studies to quantify the quality of neural sound encoding (Kraus et al., 2017). Yet questions remain about its composition and origin (Coffey et al., 2016; Tichko and Skoe, 2017; Bidelman, 2018). According to the delay-based model, the FFR is generated via the summation of individual components from successive nuclei along the auditory neuraxis with increasing latencies (Gardi et al., 1979; Tichko and Skoe, 2017). In contrast, the oscillatory model suggests that the FFR is generated by neuronal circuitry with inherently oscillatory properties even at low levels (e.g., cochlear nucleus, inferior colliculus; Lerud et al., 2019). The delay-based and oscillatory models make different predictions about auditory signal processing and perception, although they are not mutually exclusive.

In line with criteria for defining neural entrainment (Haegens and Zion Golumbic, 2018; Obleser and Kayser, 2019), we reasoned that if the representation of the auditory system of fine frequency information is oscillatory in nature, we should observe persistent activity at the stimulus frequency following stimulus offset (Notbohm et al., 2016). Furthermore, frequency representation should converge toward the stimulation frequency over time as information accrues through successive inputs to an oscillator (Thut et al., 2012; Giraud and Arnal, 2018).

Some fragmentary evidence in the literature supports an FFR aftereffect following sound offset in monkey and human cortex (Steinschneider et al., 1980; Liégeois-Chauvel et al., 2004), and even as early as the auditory nerve in animal models (Irvine, 1986), although not all authors have reported it (Wickesberg and Stevens, 1998; Gao et al., 2016). Persistent poststimulus activity has recently been documented in marmoset auditory cortex (Cooke et al., 2020), but the relationship of this phenomenon to entrainment remains to be understood. Xu and Ye (2015) systematically recorded human FFRs with EEG to stimuli with six different durations (5–18 cycles of the fundamental frequency). Although the authors did not focus on the presence of an

Figure 1. Analysis of the stimulus. *a*, Amplitude spectra for the 120 ms speech syllable /da/, calculated over a 300 ms window encompassing the entire speech stimulus. A sharp spectral peak is observed precisely at 98 Hz fundamental frequency (arrow). *b*, A magnitude scalogram of the same stimulus, which shows the change in spectral content over time. Notably, the energy at the F0 is stable at 98 Hz for the duration of the vowel. The red overlaid line indicates the results of the F0 frequency-tracking algorithm when applied to the stimulus under conditions identical to those used in the MEG and EEG FFR analyses. These control analyses confirm that the stimulus has a stable F0 and that the tracking algorithm does not introduce the frequency changes observed in the physiological data.

aftereffect, or on the relative durations of the responses, comparison of the number of cycles in the stimuli and the average responses (Xu and Ye, 2015, their Figs. 1, 2) suggests that there is indeed a poststimulus response, which persists up to three cycles, dependent on stimulus duration. For example, when the stimulus encompasses 5 cycles, the response also has 5 cycles; but when the stimulus has 18 cycles, the response appears to have 21 cycles. A neural delay model could account only for an aftereffect of fixed duration.

To our knowledge, the entrainment criterion that the FFR frequency converges to the target frequency progressively over cycles has not been tested or remarked on. However, common methods used in FFR analysis and data presentation, such as creating a single spectrum over the stable portion of the FFR period (Skoe and Kraus, 2010), selecting the peak from within a frequency range around the expected peak frequency (Lee et al., 2015), presenting results averaged over subjects, or presenting time series data, in which frequency differences are represented

Figure 2. Time series of activity in the left and right auditory cortex, compared with the audio stimulus. **a**, Regions of interest, illustrated in a sample subject. **b**, The audio signal, in original form and after the 80–450 Hz bandpass filter used on the data is applied to it (blue trace). **c**, Activity continues for several cycles following stimulus offset (arrows), which is not observed in the filtered audio signal in **b**. The gray dashed reference line indicates stimulus offset. Shaded bars represent SEM.

in subtle changes to cycle length, could easily obscure such patterns.

To evaluate evidence of oscillatory entrainment, we reanalyzed an existing dataset in which magnetoencephalography (MEG)/electroencephalography (EEG) FFRs were measured from healthy young adults listening to a speech syllable (Coffey et al., 2016) to test the following hypotheses: (1) whether a post-stimulus aftereffect exists; (2) whether the frequency representation converges toward the stimulus frequency over time; and, if so, (3) whether in the absence of continued stimulation it relaxes toward its original state (Obleser and Kayser, 2019). Finding

positive results, we conducted a new MEG experiment to test whether a preceding stimulus affects the representation by the brain of an incoming stimulus. Our approach avoids pitfalls of analysis that might produce artifactual oscillations (Gourévitch et al., 2020). Moreover, MEG allows for the separation of cortical and subcortical sources of the FFR (Coffey et al., 2016) such that we can test whether these phenomena can be observed at all levels of the auditory system or not, which the existing literature does not address.

Materials and Methods: Experiment 1

Participants. Data from 20 neurologically healthy young adults included in a previous study (Coffey et al., 2016) were used in this study (mean age, 25.7 years; SD, 4.2; 12 females; all were right handed and had normal or corrected-to-normal vision; ≤ 25 dB hearing level thresholds for frequencies between 500 and 6000 Hz assessed by pure-tone audiometry; and no history of neurologic disorders). The sample size was determined by the research questions for which the study was originally designed (Coffey et al., 2016, 2017), which robustly showed the cortical contribution to the FFR and has since been replicated several times (Bidelman, 2018; Hartmann and Weisz, 2019; Ross et al., 2020; Gorina-Careta et al., 2021). Informed consent was obtained, and all experimental procedures were approved by the Montreal Neurologic Institute Research Ethics Board.

Stimulus presentation. The stimulus for the MEG/EEG recordings was a synthetic 120 ms speech syllable, /da/, which comprised a 10 ms consonant burst, a 30 ms formant transition, and an 80 ms steady-state vowel with a fundamental frequency of 98 Hz. This syllable is favored by many FFR researchers for its acoustic properties, ecological validity in speech [human speech (fundamental frequency) F0 ranges between 80–400 Hz], and ability to produce a robust FFRs in most subjects (Skoe and Kraus, 2010). The stimulus was presented binaurally at 80 dB SPL, $\sim 14,000$ times in alternating polarity, through insert earphones with foam tips (model ER-3A, Etymotic Research). For five subjects, $\sim 11,000$ epochs were collected because of time constraints. Stimulus onset asynchrony (SOA) was randomly selected between 195 and 205 ms from a normal distribution. To control for attention and reduce fidgeting, a silent wildlife documentary (“Yellowstone: Battle for Life,” BBC, 2009) was projected onto a screen at a comfortable distance from the subject’s face. This film was selected for being continuously visually appealing; subtitles were not provided so as to minimize saccades.

Neurophysiological recording and preprocessing. Two hundred and seventy-four channels of MEG (axial gradiometers); one channel of EEG data (Cz, 10–20 International System, mathematically averaged mastoid references), EOG, and EKG; and one audio channel were simultaneously acquired using an MEG system and its in-built EEG system (model Omega 275, CTF Systems). All data were sampled at 12 kHz.

Data preprocessing was performed with Brainstorm (Tadel et al., 2011) and using custom MATLAB scripts (MathWorks). After bandpass filtering (80–450 Hz; 4306-order linear phase FIR (finite impulse response) filter with a Kaiser window and 60 dB stopband attenuation; the order is estimated using the MATLAB kaiserord function, and filter delay is compensated by shifting the sequence to effectively achieve zero phase and zero delay, as per Brainstorm default settings (Tadel et al., 2011) and down-sampling to 1000 Hz, which removes low-frequency components of the evoked response (Musacchia et al., 2008); data were

epoch (−50 to 300 ms relative to stimulus onset). Data were reanalyzed with respect to a previous analysis, to extend the time window of analysis to 300 ms after stimulus offset, as the end of the response appeared to have been truncated by the 150 ms window used previously in the study by Coffey et al. (2016); however, analyses and figures will be limited to 200 ms, which corresponds to the mean SOA. A simple threshold-based artifact rejection was applied of $\pm 35 \mu\text{V}$ on EEG channels and $\pm 1000 \text{ fT}$ on MEG channels. On average, >95% of epochs were kept. Subject averages were created by first averaging epochs of each polarity and then summing negative and positive polarity averages.

We took advantage of Brainstorm software developments since the original analysis to better standardize the location of regions of interest (ROIs). For the subcortical ROIs, the center of each region of interest was defined using standard space (MNI152) coordinates, transformed to the subjects' T1-weighted image from MRI, and visually inspected to confirm location. The coordinates of subcortical ROIs were as follows: right cochlear nucleus (rCN): 8, −36, −38; left CN (lCN): −4, −36, −38; right inferior colliculus (rIC): 6, −36, −10; left inferior colliculus (lIC): −4, −36, −10; right medial geniculate body (rMGB): 14, −30, 6; and lMGB: −10, −32, 6. Each ROI was between 0.4 and 0.5 cm^3 .

We used a distributed source modeling approach, in which the amplitudes of a large set of dipoles are used to map activity originating in multiple generator sites; these are constrained by spatial priors derived from each subject's T1-weighted anatomic MRI scan (Baillet et al., 2001; Gross et al., 2013), from which cortical sources and subcortical structures were prepared using FreeSurfer (Fischl, 2012). Anatomical data were imported into Brainstorm, and the brainstem and thalamic structures were combined with the cortex surface to form the image support of MEG distributed sources: the mixed surface/volume model included a triangulation of the cortical surface (~15,000 vertices), and brainstem and thalamus as a three-dimensional dipole grid (~18,000 points). An overlapping-sphere head model was computed for each run; this forward model explains how an electric current flowing in the brain would be recorded at the level of the sensors, with fair accuracy (Tadel et al., 2011). A noise covariance matrix was computed from 1 min empty-room recordings taken before each session. The inverse imaging model estimates the distribution of brain currents that account for data recorded at the sensors. We computed the minimum-norm estimate (MNE) source distribution with unconstrained source orientations for each run using Brainstorm default parameters. The MNE source model is simple, robust to noise and model approximations, and very frequently used in the literature (Hämäläinen, 2009). Source models for each run were averaged within subject.

The cortical ROIs were defined using the atlas of Destrieux et al. (2010) by combining the regions labeled as “S_temporal_transverse,” “G_temp_sup-Plan-tempo,” and “G_temp_sup-G_T_transv” for the left and right hemispheres, respectively. The left and right auditory cortex (AC) regions cover the posterior superior temporal gyrus. This approach yielded AC ROIs that are similar to those reported in our previous work (Coffey et al., 2016), yet the process does not require identification of AC by visual inspection, and is thus more standardized and reproducible.

We extracted a single time series of mean amplitude for each cortical ROI, and one for each of the three orientations (x , y , z) for the subcortical ROIs, which have unconstrained orientation. For analyses involving FFR F0 strength and peak frequency, time windows of interest (described for specific analyses below) were

obtained by first windowing the signal (5 ms raised cosine ramp), zero padding to 1 s to enable a 1 Hz frequency resolution with subsequent fast Fourier transform, and rescaling by the proportion of signal length to zero padding. The spectra of the three orientations were then summed in the frequency domain to obtain the amplitude of each subject's neurologic response at the peak frequency close to the fundamental frequency, which was detected by an automatic script. Note that at the level of subcortical nuclei, magnetic fields emanating from activity of neural populations in the right and left sides are likely not spatially separable using the present techniques and we do not address subcortical lateralization hypotheses herein; we therefore average them.

Data analysis. Nonparametric statistical tests are used throughout, as FFR amplitudes in a population tend to be non-normally distributed (Wilcoxon signed-rank test for within-group between-condition comparisons; Spearman's ρ (r_s) for correlations; $\alpha = 0.05$). Effect sizes for between-condition comparisons use the probability of superiority (PS_{dep}), which is calculated as the number of positive difference scores divided by the total number of paired scores (Grissom and Kim, 2012). For correlational analyses, r_s is itself a quantitative measure of the strength of the relationship between variables.

To evaluate the presence of a poststimulus aftereffect, we first compared the strength of the oscillatory brain activity in a 20 Hz frequency band encompassing the fundamental frequency (88–108 Hz) in the postoffset period (130–180 ms) with the amplitude of the response of the brain at the same frequency during the prestimulus period (−50 to 0 ms), when no signal would be expected. Comparison to a prestimulus baseline period of equivalent length accounts for possible artifactual peaks in the spectrum caused by the bandpass filter. For visual comparison, we calculated spectra averaged across subjects during the prestimulus window, two time periods of equivalent length (early, 10–60 ms; late: 80–130 ms) during stimulation, and during the poststimulus window.

As a control, based on a possible cortical FFR generator at 25 ms (Tichko and Skoe (2017)), we assessed whether there was above-baseline oscillatory activity in a window that started at 25 ms poststimulus offset (i.e., 145–195 ms), and counted the number of clear oscillations observed in the signal from the cortical ROIs relative to the number of periods in the stimulus.

The postoffset period begins 10 ms post-stimulus offset to account for approximate neural conduction delays to cortex (Tichko and Skoe, 2017). Values from subcortical structures were averaged across hemispheres, whereas data from the left and right auditory cortex were analyzed separately (using one-tailed statistical tests), for a total of five ROIs (lAC, rAC, MGB, IC, CN). Results were corrected for multiple comparisons using false discovery rate (FDR; Benjamini and Yekutieli, 2001).

Reasoning that the strength of the aftereffect should be somewhat related to the strength of the FFR, we used r_s to test for a positive relationship between the FFR amplitude in the late window and the postoffset window across all levels (one-tailed, FDR corrected). To investigate lateralization of the aftereffect at the cortical level, we compared the strength of the correlation between amplitude during versus poststimulus on the right versus the left side, using Fisher's exact test. The purpose of these exploratory analyses is to document the relationships between FFR and aftereffect amplitude at each level of the auditory system, which may be informative for future experiments.

To investigate the convergence of frequency accuracy, we first calculated the peak frequency of the averaged FFR responses by

Figure 3. Changes in FFR spectral amplitude pre, during, and post sound presentation. **a**, Schematic of the timing of the four 50 ms analysis windows relative to the auditory stimulation: prestimulation (baseline), during stimulation the first and second halves of the FFR, and poststimulation. **b**, Mean amplitude spectra of each region of interest for each time period. The gray dashed line at 98 Hz denotes the fundamental frequency of the stimulus (F₀). Arrows highlight the difference between peak frequency and stimulation frequency, which is most pronounced in the cortical regions, yet not during the late stimulation phase.

subject for each region of interest. To test whether the FFR peak frequency accuracy converges toward the frequency of the stimulus, we calculated the difference between the fundamental frequency of the stimulus (98 Hz) and each subject's FFR in a 50 ms window at the beginning (10–60 ms) and the end of the FFR (80–130 ms; see Fig. 3). Our primary interest was the MEG FFR originating in the right auditory cortex, which has a higher amplitude than that of the left auditory cortex and has been shown to correlate with individual differences in experience and perception (Coffey et al., 2016, 2017) and with top-down processes such as attentional modulation (Hartmann and Weisz, 2019);

nevertheless, we conducted one-tailed Wilcoxon signed-rank tests at each region of interest, correcting for multiple comparisons (FDR).

We then calculated peak frequencies over successive overlapping 50 ms windows (1 ms steps), and plotted mean and SE of peak frequency for subjects whose signal-to-noise ratio (SNR) was ≥ 5 (i.e., the amplitude of the fundamental was at least five times higher than the same frequency during the 50 ms silent baseline period preceding stimulus presentation), when at least 30% of subjects met the criterion. We plotted the mean amplitude for subjects and datapoints reaching criteria, after first

normalizing amplitude to that calculated during the baseline period over a ~ 50 Hz frequency range centered on 98 Hz (range, 73–122 Hz). We used amplitude-based measures for the research questions concerning frequency change because although phase measures are independent of signal amplitude fluctuations in principle, they are noisy when signal amplitude is weak.

As a control, we confirmed that when subjected to the same spectral analysis, the audio stimulus itself showed a fundamental frequency peak at precisely 98 Hz (Fig. 1*a*). We further calculated a magnitude scalogram via a continuous wavelet transform (“cwt” function in MATLAB, default parameters) to visualize the spectral content of the stimulus over time and to confirm that the fundamental frequency remained static. We filtered and downsampled the audio stimulus time series using parameters and functions identical to those used on the MEG and EEG data, and applied the tracking function on 50 ms overlapping windows, as described above for the data. The resulting tracking results are overlaid on the magnitude scalogram in Figure 1*b*. Any difference in the FFR fundamental response found in these analyses is therefore not accounted for by the frequency content of the stimulus itself.

Finally, we explored the possibility that multiple FFRs captured inadvertently within the same ROI might interact to distort the apparent frequency content of the FFR. This situation could arise if regions of interest in the MEG analysis capture information from more than one source, as might be the case with two auditory cortex FFR sources at 13 and 25 ms, as proposed in the study by Tichko and Skoe (2017). We generated a simulated FFR made from the filtered and downsampled audio stimulus as above, added to itself with delays of 13 and 25 ms. We also created a second simulated FFR that approximates six sources at different delays in the EEG FFR model of Tichko and Skoe (2017), and analyzed the frequency content of 50 ms windows in the beginning, middle, and end of the signal (similar to the approach presented in Fig. 3). We observed no frequency shift in the early or middle windows, and a small frequency shift of ~ 4 Hz when amplitude was low and spectral peak was broad and flattened. The duration and magnitude of this effect would be insufficient to account for a prolonged, pronounced frequency shift. However, more work is needed to clarify how cross talk and delays between different subcortical and cortical sources may influence the frequency content of measured signals.

When the stimulus ends, we expect an entrained oscillatory system with a preferred frequency that differs from that of the input to gradually relax toward that frequency. To test for oscillatory relaxation, we fit linear functions (least squares) to tracked frequencies during stimulation (80–120 ms post-stimulus onset) and after stimulus offset (10 ms to 50 ms post-stimulus offset) for each subject, and compared the resulting slopes for statistical differences at the group level using a Wilcoxon signed-rank test (two tailed). Such that all subjects had the same number of data points included in the linear fit, we did not exclude datapoints that did not meet an SNR criterion, as was applied in the previous analysis. Tests were conducted at each ROI, and reported p values are corrected for multiple comparisons (FDR).

Because EEG is more commonly used to measure FFR (Coffey et al., 2019), and because MEG and EEG may differ in the degree to which they pick up different sources (Bidelman, 2018; Ross et al., 2020), we repeated the analyses described above in the simultaneously recorded single EEG channel (Cz referenced to mastoids) to investigate whether evidence of oscillatory phenomena might also be observed in EEG FFR recordings.

Results: Experiment 1

There is poststimulation activity throughout the auditory system

In the MEG FFRs from the left and right auditory cortex, the group averages clearly indicate the presence of continued activity following stimulus offset, although none is observed when the audio stimulus is filtered and sampled in the same manner (Fig. 2*b*). A conservative estimate of poststimulus entrainment is used that excludes the first 10 ms following stimulus offset, during which time feedforward transmission through the subcortical auditory system is expected (Tichko and Skoe, 2017). The time series is available for the cortex (Fig. 2), which is modeled as a surface and yields a single time series per region of interest. Because the subcortical regions are modeled as a volume and yield three time series per region from which a meaningful single time series is not readily reconstructed, we do not show an equivalent figure for the subcortical regions. However, in both cases the signals from both cortical and subcortical regions are analyzed in the spectral domain.

At each level of the auditory system, there was significantly greater activity in the postoffset period at the stimulus' fundamental frequency, compared with activity at the same frequency in the prestimulus baseline period (IAC: $Z = 3.70$, $p = 0.00027$, $PS_{dep} = 0.95$; rAC: $Z = 3.73$, $p = 0.00047$, $PS_{dep} = 0.95$; MGB: $Z = 3.32$, $p = 0.00074$, $PS_{dep} = 0.85$; IC: $Z = 3.10$, $p = 0.0012$, $PS_{dep} = 0.85$; CN: $Z = 2.80$, $p = 0.0026$, $PS_{dep} = 0.85$; p values are FDR adjusted).

To control for the possibility that the observed aftereffect might have been generated by a cortical source at 25 ms (Tichko and Skoe, 2017), we confirmed that oscillatory activity in the window 145–195 ms post-stimulus offset at the cortical ROIs were significantly above baseline levels (i.e. a window that starts 25 ms after stimulus offset) (IAC: $Z = 2.91$, $p = 0.0018$, $PS_{dep} = 0.75$; rAC: $Z = 3.73$, $p = 0.0005$, $PS_{dep} = 0.80$). Furthermore, visual inspection of the FFR from the IAC and rAC suggests that it contains 15 cycles, whereas the stimulus has 11 (Fig. 2). A cortical feedforward response delayed by up to 25 ms therefore does not account for the result, nor would the summation of two cortical generators delayed by 13 and 25 (Tichko and Skoe, 2017) result in >12 cycles.

The strength of the aftereffect was significantly correlated with the strength of the FFR in the late window at the level of the rAC ($r_s = 0.73$, $p = 0.00086$) and IC ($r_s = 0.49$, $p = 0.037$); values trended in the same direction but did not reach significance at the level of the IAC ($r_s = 0.34$, $p = 0.092$), MGB ($r_s = 0.32$, $p = 0.082$), and CN ($r_s = 0.42$, $p = 0.56$). FDR-adjusted p values are reported. The correlation between FFR and aftereffect strength was marginally stronger in the right auditory cortex compared with left auditory cortex ($Z = 1.63$, $p = 0.051$).

Figure 3 illustrates changes in spectral amplitude prestimulus (baseline), during the FFR (early and late periods), and poststimulus. The frequency of averaged peak amplitude differs from that of the stimulus during the early and poststimulation periods, whereas during the late portion of the FFR, the FFR peak frequency matches that of the stimulus closely. This pattern is most pronounced at the cortical level.

Frequency accuracy increases with time

The difference between the fundamental frequency of the stimulus (98 Hz) and each subject's FFR in a 50 ms window at the

Table 1. Difference between peak frequency and the 98 Hz stimulation frequency, during the early (10–60 ms) and late (80–130 ms) portions of the FFR

Region	Early mean (SD)	Late mean (SD)	<i>p</i> Value	Effect size (PS_{dep})
IAC	9.95 (4.96)	3.95 (3.87)	0.00021*	0.75
rAC	11.85 (3.07)	2.70 (2.41)	0.00030*	1.0
MGB	8.60 (3.30)	3.05 (2.48)	0.00083*	0.95
IC	9.05 (4.01)	3.50 (3.59)	0.0010*	0.90
CN	6.60 (4.48)	4.35 (4.04)	0.051	0.60

The mean and SD values are in hertz (absolute value of difference from F0). Statistically significant comparisons are indicated with an asterisk (*), and are FDR corrected.

Figure 4. Frequency tracking in the left and right auditory cortex (top), and the corresponding normalized amplitude of the detected peaks (bottom), over time. Data are plotted for 50 ms windows in which at least 30% of subjects had peaks that reached a +5 SNR threshold with respect to baseline. Center of the window is reported relative to stimulus onset. The horizontal and vertical dashed lines represent stimulus F0 and offset, respectively. Shaded bars represent the SEM.

beginning (10–60 ms) of the response was significantly greater than at the end (80–130 ms) of the response for all regions of interest except for the CN, which was marginal (Table 1). We speculate that frequency tracking improves over time, converging toward the stimulation frequency as information is accrued.

A finer-grained analysis of peak frequency in 50 ms overlapping windows reveals a pattern of convergence toward the fundamental frequency of the stimulus, in both the left and right auditory cortex (Fig. 4). Mean peak frequencies are plotted for each window, for subjects who showed a clear peak close to the stimulus F0 (defined as five times higher than the amplitude of the same frequency found in the prestimulus period). The FFR reaches the 98 Hz stimulation frequency ~100 ms following stimulation onset (approximately nine cycles of the F0 of the vowel). After sound offset, the mean peak frequency decreases. Tracking results are unstable at the beginning and end of the analysis window, indicated by higher tracking variability, when FFR amplitude is low and few subjects showed clear spectral peaks.

In the subcortical structures, the frequency tracking accuracy also increases toward the end of the stimulus, as it does for the cortex, but as the initial frequency is closer to the stimulus F0, a less pronounced pattern of convergence is observed (Fig. 5).

Tracked frequency relaxes after external input ends

Positive relationships between time and tracked frequency before stimulus offset and the negative relationship following offset indicates first convergence toward 98 Hz, followed by a relaxation to lower frequencies. At the rAC, the mean slope of tracked frequency during stimulus presentation was 0.20 (SD, 0.14), and after stimulus offset was –0.31 (SD, 0.18). The pattern was significant at the rAC level ($Z = 3.91$, $p = 0.00044$, $PS_{dep} = 1.0$). A similar pattern of results was found at the IAC (mean during, 0.17, SD, 0.24; mean afterward, –0.21, SD, 0.30; $Z = 3.29$, $p = 0.0025$, $PS_{dep} = 0.85$). At subcortical levels, slope pre-stimulus offset versus post-stimulus offset was significantly different at the MGB (mean during, 0.21, SD, 0.31; mean afterward, –0.14, SD, 0.38; $Z = 2.31$, $p = 0.034$, $PS_{dep} = 0.75$), IC (mean during, 0.15, SD, 0.29; mean afterward, –0.15, SD, 0.41; $Z = 2.12$, $p = 0.033$, $PS_{dep} = 0.75$), and CN (mean during, 0.14, SD, 0.22; mean afterward, –0.11, SD, 0.35; $Z = 2.20$, $p = 0.035$, $PS_{dep} = 0.75$); FDR-adjusted p values are reported.

Evidence for oscillatory phenomena in the EEG FFR

In the single-channel EEG recording, as observed in MEG, there was significantly greater activity in the postoffset period close to the fundamental frequency of stimulus, compared with activity at the same frequency band in the prestimulus baseline period ($Z = 3.92$, $p = 0.00014$, $PS_{dep} = 1.0$; Fig. 6a). In contrast to the MEG data from regions of interest, the accuracy of frequency representation was not significantly better during the late FFR period (mean difference from stimulus F0, 2.95, SD, 3.02) than during the early FFR period (mean, 2.10, SD, 1.59; $Z = 1.36$, $p = 0.087$, $PS_{dep} = 0.5$); nor did the averaged spectral peak differ between the early and late period (Fig. 6b). The frequency tracking

analysis (Fig. 6c) suggests that following the stimulus offset, the frequency trends toward lower frequencies, as observed in the AC (but less so in the subcortical MEG FFRs). Qualitatively, the tracking results of the EEG signal do not resemble those of any single one of the MEG ROI signals, further supporting the proposal that the EEG FFR is a composite signal that may represent different mixtures of sources over its duration (Tichko and Skoe, 2017; Coffey et al., 2019). Because FFR amplitude in the auditory cortex peaks last, at ~60 ms (Coffey et al., 2016; but see Fig. 4), a plausible interpretation is that EEG is at first driven by the brainstem sources close to the stimulation frequency, and then EEG becomes increasingly dominated by cortical sources.

As noted in previous work (Coffey et al., 2016), harmonic components (i.e., 196, 294, and 392 Hz) are clearer in the EEG signal than in the MEG signal. Interestingly, the harmonics appear to be strongly represented in the late FFR period but do not appear in the post-FFR period, suggesting that the oscillatory phenomenon may be related only to representation of the envelope of the stimulus rather than its high-frequency fine structure. Although these results raise further questions about the interactions between structures, and the representation of subcortical

Figure 5. Frequency tracking in subcortical structures. *a–c*, Frequency tracking and normalized amplitude of detected peaks in the MGB (*a*), IC (*b*), and CN (*c*). Data are plotted for 50 ms windows in which at least 30% of subjects had peaks that reached a +2 SNR threshold with respect to baseline. The horizontal and vertical dashed lines represent stimulus F0 and offset, respectively. Shaded bars represent the SEM.

and cortical generators in the compound EEG FFR signal recorded at the scalp over time, our results generally support the involvement of oscillatory dynamics in the encoding and representation of periodic acoustical information of the brain.

Figure 6. Evidence for oscillatory phenomena in EEG. *a*, Activity in the EEG recording continues for several cycles after stimulus offset (arrow). *b*, Mean amplitude spectra for the EEG FFR for each time period (Fig. 3*a*). In contrast to the AC, EEG FFR does not exhibit a significant difference in frequency representation in early versus late FFR periods. *c*, Frequency tracking in EEG trace (top), and the corresponding normalized amplitude of the detected peaks (bottom), over time. Data are plotted for 50 ms windows in which at least 30% of subjects had peaks that reached a +5 SNR threshold with respect to baseline. The horizontal and vertical dashed lines represent stimulus F0 and offset, respectively. Shaded bars represent the SEM.

Figure 7. Audio stimulation used in experiment 2, in which three tones are presented randomly in a continuous stream. **a**, The relative pitch of the three tones, shown schematically in a representative 500 ms segment. The duration of each tone is equivalent to nine cycles of its fundamental frequency. **b**, Tone pair dyads are defined within the continuous stream for analysis, such that the relative effects of higher- or lower-pitch preceding tones may be studied. Color boxes represent tone dyads: orange, AB; green, BC; blue, CB; purple, BA.

Control analyses for analytic artifacts

Signal processing steps such as bandpass filtering and peak frequency extraction can introduce artifactual oscillations, and, potentially, frequency distortions. As described above, we tested the effects of the filtering and resampling procedures on the simultaneously recorded audio channel, and found no aftereffect; nor did the fundamental frequency stray from 98 Hz (Figs. 1*b*, 2*b*). However, broadband MEG and EEG data might behave differently because of the strong $1/f$ component in neural oscillations. To further confirm that the two main results of experiment 1 (i.e., the presence of an aftereffect and of a changing fundamental frequency representation over time) are not the result of our analytic choices, we created an unfiltered average FFR for a single subject's EEG data at the original 12 kHz sampling frequency. Visual comparison of the subject's raw and filtered averaged revealed a similar number (3) of aftereffect oscillations. Artifactual oscillations could be induced only with a very narrowband version of the same filter (e.g., 93–103 Hz or 10 Hz wide).

To observe whether the observed frequency dynamics in the F0 tracking were specific to the Fourier transform-based sliding window analysis, we applied a continuous 1-D Morlet waveform analysis, a contour analysis, and ridge extraction to the raw average (MATLAB Wavelet Toolbox; default parameters). In all analyses, we observed similar patterns of F0 change compared with the individuals' frequency tracking and the group average presented in Figure 6*c*. We can thus conclude that while signal-processing steps can distort the results, they do not account for the observations in the present work.

Experiment 2

Finding positive evidence for oscillatory phenomena, we designed a new study to test the prediction that the frequency representation of an incoming stimulus would be influenced by the frequency content of an immediately preceding stimulus. We created a stimulus sequence in which three short complex tones with different fundamental frequencies (tone A, 98 Hz; tone B, 131 Hz; and tone C, 147 Hz) were presented in a continuous stream in a pseudorandom order, with equal transitional

probabilities. This arrangement allowed us to examine how the recent exposure to the higher-pitched tones B and C affected the representation of tone A, which has the same pitch as the stimulus in experiment 1 and is robustly represented at all system levels.

In experiment 1, we had observed an unexpected phenomenon: that convergence at the cortical level occurred consistently from lower rather than random frequencies (i.e., <98 Hz; Fig. 4), when sound was presented following a short period of silence. This observation is consistent with the existence of a preferred or resonant frequency that is somewhat lower than the 98 Hz fundamental, but can be progressively entrained to 98 Hz over ~100 ms (approximately nine cycles for tone A). In experiment 2, we hypothesized that after nine cycles of entraining to a higher-pitched tone, convergence to tone A would occur from higher, rather than lower frequencies as observed in experiment 1, and that differences in sensory history would affect the representation of a stimulus.

Materials and Methods: Experiment 2

Participants. Thirty neurologically healthy young adults (mean age, 25.8 years; SD, 4.8; 15 females; all were right handed and had <25 dB hearing level thresholds for frequencies between 500 and 4000 Hz assessed by pure-tone audiometry; and no history of neurologic disorders).

Experimental procedure. Subjects were recruited to participate in a three-session experiment in which the effect of TMS to the right auditory cortex 5 min before FFR recording was assessed. Here we report only the MEG data from the sham TMS condition, in which the TMS device was discharged perpendicularly to the head as a control condition for the other sessions (i.e., no magnetic stimulation of the brain was performed, but the subjects were aware of clicks). Written informed consent was obtained, and all experimental procedures were approved by the Montreal Neurologic Institute Research Ethics Board.

Stimulus presentation. Each stimulus run included a random sequence of three tones, which were comprised of a sine wave at the fundamental frequency, added to sine waves at the frequencies of its second to fourth harmonics (Fig. 7). Tones were kept short to maximize the number of transitions between them, and transitions were immediate without a prestimulus interval to allow us to assess whether a preceding tone would affect representation of an incoming tone. Each tone includes nine cycles of its fundamental frequency, such that the sine waves of each harmonic begin and end at a zero phase; therefore, the stimuli are of different durations. The fundamentals of the tones were as follows (corresponding Western musical scale value given in parentheses): tone A (G2), 98 Hz (period, 10.2 ms; duration, 91.83 ms); tone B (C3), 131 Hz (period, 7.7 ms; duration, 68.83 ms); and tone C (D3), 147 Hz (period, 6.8 ms; duration, 61.17 ms). A silent gap, the same length as tone C, was occasionally presented between two tone C presentations for an unrelated research question. A 5 ms raised cosine ramp was applied to each stimulus to avoid clicks. As in experiment 1, the stimulus was presented binaurally at 80 dB SPL, through inserted earphones with foam tips (model ER-3A, Etymotic Research), and to control for

attention and reduce fidgeting, a silent wildlife documentary (“Yellowstone: Battle for Life,” BBC, 2009) was projected onto a screen at a comfortable distance from the subject’s face.

Neurophysiological recording and preprocessing. MEG recording was identical to that described in experiment 1, except that data were sampled at 6 kHz, sufficient to capture the fundamental frequencies that are the focus of these analyses. Preprocessing was also as described above to maximize comparability of the results; data were bandpass filtered (80–450 Hz), downsampled to 1000 Hz, and epoched in windows of –50 to 351 ms relative to the onset of the first of a pair of transition dyads (e.g., stimulus pairs AB, BC), and later trimmed to the portion of data relevant to each analysis, as will be further described in each section. A simple threshold-based artifact rejection was applied of ± 1000 fT on MEG channels, and subject averages were created by first averaging epochs of each polarity and then summing negative and positive polarity averages. ROIs were defined, distributed source models were created based on the subjects’ anatomy, MNE mixed models were created, and signals were extracted from the ROIs, as described in experiment 1. In total, each dyad (encompassing a transition) occurred >2000 times. Whole-cap EEG was simultaneously measured with an external system (BrainAmp MR, Brain Products) to address research questions not reported here.

Data analysis. All three tones have fixed pitches. If all three tones are represented with high fidelity in the brain, we would expect discrete peaks in the spectra at the fundamental frequencies of those tones (i.e., at 98, 131, and 147 Hz), to the extent that levels of the auditory hierarchy are capable of robustly representing those frequencies within the stimulation time frame. If changing between frequency representations occurs gradually over multiple cycles, as suggested by experiment 1, we expect broad peaks as the representations blur together. To observe the frequency representation across regions of interest, we first calculated mean spectra across responses to all stimuli with the same F0 (3600 epochs for each tone) at each level of the auditory system using a Fourier transform, averaged across subjects for each region of interest. Visual inspection of these averages supported the second interpretation.

In experiment 1, the right auditory cortex showed a stronger, clearer FFR than the left auditory cortex. To test whether the previously observed result was replicable, we evaluated whether the amplitude of the right AC FFR for tone A was statistically greater than that of the left AC (Wilcoxon signed-rank test, one-tailed).

We tracked the frequency representation (as described in experiment 1, using a 50 ms sliding window) during the presentation of tone A, which occurred after tones B and C. Because each tone does not have a preceding silent period that can be used as a baseline for an SNR threshold, we instead calculated the mean amplitude within the 250–450 Hz frequency range for each subject during tone presentation. Within this frequency range, the averaged spectra are relatively flat, lacking meaningful signals, and can therefore be used as a proxy of the background noise levels of the filtered data. Data points were included in the group average that exceeded the high-frequency range mean by a factor for 2. Data were smoothed with a 5 point moving average filter for visualization purposes.

To fairly test whether frequency representation convergences from higher frequencies to the fundamental of tone A, we must avoid including tracking data points that are derived from windows that overlap with the preceding stimulus. Therefore, we used a short tracking window (30 ms) to allow for multiple sliding windows, and included window centers starting at 25 ms

(i.e., window half-width of 15 ms plus the 10 ms period following the transition to avoid the possibility of contaminating the response from subcortical neural activity of the preceding stimulus). As in the slope-based analysis in experiment 1, we do not exclude data points on the basis of SNR, such that all subjects and brain regions have the same number of points. We then fit linear functions to the tracked frequencies of 25–85 ms and tested whether slopes are less than zero, indicating convergence from higher to lower frequencies, using a Wilcoxon signed-rank test (one-tailed; FDR corrected for multiple comparisons across ROIs, within each condition).

To observe the effect of preceding stimuli on phase and amplitude in the time domain, we bandpass filtered (80–150 Hz) the time series for dyads AB, CB, BA, and BC, and calculated a group average and SE aligned to the transitions between the tones for each pair, for qualitative visual inspection. It is difficult to cleanly quantify the effects of the aftereffect in the frequency domain by comparing portions of dyad pairs with different transitions, because a window of at least a few cycles is needed to accurately quantify oscillatory activity in a spectral analysis, and because the expected aftereffect lasts three to four cycles, requiring a 10 ms buffer period to account for neural delay. We therefore conducted *post hoc* analyses on the amplitude of tone B, focusing on the AB versus CB comparison, testing the hypothesis that the amplitude of B would be higher when preceded by A as opposed to C. We defined two discrete 30 ms time windows that capture information about the brain’s response to sound starting 10 and 40 ms after the transition to B (“early” and “late”) for further analysis (see Fig. 10). Instead of using the amplitude of spectral peaks, which can be less reliably detected in such short time windows particularly for deeper regions, we averaged the amplitude values within a frequency band covering all tone fundamental frequencies (i.e., 80–160 Hz) and computed the relative difference between the amplitude of tone B when it occurred following tone A versus tone C. This measure captures the relative difference in amplitude of frequency representation across conditions, even if the response of the brain has not fully converged to the fundamental frequency of the stimulus.

Results: Experiment 2

Replication of right lateralization of cortical FFR

We confirmed that the amplitude of the right auditory cortex for tone A (98 Hz, comparable to that used in our previous study) was statistically greater (mean, 0.060; SD, 0.031) than that of the left (mean, 0.037; SD, 0.016; $Z = 4.1$, $p < 0.0001$, $PS_{dep} = 0.87$; Fig. 8), replicating the finding of strong lateralization of amplitude in the cortical MEG FFR, in this larger sample. We hereafter focus on the right auditory cortex response based on the observation from experiment 1 that patterns of convergence and the duration of the aftereffect in left and right auditory cortex were similar (Fig. 2).

Frequency tracking is modulated by preceding frequencies

Results of frequency tracking in the right auditory cortex for tone A following either tone C or B, which are both higher frequency, are presented in Figure 9. Tracking results at the end of tones C and B imply an incomplete convergence to their stimulation frequencies by tone offset (62 and 69 ms, respectively). This result is in agreement with the results from experiment 1, in which convergence was observed only after ~100 ms of steady periodic input. Importantly, frequency tracking convergences in

Figure 8. Distribution of normalized left–right amplitude differences in the auditory cortex ROIs for each individual ($n = 30$) using the mixed surface-volume MNE model shows strong right-sided asymmetry. Results are normalized between -1 and 1 within subject.

both cases from higher frequencies indicating an effect of the preceding stimulus on tone A frequency representation.

We evaluated the hypothesis that tone A representation was affected by having been preceded by the tone B, by testing whether the slope of frequency tracking was less than zero, indicating convergence from higher frequencies to the 98 Hz fundamental of tone A. At the rAC, the mean slope of tracked frequency during tone A stimulus presentation was -0.21 (SD, 0.36), which was significantly below zero ($Z = -3.24$, $p = 0.0030$, $PS_{dep} = 0.83$) as were results at the lAC (mean slope, -0.30 ; SD, 0.36 ; $Z = -3.67$, $p = 0.00030$, $PS_{dep} = 0.83$). At subcortical levels, the signals of which have lower SNR, the success of tracking results appeared unstable across participants, suggesting that the design and data did not allow for the reliable estimate of frequency needed for this analysis. Results are reported for completeness: MGB (mean slope, 0.12 ; SD, 0.37 ; $Z = 1.49$, $p = 0.43$, $PS_{dep} = 0.43$), IC (mean slope, -0.12 ; SD, 0.39 ; $Z = -1.49$, $p = 0.085$, $PS_{dep} = 0.63$), and CN (mean slope, 0.14 ; SD, 0.35 ; $Z = -2.05$; $p = 0.034$; $PS_{dep} = 0.60$).

We repeated the process for tone A when preceded by tone C, finding a similar pattern of results at the cortical level; at the rAC, the mean slope of tracked frequency during tone A stimulus presentation was -0.31 (SD, 0.44 ; $Z = -3.36$; $p = 0.0019$, $PS_{dep} = 0.80$), and at the lAC (mean slope, -0.22 ; SD, 0.41 ; $Z = -2.66$; $p = 0.0098$; $PS_{dep} = 0.77$). Subcortical results are reported for completeness, as follows: MGB (mean slope, 0.06 ; SD, 0.38 ; $Z = 0.46$; $p = 0.22$; $PS_{dep} = 0.43$), IC (mean slope, 0.06 ; SD, 0.38 ; $Z = 0.46$; $p = 0.28$, $PS_{dep} = 0.43$), and CN (mean slope, -0.037 ; SD, 0.41 ; $Z = -0.79$; $p = 0.21$; $PS_{dep} = 0.60$). Reported p values are FDR-adjusted within each analysis.

Effect of preceding events on sound representation

To further investigate whether and where convergence toward stimulation frequency occurs progressively in a process of oscillatory entrainment, we examined the mean time series of the transition point between converging dyads (AB and CB) and diverging dyads (BA vs BC). Qualitatively, it appears that after the AB and CB transitions, the phase converges somewhat, whereas in the BA versus BC comparison, the signal appears similar post-transition, suggesting a lingering effect of B. However, although phase differences are suggestive in the grand average, they appear highly variable at the subject level. The grand averages suggest that persistent differences may be more clearly manifested in measures of amplitude; namely, the amplitude of the FFR to tone B appears higher after the higher-amplitude A FFR than after the lower-amplitude C FFR (Fig. 10, top). Conversely, the response of the brain to tones A and C appears

Figure 9. Frequency tracking in the right auditory cortex for tone dyads BA and CA, which both start with a higher-frequency tone. The pattern of results after transition to tone A (98 Hz, dashed horizontal line) demonstrates frequency tracking convergence from higher frequencies. Data are plotted for 50 ms windows for subjects who had peaks that reached a $+2$ SNR threshold in each window (relative to the mean amplitude averaged in a 250–450 Hz frequency band). Time is reported relative to the onset of tone A, and vertical lines mark tone transitions to and from tone A. Shaded bars represent the SEM.

Figure 10. Activity in the right auditory cortex in response to pairs of tones, aligned at the transition point between tone dyads (dashed lines). In the transition to the same stimulus (top), the cortical response to tone B appears to differ, whereas after transition from tone B (bottom), both traces resemble one another despite a 49 Hz difference in stimulus frequency. Early (10–40 ms) and late (40–70 ms) windows are marked with gray boxes on the first dyad (top) for further analysis. Shaded areas represent the SEM.

highly similar when both are preceded by tone B, despite that the stimuli differ in frequency by 49 Hz (Fig. 10, bottom).

Results for the amplitude comparison of AB versus CB are presented in Figure 11, and positive difference values indicate that amplitude was greater following tone A than tone C. During the early window (10–40 ms post-transition), tone B mean response amplitude difference (reported in mean amplitude in $pAm^{-1} \times 10^4$) is higher after tone A than after tone C at the level of the left AC (mean, 9.55 ; SD, 21.43 ; $Z = 2.06$; $p = 0.025$; $PS_{dep} = 0.60$), right AC (mean, 31.33 ; SD, 39.46 ; $Z = 3.78$; $p = 0.00038$; $PS_{dep} = 0.80$), IC (mean, 7.05 ; SD, 11.48 ; $Z = 2.96$; $p = 0.0038$; $PS_{dep} = 0.70$), and CN (mean, 5.22 ; SD, 10.89 ; $Z = 2.63$; $p = 0.0070$; $PS_{dep} = 0.73$). It was not significant at the MGB (mean, 1.13 ; SD, 6.53 ; $Z = 1.32$; $p = 0.24$; $PS_{dep} = 0.47$).

A difference in amplitude as a function of the preceding tone was also observable within the later window (40–70 ms post-transition), but only at cortical levels: left AC (mean, 12.85 ; SD, 28.22 ; $Z = 2.88$; $p = 0.01$; $PS_{dep} = 0.67$), right AC (mean, 12.85 ;

Figure 11. Effect of preceding stimulus on FFR amplitude. The difference in amplitude of the FFR to tone B when it is preceded by tone A versus tone C is plotted for each subject, for each region of interest (MEG). An effect is found during the early window (10–40 ms) for most levels of the auditory system, but only the cortex shows a sustained effect into the late window (40–70 ms). Dashed gray reference lines indicate zero difference between response amplitudes when tone B is preceded by tone A versus tone C. Solid black lines show the group mean. Statistically significant comparisons are indicated with an asterisk (*), and are FDR corrected for multiple comparisons.

SD, 28.22; $Z = 2.20$; $p = 0.035$; $PS_{dep} = 0.67$). Subcortical levels did not show significant differences: MGB (mean, 2.68; SD, 8.41; $Z = 1.32$; $p = 0.12$; $PS_{dep} = 0.53$), IC (mean, 1.89; SD, 11.80; $Z = 0.80$; $p = 0.21$; $PS_{dep} = 0.63$), and CN (mean, 3.06; SD, 9.25; $Z = 1.42$; $p = 0.13$; $PS_{dep} = 0.57$). Reported p values are FDR adjusted within each analysis. As a control, we confirmed that the pattern of results holds when amplitudes are normalized within subjects, demonstrating that the analysis is not sensitive to between-subject differences in overall amplitude.

Discussion

The main aim of this work was to establish evidence for oscillatory entrainment in the auditory system. Consistent with this goal, experiment 1 showed that MEG- and EEG-based FFR frequencies displayed an oscillatory aftereffect for several cycles post-stimulus offset at each level of the auditory system (CN, IC, MGB, and left and right AC, based on MEG localization; Figs. 2, 3). Also consistent with the oscillation hypothesis, tracking of the FFR fundamental frequency showed that it converged to the

stimulus frequency from a lower value over ~ 100 ms (corresponding to approximately nine cycles of the F0), and then progressively diverged back to a lower value after stimulus offset (Figs. 4, 5, Table 1). Experiment 2 provided further evidence for oscillatory interactions between the offset of one stimulus and the onset of the next. Tracking the FFR frequency of the lowest tone used, 98 Hz, showed that it was systematically higher for the first 30–40 ms when preceded by higher-frequency tones (Fig. 9). Finally, transitions to the same tone from a higher- or lower-frequency preceding tone influenced the FFR amplitude of the target tone for the first 30 ms, such that amplitude was higher when the preceding FFR amplitude was also higher; this effect was seen at both cortical and subcortical levels but appeared to persist longer at cortical levels (Fig. 11).

Declaring the existence of neural entrainment phenomena must be approached with caution because many apparent oscillatory phenomena can occur because of other factors (Haegens and Zion Golumbic, 2018; Obleser and Kayser, 2019; Gourévitch et al., 2020). Haegens and Zion Golumbic (2018) propose three stringent criteria. The first criterion is the continuation of oscillatory activity for at least a number of cycles beyond external input (i.e., reverberation); results from experiment 1 clearly support the existence of a reverberation or aftereffect, and results from experiment 2 show that it interacts with incoming stimuli, leading to measurable differences in the frequency and amplitude of their representation. The search for oscillatory entrainment in the auditory system and its role has been confounded by methodological issues in the past (Gourévitch et al., 2020). In the present work, we studied a well characterized auditory evoked brain response, and selected analyses like frequency tracking that are unlikely to introduce artifactual oscillations (as might narrow-band filtering or broad spectral analysis). Our data are consistent with prior reports of poststimulus FFR aftereffects in EEG data (Xu and Ye, 2015), and with recent data also documenting similar aftereffects from intracerebral electrode cortical recordings and cortical MEG in humans in the 60–80 Hz range, similar to the stimuli used here (Lerousseau et al., 2019; Ross et al., 2020). Together, these various findings provide strong evidence for entrainment, and our data furthermore indicate that it is a pervasive feature of auditory processing in subcortical nuclei as well as cortex. The lack of clear evidence for convergence in the EEG data points to the importance of separating the different sources of FFR throughout the auditory system, the complementary nature of the MEG and EEG signals, and the value of comparing their results with those acquired using methods with more direct access to neural tissue (Ross et al., 2020).

The second criterion for neural entrainment according to Haegens and Zion Golumbic (2018) is phase alignment, which occurs when external information is introduced that contains frequencies close to the intrinsic rate of the oscillator, after which dampening or relaxing to the initial state should occur. Frequency-tracking results from both experiments support a gradual phase alignment or convergence of the frequency representation of the brain toward the stimulation frequency, and in experiment 1, postoffset relaxation. This finding fits with the concept that incoming stimuli engage an endogenous oscillatory mechanism, as has been suggested in lower frequency ranges (Thut et al., 2012).

The third criterion for entrainment is endogenous oscillatory activity in the absence of stimulation. Although we cannot observe endogenous activity in the absence of stimulation using these time-locked averaging-based techniques, the onset frequency may indicate subthreshold nonaligned activity, at the

preferred frequency of the system. In experiment 1, in which silence preceded the tone, tracking frequencies in the cortex started well below stimulation frequency, which we speculate could be caused by an oscillator with a lower preferred frequency. More work is needed to establish whether there are multiple preferred oscillatory frequencies within the auditory system, and how they might interact with peaks in FFR amplitude response at certain frequencies introduced by constructive interference of delayed feedforward signals coming from different structures (Tichko and Skoe, 2017).

Important open questions are whether and how the entrainment of oscillations plays a functional role in the auditory system, and how the underlying mechanism involved in the representation of higher-frequency oscillations relates to that used in lower ones (Thut et al., 2012; Lakatos et al., 2013; Haegens and Zion Golumbic, 2018; Gourévitch et al., 2020). Entrainment in lower frequency ranges (i.e., delta, theta bands) has been proposed as a mechanism during active stimulus processing, for example to define the parsing window of linguistic segments from continuous speech (Kösem et al., 2018; Gourévitch et al., 2020), and relationships between entrained oscillations and perceptual performance are present in the visual system (Mathewson et al., 2012; Spaak et al., 2014). The FFR is an evoked response to a periodic signal, but it is not known whether there is some advantage to an oscillatory representation as opposed to a veridical feedforward one. The FFR clearly carries periodicity information, and its strength and quality are associated with performance on perceptual tasks with a pitch component, including hearing-in-noise tasks (Anderson et al., 2010; Li and Jeng, 2011; Song et al., 2011; Marmel et al., 2013; Bharadwaj et al., 2015; Coffey et al., 2017; Presacco et al., 2019), suggesting that resistance to noise or sound degradation could be one of its functions. The oscillatory aspect might act as a very fine-scale temporal predictor, such that an incoming impulse of sound falling on the next cycle would contribute to its strength, and small timing differences could entrain the oscillation as evidence of the external cycle length accumulates over cycles. The apparatus that produces periodic sounds (e.g., vibrating vocal chords, strings, or columns of air) is constrained by physical properties and generally continues for multiple cycles, making frequency stability ecologically relevant.

The frequency tracking data indicate that the FFR is surprisingly sluggish, at least at the frequency tested, to reach the veridical value corresponding to the fundamental frequency of the stimulus. However, psychophysical studies show that pitch discrimination and identification improves as a function of duration up to and even beyond a range of 100 ms, similar to the duration required for the FFR to reach the nominal frequency of the fundamental recorded in the present study (for review, see Freyman and Nelson, 1986; Robinson and Patterson, 1995; Micheyl et al., 2012). It is therefore possible that the relatively slow frequency tracking is related to the fact that pitch processing improves with more cycles of input, but different frequencies need to be examined to test the generality of this phenomenon. There were large individual differences in the accuracy of the FFR tracking in our data, however, as well as differences between cortical and subcortical FFR sources; it will therefore be interesting to see how this variability is related to individual differences in pitch perception, which can also be quite large (Micheyl et al., 2006).

In contrast to models for syllabic processing, for example, in which a distinction is made between predictive coding (“what”) and predictive timing (“when”; Arnal and Giraud, 2012); in the case of frequency coding, the distinction may not apply. We therefore speculatively propose a second function: that the oscillator’s ongoing activity might act as an internal template to which incoming information can be compared; in effect, the “what” predicted by the system is pitch continuity. A low-level prediction mechanism that could rapidly generate error signals when stimulus periodicity changes could be useful, because encoding rapid changes in auditory information is critically important in complex sound processing, as in speech or music. Future work is needed to determine whether differences between incoming information and the FFR oscillatory entrainment produce brain responses consistent with predictive coding signals (Chao et al., 2018; Gourévitch et al., 2020).

If the FFR frequency and amplitude are affected by an oscillatory entrainment mechanism, as shown here (in both the MEG and EEG FFR; Fig. 6), the future challenge will be to understand its relationship to behavioral variables. Despite being only clearly measurable within a restricted range of human hearing (Tichko and Skoe, 2017); FFR amplitude and fidelity are known to correlate with many behavioral functions including fine pitch discrimination, hearing-in-noise perception, reading, and musicianship (for reviews, see Coffey et al., 2019; Kraus et al., 2017). Because the strength of the aftereffect is strongly correlated with that of the FFR steady state, and because the FFR includes an oscillatory component, observed behavioral correlations are suggestive of a functional role of entrainment. However, attempting to relate the strength of an aftereffect to a behavioral variable directly would be confounded by the strong correlation between FFR and aftereffect amplitude. Put another way, correlations found in the literature between FFR amplitude and behavior would also likely appear between aftereffect amplitude and behavior, but would not necessarily imply a functional role of the aftereffect. Neither can the longevity of the aftereffect be necessarily used as a metric of strength of oscillatory entrainment, as protective mechanisms that prevent the development of hypersynchronicity may be independent of the oscillatory mechanism itself. For example, cross-frequency interactions with downstream regions may contribute to adjusting the phase and frequency of local oscillators flexibly, depending on context and internal prediction models (Baillet, 2017)—a mechanism that appears to be impaired in epilepsy (Samiee et al., 2018). Clever approaches will be needed to investigate the perceptual function; for example, Baltus and Herrmann (2015) showed that the ability to perceive short gaps in tones is linked to an individual’s preferred auditory steady-state response (~40 Hz), suggesting that an oscillatory response in lower frequency ranges might influence perception. Similar approaches could be used to clarify the perceptual relevance of individual differences in FFR entrainment mechanisms.

Our results establish that frequency coding in the human brain includes a mechanism of oscillatory entrainment, with the strongest effects at the cortical level, but found throughout the system. More work will be needed to clarify how delays between subcortical and cortical sources may influence frequency content of measured signals; to explore the precise location and the frequency range over which it operates, and its function (if any); and to establish how this mechanism arises as a cumulative product of circuits and even cells with resonant properties at or between different subcortical and cortical levels, perhaps through intracranial recordings and computational modeling (Tichko

and Skoe, 2017; Lerud et al., 2019). This work may lead to a better understanding of neural signal processing principles underlying both low-level behavioral properties such as pitch discrimination, as well as higher-order cognitive mechanisms of speech and music perception.

References

- Anderson S, Skoe E, Chandrasekaran B, Zecker S, Kraus N (2010) Brainstem correlates of speech-in-noise perception in children. *Hear Res* 270:151–157.
- Arnal LH, Giraud A-L (2012) Cortical oscillations and sensory predictions. *Trends Cogn Sci* 16:390–398.
- Baillet S (2017) Magnetoencephalography for brain electrophysiology and imaging. *Nat Neurosci* 20:327–339.
- Baillet S, Mosher JC, Leahy RM (2001) Electromagnetic brain mapping. *IEEE Signal Process Mag* 18:14–30.
- Baltus A, Herrmann CS (2015) Auditory temporal resolution is linked to resonance frequency of the auditory cortex. *Int J Psychophysiol* 98:1–7.
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29:1165–1188.
- Bharadwaj HM, Masud S, Mehraei G, Verhulst S, Shinn-Cunningham BG (2015) Individual differences reveal correlates of hidden hearing deficits. *J Neurosci* 35:2161–2172.
- Bidelman GM (2015) Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient abr. *Hear Res* 323:68–80.
- Bidelman GM (2018) Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage* 175:56–69.
- Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. *Science* 304:1926–1929.
- Chao ZC, Takaura K, Wang L, Fujii N, Dehaene S (2018) Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. *Neuron* 100:1252–1266.
- Coffey EB, Herholz SC, Chapesiuk AM, Baillet S, Zatorre RJ (2016) Cortical contributions to the auditory frequency-following response revealed by meg. *Nat Commun* 7:11070.
- Coffey EB, Chapesiuk AM, Herholz SC, Baillet S, Zatorre RJ (2017) Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Front Neurosci* 11:479.
- Coffey EB, Nicol T, White-Schwoch T, Chandrasekaran B, Krizman J, Skoe E, Zatorre RJ, Kraus N (2019) Evolving perspectives on the sources of the frequency-following response. *Nat Commun* 10:5036.
- Cooke JE, Lee JJ, Bartlett EL, Wang X, Bendor D (2020) Post-stimulatory activity in primate auditory cortex evoked by sensory stimulation during passive listening. *Sci Rep* 10:13885.
- Destrieux C, Fischl B, Dale A, Hagren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53:1–15.
- Fischl B (2012) Freesurfer. *Neuroimage* 62:774–781.
- Freyman RL, Nelson DA (1986) Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners. *J Acoust Soc Am* 79:1034–1044.
- Gao L, Kostlan K, Wang Y, Wang X (2016) Distinct subthreshold mechanisms underlying rate-coding principles in primate auditory cortex. *Neuron* 91:905–919.
- Gardi J, Merzenich M, McKean C (1979) Origins of the scalp-recorded frequency-following response in the cat. *Int J Audiol* 18:353–380.
- Giraud A-L, Arnal LH (2018) Hierarchical predictive information is channeled by asymmetric oscillatory activity. *Neuron* 100:1022–1024.
- Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15:511–517.
- Gorina-Careta N, Kurkela JL, Hämäläinen J, Astikainen P, Escera C (2021) Neural generators of the frequency-following response elicited to stimuli of low and high frequency: a magnetoencephalographic (meg) study. *Neuroimage* 231:117866.
- Gourévitch B, Martin C, Postal O, Eggermont JJ (2020) Oscillations in the auditory system and their possible role. *Neurosci Biobehav Rev* 113:507–528.
- Grissom RJ, Kim JJ (2012) Effect sizes for research: univariate and multivariate applications. New York, USA: Routledge.
- Gross J, Baillet S, Barnes GR, Henson RN, Hillebrand A, Jensen O, Jerbi K, Litvak V, Maess B, Oostenveld R, Parkkonen L, Taylor JR, van Wassenhove V, Wibral M, Schoffelen J-M (2013) Good practice for conducting and reporting meg research. *Neuroimage* 65:349–363.
- Haegens S, Zion Golumbic E (2018) Rhythmic facilitation of sensory processing: a critical review. *Neurosci Biobehav Rev* 86:150–165.
- Hämäläinen M (2009) Mne software user's guide v2. 7. Charlestown, MA: Athinoula A. Martinos Center for Biomedical Imaging Massachusetts General Hospital.
- Hartmann T, Weisz N (2019) Auditory cortical generators of the frequency following response are modulated by intermodal attention. *Neuroimage* 203:116185.
- Irvine D (1986) The auditory brainstem: a review of the structure and function of auditory processing mechanisms. In: *Progress in sensory physiology* (Autrum H, Ottoson D, Perl E, Schmidt R, Shimazu H, Willis W, eds), pp 71–279. Berlin: Springer.
- Kösem A, Bosker HR, Takashima A, Meyer A, Jensen O, Hagoort P (2018) Neural entrainment determines the words we hear. *Curr Biol* 28:2867–2875.
- Kraus N, Anderson S, White-Schwoch T (2017) The frequency-following response: a window into human communication. In: *The frequency-following response* (Kraus N, Anderson S, White-Schwoch T, Fay RR, Popper AN, eds), pp 1–15. Cham, Switzerland: Springer.
- Krizman J, Kraus N (2019) Analyzing the ffr: a tutorial for decoding the richness of auditory function. *Hear Res* 382:107779.
- Lakatos P, Musacchia G, O'Connell MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77:750–761.
- Lee KM, Skoe E, Kraus N, Ashley R (2015) Neural transformation of dissonant intervals in the auditory brainstem. *Music Percept* 32:445–459.
- Lerousseau JP, Trébuchon A, Morillon B, Schon D (2019) Persistent neural entrainment in the human cortex is frequency selective. *bioRxiv* 834226. doi: 10.1101/834226.
- Lerud KD, Kim JC, Almonte FV, Carney LH, Large EW (2019) A canonical oscillator model of cochlear dynamics. *Hear Res* 380:100–107.
- Li X, Jeng F-C (2011) Noise tolerance in human frequency-following responses to voice pitch. *J Acoust Soc Am* 129:EL21–EL26.
- Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J, Chauvel P (2004) Temporal envelope processing in the human left and right auditory cortices. *Cereb Cortex* 14:731–740.
- Marmel F, Linley D, Carlyon R, Gockel H, Hopkins K, Plack C (2013) Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently. *J Assoc Res Otolaryngol* 14:757–766.
- Mathewson KE, Prudhomme C, Fabiani M, Beck DM, Lleras A, Gratton G (2012) Making waves in the stream of consciousness: entraining oscillations in EEG alpha and fluctuations in visual awareness with rhythmic visual stimulation. *J Cogn Neurosci* 24:2321–2333.
- Micheyl C, Delhommeau K, Perrot X, Oxenham AJ (2006) Influence of musical and psychoacoustical training on pitch discrimination. *Hear Res* 219:36–47.
- Micheyl C, Xiao L, Oxenham AJ (2012) Characterizing the dependence of pure-tone frequency difference limens on frequency, duration, and level. *Hear Res* 292:1–13.
- Musacchia G, Strait D, Kraus N (2008) Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear Res* 241:34–42.
- Neymotin SA, Barczak A, O'Connell MN, McGinnis T, Markowitz N, Espinal E, Griffith E, Anwar H, Dura-Bernal S, Lytton WW, Jones SR, Bickel S, Lakatos P (2020) Taxonomy of neural oscillation events in primate auditory cortex. *bioRxiv*. doi: 10.1101/2020.04.16.045021.
- Notbohm A, Kurths J, Herrmann CS (2016) Modification of brain oscillations via rhythmic light stimulation provides evidence for entrainment but not for superposition of event-related responses. *Front Hum Neurosci* 10:10.
- Nozaradan S, Peretz I, Missal M, Mouraux A (2011) Tagging the neuronal entrainment to beat and meter. *J Neurosci* 31:10234–10240.
- Obleser J, Kayser C (2019) Neural entrainment and attentional selection in the listening brain. *Trends Cogn Sci* 23:913–926.
- Presacco A, Simon JZ, Anderson S (2019) Speech-in-noise representation in the aging midbrain and cortex: effects of hearing loss. *PLoS One* 14:e0213899.

- Robinson K, Patterson RD (1995) The stimulus duration required to identify vowels, their octave, and their pitch chroma. *J Acoust Soc Am* 98:1858–1865.
- Ross B, Tremblay KL, Alain C (2020) Simultaneous eeg and meg recordings reveal vocal pitch elicited cortical gamma oscillations in young and older adults. *Neuroimage* 204:116253.
- Samiee S, Lévesque M, Avoli M, Baillet S (2018) Phase-amplitude coupling and epileptogenesis in an animal model of mesial temporal lobe epilepsy. *Neurobiol Dis* 114:111–119.
- Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci* 32:9–18.
- Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci* 12:106–113.
- Skoe E, Kraus N (2010) Auditory brainstem response to complex sounds: a tutorial. *Ear Hear* 31:302–324.
- Song JH, Skoe E, Banai K, Kraus N (2011) Perception of speech in noise: neural correlates. *J Cogn Neurosci* 23:2268–2279.
- Spaak E, de Lange FP, Jensen O (2014) Local entrainment of alpha oscillations by visual stimuli causes cyclic modulation of perception. *J Neurosci* 34:3536–3544.
- Steinschneider M, Arezzo J, Vaughan HG Jr (1980) Phase-locked cortical responses to a human speech sound and low-frequency tones in the monkey. *Brain Res* 198:75–84.
- Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM (2011) Brainstorm: a user-friendly application for meg/eeg analysis. *Comput Intell Neurosci* 2011:879716.
- Thut G, Miniussi C, Gross J (2012) The functional importance of rhythmic activity in the brain. *Curr Biol* 22:R658–R663.
- Tichko P, Skoe E (2017) Frequency-dependent fine structure in the frequency-following response: the byproduct of multiple generators. *Hear Res* 348:1–15.
- Wickesberg RE, Stevens HE (1998) Responses of auditory nerve fibers to trains of clicks. *J Acoust Soc Am* 103:1990–1999.
- Xu Q, Ye D (2015) Temporal integration reflected by frequency following response in auditory brainstem. *BME* 26:S767–S778.
- Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron* 77:980–991.