Neural Correlates of Early Sound Encoding and their Relationship to Speech-in-Noise Perception

Emily B. J. Coffey [1,2,3*], Alexander M. P. Chepesiuk [1], Sibylle C. Herholz [1,2,3,4], Sylvain Baillet [3,5] and Robert J. Zatorre [1,2,3*]

[1] Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montréal, QC, Canada, [2] Laboratory for Brain, Music and Sound Research, Montréal, QC, Canada, [3] Centre for Research on Brain, Language and Music, Montréal, QC, Canada, [4] German Center for Neurodegenerative Diseases, Bonn, Germany, [5] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montréal, QC, Canada

Speech-in-noise (SIN) perception is a complex cognitive skill that affects social, vocational, and educational activities. Poor SIN ability particularly affects young and elderly populations, yet varies considerably even among healthy young adults with normal hearing. Although SIN skills are known to be influenced by top-down processes that can selectively enhance lower-level sound representations, the complementary role of feed-forward mechanisms and their relationship to musical training is poorly understood. Using a paradigm that minimizes the main top-down factors that have been implicated in SIN performance such as working memory, we aimed to better understand how robust encoding of periodicity in the auditory system (as measured by the frequency-following response) contributes to SIN perception. Using magnetoencephalography, we found that the strength of encoding at the fundamental frequency in the brainstem, thalamus, and cortex is correlated with SIN accuracy. The amplitude of the slower cortical P2 wave was previously also shown to be related to SIN accuracy and FFR strength; we use MEG source localization to show that the P2 wave originates in a temporal region anterior to that of the cortical FFR. We also confirm that the observed enhancements were related to the extent and timing of musicianship. These results are consistent with the hypothesis that basic feed-forward sound encoding affects SIN perception by providing better information to later processing stages, and that modifying this process may be one mechanism through which musical training might enhance the auditory networks that subserve both musical and language functions.

Keywords: frequency-following response, speech-in-noise, magnetoencephalography, electroencephalography, neuroplasticity, auditory perception, inter-individual variability, musical training

INTRODUCTION

Understanding the neural bases of good speech-in-noise (SIN) perception during development, adulthood, and into old age is both clinically and scientifically important. However, it is challenging due to the complexity of the skill, which can be considered as a special case of auditory scene analysis, and can involve multiple cognitive processes depending on the information that is offered including spatial location, spectral and temporal regularity, and modulation (Moore and Gockel, 2002; Pressnitzer et al., 2011), and can be aided by visual cues (Suied et al., 2009), by predictions formed with the motor system (Du et al., 2014), and based on prior knowledge such as of language

(Pickering and Garrod, 2007; Golestani et al., 2009) that can be used to constrain the interpretation of noisy information (Bendixen, 2014). The contribution of the fidelity with which an individual encodes various sound properties, including periodicity, which varies according to a variety of life experiences (e.g., musical experience Anderson et al., 2013a) is not yet clear.

One means of observing the inter-individual differences in how people encode periodic characteristics of sound is the frequency-following response (FFR), an evoked response that is an index of the temporal representation of periodic sound in the brainstem (Chandrasekaran and Kraus, 2010; Skoe and Kraus, 2010), thalamus, and auditory cortex (Coffey et al., 2016b,c). Differences in the strength and fidelity of the fundamental frequency (f0) of the FFR have been linked to SIN perception such that increased FFR amplitude is associated with better performance (reviewed in Du et al., 2011, see also Parbery-Clark et al., 2009a; Anderson et al., 2012). However, enhancements and deficits of neural correlates that are related to SIN perception are most consistently observed either when the FFR is measured in very challenging listening conditions (e.g., Parbery-Clark et al., 2009b), in the degree of degradation of the FFR signal between quiet and noisy conditions (e.g., Cunningham et al., 2001; Parbery-Clark et al., 2011a; Song et al., 2011), or in the magnitude of enhancement that is conferred by predictability within a sound stream (e.g., Parbery-Clark et al., 2011c). f0 representation in the FFR may be enhanced by training (Song et al., 2008, 2012) and is often observed to be stronger among musicians even to sounds presented in silence (e.g., Musacchia et al., 2007), suggesting that learning mechanisms related to identifying task-relevant features and possibly attention might act to bias and enhance incoming acoustic information and suppress noise (Suga, 2012).

However, a clear picture has not yet emerged (Coffey et al., 2017). It is unclear if the sometimes-observed relationship between the FFR and SIN is due *only* to better top-down mechanisms such as better stream segregation (Başkent and Gaudrain, 2016) or selective auditory attention (Parbery-Clark et al., 2011b; Song et al., 2011; Lehmann and Schönwiesner, 2014), or if enhanced feed-forward stimulus encoding also plays a role. Here, we aimed to better understand the neural bases of periodicity coding in the brain under optimal listening conditions to understand its relevance to SIN; if basic encoding of sound quality in silence is important for more complex tasks such as hearing in noise, then we predict that there should be a relationship between FFR measured in silence and SIN performance. A secondary question we address is whether this relationship might be enhanced by musicianship.

Musicians are thought to have both enhanced bottom-up (Musacchia et al., 2007; Bidelman and Weiss, 2014) and top-down (Strait et al., 2010; Kraus et al., 2012) processing of sound. Because SIN perception and measures of basic sound encoding are related to musicianship, musical training has been proposed as a means of ameliorating poor SIN performance (reviewed in: Alain et al., 2014). Musical training places high demands on sensory, motor, and cognitive processing mechanisms that overlap between music and speech perception, and offers extensive repetition and emotional reward, which could stimulate auditory system enhancements that in turn impact speech

processing (Patel, 2014). Several longitudinal studies support a causal relationship between musical training and SIN skills (Tierney et al., 2013; Kraus et al., 2014; Slater et al., 2015), although it is difficult to maintain full, experimental control over naturalistic training studies (Evans et al., 2014). A number of cross-sectional studies have also reported a musician advantage in SIN perception (Parbery-Clark et al., 2009a,b, 2011b, 2012b; Strait et al., 2012; Zendel and Alain, 2012; Swaminathan et al., 2015); however, other studies have not found significant group differences (Ruggles et al., 2014; Boebinger et al., 2015) or have found the musicianship effect to be dependent upon the specific SIN task variations, such as the degree of information masking (Swaminathan et al., 2015; Başkent and Gaudrain, 2016) or the degree of reliance on pitch cues (Fuller et al., 2014). A recent review of SIN perception among musicians concluded that on balance there is good evidence for musician enhancement of SIN, but also highlighted the diversity of study designs used to study hearing in noisy conditions, which may contribute to inconsistent findings (Coffey et al., 2017). However, it remains uncertain to which aspects of cognition any musician advantage is owed: top-down processes such as selective attention and working memory that modulate early levels (Rinne et al., 2008) to filter and temporarily store incoming information (Strait and Kraus, 2011; Kraus et al., 2012), relatively immutable factors such as non-verbal IQ (Boebinger et al., 2015) that might affect multiple cognitive processes, or differences in basic sound encoding (reviewed in Anderson and Kraus, 2010b; Du et al., 2011; Alain et al., 2014, see also Weiss and Bidelman, 2015).

In the present study, we first aimed to clarify whether robust f0 encoding in the auditory system, which is known to be enhanced in musicians (Musacchia et al., 2007; Bidelman et al., 2011a,b), influences SIN perception in a feed-forward fashion. Rather than relating fundamental encoding recorded in the presence of noise to later performance (which has previously been shown, described above) and might include influences from top-down processes that spontaneously act to separate speech and noise streams, here we reduce the similarity between the conditions of the electrophysiological recording and the offline SIN behavioral task to a single overlapping feature: the presence of pitch-related information. Although several studies have not found a significant relationship between FFR-f0 measured in conditions of silence and SIN performance (e.g., Parbery-Clark et al., 2009a), such relationships may be obscured by EEG-based FFR recordings which likely blend responses coming from different sources (Zhang and Gong, 2016; Tichko and Skoe, 2017). Despite the relative insensitivity of magnetoencephalography (MEG) to deep sources (which approximate radial sources, Baillet et al., 2001), sufficient information is preserved in the MEG signal for accurate localization of deeper structures such as the hippocampus, amygdala and thalamus (Attal and Schwartz, 2013; Dumas et al., 2013), and for the contributions from subcortical and cortical FFR generator sites to be separated (Coffey et al., 2016b), which may increase the sensitivity of the experimental design to behavioral relationships. Therefore, an additional novel aspect of the present study is to use MEG to determine how FFR signals coming from distinct anatomical structures may be contributing to the putative relationship to SIN performance. We

thus extend previous investigations of the anatomical origins of FFR signals to their behavioral meaning in the context of SIN perception.

If enhanced encoding is partly responsible for better auditory skills because a better quality signal is encoded from incoming sound and passed to higher-order cognitive processes and networks (Irvine, 1986; Musacchia et al., 2008), we would expect that the relationship between SIN and sound encoding would persist even under optimal listening conditions when the system is not challenged, and when the listener's attention is otherwise engaged. We therefore first measured SIN perception behaviorally, then in a separate session simultaneously recorded EEG and MEG data while listeners were presented with a speech sound in quiet as they watched a silent film. Secondarily, we evaluate correlations of measures of musical experience with FFR-f0 and SIN within our sample to evaluate the possible influence of musicianship on performance.

In addition to FFR-f0, which is derived from the higher-frequency EEG activity (Skoe and Kraus, 2010), other lower-frequency cortical potential measures covary with SIN performance, in particular the ERP P2 component (~200 ms post stimulus onset; Cunningham et al., 2001), which is also known to be related to speech processing and is sensitive to training effects (Key et al., 2005; Musacchia et al., 2008; Bidelman and Weiss, 2014; Tremblay et al., 2014). If the two signals represent sequential processes in the same processing stream, we would expect enhancements in FFR-f0 to be paralleled by enhancements in the strength of the ERP P2 component, and for each of these measures to be related to SIN accuracy. To test this hypothesis, we used distributed source modeling of the magnetic signals to localize the neural origins of the MEG FFR-f0 and the P2, and examined their spatial and statistical relationships to each other (as well as spatial relationships to preceding and following ERP components) for the first time in order to explore how these signals may be related. Collectively, these data should help us to understand the neural basis of inter-individual differences in sound encoding and its effects on the important real-world function of SIN perception.

METHODS AND MATERIALS

The experimental procedures concerning the MEG and (single channel, Cz) EEG recordings of the brain's response to the speech syllable /da/, and much of the pre-processing, have previously been reported in the context of determining their neural origins and will be discussed only briefly here (please see Coffey et al., 2016b "Methods" for details). The correlations between FFR-f0 strength and musicianship that are included in the summary of musical enhancements in **Table 1** have been reported in Coffey et al.; all other findings have not been reported previously. Behavioral testing took place in a sound-attenuated room on different day prior to the MEG recording session.

Participants

Data from the same 20 neurologically healthy young adults included in the previous study (Coffey et al., 2016b) were included in this study (mean age: 25.7 years; $SD = 4.2$; 12

TABLE 1 | Summary of evidence for musicianship-related behavioral and neurophysiological enhancements ($N = 12$).

Measure	Age of start	Practice hours
SIN	rs = −0.70, p = 0.006*	rs = 0.39, p = 0.10
Fine pitch discrimination	rs = 0.45, p = 0.07	rs = −0.67, p = 0.008*
FFR-f0 (right AC)	rs = −0.53, p = 0.05*	rs = 0.57, p = 0.04*
P2 amplitude	rs = −0.21, p = 0.25	rs = 0.59, p = 0.05*

Asterisks (*) indicate significant rank correlations (alpha < 0.05, one tailed). In general, earlier start ages and a larger number of practice hours are associated with enhancements, suggesting an influence of musical training. Note that lower fine pitch discrimination scores indicate better performance; therefore correlations in opposite directions are expected.

female; all were right-handed and had normal or corrected-to-normal vision; $< = 25$ dB hearing level thresholds for frequencies between 500 and 4,000 Hz assessed by pure-tone audiometry; and no history of neurological disorders). All but three subjects were native English speakers; the other three (one Korean, two French speakers) were highly proficient in English and all scored within the range of the native speakers on the HINT task, thus ruling out that any of our findings were due to second-language effects. Informed consent was obtained and all experimental procedures were approved by the Montreal Neurological Institute Research Ethics Board.

Speech-in-Noise Assessment

SIN was measured using a custom computerized implementation of the hearing in noise test (HINT; Nilsson, 1994) that allowed us to obtain a relative measure of SIN ability using a portable computer, without specialized equipment. In the standard HINT task, speech-spectrum noise is presented at a fixed level and sentences are varied in a staircase procedure to obtain a (single-value) SIN perceptual threshold (Nilsson, 1994). Our modified HINT task used a subset of the same sentence lists (Bench et al., 1979) and speech-spectrum noise, but presented thirty sentences in three empirically determined difficulty levels in randomized order: easy (2 dB SNR; i.e., target speech was 2 dB louder than noise), medium (−2 dB SNR), and difficult (−6 dB SNR). The sentences and noise were combined using sound processing software (Audacity, version 1.3.14-beta, http://audacity.sourceforge.net/; 44100Hz sampling frequency). Stimuli were presented diotically (i.e., identical speech and masker in each ear) via headphones (JVC HA-M5X) with the noise adjusted to a loud but not uncomfortable sound level on pilot subjects (~75 db SPL) and thereafter held constant. No verbal or visual feedback was given. A single overall accuracy score as the proportion of sentences correctly repeated back to experimenter was calculated by averaging the accuracy across all three levels; however, the score distributions showed a clear ceiling effect in the easiest level, with 12 out of 20 participants scoring over 95% (mean of subset of easy trials: 93.5%, $SD = 7.5$; medium trials: 79.9%, $SD = 14.0$; hard trials: 32.6%, $SD = 12.1$). We therefore excluded the easiest trials from the mean accuracy score in order to obtain a cleaner estimate of inter-individual variability; control analyses were conducted for the main research questions by assessing correlations between SIN scores at each level of

difficulty and the FFR-f0 strength to ensure that the pattern of results was robust to this exclusion.

Fine Pitch Discrimination

Fine pitch discrimination thresholds were measured as described in Coffey et al. (2016b), using a two-interval forced-choice task and a two-down one-up rule to estimate the threshold at 79% correct point on the psychometric curve (Levitt, 1971). The reference tone, which was presented once per trial, had a frequency of 500 Hz. The adaptive procedure was stopped after 15 reversals and the geometric mean of the last eight trials was recorded. Thresholds were derived from the average of five task repetitions.

Stimulus Presentation

The stimulus for the MEG/EEG recordings was a 120-ms synthesized speech syllable (/da/) with a fundamental frequency in the sustained vowel portion of 98 Hz. The stimulus was presented binaurally at 80 dB SPL, ∼14,000 times in alternating polarity, through Etymotic ER-3A insert earphones with foam tips (Etymotic Research). For five subjects, ∼11,000 epochs were collected due to time constraints. Stimulus onset synchrony (SOA) was randomly selected between 195 and 205 ms from a normal distribution. A separate run was collected of ∼600 stimulus repetitions spaced ∼500 ms apart, to record later waves of the slower cortical responses. 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 in order 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, averaged mastoid references), EOG and ECG, and one audio channel were simultaneously acquired using a CTF MEG System and its in-built EEG system (Omega 275, CTF Systems Inc.). All data were sampled at 12 kHz. Data preprocessing was performed with Brainstorm Tadel et al. (2011) and using custom Matlab scripts (The Mathworks Inc., MA, USA) as described in Coffey et al. (2016b), and in brief, below.

FFR Correlates of SIN Accuracy

FFR-f0 strength was extracted from regions of interest (ROIs) in the auditory system (AC: auditory cortex, MGB: medial geniculate body of the thalamus, IC: inferior colliculus and CN: cochlear nucleus) using the MEG distributed source modeling approach described previously (for the specifications of each ROI, please see Methods in Coffey et al., 2016b). Using this approach, the amplitude 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 anatomical MRI scan (Baillet et al., 2001; Gross et al., 2013), from which cortical sources and subcortical structures

were prepared using FreeSurfer (Fischl, 2012). As reported in Coffey et al., anatomical data were imported into Brainstorm (Tadel et al., 2011), 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 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 literature (Hämäläinen, 2009). Source models for each run were averaged within subject. We extracted a timeseries of mean amplitude for each ROI and for each of the three orientations in the unconstrained orientation source model, from which three spectra 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 neurological response at the fundamental frequency, which was detected by an automatic script; this is referred hereafter as the FFR-f0 strength.

We first evaluated correlations between SIN accuracy scores and FFR-f0 strength averaged across bilateral pairs of structures, using Spearman's rho (rs; one-tailed). Non-parametric statistics were used throughout as FFR-f0 measures were generally not normally distributed (using Shapiro-Wilk's parametric hypothesis test of composite normality, the null hypothesis was rejected for AC, CN, and IC bilateral averages), and one-tailed tests were used as our goal was to test the specific hypothesis that higher amplitude FFR-would be related only to better behavioral performance, as stronger or less degraded FFRs in the presence of noise have been reported consistently in the EEG literature (Cunningham et al., 2001; Parbery-Clark et al., 2009a,b, 2012a; Song et al., 2011), see also (Du et al., 2011). To correct for multiple comparisons, the false discovery rate (FDR) was controlled at δ = 0.05 where tests on multiple ROIs are used (Benjamini and Hochberg, 1995). The EEG equivalent of the FFR-f0 was also computed for comparison of sensitivity to behavioral measures. Correlations were computed between SIN accuracy and the left and right auditory cortex ROIs separately, as a lateralization effect in FFR-f0 strength and its relationship to measures of musicianship and fine pitch discrimination had been observed previously (see Figure 5c–e in Coffey et al., 2016b). We tested for a stronger correlation on the right than left side using Fisher's r-to-Z transformation (one-tailed, alpha= 0.05).

Later Cortical Evoked Responses

Event-related potentials (ERPs) within the 2–40 Hz band-pass filtered single-channel EEG data were obtained in order to establish a connection between previous FFR-ERP research that showed SIN sensitivity at ERP components P2 and N2 (Cunningham et al., 2001; Parbery-Clark et al., 2011a) and the MEG data. In order to maximize the interpretability of the results with respect to a large body of work that has used EEG-based measures of P2 amplitude (which may not be entirely equivalent to their magnetic counterparts due to differences in each technique's sensitivity to source orientation), the primary measure of ERP amplitude is based on EEG rather than MEG; MEG sources were localized for the same time period.

Although, SIN perception has also been found to be correlated with N1 latency and amplitude measures (e.g., Parbery-Clark et al., 2011a; Billings et al., 2013; Bidelman and Howell, 2016), N1 is strongly affected by the characteristics of the stimulus and its stimulation paradigm (Billings et al., 2011). In this paradigm, and using a single-EEG channel positioned at the vertex (Cz), we did not observe a clear N1 nor N2 from all subjects. We therefore took the amplitude of only P2 as a measure; this simpler metric occurs at a single time point and also allowed for a more straightforward comparison to and interpretation of the MEG equivalent. A researcher who was blinded to the subjects' FFR-f0 amplitudes and behavioral results at the time of measurement selected P2 wave peaks individually on ERP waves averaged across epochs for each subject (cortically processed; 2–40 Hz with −50 to 0 ms DC baseline correction; P2 was considered to be the strongest positive deflection within a ∼40 ms window centered on the group grand average P2 at 183 ms). Amplitudes of these custom peaks were then correlated with SIN accuracy and FFR-f0 strength.

MEG evoked response fields (ERFs) on simultaneously recorded data were obtained in order to extend this work using distributed source modeling. The EEG cortical evoked response complex (ERP) elicited by the speech syllable /da/ consists of two positive waves at about 50–90 ms ("P1") and between 170 and 200 ms ("P2" or "P1 prime") and two negative waves at about 110 ms ("N1") and after 200 ms ("N2" or "N1 prime") (reviewed in Key et al., 2005); see also Cunningham et al., Figure 6 (Cunningham et al., 2001) and Musacchia et al., Figure 2 (Musacchia et al., 2008). For the purposes of this study we identified wave peaks in the ERP and ERF average at the group level for the SIN-sensitive P2 peak (183 ms), and at the earlier P1 component that has a well-known physiological origin in order to confirm the quality of data and validity of the analysis (60 ms; see **Figures 3A,C**).

Origins of Later Cortical ERP Components

To confirm that the MEG data could be used to localize areas that showed above-baseline activity at the group level, and to observe the origins of the SIN-sensitive P2 wave in relation to preceding and following ERP waves, we first computed cortical volume MNE models based on each subject's T1-weighted MRI scan in which the orientation of sources was unconstrained, but their location was constrained within the volume encompassed by the cortical surface. These models were normalized to the baseline period (−50 to 0 ms). We exported 10 ms time windows

around each peak of interest (mean-rectified signal amplitude) and for the baseline (−50 to 0 ms) for statistical analysis in the neuroimaging software package FSL (Smith et al., 2004; Jenkinson et al., 2012). These source volume maps were co-registered to the subject's high-resolution T1 anatomical MRI scan (FLIRT, 6 parameter linear transformation), and then to the 2 mm MNI152 template (12 parameter linear transformation, Evans et al., 2012). Normalized difference images were created by subtracting the baseline images from those of the peaks of interest and calculating z-scores within each image (P1 > Baseline, P2 > Baseline). Permutation testing was used to reveal locations where the magnetic signal was greater during peaks of interest as compared with baseline [non-parametric one-sample t-test (Winkler et al., 2014); 10,000 permutations]. The family-wise error rate was controlled using threshold-free cluster enhancement as implemented in FSL ($p < 0.01$), after applying a cortical mask of the MNI 152 template with the brainstem and cerebellum removed (these latter structures were not included in the MEG source model).

Comodulation of Low and High Frequency Activity

We considered the spatial relationship between FFR-f0 generators and the source of the SIN-sensitive P2 wave by inspecting the FFR-f0 > Baseline and P2 > Baseline maps in the MEG data, and calculated Spearman's correlations between the FFR-f0 strength from each auditory cortex ROI (MEG) and the amplitude of the P2 wave measured with EEG.

Musicianship Enhancements

Twelve subjects reported varying levels of musical experience on a range of musical instruments [primary instruments: piano (6), guitar (2), flute (1), saxophone (1), trumpet (1), voice (1)], as obtained by self-report using the Montreal Music History Questionnaire (Coffey et al., 2011). Start ages ranged from 5 to 12 years, and total cumulative practice hours ranged from 1,000 to 16,000 h. We assessed correlations between SIN accuracy and total music practice hours and age of training start. We then evaluated the relationship between P2 amplitude in the EEG recording and musicianship, and between fine pitch discrimination skills and musicianship.
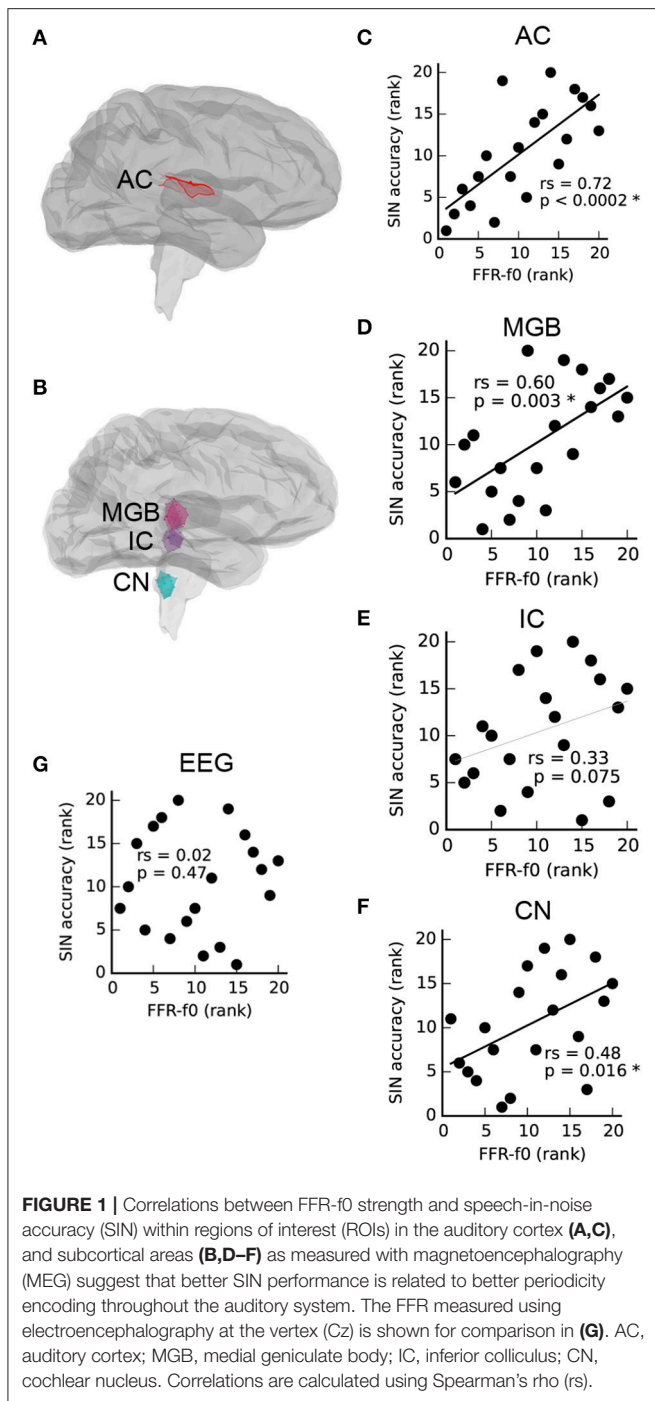
RESULTS

Behavioral Scores

The mean averaged SIN score was 56.3% ($SD = 12.0$). Subjects with finer pitch discrimination ability had statistically better SIN accuracy (one-tailed rs $= -0.47$, $p = 0.018$).

MEG FFR-f0 Strength Is Related to SIN throughout the Auditory System

As described by Coffey et al. (2016b), MEG is able to separate FFR activity arising from auditory cortex (**Figure 1A**), as well as brainstem and thalamus (**Figure 1B**). Relationships between FFR values measured via MEG from each ROI (averaged across left and right pairs) and SIN accuracy scores are presented in **Figures 1C–F**. A positive correlation between SIN accuracy and

FIGURE 1 | Correlations between FFR-f0 strength and speech-in-noise accuracy (SIN) within regions of interest (ROIs) in the auditory cortex **(A,C)**, and subcortical areas **(B,D–F)** as measured with magnetoencephalography (MEG) suggest that better SIN performance is related to better periodicity encoding throughout the auditory system. The FFR measured using electroencephalography at the vertex (Cz) is shown for comparison in **(G)**. AC, auditory cortex; MGB, medial geniculate body; IC, inferior colliculus; CN, cochlear nucleus. Correlations are calculated using Spearman's rho (rs).

FFR-f0 strength was found at each of the four structures tested, statistically significant (FDR-corrected for multiple comparisons) in all but the inferior colliculus, where a similar trend was nonetheless noted. We did not find evidence of a relationship between the EEG-derived FFR-f0 and SIN accuracy (**Figure 1G**), nor did a relationship appear with the inclusion of age as a covariate (rs = 0.08, $p = 0.37$).

To confirm that the precautionary exclusion of the easiest SIN trials in which a ceiling effect was found was inconsequential with

respect to the observed relationships between SIN and FFR-f0 strength, we recalculated the correlation between rAC FFR-f0 and SIN accuracy including all items (rs = 0.71, $p = 0.0003$; compare with the reported value with the exclusion, which is rs = 0.72, $p < 0.0002$). The general pattern of a positive correlation between rAC FFR-f0 and SIN accuracy was even replicated within the small subsets of easy (rs = 0.65, $p = 0.001$), medium (rs = 0.80, $p < 0.0001$), and hard items (rs = 0.57, $p = 0.004$), suggesting a robust relationship that is not highly sensitive to how the overall SIN accuracy score is calculated.

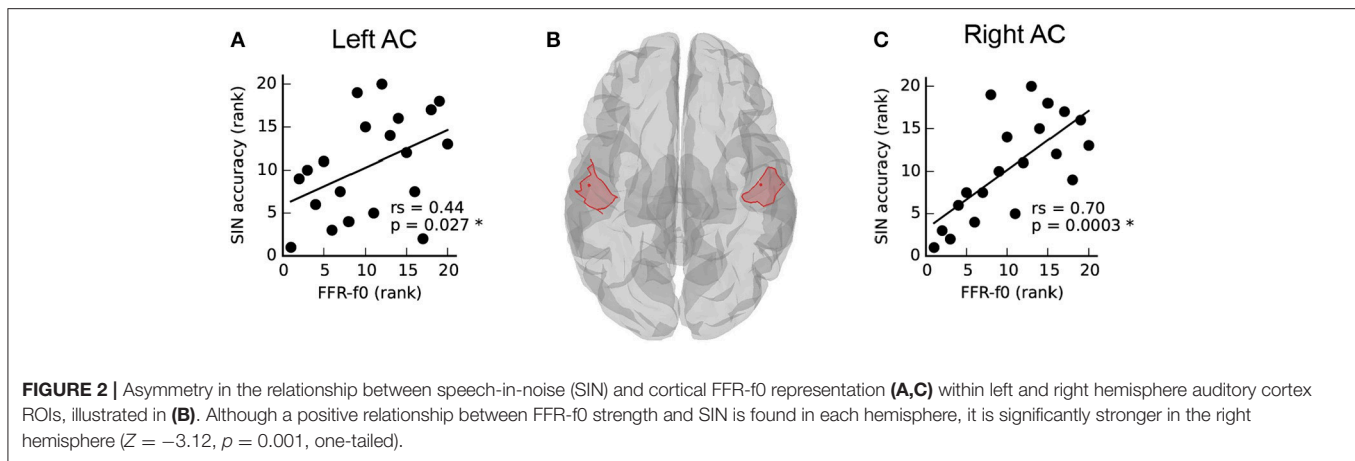## The Relationship between SIN and Cortical FFR-f0 Is Lateralized

The relationship between SIN and FFR-f0 strength from auditory cortical ROIs in each hemisphere is depicted in **Figure 2**. SIN accuracy was related to the strength of the FFR-f0 in both hemispheres, but was numerically larger on the right. We directly compared the strength of these correlations using Fisher's r-to-Z-transformation (one-tailed), and found it to be stronger in the right hemisphere ($Z = -3.12$, $p = 0.001$; the correlation between the FFR-f0 strength across two hemispheres, which is used for statistical comparison of correlation strength, was rs = 0.89).

Origins of Later Cortical ERP Components

We confirmed that the MEG data analysis used here is suitable for localizing temporal lobe auditory activity at the group level using P1, which was known to originate in the primary auditory areas bilaterally (**Figure 3D**). Note that while we had selected the earliest maximum in the P1 wave in the EEG signal in order to capture primary auditory cortex activity (mean latency: 60 ms), the peak energy in the MEG signal is slightly later (~15 ms); nonetheless, visual inspection of the same analysis performed on a 10 ms window centered on 75 ms indicates that this analysis is not sensitive to minor variations in P1 window selection. The mean latency of P2, the second prominent positive EEG wave, was 183 ms ($SD = 11$ ms), and its mean amplitude was 4.1 uV ($SD = 1.7$). We confirmed, as previously reported by Cunningham et al. (2001) using a pediatric sample, that P2 amplitude was related to SIN accuracy in the current sample (**Figure 3B**) thus providing a basis for further investigating FFR-f0 and P2 relationships. P1 amplitude was not related to SIN accuracy (rs = 0.11, $p = 0.33$). We then identified the sources of magnetic activity that was concurrent with the EEG-derived P2 wave, which proved to be relatively more anterior, and right-lateralized (**Figure 3D**; colored areas depict significant clusters corrected for multiple comparisons; maps are thresholded to best expose the areas of strongest signal).

Low and High Frequency Activity Covary

Right but not left AC FFR-f0 strength was significantly related to P2 amplitude (**Figures 4C,E**). For completeness, we also calculated the correlation between the EEG FFR-f0 and P2 amplitude but it was not significant: rs = 0.23, $p = 0.16$. The magnetic equivalent of the P2 wave overlapped considerably with the FFR-f0 regions using a corrected significance-based threshold of $p < 0.05$. However, inspection of the centroid of each map showed that whereas the FFR-f0 sources were distributed in the

FIGURE 2 | Asymmetry in the relationship between speech-in-noise (SIN) and cortical FFR-f0 representation **(A,C)** within left and right hemisphere auditory cortex ROIs, illustrated in **(B)**. Although a positive relationship between FFR-f0 strength and SIN is found in each hemisphere, it is significantly stronger in the right hemisphere ($Z = -3.12$, $p = 0.001$, one-tailed).

posterior section of the superior temporal gyrus, the P2 wave's foci were more anterior (**Figure 4D**).

Measures of Musicianship

Twelve out of 20 subjects reported some level of musical training. We previously showed that FFR-f0 strength in the right AC is related to hours of musical training and age of training start in this sample (Coffey et al., 2016b). Here, we first confirmed that the correlation between AC FFR-f0 strength and SIN scores that we observed within the whole group was present within both the musician subgroup (averaged AC ROIs: rs = 0.67, $p = 0.012$) and the non-musical subgroup (averaged AC ROIs: rs = 0.86, $p = 0.005$); these results represents an internal replication in independent groups and suggest that the sample size used in this study is sufficiently large to be sensitive to the statistical relationships of interest. Musical enhancements in behavioral and neurophysiological measures reported in this study are summarized in **Table 1**. Among those with musical experience, earlier start ages significantly correlated with better SIN scores, but the correlation between total practice hours and SIN showed only a non-significant trend. P2 wave amplitude correlated with practice hours, but not the age of start. We also found a significant correlation between hours of musical training and fine pitch discrimination ability, as expected, and a non-significant trend between start age and fine pitch discrimination ability.
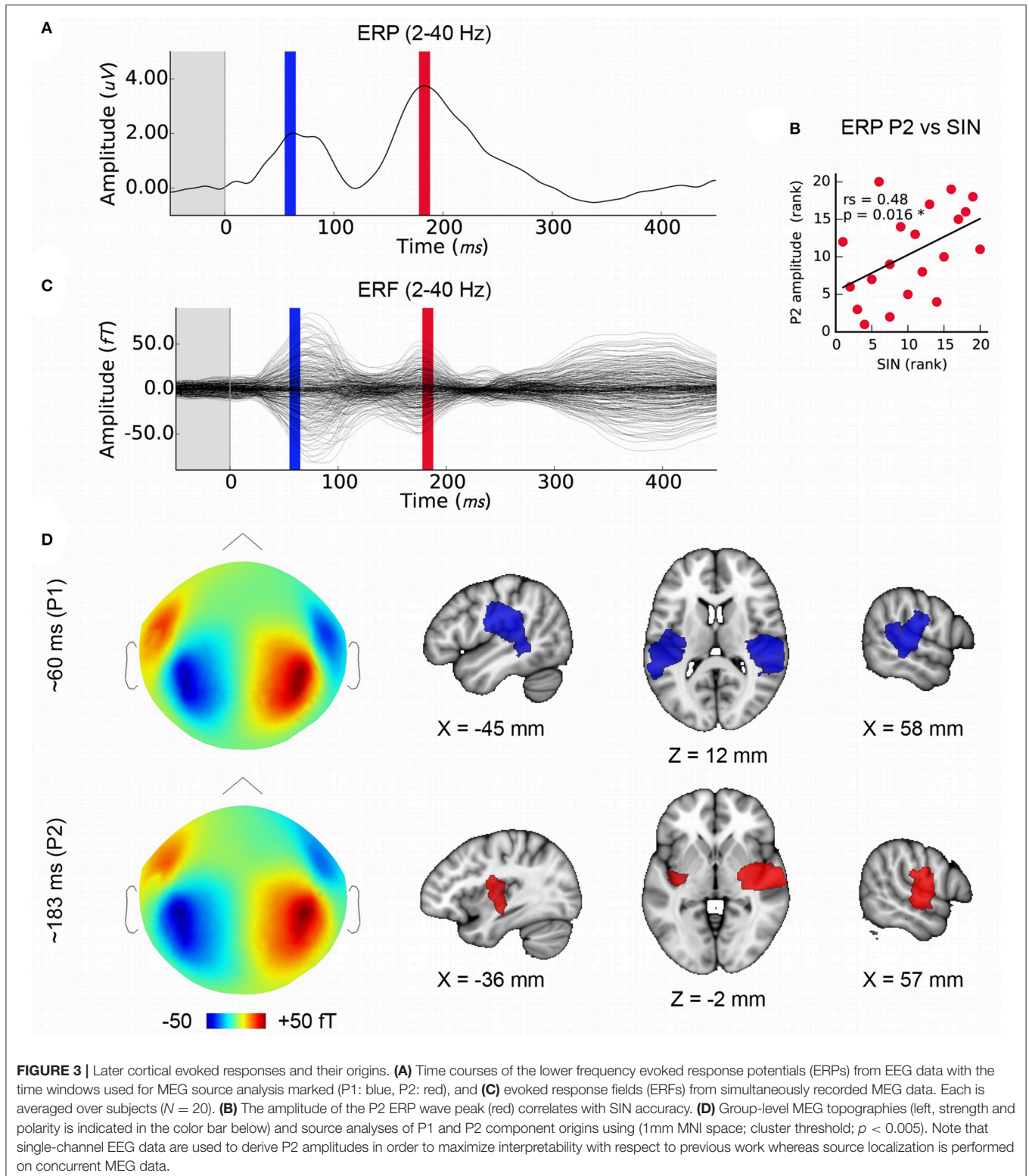
DISCUSSION

In this study, we aimed to clarify whether individual differences observed in fundamental frequency (f0) encoding in the auditory system of normal-hearing adults is related to SIN perception. Toward this end, we filtered the neural responses in two frequency bands in order to isolate the higher-frequency 98 Hz f0 within the FFR, and the lower-frequency cortical responses (2–40 Hz), and we compared their strength, spatial origins, and relationships to behavioral and musical experience measures.

We first showed that the strength of the MEG-based FFR-f0 attributed to structures throughout the ascending auditory neuraxis, including the auditory cortex in each hemisphere,
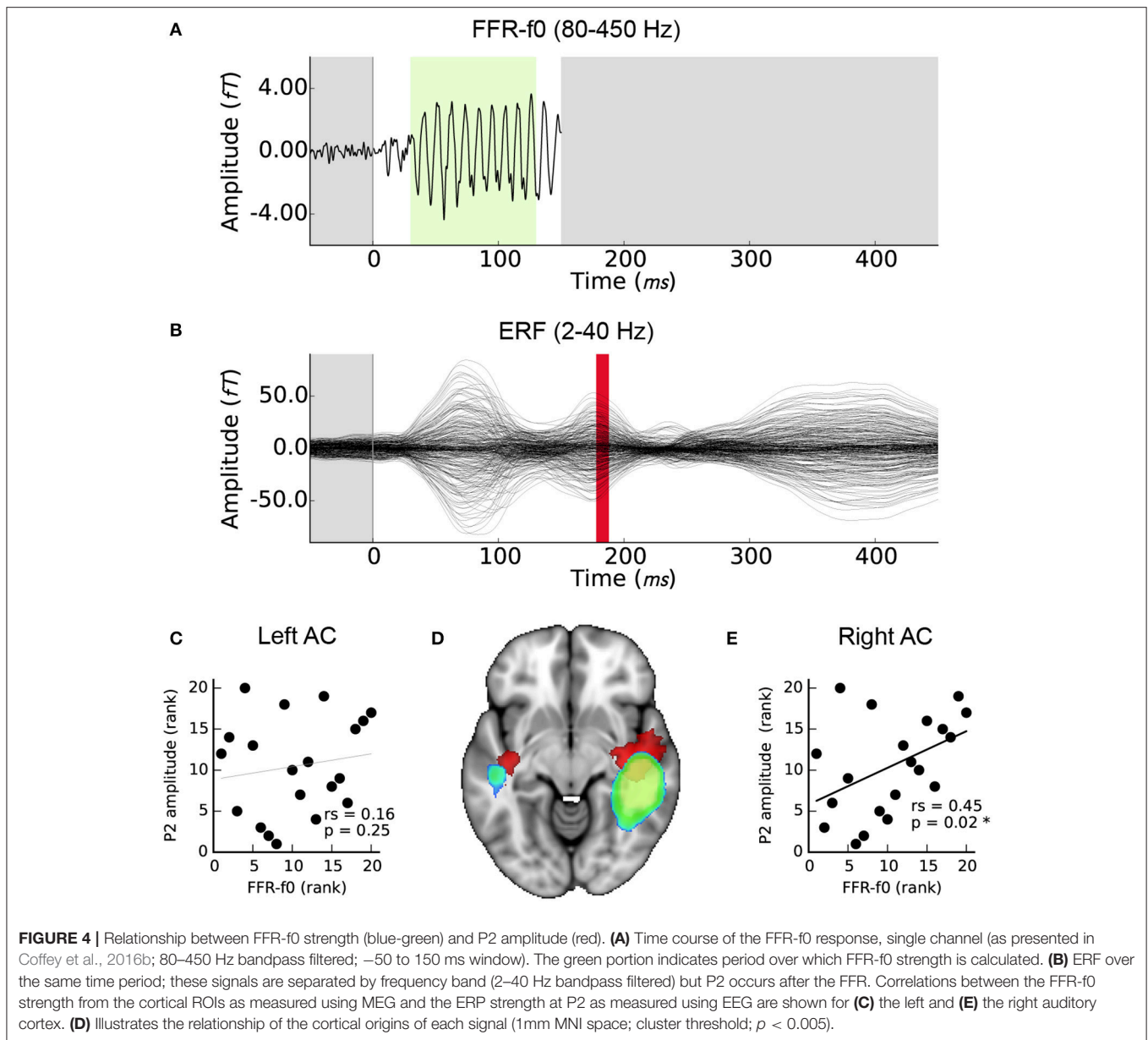
is positively correlated with SIN accuracy (**Figures 1C–F**), suggesting that basic periodic encoding is enhanced throughout the auditory system in people with better ability to perceive speech under challenging noise conditions. Although, the IC vs. SIN relationship falls short of significance, the trend is in the predicted direction (rs = 33, $p = 0.075$); it is therefore most parsimonious to assume that we failed to observe the relationship strongly here for reasons related to noise in the data rather than that the strength of the pitch representation is uncorrelated at a midpoint location between nuclei through which the same information passes.

Importantly, the principal similarity between the conditions under which the neurophysiological measurement was made (i.e., passive listening in silence) and the behavioral measure [i.e., deciphering speech in sentences, similar to the clinical HINT task (Nilsson, 1994)] was the presentation of an auditory speech stimulus. Naturalistic SIN situations offer multiple cues, many of them redundant, that can result in flexibility in how the task is solved. This is true to a lesser degree of clinical tests of SIN (including the HINT variant used here), which approximate the naturalistic experience of perceiving speech masked by sound, but lack factors that are important in real-life conversations such as familiarity with the talker (Nygaard et al., 1994; Souza et al., 2013), visual cues (Zion Golumbic et al., 2013), and context- and listener-dependent adaptations of the speaker (Lombard, 1911). Tasks that have been used to study the neural correlates of SIN perception range in cue-richness from natural language comprehension in daily life, to intermediate tasks like the sentence-in-noise and word-in-noise measures, to discrimination of phonemes from among a restricted set of possibilities (Du et al., 2014), and finally to passively listening to single sounds with or without masking noise; different types of noise such as energetic maskers and informational maskers have also been investigated (Swaminathan et al., 2015). Because the effects of variations in paradigm design are unclear (Wilson et al., 2007), [for example the HINT may be presented with or without spatial cues (Nilsson, 1994)], this diversity may be contributing to confusion in SIN literature, for example regarding the possibility of a using musical training to improve SIN skills (Coffey et al., 2017). A systematic analysis of SIN task

FIGURE 3 | Later cortical evoked responses and their origins. **(A)** Time courses of the lower frequency evoked response potentials (ERPs) from EEG data with the time windows used for MEG source analysis marked (P1: blue, P2: red), and **(C)** evoked response fields (ERFs) from simultaneously recorded MEG data. Each is averaged over subjects ($N = 20$). **(B)** The amplitude of the P2 ERP wave peak (red) correlates with SIN accuracy. **(D)** Group-level MEG topographies (left, strength and polarity is indicated in the color bar below) and source analyses of P1 and P2 component origins using (1mm MNI space; cluster threshold; $p < 0.005$). Note that single-channel EEG data are used to derive P2 amplitudes in order to maximize interpretability with respect to previous work whereas source localization is performed on concurrent MEG data.

requirements and differences in their neural correlates may be helpful to clarify these issues [e.g., by using task decomposition (Coffey and Herholz, 2013)].

According to such an approach, our neurophysiological recording paradigm did not explicitly engage auditory working memory and it did not require attention, which instead was

FIGURE 4 | Relationship between FFR-f0 strength (blue-green) and P2 amplitude (red). **(A)** Time course of the FFR-f0 response, single channel (as presented in Coffey et al., 2016b; 80–450 Hz bandpass filtered; −50 to 150 ms window). The green portion indicates period over which FFR-f0 strength is calculated. **(B)** ERF over the same time period; these signals are separated by frequency band (2–40 Hz bandpass filtered) but P2 occurs after the FFR. Correlations between the FFR-f0 strength from the cortical ROIs as measured using MEG and the ERP strength at P2 as measured using EEG are shown for **(C)** the left and **(E)** the right auditory cortex. **(D)** Illustrates the relationship of the cortical origins of each signal (1mm MNI space; cluster threshold; $p < 0.005$).

directed to a silent film. Although, top-down processes might conceivably act spontaneously even in the presence of such a distraction, the speech sounds were presented in perfect clarity and did not offer more than one stream of information upon which top-down mechanisms of auditory stream analysis such as stream segregation (Başkent and Gaudrain, 2016) or selective auditory attention (Song et al., 2011; Lehmann and Schönwiesner, 2014) might act.

Because we have eliminated cues that might be used by these higher-level processes that are known to affect SIN, either by enhancing the incoming signal (e.g., Lehmann and Schönwiesner, 2014) or at later linguistic/cognitive processing stages, any residual relationship between basic sound encoding (measured in silence) and scores on a high-level SIN task is most parsimoniously explained by the benefits of better lower-level

sound encoding. Thus, we believe that the correlations we observed at brainstem, thalamic and cortical levels are best interpreted as reflecting these low-level enhancements.

Amplitude differences in the P2 cortical component and in FFR-f0 strength have been related to SIN perception differences between normal and learning-disordered children when recorded in noise (Cunningham et al., 2001). P2 amplitude also appears to be stronger in musicians (Shahin et al., 2003; Kuriki et al., 2006; Bidelman and Weiss, 2014; though this has not been observed in all studies Musacchia et al., 2008). P2, along with the P1 and the intervening negativity, is affected by stimulus parameters including frequency, location, duration, intensity, and presence of noise (reviewed in Alain et al., 2013; see also Ross and Fujioka, 2016), is affected by short-term training (Lappe et al., 2011; Tremblay et al., 2014), and is correlated

with language-related performance measures such as categorical speech perception (Bidelman and Weiss, 2014). While P2 amplitude is affected by repetition and predictability (Näätänen and Picton, 1987; Tremblay et al., 2014), the relationships we observe here between the P2 amplitude and SIN performance are unlikely to be unique to the high number of repetitions used in this experiment, as the relationship between P2 and SIN has been previously observed with fewer trials (Cunningham et al., 2001). In any case, any fatigue effects on P2 amplitude would be more likely to decrease rather than be the cause of a correlation with the FFR-f0 amplitude, which is known to be comparatively resistant to cognitive manipulation (Varghese et al., 2015) except under very specific conditions (Lehmann and Schönwiesner, 2014; Coffey et al., 2016a). Our findings therefore suggest that the underlying processes are critical to sound representation generally, though the nature and roles of component processes represented in the P2 and their relationships to oscillatory brain networks are still being clarified (e.g., Ross et al., 2012; Ross and Fujioka, 2016).

We found that variability in P2 amplitude correlated with inter-individual differences in SIN ability (**Figure 3B**) and in FFR-f0 strength (**Figures 4C,E**), despite that the responses were measured only in quiet conditions and in a normal healthy adult population. These results suggest that previously reported relationships may be present as a continuum in the population and even in optimal listening conditions. The MEG-FFR technique may allow us to more consistently observe behavioral and experience-related relationships with FFR-f0 strength in less challenging listening conditions as compared with the EEG-FFR. The EEG-FFR is likely a composite from several subcortical and cortical sources (Herdman et al., 2002; Kuwada and Anderson, 2002; Coffey et al., 2016b; King et al., 2016; Zhang and Gong, 2016; Tichko and Skoe, 2017). In recent work, we compared two common single-channel EEG montages (Cz-mastoids and Fz-C7) and found that while FFR-f0 strength in each montage (measured simultaneously) was moderately correlated, a large proportion of variability was unaccounted for and the two methods differed in their sensitivity to a behavioral measure of interest (Coffey et al., 2016a). This observation suggests that differences in individuals' head and brain geometry may sometimes obscure EEG-FFR vs. behavioral relationships, possibly due to interference from source summation at a given point of measurement.

The spatial resolution of MEG source imaging may help to clarify the auditory processes that generate the ERP and ERF components, which have a long history in auditory neuroscience yet have predominantly been studied at the sensor level or using simpler models. We used distributed source modeling based on individual anatomy to localize the sources of each ERP/ERF wave (**Figure 3D**) with a view to confirming the localization of the wave of interest. The P1 wave originated bilaterally in the primary auditory areas, as expected (Liégeois-Chauvel et al., 1994; Key et al., 2005). P2 appeared to be right-lateralized and comparatively more anterior along the superior temporal plane as compared to the P1 signal. This observation is generally consistent with previous work (Alain et al., 2013), but contrasts with an analysis of equivalent current dipoles

that suggested a more posterior and medial source for the P2 (Shahin et al., 2003). However, both the stimulation and analysis (e.g., use of the standard brain rather than individual anatomy) vary considerably between these studies, making this difference difficult to interpret.

The relatively more anterior location of the P2 compared to the FFR generators (**Figure 4D**) could be explained by a right-lateralized anterior flow of pitch-relevant information that supports SIN processing, although future work will be needed to clarify whether the relationship between periodic encoding and later waves is causal in nature, or if these different frequency bands represent neural activity in parallel processing streams in neighboring neural populations. Furthermore, as mentioned previously, earlier ERP components such as N1 have previously been related to SIN perception (Billings et al., 2013; Bidelman and Howell, 2016). Although, we were not able to measure it with the experimental design used here, it is possible that a relationship exists between FFR and slower cortical activity earlier than P2. A combination of multichannel EEG and MEG, and a stimulation paradigm that is optimized to clearly and consistently evoke each ERP waveform may be highly informative as to how basic auditory information is separated and streamed to other cortical areas in order to accomplish different auditory tasks, using the spatial information in MEG data or a combinations of EEG and fMRI data.

The strength of signal generators in the right hemisphere was stronger for both the FFR-f0 and P2 waves (**Figures 4A,B,D**). We found a positive correlation between P2 amplitude (measured with EEG at Cz) and MEG FFR-f0 in the right hemisphere (but not left; **Figures 4C,E**). These results corroborate previous work suggesting that the right auditory cortex is relatively specialized for pitch and tonal processing (Zatorre et al., 1994; Patel and Balaban, 2001; Patterson et al., 2002; Schneider et al., 2002; Hyde et al., 2008; Mathys et al., 2010; Albouy et al., 2013; Andoh et al., 2015; Herholz et al., 2015; Cha et al., 2016), as well as reports of experience-sensitive relationships between FFR-f0 and lower frequency auditory evoked responses (Musacchia et al., 2008; Bidelman et al., 2014). It has been proposed that subtle differences in neural responses early in the cortical processing stream may lead to distinct functional roles for higher level processes out of a need for optimization, in particular a right-hemisphere bias for periodicity and left-hemisphere bias for fine temporal resolution (Zatorre et al., 2002). It is likely that other aspects of SIN processing, particularly those related to linguistic cues, are also relatively more left-lateralized, and that overall laterality of SIN-related processing may depend on the degree to which a given paradigm engages combinations of neural systems (Shtyrov et al., 1998, 1999; Laine et al., 1999; Du et al., 2014, 2016; Bidelman and Howell, 2016). The relatively stronger temporal representation of periodicity in the right hemisphere may be the underlying reason for which we had previously observed a hemispheric asymmetry in the FFR-f0 and its relationship to fine pitch discrimination ability, hours of musical practice, and age of training onset (see Figure 5c–e in (Coffey et al., 2016b); statistics are reported in the present work in **Table 1**). The present results are congruent with these hypotheses.

Periodicity encoding is related to pitch information (Gockel et al., 2011), which is one of several cues that the brain can use to separate streams of auditory information (Bregman, 1994; Moore and Gockel, 2002) and which is useful to help distinguishing one vowel sound from another (Chalikia and Bregman, 1989; Summerfield and Assmann, 1999) and can help speech segregation at the sentence level (Brokx and Nooteboom, 1982). We previously showed that fine pitch discrimination skills are correlated with FFR-f0 strength in the right auditory cortex (Coffey et al., 2016b). Here, we add that discrimination thresholds correlate with SIN accuracy, and that SIN accuracy is related to periodic encoding in the auditory cortex. Together, these results support a mechanistic explanation for SIN enhancement via better pitch processing leading to better stream segregation (Anderson and Kraus, 2010a). This explanation can also account for musician advantages in SIN, as well as the inconsistency with which it is observed across studies, whose design may emphasize other, non-periodic SIN cues. In the subset of subjects who reported having had musical training, measures of the extent and timing (age of start) of musical practice were related to behavioral measures of SIN accuracy and fine pitch discrimination, and were paralleled in physiology by relationships to FFR-f0 and P2 amplitude. Experience-dependent plasticity therefore likely tunes FFR-f0 strength and tracking ability as suggested by several prior studies (Musacchia et al., 2007; Song et al., 2008; Bidelman et al., 2011a; Carcagno and Plack, 2011). Stronger periodicity encoding might thereby account in part for a musician advantage. However, other top-down factors are also at play in SIN perception, including auditory working memory, long-term memory, and selective attention. Each of these may be influenced by experience, and other peripheral and central factors (Anderson et al., 2013b). These latter factors would likely be related to top-down effects originating in extra-auditory cortical areas such as motor and frontal cortices (Du et al., 2014) whereas the feed-forward mechanisms we emphasize here likely represent neural modulations within ascending neural pathways including brainstem nuclei, thalamus, and cortex. Any or all of these mechanisms may be enhanced by musical training.

We propose that whether a musician advantage in SIN perception is observed or not in a given study may depend on interactions between the current state of the auditory system and the specific cognitive demands of the SIN task used in the study. Specifically, performance can depend on (1) the cues offered to the listener in the SIN paradigm [e.g., spatial cues and degree of information masking; (Swaminathan et al., 2015)]; (2) the degree to which an individual's experience has enhanced representations and mechanisms related to the available cues and caused them to be more strongly weighted; and (3) how well individuals can adapt to use alternative cues and mechanisms when one or more cues becomes less useful either through task differences like levels of noise (e.g., Du et al., 2014) or due to physiological deterioration (Anderson et al., 2013b).

CONCLUSION

In this study we present novel evidence that the quality of basic feed-forward periodicity encoding is related to the

clinically relevant problem of separating speech from noise signals, and musical training. Specifically, in the absence of contextual cues and task demands and given a measurement tool that is sensitive to signal sources (i.e., MEG), enhancements in periodic sound encoding throughout the auditory neuraxis were correlated with better SIN ability in an offline task of sentence perception. This effect was observed to be stronger in the FFR signal localized to the right auditory cortex, and was related to slower cortical P2 wave amplitude measured by EEG, which is concurrent with activity in the right secondary auditory cortex measured with MEG and suggests an anterior flow of pitch-related information. Musicians show an advantage related to FFR strength, suggesting a possible role of experience. Our results suggest that inter-individual differences in neural correlates of basic periodic sound representation observed within the normal-hearing population (Ruggles et al., 2011; Coffey et al., 2016a) may in part be responsible for the surprising variability in SIN perception observed inter-individually. This work sketches in the anatomical and temporal properties of a stream of pitch-relevant information from subcortical areas up to and beyond right primary auditory cortex. More work is needed to explore exactly how this information is routed and subsequently used by higher-level networks. We conclude that better sound encoding likely improves SIN perception through better representation of periodicity, which in turn leads to better stream segregation. Importantly, these findings may have practical implications for people suffering from difficulty with SIN perception because basic sound encoding fidelity has been shown to be malleable via training; these results therefore support efforts to develop training-based treatment strategies (Bidelman and Alain, 2015) to improve people's participation in social, vocational, and educational activities (Anderson and Kraus, 2010b).

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Montreal Neurological Institute Research Ethics Board with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Montreal Neurological Institute Research Ethics Board.

AUTHOR CONTRIBUTIONS

EC, SH, SB, and RZ designed the experiment, AC and EC collected the data, EC analyzed the data, and EC, SH, SB, and RZ wrote the paper.

ACKNOWLEDGMENTS

REFERENCES

Alain, C., Roye, A., and Arnott, S. R. (2013). "Middle- and long-latency auditory evoked potentials: what are they telling us on central auditory disorders," in *Handbook of Clinical Neurophysiology: Disorders of Peripheral and Central Auditory Processing, Vol. 10,* ed. G. G. Celesia (Amsterdam: Elsevier B.V), 177–199.

Alain, C., Zendel, B. R., Hutka, S., and Bidelman, G. M. (2014). Turning down the noise: the benefit of musical training on the aging auditory brain. *Hear. Res.* 308, 162–173. doi: 10.1016/j.heares.2013.06.008

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.-E., et al. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain* 136, 1639–1661. doi: 10.1093/brain/awt082

Anderson, S., and Kraus, N. (2010a). Objective neural indices of speech-in-noise perception. *Trends Amplif.* 14, 73–83. doi: 10.1177/1084713810380227

Anderson, S., and Kraus, N. (2010b). Sensory-cognitive interaction in the neural encoding of speech in noise: a review. *J. Am. Acad. Audiol.* 21, 575–585. doi: 10.3766/jaaa.21.9.3

Anderson, S., Parbery-Clark, A., White-Schwoch, T., and Kraus, N. (2012). Aging affects neural precision of speech encoding. *J. Neurosci.* 32, 14156–14164. doi: 10.1523/JNEUROSCI.2176-12.2012

Anderson, S., White-Schwoch, T., Choi, H. J., and Kraus, N. (2013a). Training changes processing of speech cues in older adults with hearing loss. *Front. Syst. Neurosci.* 7:97. doi: 10.3389/fnsys.2013.00097

Anderson, S., White-Schwoch, T., Parbery-Clark, A., and Kraus, N. (2013b). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hear. Res.* 300, 18–32. doi: 10.1016/j.heares.2013.03.006

Andoh, J., Matsushita, R., and Zatorre, R. J. (2015). Asymmetric interhemispheric transfer in the auditory network: evidence from TMS, resting-state fMRI, and diffusion imaging. *J. Neurosci.* 35, 14602–14611. doi: 10.1523/JNEUROSCI.2333-15.2015

Attal, Y., and Schwartz, D. (2013). Assessment of subcortical source localization using deep brain activity imaging model with minimum norm operators: a MEG study. *PLoS ONE* 8:e59856. doi: 10.1371/journal.pone.0059856

Baillet, S., Mosher, J. C., and Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Process. Mag.* 18, 14–30. doi: 10.1109/79.962275

Başkent, D., and Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *J. Acoust. Soc. Am.* 139, EL51–EL56. doi: 10.1121/1.4942628

Bench, J., Kowal, A., and Bamford, J. (1979). The BKB. (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br. J. Audiol.* 13, 108–112. doi: 10.3109/03005367909078884

Bendixen, A. (2014). Predictability effects in auditory scene analysis: a review. *Front. Neurosci.* 8:60. doi: 10.3389/fnins.2014.00060

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57, 289–300.

Bidelman, G. M. Krishnan, A., and Gandour, J. T. (2011a). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *Eur. J. Neurosci.* 33, 530–538. doi: 10.1111/j.1460-9568.2010.07527.x

Bidelman, G. M., and Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J. Neurosci.* 35, 1240–1249. doi: 10.1523/JNEUROSCI.3292-14.2015

Bidelman, G. M., and Howell, M. (2016). Functional changes in inter- and intra-hemispheric cortical processing underlying degraded speech perception. *Neuroimage* 124, 581–590. doi: 10.1016/j.neuroimage.2015.09.020

Bidelman, G. M., Gandour, J. T., and Krishnan, A. (2011b). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain Cogn.* 77, 1–10. doi: 10.1016/j.bandc.2011.07.006

Bidelman, G. M., Villafuerte, J. W., Moreno, S., and Alain, C. (2014). Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiol. Aging* 35, 2526–2540. doi: 10.1016/j.neurobiolaging.2014.05.006

Bidelman, G., and Weiss, M. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2672. doi: 10.1111/ejn.12627

Billings, C. J., Bennett, K. O., Molis, M. R., and Leek, M. R. (2011). Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. *Ear Hear.* 32, 53–60. doi: 10.1097/AUD.0b013e3181ec5c46

Billings, C. J., McMillan, G. P., Penman, T. M., and Gille, S. M. (2013). Predicting perception in noise using cortical auditory evoked potentials. *J. Assoc. Res. Otolaryngol.* 14, 891–903. doi: 10.1007/s10162-013-0415-y

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., and Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *J. Acoust. Soc. Am.* 137, 378–387. doi: 10.1121/1.4904537

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound.* Boston, MA: MIT Press. Available online at: https://books.google.com/books?hl=en&lr=&id=jI8muSpAC5AC&pgis=1

Brokx, J. P. L., and Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *J. Phon.* 10, 23–36.

Carcagno, S., and Plack, C. J. (2011). Subcortical plasticity following perceptual learning in a pitch discrimination task. *J. Assoc. Res. Otolaryngol.* 12, 89–100. doi: 10.1007/s10162-010-0236-1

Cha, K., Zatorre, R. J., and Schönwiesner, M. (2016). Frequency selectivity of voxel-by-voxel functional connectivity in human auditory cortex. *Cereb. Cortex* 26, 211–224. doi: 10.1093/cercor/bhu193

Chalikia, M. H., and Bregman, A. S. (1989). The perceptual segregation of simultaneous auditory signals: pulse train segregation and vowel segregation. *Percept. Psychophys.* 46, 487–496. doi: 10.3758/BF03210865

Chandrasekaran, B., and Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47, 236–246. doi: 10.1111/j.1469-8986.2009.00928.x

Coffey, E. B. J., and Herholz, S. C. (2013). Task decomposition: a framework for comparing diverse training models in human brain plasticity studies. *Front. Hum. Neurosci.* 7:640. doi: 10.3389/fnhum.2013.00640

Coffey, E. B. J., Colagrosso, E. M. G., Lehmann, A., Schönwiesner, M., and Zatorre, R. J. (2016a). Individual differences in the frequency-following response: relation to pitch perception. Ed. Frederic Dick. *PLoS ONE* 11:e0152374. doi: 10.1371/journal.pone.0152374

Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., and Zatorre, R. J. (2016b). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.* 7:11070. doi: 10.1038/ncomms11070

Coffey, E. B. J., Herholz, S. C., Scala, S., and Zatorre, R. J. (2011). "Montreal Music History Questionnaire: a tool for the assessment of music-related experience in music cognition research," in *The Neurosciences and Music IV: Learning and Memory, Conference* (Edinburgh).

Coffey, E., Mogilever, N., and Zatorre, R. (2017). Speech-in-noise perception in musicians: a review. *Hear. Res.* 352, 49–69. doi: 10.1016/j.heares.2017.02.006

Coffey, E. B. J., Musacchia, G., and Zatorre, R. J. (2016c). Cortical correlates of the auditory frequency-following and onset responses: EEG and fMRI evidence. *J. Neurosci.* 37, 830–838. doi: 10.1523/JNEUROSCI.1265-16.2016

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., and Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin. Neurophysiol.* 112, 758–767. doi: 10.1016/S1388-2457(01)00465-5

Du, Y., Buchsbaum, B. R., Grady, C. L., and Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7126–7131. doi: 10.1073/pnas.1318738111

Du, Y., Buchsbaum, B. R., Grady, C. L., and Alain, C. (2016). Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nat. Commun.* 7:12241. doi: 10.1038/ncomms12241

Du, Y., Kong, L., Wang, Q., Wu, X., and Li, L. (2011). Auditory frequency-following response: a neurophysiological measure for studying the "cocktail-party problem". *Neurosci. Biobehav. Rev.* 35, 2046–2057. doi: 10.1016/j.neubiorev.2011.05.008

Dumas, T., Dubal, S., Attal, Y., Chupin, M., Jouvent, R., Morel, S., et al. (2013). MEG evidence for dynamic amygdala modulations by gaze and facial emotions. Ed. Andreas Keil. *PLoS ONE* 8:e74145. doi: 10.1371/journal.pone.0074145

Evans, A. C., Janke, A. L., Collins, D. L., and Baillet, S. (2012). Brain templates and atlases. *Neuroimage* 62, 911–922. doi: 10.1016/j.neuroimage.2012.01.024

Evans, S., Meekings, S., Nuttall, H. E., Jasmin, K. M., Boebinger, D., Adank, P., et al. (2014). Does musical enrichment enhance the neural coding of syllables? Neuroscientific interventions and the importance of behavioral data. *Front. Hum. Neurosci.* 8:964. doi: 10.3389/fnhum.2014.00964

Fischl, B. (2012). FreeSurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., and Başkent, D. (2014). The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? *Front. Neurosci.* 8:179. doi: 10.3389/fnins.2014.00179

Gockel, H., Carlyon, R., Mehta, A., Plack, C., and Lack, C. H. J. P. (2011). The frequency following response. (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *J. Assoc. Res. Otolaryngol.* 782, 767–782. doi: 10.1007/s10162-011-0284-1

Golestani, N., Rosen, S., and Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: the contribution of semantics. *Biling. Cogn.* 12, 385–392. doi: 10.1017/S1366728909990150

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., et al. (2013). Good practice for conducting and reporting MEG research. *Neuroimage* 65, 349–363. doi: 10.1016/j.neuroimage.2012.10.001

Hämäläinen, M. S. (2009). *MNE Software User's Guide v2.7.* MGH/HMS/MIT Athinoula, A. Martinos Center for Biomedical Imaging Massachusetts General Hospital, Charlestown, MA.

Herdman, A. T., Lins, O., Roon, P., Van Stapells, D. R., Scherg, M., and Picton T. W. (2002). Intracerebral sources of human auditory steady-state responses. *Brain Topogr.* 15, 69–86. doi: 10.1023/A:1021470822922

Herholz, S. C., Coffey, E. B. J., Pantev, C., and Zatorre, R. J. (2015). Dissociation of neural networks for predisposition and for training-related plasticity in auditory-motor learning. *Cereb. Cortex* 26, 3125–3134. doi: 10.1093/cercor/bhv138

Hyde, K. L., Peretz, I., and Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* 46, 632–639. doi: 10.1016/j.neuropsychologia.2007.09.004

Irvine, D. R. F. (1986). "The Auditory Brainstem: a review of the structure and function of the auditory brainstem processing mechanisms," in *Progress in Sensory Physiology, Vol. 7*, eds H. Autrum, D. Ottoson, E. R. Perl, R. F. Schmidt, H. Shimazu, W. D. Willis, and D. Ottoson (Berlin: Springer), 22–39.

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Key, A. P. F., Dove, G. O., and Maguire, M. J. (2005). Linking brainwaves to the brain: an ERP primer. *Dev. Neuropsychol.* 27, 183–215. doi: 10.1207/s15326942dn2702_1

King, A., Hopkins, K., and Plack, C. J. (2016). Differential group delay of the frequency following response measured vertically and horizontally. *J. Assoc. Res. Otolaryngol.* 17, 133–143. doi: 10.1007/s10162-016-0556-x

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., et al. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *J. Neurosci.* 34, 11913–11918. doi: 10.1523/JNEUROSCI.1881-14.2014

Kraus, N., Strait, D. L., and Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. *Ann. N. Y. Acad. Sci.* 1252, 100–107. doi: 10.1111/j.1749-6632.2012.06463.x

Kuriki, S., Kanda, S., and Hirata, Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *J. Neurosci.* 26, 4046–4053. doi: 10.1523/JNEUROSCI.3907-05.2006

Kuwada, S., and Anderson, J. (2002). Sources of the scalp-recorded amplitude-modulation following response. *J. Am. Acad. Audiol.* 13, 188–204.

Laine, M., Rinne, J. O., Krause, B. J., Teräs, M., and Sipilä, H. (1999). Left hemisphere activation during processing of morphologically complex word forms in adults. *Neurosci. Lett.* 271, 85–88. doi: 10.1016/S0304-3940(99)00527-3

Lappe, C., Trainor, L. J., Herholz, S. C., and Pantev, C. (2011). Cortical plasticity induced by short-term multimodal musical rhythm training. *PLoS ONE* 6:e21493. doi: 10.1371/journal.pone.0021493

Lehmann, A., and Schönwiesner, M. (2014). Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. *PLoS ONE* 9:e85442. doi: 10.1371/journal.pone.0085442

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375

Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., and Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr. Clin. Neurophysiol.* 92, 204–214. doi: 10.1016/0168-5597(94)90064-7

Lombard, E. (1911). Le signe de l'élévation de la voix. *Ann Mal Oreille Larynx Nez Pharynx* 37, 101–119.

Mathys, C., Loui, P., Zheng, X., and Schlaug, G. (2010). Non-invasive brain stimulation applied to Heschl's gyrus modulates pitch discrimination. *Front. Psychol.* 1:193. doi: 10.3389/fpsyg.2010.00193

Moore, B. C. J., and Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acust. United Acust.* 88, 320–332. Available online at: http://www.ingentaconnect.com/content/dav/aaua/2002/00000088/00000003/art00004

Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894–15898. doi: 10.1073/pnas.0701498104

Musacchia, G., Strait, D. L., and 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. doi: 10.1016/j.heares.2008.04.013

Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425. doi: 10.1111/j.1469-8986.1987.tb00311.x

Nilsson, M. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.* 95, 1085–1099. doi: 10.1121/1.408469

Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychol. Sci.* 5, 42–46. doi: 10.1111/j.1467-9280.1994.tb00612.x

Parbery-Clark, A., Anderson, S., Hittner, E., and Kraus, N. (2012a). Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. *Front. Aging Neurosci.* 4:30. doi: 10.3389/fnagi.2012.00030

Parbery-Clark, A., Anderson, S., Hittner, E., and Kraus, N. (2012b). Musical experience offsets age-related delays in neural timing. *Neurobiol Aging* 33, 1483.e1–1483.e4. doi: 10.1016/j.neurobiolaging.2011.12.015

Parbery-Clark, A., Marmel, F., Bair, J., and Kraus, N. (2011a). What subcortical-cortical relationships tell us about processing speech in noise. *Eur. J. Neurosci.* 33, 549–557. doi: 10.1111/j.1460-9568.2010.07546.x

Parbery-Clark, A., Skoe, E., and Kraus, N. (2009a). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107. doi: 10.1523/JNEUROSCI.3256-09.2009

Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009b). Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661. doi: 10.1097/AUD.0b013e3181b412e9

Parbery-Clark, A., Strait, D. L., and Kraus, N. (2011c). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia* 49, 3338–3345. doi: 10.1016/j.neuropsychologia.2011.08.007

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., and Kraus, N. (2011b). Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS ONE* 6:e18082. doi: 10.1371/journal.pone.0018082

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* 308, 98–108. doi: 10.1016/j.heares.2013.08.011

Patel, A. D., and Balaban, E. (2001). Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nat. Neurosci.* 4, 839–844. doi: 10.1038/90557

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36, 767–776. doi: 10.1016/S0896-6273(02)01060-7

Pickering, M. J., and Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.* 11, 105–110. doi: 10.1016/j.tics.2006.12.002

Pressnitzer, D., Suied, C., and Shamma, S. (2011). Auditory scene analysis: the sweet music of ambiguity. *Front. Hum. Neurosci.* 5:158. doi: 10.3389/fnhum.2011.00158

Rinne, T., Balk, M. H., Koistinen, S., Autti, T., Alho, K., and Sams, M. (2008). Auditory selective attention modulates activation of human inferior colliculus. *J. Neurophysiol.* 100, 3323–3327. doi: 10.1152/jn.90607.2008

Ross, B., and Fujioka, T. (2016). 40-Hz oscillations underlying perceptual binding in young and older adults. *Psychophysiology* 53, 974–990. doi: 10.1111/psyp.12654

Ross, B., Miyazaki, T., and Fujioka, T. (2012). Interference in dichotic listening: the effect of contralateral noise on oscillatory brain networks. *Eur. J. Neurosci.* 35, 106–118. doi: 10.1111/j.1460-9568.2011.07935.x

Ruggles, D. R., Freyman, R. L., and Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PLoS ONE* 9:e86980. doi: 10.1371/journal.pone.0086980

Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15516–15521. doi: 10.1073/pnas.1108912108

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694. doi: 10.1038/nn871

Shahin, A., Bosnyak, D. J., Trainor, L. J., and Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J. Neurosci.* 23, 5545–5552. Available online at: http://www.jneurosci.org/content/23/13/5545

Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R. J., et al. (1998). Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: magnetic mismatch negativity study. *Neurosci. Lett.* 251, 141–144. doi: 10.1016/S0304-3940(98)00529-1

Shtyrov, Y., Kujala, T., Ilmoniemi, R. J., Na, R., and Nen, È. (1999). *Noise Affects Speech- Signal Processing Differently in the Cerebral Hemispheres.* Available online at: https://insights.ovid.com/neuroreport/nerep/1999/07/130/noise-affects-speech-signal-processing-differently/34/00001756.

Skoe, E., and Kraus, N. (2010). Auditory brain stem response to complex sounds: a tutorial. *Ear Hear.* 31, 302–324. doi: 10.1097/AUD.0b013e3181cdb272

Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., and Kraus, N. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behav. Brain Res.* 291, 244–252. doi: 10.1016/j.bbr.2015.05.026

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23(Suppl. 1), S208–S219. doi: 10.1016/j.neuroimage.2004.07.051

Song, J. H., Skoe, E., Banai, K., and Kraus, N. (2011). Perception of speech in noise: neural correlates. *J. Cogn. Neurosci.* 23, 2268–2279. doi: 10.1162/jocn.2010.21556

Song, J. H., Skoe, E., Banai, K., and Kraus, N. (2012). Training to improve hearing speech in noise: biological mechanisms. *Cereb. Cortex* 22, 1180–1190. doi: 10.1093/cercor/bhr196

Song, J., Skoe, E., Wong, P., and Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cogn. Neurosci.* 20, 1892–1902. doi: 10.1162/jocn.2008.20131

Souza, P., Gehani, N., Wright, R., and McCloy, D. (2013). The advantage of knowing the talker. *J. Am. Acad. Audiol.* 24, 689–700. doi: 10.3766/jaaa.24.8.6

Strait, D. L., and Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front. Psychol.* 2:113. doi: 10.3389/fpsyg.2011.00113

Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from

masking and auditory attention performance. *Hear. Res.* 261, 22–29. doi: 10.1016/j.heares.2009.12.021

Strait, D. L., Parbery-Clark, A., Hittner, E., and Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain Lang.* 123, 191–201. doi: 10.1016/j.bandl.2012.09.001

Suga, N. (2012). Tuning shifts of the auditory system by corticocortical and corticofugal projections and conditioning. *Neurosci. Biobehav. Rev.* 36, 969–988. doi: 10.1016/j.neubiorev.2011.11.006

Suied, C., Bonneel, N., and Viaud-Delmon, I. (2009). Integration of auditory and visual information in the recognition of realistic objects. *Exp. Brain Res.* 194, 91–102. doi: 10.1007/s00221-008-1672-6

Summerfield, Q., and Assmann, P. F. (1999). Perception of concurrent vowels: effects of harmonic misalignment and pitch-period asynchrony. *J. Acoust. Soc. Am.* 89, 1364–1377.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G., and Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Sci. Rep.* 5, 1–10. doi: 10.1038/srep11628

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 2011, 1–13. doi: 10.1155/2011/879716

Tichko, P., and Skoe, E. (2017). Frequency-dependent fine structure in the frequency-following response: the byproduct of multiple generators. *Hear. Res.* 348, 1–15. doi: 10.1016/j.heares.2017.01.014

Tierney, A., Krizman, J., Skoe, E., Johnston, K., and Kraus, N. (2013). High school music classes enhance the neural processing of speech. *Front. Psychol.* 4:855. doi: 10.3389/fpsyg.2013.00855

Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., and Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Front. Syst. Neurosci.* 8:28. doi: 10.3389/fnsys.2014.00028

Varghese, L., Bharadwaj, H. M., and Shinn-Cunningham, B. G. (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Res.* 1626, 146–164. doi: 10.1016/j.brainres.2015.06.038

Weiss, M. W., and Bidelman, G. M. (2015). Listening to the brainstem: musicianship enhances intelligibility of subcortical representations for speech. *J. Neurosci.* 35, 1687–1691. doi: 10.1523/JNEUROSCI.3680-14.2015

Wilson, R. H., McArdle, R. A., and Smith, S. L. (2007). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *J. speech Lang. Hear. Res.* 50, 844–856. doi: 10.1044/1092-4388(2007/059)

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage* 92, 381–397. doi: 10.1016/j.neuroimage.2014.01.060

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46. doi: 10.1016/S1364-6613(00)01816-7

Zatorre, R. J., Evans, A., and Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* 14, 1908–1919.

Zendel, B. R., and Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychol. Aging* 27, 410–417. doi: 10.1037/a0024816

Zhang, X., and Gong, Q. (2016). Correlation between the frequency difference limen and an index based on principal component analysis of the frequency-following response of normal hearing listeners. *Hear. Res.* 344, 255–264. doi: 10.1016/j.heares.2016.12.004

Zion Golumbic, E., Cogan, G. B., Schroeder, C. E., and Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". *J. Neurosci.* 33, 1417–1426. doi: 10.1523/JNEUROSCI.3675-12.2013