BASIC NEURAL PROCESSING OF SOUND IN ADULTS IS INFLUENCED BY BILINGUAL EXPERIENCE

ERIKA SKOE, [a,b,d*] EMILY BURAKIEWICZ, [a,d]
MICHAEL FIGUEIREDO [a,d] AND MARGARET HARDIN [c,d]

[a] *Department of Speech, Language, and Hearing Sciences, Storrs, CT 06269, USA*

[b] *Affiliate of the Department of Psychological Sciences, Cognitive Sciences Program, Connecticut Institute for Brain and Cognitive Sciences, Storrs, CT 06269, USA*

[c] *Department of Physiology and Neurobiology, Storrs, CT 06269, USA*

[d] *University of Connecticut, Storrs, CT 06269, USA*

Abstract—The central auditory nervous system (CANS) undergoes language-dependent tuning to enhance linguistically relevant features of sound. However, less is known about how dual-language exposure affects the CANS. Recent reports indicate that Spanish-English bilingual children and adolescents have larger neural responses to the fundamental frequency (F0) of vowels, as measured by the frequency-following response (FFR), a phase-locked response to sound. Given the cross-language significance of F0, this led us to hypothesize that enhanced neural responses to the F0 are not unique to Spanish-English bilingual children and adolescents but are instead a common feature of a CANS with significant early dual language experience. In support of this hypothesis, we found that early bilingual adults, representing more than a dozen languages, had more robust FFRs to the F0 compared to English-language monolinguals suggesting that bilingual experience imprints on the CANS in a similar fashion regardless of the languages of exposure. Taken together, our results suggest that early exposure to two linguistic sound systems primes the brain to respond to the F0, a basic feature of all speech sounds that signals important indexical information for vowel, talker, and language identification. © 2017 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: bilingual, frequency-following response (FFR), central auditory nervous system (CANS).

INTRODUCTION

There is a rich literature on the neurobiological correlates of bilingualism (reviewed in Costa and Sebastián-Gallés, 2014; Burgaleta et al., 2016). Yet, despite this wealth of knowledge and the prevalence of bilingualism worldwide, comparatively little is known about how extensive exposure to more than one language affects basic sound processing. Basic sound processing refers to the auditory neuroaxis and its ability to process auditory information that is not speech-specific, such as the fundamental frequency (F0) (Corriveau et al., 2007; Hakvoort et al., 2015). A variety of evidence leads us to propose that multilingualism might have an influence on basic auditory processing within the central auditory nervous system (CANS). First, nascent language learning, whether it be a single language or multiple languages, proceeds in tandem with the development of the central auditory system, setting the stage for the reciprocal development of spoken language and the CANS (Hepper and Shahidullah, 1994; Moore et al., 1995; Moore, 2002; Byers-Heinlein et al., 2010; Sanes and Woolley, 2011; Gervain, 2015). Moreover, although our understanding of how bilingualism affects CANS function and development is limited (Vihla et al., 2002; Krizman et al., 2012, 2014, 2015b), there is ample evidence from studies of acoustic deprivation, impoverishment, and enrichment that the CANS is influenced by the acoustic make-up of the environment in which a human or non-human animal is raised (Fujioka et al., 2006; Sanes and Bao, 2009; Kral and Sharma, 2012; Skoe and Kraus, 2013; Webb et al., 2015). For instance, in laboratory animals, auditory enrichment is associated with increased synthesis of CANS synaptic proteins (Alladi et al., 2002) and stronger responses from auditory cortex (Engineer et al., 2004), among other outcomes (Xu et al., 2009; Zhu et al., 2014). Consistent with this, Ressel and colleagues reported that auditory cortex is larger in adults who learned a second language (L2) earlier in life (Ressel et al., 2012), a finding that echoes previous work in late L2 learners (Wong et al., 2008). Functional differences within the CANS have also been noted between bilinguals and monolinguals. In a magnetoencephalographic study, it was reported that Swedish-Finnish adult bilinguals, compared to Finnish monolinguals, have increased auditory cortical activation to pure tones and speech-like complex sounds (Vihla et al., 2002).

A more recent line of work on basic auditory processing in bilinguals (Krizman et al., 2012, 2014, 2015a) has focused on the frequency-following response (FFR), a sustained auditory-evoked response to periodic and quasi-periodic auditory stimuli, such as vowels, pitch contours, and amplitude-modulated noise (Moushegian

*Correspondence to: E. Skoe, University of Connecticut, 850 Bolton Rd, U-1085, Storrs, CT, USA.
E-mail address: erika.skoe@uconn.edu (E. Skoe).*

Abbreviations: AoA, age of acquisition; CANS, central auditory nervous system; DPOAE, distortion product otoacoustic emission; F0, fundamental frequency; FFR, Frequency Following Response; FFT, fast Fourier transform; L1, first language; L2, second language; MEG, magnetoencephalography.

et al., 1973). The FFR, an electrical potential recorded at the scalp, is a phase-locked response to sound that reflects a complex composite of multiple generators within the auditory neuroaxis including cochlear nucleus, inferior colliculus and auditory cortex, with the relative contribution of each generator being dependent on the frequency of stimulation, among other variables (Gardi et al., 1979; Hoormann et al., 1992; Kuwada et al., 2002; Coffey et al., 2016; Tichko and Skoe, 2017). For frequencies in the range of human vocal F0 (∼85–250 Hz), the FFR is thought to reflect predominantly subcortical generators (Kuwada et al., 2002; Coffey et al., 2016). In contrast to the early cortical-evoked potentials that reflect sound onset (e.g., P1, N2, P2, etc.), the FFR bears a close resemblance to the stimulus spectrum. In fact, in many cases, the stimulus can be recognized when the response is sonified (Galbraith et al., 1995; Weiss and Bidelman, 2015), making the FFR a valuable tool for studying the fidelity of basic auditory processing. The F0, an acoustic correlate of pitch, is one of the more salient stimulus features that is captured in the FFR (Skoe and Kraus, 2010). However, there are gradients in the neurophonic quality of the response (Weinberger et al., 1970) that reflect individual differences in environmental exposure to sound, including the statistical properties of the auditory input during development. For example, studies of speakers from different languages (e.g., Mandarin, Vietnamese, German, Finnish, French, and English) suggest that the CANS undergoes language-dependent tuning to enhance linguistically relevant features of sound (Krishnan et al., 2009; Dawson et al., 2016; Intartaglia et al., 2016). In native (adult) speakers of tonal languages, the FFR shows increased fidelity to complex pitch patterns, especially those that closely approximate the spectrotemporally dynamic pitch contours found in their native language (Krishnan et al., 2009, 2011) and that are paired with native timbres (Krishnan et al., 2011). This increased fidelity in encoding pitch, however, does not generalize to all auditory stimuli (e.g., linearly changing pitch contours) (Krishnan et al., 2009) nor is it evident immediately at birth (Jeng et al., 2011), suggesting a language-specific developmental tuning of the central auditory system that translates to heightened basic processing of sound in non-linguistic contexts under some but not all conditions (Bidelman et al., 2011; Krishnan et al., 2011).

If language-specific tuning of the CANS is evident in the FFR, this sets up the possibility that dual language experience might also influence this phase-locked neural response to sound. The first study to examine this possibility focused on Spanish-English bilingual adolescents (ages 14–15 years) living in the United States (Krizman et al., 2012); this study revealed that bilinguals had larger FFR-F0 responses to a language-neutral, speech syllable "da" than age- and demographically matched English-speaking monolinguals (Krizman et al., 2012). Additionally, in the bilingual adolescent group (but not the monolingual group), increased neural representation of the F0 related to increased performance on a test of sustained auditory and visual attention, especially when the FFR speech stimulus was presented in background noise. A subsequent study, using "ga" and

"ba" stimuli, compared Spanish-English bilingual children (ages 8–9 years) who learned English and Spanish either simultaneously or sequentially, and therefore differed in the amount of bilingual experience (Krizman et al., 2015a). This follow-up study found that simultaneous bilinguals had larger FFR-F0 responses than sequential bilinguals, when considering children of roughly the same age (Krizman et al., 2015a). Like the adolescents, this enhancement was specific to the F0 and was not seen for the harmonics of the F0. There was also positive correlation between the magnitude of the neural response to F0 and years of bilingual experience; however, because age of acquisition AoA and duration of second language (L2) experience were highly inter-correlated in this study, the underlying driving factor (age vs. duration) could not be deduced.

These findings reviewed above converge to show that the phase-locked response to F0 is enhanced in Spanish-English bilinguals, but they raise the question of *Why the F0?* That is, what about bilingual experience might endow heightened neural responses to the F0? Human spoken languages vary with respect to their acoustic phonetic inventories; however, one common, perhaps even universal feature of human language, is that all speech sounds produced by a human vocal tract have a salient F0 that provides an important cue to talker separation (Darwin et al., 2003) and talker identification (Xu et al., 2013). Beyond that, there is the seemingly language-universal phenomenon for close vowels (vowels in which the tongue is close to the roof of the mouth) to have higher F0s than open vowels (Whalen and Levitt, 1995). Thus, across languages, increased neural sensitivity to the F0, as previously observed in Spanish-English bilinguals, may serve to improve vowel detection, independent of the language being spoken. More recent cross-linguistic work suggests that there are also language-specific differences in voice F0 (both its median value and range) that are unrelated to physiological differences in vocal production across speakers (Majewski et al., 1972; Dolson, 1994; Mennen et al., 2012; Andreeva et al., 2014). Based on this evidence, it can be argued that increased neural sensitivity to the F0 may aid in discriminating languages (Andreeva et al., 2014), something that bilinguals are particularly good at (Kuipers and Thierry, 2010). In addition, bilingual speakers have been shown to modulate their vocal F0 depending on which of their two languages they are speaking (Altenberg and Ferrand, 2006), making the F0 an important cue for monitoring within-speaker language switches. Moreover, on speech sound categorization tasks, bilinguals have been shown to place greater weight on the F0 compared to monolinguals (Llanos et al., 2013). The saliency of low-frequency acoustics, such as the F0 in the prenatal environment and the saliency of F0 during bilingual infant directed speech (Kyle Danielson et al., 2014), also suggest that the F0 is an acoustic cue that bilinguals may latch onto from an early age. Finally, within the FFR, the response to the F0, especially for harmonically complex speech stimuli, dominates over other harmonic components (Musacchia et al., 2007; Skoe and Kraus, 2010), suggesting that experience-dependent effects may be more readily observable for

the F0 compared to the other spectral components of the phase-locked response.

This collective evidence leads us to hypothesize that enhanced FFR-F0 responses are a characteristic feature of an auditory system that has undergone significant bilingual experience. If this hypothesis holds, we would expect the previous findings in children and adolescent Spanish-English bilinguals to generalize to adults with diverse language backgrounds. To test this hypothesis, we recruited a diverse group of college-age bilinguals, representing a total of 21 different languages (including English), and a wide range of age and duration of L2 exposure.

EXPERIMENTAL PROCEDURES

Participants

Participants included 53 young adults (10 males), ranging in age from 18 to 34 years. All participants were students at the University of Connecticut. Prior to any testing, participants signed an informed consent document approved by the institutional review board at the University of Connecticut. Consent, as well as all testing materials and instructions, were delivered in English. None of the participants had a (self-reported) history of chronic ear infections, premature birth, attention or neurological problems. For their participation, participants received either monetary compensation or course credit (their choice). Three additional participants were excluded from the final dataset because of excessively noisy FFR recordings due to electrical noise or high levels of myogenic noise.

All participants exhibited clinically normal auditory function with normal otoscopy, normal bilateral air conduction thresholds < 25-dB HL for octaves from 125 to 8000 Hz, and auditory brainstem response (ABR) wave latencies within normal limits for a 70-dB nHL click presented at 31.25 Hz (Skoe et al., 2015). Normal outer hair cell function was confirmed using a distortion product otoacoustic emissions (DPOAE) screening performed using a handheld screener (Madsen Alpha OAE + Screener, GN Otometrics). The F2 test frequencies were 1.5, 2, 3, 4, 5, and 6 kHz with an F2/F1 frequency ratio of 1.2 and an L1/L2 intensity ratio of 65/55 dB SPL. The passing criterion for the DPOAE screening was a DP signal-to-noise ratio > 6 dB for four of the six test frequencies.

Participants completed a survey of their medical history, musical training history, and language learning history. On the survey, participants rated the degree to which they were exposed to their first language (L1) vs. their L2 at different points in their life, broken down in increments of three years (i.e. 0–3, 3–6, 6–9, etc.), using a rating of 0, 25, 50, 75, or 100%, with 100% indicating exposure to L1 only and 0% indicating exposure to L2 only. In this study, we operationally define L2 AoA as the earliest age at which the participant was exposed to their L2 for at least 25% of the day. Based on this calculated AoA, participants were split into one of three categories: simultaneous

bilinguals, sequential bilinguals, or monolinguals. Total years of bilingual experience was then calculated by subtracting the calculated AoA from the age of the participant.

The "simultaneous bilinguals" reported that they were first exposed to their L2 between ages 0–3 ($n = 16$, two males). For this group, both L1 and L2 were spoken at home, and the current self-rated proficiency was 9.5/10 (on average) for L1, compared to 8.13/10 for L2, with 75% of their current language exposure being in English. The L1 languages included English ($\times$12), Mandarin, Spanish, Tamil, and Telugu. The L2 languages included English ($\times$4), French, Mongolian, Portuguese ($\times$2), Punjabi ($\times$2), Runyankore, Spanish ($\times$4), and the Fuzhou dialect of Chinese. Nine of the 16 simultaneous bilingual participants reported having lived outside the U.S. at some point during their life.

The "sequential bilinguals" reported that they were first exposed to their L2 after age 3 ($n = 21$, five males). For the sequential bilinguals, the age of L2 acquisition spanned from 3 to 18 years of age, with all but four participants having been exposed to their L2 before age 9. The average L1 self-rated proficiency was 8.74/10, relative to an L2 self-rated proficiency of 8.48/10. On average, 78.5% of their current language exposure was in English. Like the simultaneous bilinguals, many of the sequential bilinguals had lived outside the U.S. ($n = 13$); however the manner of L2 acquisition varied more than the simultaneous bilinguals. Most of the sequential bilingual participants learned L2 upon entrance into an English-speaking school system ($n = 9$) or spoke both L1 and L2 at home ($n = 8$). The others reported learning their L2 when they first moved to an English speaking country ($n = 3$), as part of a language class ($n = 2$) or through media exposure (e.g., TV, radio) ($n = 2$). The L1 languages included Albanian, English ($\times$6), Japanese, Korean, Lithuanian, Polish, Portuguese ($\times$4), Russian, Serbian, Sinhalese, and Spanish ($\times$3). The L2 languages included English ($\times$15), Hindi, Malayalam, Mandarin, and Spanish ($\times$3).

The English-speaking monolinguals all self-reported as such ($n = 16$, three males). All but four of the monolinguals had had some formal L2 instruction as part of their primary or secondary education, although their average self-rated proficiency was low (2.92/10). In contrast to the two bilingual groups, none of the monolinguals reported having lived outside the U.S., and the average amount of current English exposure was higher, at 95.31% (on average).

The three groups were statistically matched with respect to pure tone audiometric averages (PTA), using a clinical metric that averages across 0.5, 1, and 2 kHz (right ear: $F(2,50) = 1.25$, $p = 0.30$, left ear: $F(2,50) = 1.31$, $p = 0.28$) (Table 1). The groups were also matched on three measures of cognitive function administered through the NIH toolbox (http://www.nihtoolbox.org/) (Table 1). The NIH Toolbox test battery included the Pattern Comparison Processing Speed Test, the Dimensional Change Card Sort Test, and the Flanker Inhibitory Control and Attention Test (Table 1). These tests were administered according to the

Table 1. Group means (bold) and standard deviations for age, age of acquisition (AoA) for their second language (L2), self-rated current proficiency of L1 and L2, current English exposure, years of musical training, pure tone audiometric thresholds (pure tone averages (PTA)), and cognitive measures obtained with the NIH toolbox and IVA tests for the simultaneous (Sim) bilingual, sequential (Seq) bilingual, and monolingual (Mon) groups. Significant group effects are flagged with [*].

	Sim	Seq	Mono	*F, p*
Age (years)	**20.94**	**22.19**	**20.13**	2.40
	2.54	3.86	1.26	0.10
L2 AoA (years)[*]	**0.00**	**6.14**	**8.60**	17.07
	0.00	4.60	5.62	< 0.005
L1 Proficiency (/10)[*]	**9.50**	**8.74**	**9.81**	5.39
	0.89	1.36	0.54	0.01
L2 Proficiency (/10)[*]	**8.13**	**8.48**	**2.92**	51.91
	1.89	1.54	1.24	< 0.005
Current English Exposure (%)[*]	**75.00**	**78.57**	**95.31**	8.10
	12.91	19.82	10.08	< 0.005
Musical Training (years)	**6.81**	**5.05**	**2.69**	2.83
	6.33	4.74	3.28	0.07
Right Ear PTA (dB HL)	**8.54**	**10.16**	**10.94**	1.25
	4.83	4.11	4.30	0.30
Left Ear PTA (dB HL)	**7.29**	**9.68**	**8.54**	1.31
	4.34	4.04	5.05	0.28
NIH Toolbox Pattern Comparison (percentile)	**78.13**	**65.90**	**66.04**	0.70
	26.43	34.37	40.42	0.50
NIH Toolbox Dimensional Card Sort (percentile)	**67.66**	**61.90**	**73.49**	1.61
	15.84	24.12	15.61	0.21
NIH Toolbox Flanker (percentile)	**64.80**	**66.82**	**65.48**	0.04
	21.69	18.28	24.03	0.96
NIH Toolbox English Vocabulary Test (percentile)	**62.88**	**65.71**	**70.88**	0.39
	31.25	26.86	18.70	0.68
IVA Attention Quotient (standard score)	**98.63**	**91.40**	**90.88**	0.43
	15.32	30.77	29.97	0.65

guidelines of the NIH Toolbox. The groups also did not differ on the IVA + Plus Test (Brain Train, Inc), a test of integrated visual and auditory continuous performance, for the attention quotient or any of the other subscales (Table 1).

The groups, however, differed on their self-reported L1 proficiency, with the sequential group rating themselves lower than the other two groups (Table 1), likely because for many of these participants, L1 was no longer their primary language of communication. However, on an English vocabulary test (Picture Vocabulary Test, NIH Toolbox) the groups performed similarly ($F(2,50) = 0.39$, $p = 0.68$). There was also a trend for the groups to differ with respect to the duration of musical training, with the monolinguals having (on average) fewer years of musical training than the two bilingual groups (Table 1). Previous work utilizing the same "da" stimulus as the current investigation found that young adult musicians did not differ from non-musicians with respect to the FFR-F0 response (Parbery-Clark et al., 2009). However, because of the previously established relationship between FFR-F0 amplitude and musical training for other stimuli (Bidelman and Krishnan, 2010; Skoe and Kraus, 2012; Weiss and Bidelman, 2015), we included the duration of musical training (years) as a covariate in all analyses.

Electrophysiological recording

FFRs were recorded to a synthesized speech syllable /da/. To facilitate comparisons with previous FFR investigations, the stimulus was identical to that used in Krizman et al., 2012, with similar recording parameters. The stimulus, which has been described in detail elsewhere (e.g., Johnson et al., 2008; Song et al., 2012; Coffey et al., 2016), is a 170-ms six-formant sound, with a stable F0 at 100 Hz. During the first 50 ms, the first formant (F1) ramps up in frequency from 400 to 720 Hz, and F2 and F3 ramp down (1700–1240 and 2850–2500 Hz, respectively). This formant transition period is followed by a 120-ms vowel period, during which the F0 and all formants are constant. For the entire syllable, F4–F6 are held constant at 3330, 3750, and 4900 Hz.

The phonemes /d/ and /a/ are included in the phonetic inventories of many languages (Maddieson, 1984), making them nearly language universal. Among the languages spoken by our bilingual participants, all have some variant of a voiced alveolar/dental stop (/d/) as well as an open central vowel (/a/) (Maddieson, 1984). The particular synthesized [da] stimulus token used in our study was created to be language neutral; it does not strongly invoke a particular language, especially when played in fast repetition under conditions of passive listening, as was done here.

Stimulus delivery was controlled in MATLAB R2015a using Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). The stimulus was outputted by a USB-powered external sound card (M-Audio M-Track) at an LAeq of 75.2 dB and a rate that varied between 4.5–4.8 Hz. Stimuli were presented to the right ear in alternating polarities through a Mu-metal electromagnetically (EM) shielded transducer (ER-3), with custom EM

shielding provided by Intelligent Hearing Systems Miami, Florida, USA. The sound outputted by the transducer was delivered into the ear canal via 10 cm of plastic tubing that had a foam ear insert (ER3-14, Etymotic Research Inc., Elk Grove Village, Illinois, USA) secured to its terminus. The audio stimulus was also routed from the soundcard through a StimTrak box (Brain Products GMBH) into an auxiliary channel of the EEG amplifier (actiCHamp amplifier, Brain Products GMBH). This auxiliary recording provided accurate stimulus timing information when averaging the responses offline.

A three-electrode vertical, ipsilateral montage was used, measuring from Cz to the right ear, with the ground electrode placed on the forehead. Prior to placing the Ag–AgCl multitrodes on the scalp, the electrode sites were mildly exfoliated to produce impedances under 5 kOhms. The electrodes were plugged into a bipolar amplifier that provided a gain factor of 50 (EP-preamp module, Brain Products, GmbH, Gilching, Germany). The bipolar amplifier interfaced with the actiCHamp amplifier (Brain Products GMBH) through a splitter box. Recordings were made with a 25-kHz sampling rate in the Recorder software (Brain Products, GmbH), with no on-line filters applied.

All recordings were made passively with the participant sitting reclined comfortably in an EM-shielded double-walled sound booth (IAC Acoustics, Winchester, United Kingdom) while watching a self-selected, muted video with English captions. The video was projected onto the wall of the booth, about five feet from the participant's head, using a ceiling mounted LCD projector placed outside the booth window. As is common for FFR experiments, participants were instructed to sit quietly in the dimmed room and watch the video (Skoe and Kraus, 2010).

Data averaging and analysis

Data averaging and analysis utilized custom routines developed by the first author. In MATLAB, the electrophysiological responses were filtered from 80 to 950 Hz using a third order, one dimensional Butterworth filter. The recording was then epoched using a 230-ms window that spanned 30 ms prior to stimulus onset, and extended 200 ms after stimulus onset. The timing of each stimulus onset was derived from the auxiliary channel recording. After excluding epochs in which the amplitude exceeded ±140 μV, 2000 trials of each stimulus polarity (4000 trials total) were averaged. The process of averaging responses to alternating stimulus polarities accentuates the lower frequency components of the FFR, including the response to the F0 (Chimento and Schreiner, 1990; Aiken and Picton, 2008).

As in Krizman et al., 2012, the FFR amplitude was calculated separately for the response to the formant transition region and the response to the steady-state (vowel) region. These time regions (20–60, 60–180 ms), take into account the lag between the stimulus onset and the emergence of the phase-locked FFR (~20 ms) in addition to the time it takes for the phase-locked response to ramp down after stimulus offset. Although the F0 is stable throughout, the amplitude of the response to the F0 varies

between these two time regions as a consequence of the unique spectrotemporal make-up of the formant frequencies of the stimulus during the formant transition versus the steady-state region (Skoe et al., 2011). For each time region, the long-term spectral profile of the FFR was derived using a discrete fast Fourier transform (FFT), after first applying a Hanning ramp. Zero-padding was applied as part of the FFT to increase the sampling rate of the spectral estimate, yielding a spectral resolution of 1 Hz. The output was scaled to microvolts by taking the absolute value of the FFT and then multiplying the output by two divided by the signal length. To derive the amplitude of the FFR-F0 response, the spectral amplitudes were averaged between 80 and 120 Hz, as in Krizman et al., 2012. To derive the amplitude of the FFR-harmonic response, we adopted a similar approach to previous reports (Parbery-Clark et al., 2009; Krizman et al., 2015a): first, the response amplitude to each harmonic (between H2 and H8) was found by averaging over 40-Hz bins surrounding integer multiples of 100 Hz, then the responses were averaged across the seven harmonics to derive a composite measure of harmonic encoding.

In addition to calculating the spectral amplitude of the response to the F0 (as was done in (Krizman et al., 2012)), we probed other dimensions of the phase-locked neural response to F0 beyond its magnitude. These additional measures of phase-locking precision included: (1) the bandwidth of phase-locked response to F0 (*Q-3 Bandwidth*); (2) The accuracy of the FFR to capture 100 Hz, the driving frequency (*Frequency Error*); and (3) the "*Pitch Strength*" of the phase-locked response (Krishnan et al., 2005, 2011). To calculate the bandwidth (Q-3), we first identified the dominant spectral peak in the 80–120-Hz range of each long-term response spectrum, to extract the center frequency, fc. From there, we calculated the frequency range or Δf that was 3 dB down from either side of fc, after first converting the signal to dB. Q-3 *Bandwidth* was defined as Δf, with higher values reflective of wider bandwidths. *Frequency Error* was defined as the absolute value of 100 Hz minus fc (Song et al., 2008; Li and Jeng, 2011). *Pitch Strength* was derived from the long-term autocorrelation function of the time–amplitude response waveform by finding the height of the first peak in the autocorrelogram (Krishnan et al., 2005, 2011). For a signal with a F0 of 100 Hz, the first peak occurs at a time lag of 10 ms. Larger *Pitch Strength* values (closer to one) are reflective of responses that are more periodic in nature.

Verifying neural origin of the recordings

Stimulus artifact is a concern when measuring FFRs, given the early latency of the response and its close resemblance to the stimulus (Akhoun et al., 2008; Skoe and Kraus, 2010; Campbell et al., 2012). To safeguard against stimulus artifact, a number of precautions were taken including (1) using Mu-metal shielded transducers, (2) using tubephones to create distance between the transducer and the electrodes placed on the scalp, and (3) using alternating polarity stimulation. As a demonstration of the combined effectiveness of these techniques, FFRs were recorded in a single participant using the standard testing procedures and then again with the same

procedures except with tape covering the tubephones that blocked the participant from hearing the sound stimulus (Fig. 1). The two recordings were processed in the same fashion and then plotted in the frequency domain, using a running window spectrogram to illustrate spectral amplitudes over time. In Fig. 1, note that the spectral peaks that are present at 100 Hz (F0) and its harmonics when the participant can hear the stimulus are absent when the sound is playing using the same hardware and software but is inaudible to the participant. From this we conclude that our measurement of the FFR reflects neurogenic activity and that the electrical activity recorded at the F0 and its harmonics is not an artifact produced by our equipment.

RESULTS

To test whether the previous findings generalized to adult bilinguals (Krizman et al., 2012), we began by comparing the magnitude of the phase-locked response to F0 between monolinguals and simultaneous bilinguals. The groups were compared using a multivariate analysis of covariance (ANCOVA), which like all other subsequent analysis covaried for the duration of musical training (Table 1). As in the Krizman et al., 2012 study, the groups did not differ for the transition region of the stimulus (20–60 ms time window ($F(1,29) = 1.85$, $p = 0.19$)), but they did differ for the steady-state region corresponding to the vowel ($F(1,29) = 5.57$, $p = 0.025$), where the simultaneous bilinguals were found to have more robust phase-locked responses to F0 (Fig. 2). As a further replication of the previous work, no main effect of group was evident for the phase-locked response to the harmonics (composite measure H2–H8, $F(1,29) = 0.11$, $p = 0.74$) (Fig. 2). In Fig. 2 (top right panel), the group-average amplitudes are plotted for the response to F0 and its harmonics (up to H8), to illustrate that the effect is limited to 100 Hz (the F0) and does not extend to the higher speech harmonics.

The next set of analyses examined whether the age of L2 acquisition (L2-AoA) influenced the magnitude of the

phase-locked response to F0. This was done by introducing the simultaneous bilingual group into the analysis. Previous work in children (ages 8–9) found that the magnitude of the phase-locked response to F0 was larger in children who learned L1 and L2 simultaneously compared to those who learned L1 and L2 sequentially. In our dataset, the age of L2 exposure spanned from 0 to 18 years, with a heavier sampling on the younger than the older end of the spectrum. Fig. 3 shows the mean, along with the distribution of the FFR-F0 responses, for each L2-AoA sub-grouping compared to the monolinguals. For the small number of participants who learned L2 after age 9, Fig. 3 reveals that the magnitude of the FFR-F0 response falls on the lower end of the range of values expressed in the other groups; however, due to the limited sample, these four data points were not included in any statistical treatments. To examine the effect of AoA on the magnitude of the FFR-F0 response during the steady-state region, we focused only on those who learned L2 before age 9 ($n = 33$), and we refer to this subset of bilinguals as "early bilinguals". Within this group of early bilinguals, no main effect of L2-AoA was found when considering the three AoA groupings ($F(2,29) = 0.08$, $p = 0.93$), even when co-varying for the current age of the participants (in addition to the duration of musical training) ($F(2,28) = 0.13$, $p = 0.88$). As a complement to this analysis, we examined the relationship between the FFR-F0 amplitude (during the steady-state period) and total years of bilingual experience. For this, we again covaried for the duration of musical training. The outcomes of this partial correlational analysis revealed that years of total bilingual experience, which spanned from 15 to 28 years in this group of early bilinguals, did not correlate with the FFR-F0 amplitude ($r = -0.17$, $p = 0.36$).

Although the three AoA groups did not differ from each other, they (collectively as a composite group of 33 early bilinguals) differed from the monolinguals with respect to the FFR-F0 amplitude for the vowel region ($F(1,46) = 5.89$, $p = 0.02$) (Fig. 4). Moreover, although the monolinguals had less musical training on average than the bilinguals (Table 1), FFR-F0 amplitude did not vary as a function of the duration of musical experience ($r(49) = 0.11$, $p = 0.45$). The group difference for FFR-F0 held even after removing the tonal language speakers ($n = 6$; Punjabi × 2, Mandarin × 2, Mandarin × 1, Fuzhou dialect of Chinese) from the sample of the early bilinguals ($F(1,40) = 5.05$, $p = 0.03$). For all subsequent analyses, the data were pooled across the early bilinguals, including both tonal and non-tonal speakers.

After observing that the early adult bilinguals had larger phase-locked neural responses to the F0 than their
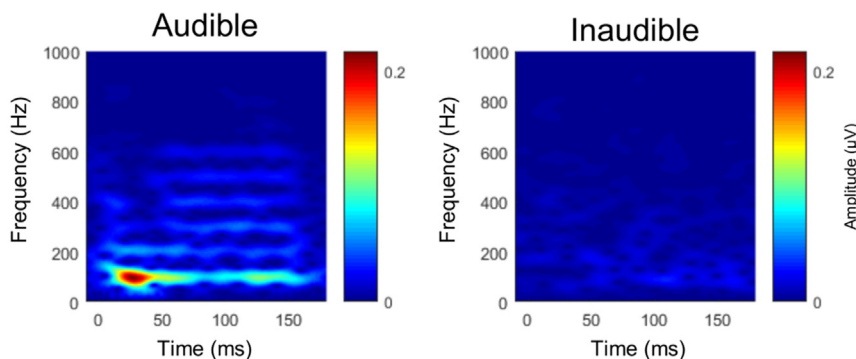
Fig. 1. The frequency-following response (FFRs) to the fundamental frequency (F0) is neurogenic not artifact. "FFRs" were first measured with the insert earphone placed in the ear canal with the sound output at a LAeq of 75.2 dB (left). For comparison, the sound was then played through the same hardware-software setup but with the output of the tubephones mechanically blocked, making the sound inaudible to the listener (right). Recordings are plotted using a running window spectrogram to illustrate the spectral profile of the recording over time. In the standard condition (left), there are clear energy bands at 100 Hz, the fundamental frequency (F0) of the speech stimulus, as well as its harmonics. When the stimulus is played in the same fashion (same equipment, same software, and same intensity) but the sound is made inaudible to the listener, there is no discernable information present at the F0 or harmonics.
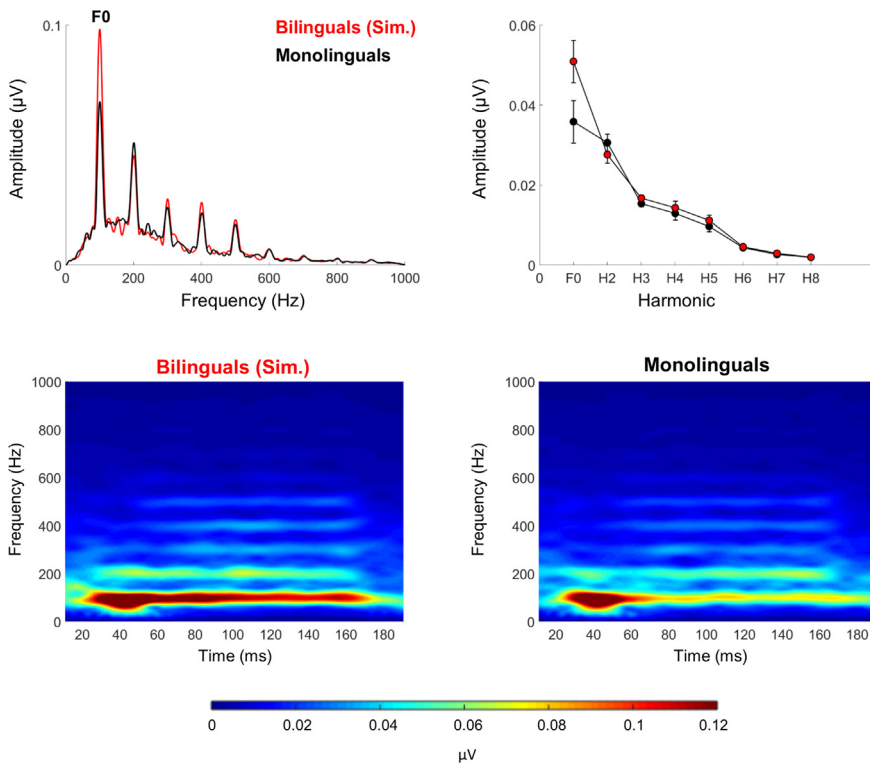
Fig. 2. Bilinguals show larger neural response to the fundamental frequency (F0) in the frequency following response (FFR) to the vowel component of the /da/ speech stimulus (60–180 ms). In the top left, the group average response spectra are plotted for the simultaneous bilinguals (red) and monolinguals (black) for the vowel-region of the response. Note that the response at 100 Hz, the F0 of the stimulus, is larger in the simultaneous bilinguals and that the response to the harmonics (integer multiples of F0) is more comparable for the two groups. In the top right, the group mean ±1 standard error of the mean are plotted for the average response amplitude of 40-Hz bins surrounding the F0 (80–120 HZ) and each of its harmonics (H2: 180–220 Hz, H3: 280–320 Hz, etc.). In the bottom left, the group average response spectrogram is plotted for the simultaneous bilinguals to illustrate how the response to the F0 varies over the duration of the stimulus. Compared to the simultaneous bilinguals, the monolinguals (bottom right) have a reduced response to the F0 during the vowel portion of the response (60–180 ms). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
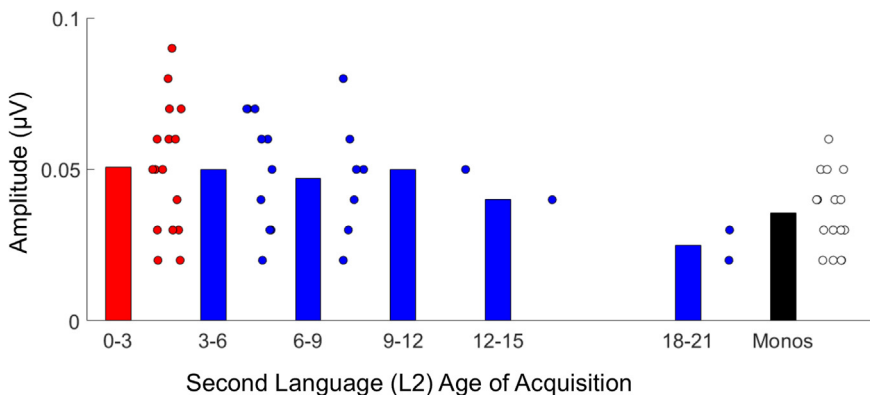

Fig. 3. The amplitude of the frequency following response (FFR) to the fundamental frequency (F0) as a function of the age of language acquisition (AoA) for the second language (L2). For each AoA subgroup, the bars reflect the group average amplitude and dots reflect individual data points for the response, averaged over the 80–120-Hz bin. For the three subgroups that learned L2 early in life (0–3, 3–6, 6–9 years), the amplitude of the F0 does not display AoA-dependent effects. The group plotted in red represents the simultaneous bilinguals (AoA = 0–3 years), whereas the groups plotted in blue represent the sequential bilinguals in the dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

monolingual counterparts, we then probed other dimensions of the neural response. Within the response to the steady-state region of the stimulus, we assessed the accuracy of the FFR to capture the driving frequency (*Frequency Error*), the bandwidth of the FFR-F0 (*Q-3 Bandwidth*), and the *Pitch Strength* of the phase-locked response. On average, the early bilingual group had more precise phase-locking compared to the monolingual group, as indicated by lower *Frequency Error* (0.88 Hz ± 5.53 vs. 2.50 Hz ± 4.57), narrower *Q-3 Bandwidth* (15.93 ± 4.34 Hz vs. 18.50 ± 17.00 Hz) and higher *Pitch Strength* (0.60 ± 0.18 vs. 0.54 ± 0.21) (Fig. 4). However, none of these differences were statistically significant ($F(1,46) = 0.67$, $p = 0.42$; $F(1,46) = 0.23$, $p = 0.631$; $F(1,46) = 0.49$, $p = 0.49$ for *Frequency Error*, *Q-3 Bandwidth*, and *Pitch Strength*, respectively). Thus, the primary difference between the early bilinguals and monolinguals appears to be that the early bilinguals have larger phase-locked responses to the F0.

DISCUSSION

We compared differences in basic auditory processing between monolinguals and bilinguals of diverse language backgrounds using the FFR, a phase-locked neural response, and found that early bilingualism is associated with more robust neural responses to the F0, an important acoustic cue for linguistic and non-linguistic communication. This finding of enhanced (i.e., larger) neural responses to the F0 in adult early bilinguals replicates previous work in Spanish-English bilingual adolescents that used the same stimulus, with similar recording parameters. This replication of findings is even more striking when you consider that outside of having been exposed to two phonetic systems, "no two bilinguals, or groups of bilinguals are the same" (Hartsuiker, 2015). By demonstrating that the magnitude of the neural response to the F0 is larger in early bilinguals of various ages and back-
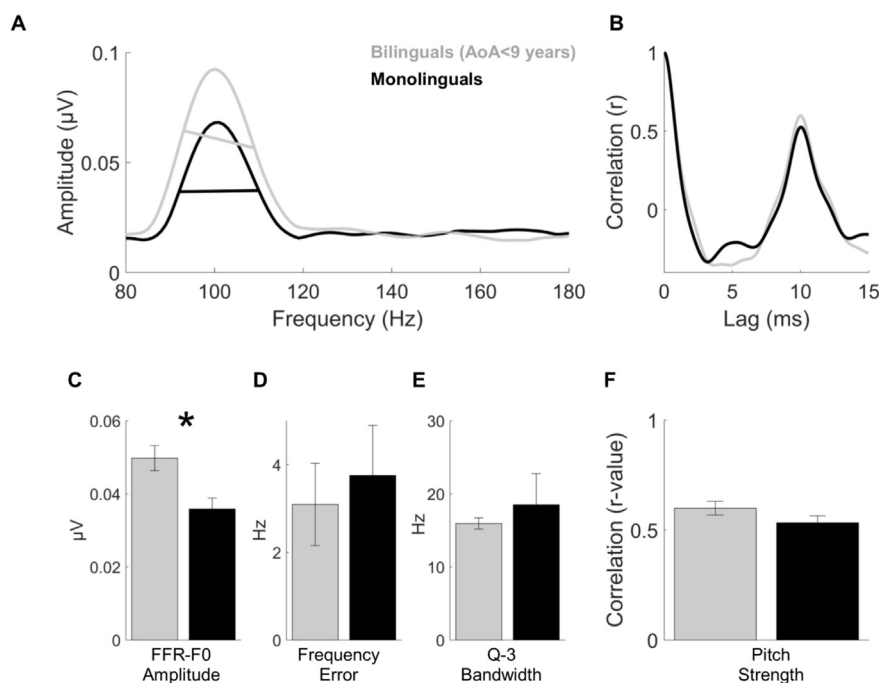
Fig. 4. Comparisons between early bilinguals (gray, *n* = 33) and monolinguals (black, *n* = 16) on various dimensions of the frequency following response (FFR) to the fundamental frequency (F0). Early bilinguals are defined based on having learned their second language before age 9. In panel (A), the group average response spectra are plotted for the vowel-region of the response. The spectrum is plotted to highlight the neural response at 100 Hz, the F0 of the speech stimulus. For illustrative purposes, the *Q-3 Bandwidth* of each group average spectrum is plotted in panel A using horizontal lines. In panel (B), the group average autocorrelation functions are plotted. *Pitch Strength* is defined as the magnitude of the first peak of the autocorrelation function. In the lower panels (C-F), group means for the FFR-F0 amplitude (C), *Frequency Error* (D), *Q-3 Bandwidth* (E), and *Pitch Strength* (F) measures are plotted. The FFR-F0 amplitude measure represents the average response between 80 and 120 Hz, following the procedures described in Krizman et al., 2012 (Krizman et al., 2012). Errorbars represent one standard error of the mean. This graph illustrates that the early bilinguals have larger, but not more precise, phase-locked responses to a speech stimulus with a steady-state F0. * < 0.05

grounds, this serves as compelling support for our hypothesis that enhanced FFR-F0 responses is a hallmark feature of early exposure to multiple languages and not a unique characteristic of Spanish-English bilinguals. Moreover, by showing that the effect of early bilingualism is specific to the magnitude of the FFR-F0 response (and not the magnitude of the response to the harmonics nor measures of phase-locking precision), our findings, together with the previous work, suggest that bilingualism is associated with increased neural gain that is specific to the F0 of auditory input.

Bilingualism, the ability to speak more than one language, can be defined across different dimensions, including, but not limited to, language proficiency, daily usage of each language, the extent of code-switching, and age of language acquisition (Hamers and Blanc, 2000; Hapsburg and Peña, 2002). The current study focused on the timeline of language acquisition, with an emphasis on the age of first contact with the L2 (AoA). In children (ages 8–9 years), the magnitude of the FFR-F0 response has been shown to decrease as a function of L2 AoA (Krizman et al., 2015a); although the lack of a monolingual group in this previous work limited the scope of how this finding could be interpreted. In the cur-

rent study, where the participants were more than a decade older (in addition to representing a more diverse set of languages), we found that bilinguals who learned L2 before age 9 had larger phase-locked responses to the F0 than monolinguals, but that there was no AoA-graded effect among these bilinguals. However, we concede that the non-continuous nature of our AoA index may be blurring more fine-grained effects related to AoA and years of bilingualism effects. With this caveat in mind, our findings, when interpreted together with previous work (Krizman et al., 2012, 2015a), suggest that any differences observed between simultaneous and sequential bilinguals during childhood with respect to the FFR, may level off over time as both groups accrue more experience with L2. Thus, early bilinguals who learn L1 and L2 sequentially may initially "lag" behind those who learned both simultaneously, but eventually the sequential learners may "catch-up" and converge with the simultaneous bilinguals to reach comparable FFR-F0 responses. The timeline of this convergence, and the experiential and neurodevelopmental factors that contribute to this putative convergence warrants further investigation. On that note, Fig. 3 reveals individual variability in the neural magnitude of the response within the different AoA sub-groups, as well as the monolinguals. This variability is a reminder that the FFR, like other neural phenomena, is overdetermined. That is, multiple different endogenous and exogenous factors can influence the FFR, with bilingualism being just one of those factors (reviewed in Skoe and Kraus, 2010, 2014; Skoe and Chandrasekaran, 2014). Understanding how different experiential and genetic factors interact to shape the auditory system, we believe, is the key to deciphering the full-range of individual differences in the FFR that are observed in typically developing populations (Chandrasekaran et al., 2012; Hairston et al., 2013; Skoe et al., 2013, 2017; Skoe and Chandrasekaran, 2014). Our study makes a valuable contribution to this effort by providing evidence that bilingual experience is one such factor that must be taken into consideration when examining neural indices of basic auditory processing in adults. However, our study cannot adequately address the impact of learning an L2 after age 9, given the limited number of late bilinguals in our sample. Future examinations of late bilinguals will need to take into account age-related changes to the FFR that have recently been documented (Skoe et al., 2015; Krizman et al., 2015b). The phase-locked response to the F0

increases in amplitude between infancy and early child-hood, reaching its maximum amplitude between ages 5–8, after which the response amplitude declines progressively. Drawing on work in musicians (Skoe and Kraus, 2013), we hypothesize that the potential for bilingual-related enhancements of the FFR is greater during the developmental upswing than at later developmental time points when bilingual experiences must compete with developmental processes that contribute to diminished FFR amplitudes. Coincidentally, the developmental upswing period coincides with the age window when (spoken) language learning occurs with relative ease, compared to later in life (Johnson and Newport, 1989). Longitudinal studies are needed to test this developmentally constrained hypothesis of experience-dependent auditory plasticity and to understand its full ramifications.

Our recruitment efforts yielded a diverse sampling of languages, allowing us to make statements about the cross-linguistic generalizability of our findings. However, our sampling was by no means exhaustive, leaving questions open for future work. For instance, because all of the individuals in this and the previous FFR bilingual studies spoke English in combination with another language, more extensive cross-linguistic analyses are needed to rule out the unlikely possibility that English is a driving factor. Follow-up studies should also explore how different languages act in isolation and combination to affect the FFR. This issue is of particular importance for tonal languages given the role of F0 in establishing lexical pitch. Previous investigations of Mandarin and Vietnamese speakers suggest that experience with a tonal language endows more robust and faithful FFR tracking of pitch in linguistic and non-linguistic contexts (Krishnan et al., 2009). Much of this work has focused on the precision with which vocal F0 is tracked within a dynamic pitch contour, using measures of autocorrelation (*Pitch Strength*) and Frequency Error (in Hz), with less focus on the spectral magnitude of the FFR to the F0 (Krishnan et al., 2005, 2009; Krishnan and Gandour, 2009; Jeng et al., 2011). The current study found that early bilinguals had larger neural phase-locked responses to the F0 than monolinguals for a speech stimulus with a flat F0 contour but that the precision of the phase-locked response was not different between groups, as assessed by the *Q-3 Bandwidth*, *Frequency Error*, and *Pitch Strength* of the response to F0. The few studies of tonal language speakers that have examined the spectral magnitude of the FFR report a variety of outcomes, including: (1) that the magnitude of the neural response to the second harmonic is consistently larger for Mandarin speakers relative to English speakers across the four Mandarin pitch contours (Krishnan et al., 2005); (2) that there is a gradation of the F0 response as a function on the linguistic-status of the stimulus timbre (Krishnan et al., 2011); and (3) that tonal language speakers have larger FFR-F0 responses to musical notes compared to non-musician, non-tonal language speakers but not compared to musicians without tonal language experience (Bidelman et al., 2011). Thus, with respect to the magnitude of the FFR-F0 response, there are apparent similarities between our findings and previous work in

tonal language speakers, suggesting a possible confounding effect of bilingualism in this previous work. This opens the previous findings on tonal language speakers up for reinterpretation (Bidelman et al., 2011; Krishnan et al., 2011), especially given that our primary finding of enhanced FFR-F0 responses held even after removing the tonal language speakers from the sample, and also when accounting for differences in musical training. One aspect that goes undiscussed, or at best under discussed, in the previous work is that the tonal language speakers were typically college-students at an American university, implying that they had sufficient English skills, and by extension were sufficiently bilingual, to be considered a native speaker of a tonal language and to be enrolled at an English-speaking university. While we do not go so far as to claim that all demonstrations of enhanced FFRs in tonal language speakers are due to their bilingual status, we do raise the possibility that the combination of being a tonal language speaker and being bilingual might accentuate the degree to which tonal language speakers differ from monolinguals, especially with respect to the magnitude of the phase-locked response to the F0. In a similar vein, being bilingual in combination with having extensive musical training, might likewise compound the differences between bilinguals and monolinguals with respect to the FFR-F0 response (Skoe and Chandrasekaran, 2014). Although we controlled for differences in musical training in our statistical comparisons, we recognize the need to further investigate how bilingualism may act in concert with other acoustically enriched or deprived conditions to influence the neural response to the F0 (Skoe and Chandrasekaran, 2014; Krizman et al., 2016b).

Neural mechanisms

The FFR is a far-field neural response resulting from phase-locked activity to the stimulus along the auditory neuroaxis (Moushegian et al., 1973; Krishnan, 2006). There are ample demonstrations of the plastic, experience-dependent nature of the FFR (with respect to both its magnitude and precision) (e.g., Musacchia et al., 2007; Song et al., 2008; Krishnan et al., 2009), giving us firm footing to conclude that the differences observed between monolinguals and early bilinguals reflect experience-driven neural mechanisms that lead to increased neural gain. Drawing on studies of auditory deprivation and enrichment in animal models, we speculate that the group differences are the result of the early bilinguals being exposed to increased acoustic diversity during CANS development, a necessary consequence of listening to and speaking multiple languages early in life. Based on our pattern of results, these experience-dependent processes appear to be highly specific, given that they do not generalize to the higher harmonic components of the vowel nor the initial formant transition of the syllable. With respect to the F0, we observed FFRs of higher magnitude; yet, this increased gain was not accompanied by increased phase-locking precision (as measured by *Q3-Bandwidth*, *Frequency Error*, and *Pitch Strength* measures). However, we acknowledge that by

using a stimulus with a flat F0 contour instead of a varying the F0, that this may not have maximally taxed the auditory system and its phase locking precision. Thus, we do not rule out the possibility that more precise phase-locking could be observed in early bilinguals compared to monolinguals if the F0 contour were more complex.

Phase-locking is a general principle of the central auditory system. Throughout the auditory system, neurons synchronize their spiking patterns to, by firing at a particular phase of the stimulus (Langner, 1992; Joris et al., 2004; Liu et al., 2006; Middlebrooks, 2008). The upper frequency limit of phase-locking decreases at each stage along the CANS pathway, with progressively longer latencies also being observed. In humans, the upper limits of phase-locking of various auditory structures have been difficult to precisely characterize because of the invasive, intracranial techniques that are required (Verschooten et al., 2015). Moreover, the upper limits of phase-locking appear to be highly species-dependent, which prevents generalization from experimental to human models (Steinschneider et al., 1980; Langner and Schreiner, 1988; Kuwada et al., 2002; Liu et al., 2006; Verschooten et al., 2015). As a result, much of what we know about the phase-locking properties of the human auditory system has come from non-invasive techniques such as the scalp-recorded FFR. For any given stimulus frequency, the FFR is a composite response, reflecting the aggregation of volume-conducted activity from different auditory structures that are phase-locking to the stimulus. Depending on the frequency of the stimulus, and the relative latencies of the underlying generators, the volume-conducted signals from difference sources will either sum constructively or destructively at the scalp (Gardi et al., 1979; Tichko and Skoe, 2017). This composite multi-generator nature of the FFR has led to much debate over the sources of the FFR, especially for vocal pitch frequencies (∼80–225 Hz), which span a frequency range where cortical phase-locking transitions from being weak to being absent (Herdman et al., 2002; Kuwada et al., 2002). In the case of the current study, we can gain insight into the putative neural mechanisms by referring to a recent investigation by Coffey et al. which recorded magnetoencephalography (MEG) in young adults using the same speech stimulus (Coffey et al., 2016). Coffey et al. found that the FFR to the 100 Hz F0 (as measured by MEG) reflected activity from cochlear nucleus, inferior colliculus, medial geniculate, and also right auditory cortex, whereas the phase-locked response to the harmonics (200 Hz and above) were determined to be primarily subcortical in origin (Coffey et al., 2016). Although the weighting of the FFR generators may be different between EEG and MEG techniques and between individual brains for a given technique, this recent study offers evidence that the human auditory cortex can phase lock to frequencies up to 100 Hz, and that auditory cortex may, therefore, contribute to the FFRs that we recorded. Moreover, although details about the participants' language backgrounds were not provided and so the bilingual status and native language of the participants cannot be factored into the interpretation of the results, it's interesting to note that this study by Coffey et al. took place in Montreal, a city that boasts a large bilingual population. With these findings and details in mind, we put forward several possible mechanistic scenarios that could give rise to larger FFRs to an F0 of 100 Hz in early bilinguals. One possibility is that early bilinguals display a system-wide gain in phase-locking to the F0, which leads to more robust phase-locking in multiple generators of the FFR, as a consequence of either more neurons contributing to the FFR in early bilinguals compared to monolinguals, or the same set of neurons firing more synchronously in the early bilingual group. A system-wide gain could also potentially emerge as a consequence of increased top-down corticofugal modulation of the subcortical auditory system by the auditory cortex (for a review see: (Krishnan et al., 2010)). A second, but not mutually exclusive, possibility is that the FFR from early bilinguals includes more cortical activity for this F0 compared to monolinguals. This scenario is in line with the findings of Ressel et al. (2012) who found that auditory cortex is larger in early bilinguals compared to monolinguals (Ressel et al., 2012) and findings by Vihla et al., (2002) who found that bilinguals have increased auditory cortical activation to pure tones and speech-like complex sounds. Because cortical generators have longer ramp-up times (i.e., longer latencies) than subcortical ones, this may potentially explain why the findings were limited to the latter half of the stimulus. Although the F0 was stable throughout our stimulus, the spectrotemporal profiles of the upper harmonics were dynamic during the first 50 ms of the stimulus and then stable thereafter. Previous work has highlighted that the phase of the phase-locked response to F0 of this stimulus can be influenced by the dynamics of the frequencies that exceed the phase-locking limits of the auditory system (Skoe et al., 2011), suggesting that the phase-locked response at the F0 reflects more than just the response *to* the stimulus F0. By adopting a wider array of F0s (that will tap into cortical vs. subcortical generators to varying degrees), by using stimuli in which the steady-state and dynamic portions are counterbalanced (e.g., /da/ vs. /ad/ vs /ada/), and by manipulating the linguistic status of the stimulus (e.g., native vs. non-native, speech vs. non-speech), a more nuanced picture of the bilingual auditory system, as well as the neuromechanistic forces behind the enhancements observed here will likely emerge. In addition, the use of FFRs, in combination with other imaging techniques, will allow for further evaluation of whether the neural gain is local to one generator of the FFR or more system-wide.

Potential behavioral advantages of enhanced phase-locked responses to the F0

At this point it remains unclear whether the enhanced neural response is simply a byproduct of early auditory enrichment or whether it provides a behavioral advantage to the bilinguals as adults. In this section we explore the potential (non-mutually exclusive) implications of having an enhanced neural response to the F0. The first possibility is that a heightened neural response to the F0 might give bilinguals an advantage on tasks where different F0s must be discriminated,

such as a pitch discrimination task or an auditory stream segregation task. It has previously been shown that in monolingual listeners (ranging in age and hearing thresholds) that the FFR-F0 magnitude (measured using a signal to noise measure) relates to frequency discrimination (Marmel et al., 2013), supporting the possibility that larger neural responses to the F0 in bilinguals might be associated with more refined sensory acuity for perceptual tasks that involve the F0, such as frequency/pitch discrimination. In further support of this possibility, Montagni and Peru (2011) provide evidence that early exposure to an L2 confers an advantage to pitch processing tasks for both linguistic and musical stimuli (Montagni and Peru, 2011). Follow-up studies should also explore whether bilinguals have better F0 discrimination on speech and non-speech tasks by using behavioral and neural tests in which the stimulus token differ minimally in pitch. A second possibility is that a heightened neural response to the F0 is associated with more fluid language switching, as well as improved talker learning and identification, given the importance of F0 in signalling who is speaking and what language is being spoken (Kuipers and Thierry, 2010; Bregman et al., 2012). Third, in monolingual speakers, heightened neural responses to the F0 have also been associated with better speech perception in noise (Anderson et al., 2010; Song et al., 2010). This has led to the speculation (Krizman et al., 2016a) that sensory enhancements of basic auditory cues, such as the F0, are the outcome of a compensatory mechanism aimed as redressing the issues that bilinguals face when listening to speech in noise (Mayo et al., 1997; Shi, 2012). Another question that remains to be addressed is whether these sensory enhancements generalize across linguistic and non-linguistic stimuli as well as static and dynamic F0 contours. Understanding the generalizability of our findings to other stimuli, would provide further insight into the potential real-world benefits of having a heightened neural response to the F0. Bidelman and colleagues found that tonal language speakers and musically trained individuals had similar FFR enhancements to musical stimuli relative to monolinguals, but that unlike the musicians, these enhancements did not translate into heightened musical pitch discrimination abilities in the tonal language speakers (Bidelman et al., 2011). This finding, which may be confounded by the bilingual status of the tonal-language participants (a point raised above), emphasizes that sensory-level enhancements do not necessarily translate to overt behavioral advantages, especially for tasks that are either not behaviorally relevant to the listener or for which the listener has limited experience. This does not rule out the possibility that sensorineural gains in bilinguals have behavioral correlates but it suggests that such correlates might be confined to specific auditory/communicative tasks.

CONCLUSION

In conclusion, our findings converge with previous work to suggest that early bilinguals from diverse (spoken) language backgrounds, have enhanced basic processing of sound that emerges even during pre-

attentive listening. This enhancement gives rise to a more robust neural response to the F0, not harmonics, of harmonically complex sounds. The generalizability of the findings is suggestive of a neural signature of bilingualism within the CANS (Kraus and Nicol, 2014).

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

REFERENCES

Aiken SJ, Picton TW (2008) Envelope and spectral frequency-following responses to vowel sounds. Hear Res 245:35–47.

Akhoun I, Moulin A, Jeanvoine A, Menard M, Buret F, Vollaire C, Scorretti R, Veuillet E, Berger-Vachon C, Collet L, Thai-Van H (2008) Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status: an experimental parametric study. J Neurosci Methods 175:196–205.

Alladi PA, Wadhwa S, Singh N (2002) Effect of prenatal auditory enrichment on developmental expression of synaptophysin and syntaxin 1 in chick brainstem auditory nuclei. Neuroscience 114:577–590.

Altenberg EP, Ferrand CT (2006) Fundamental frequency in monolingual English, bilingual English/Russian, and bilingual English/Cantonese young adult women. J Voice 20:89–96.

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.

Andreeva B, Demenko G, Wolska M, Möbius B, Zimmerer F, Jügler J, Jastrzebska M, Trouvain J (2014) Comparison of pitch range and pitch variation in slavic and germanic languages. Proceedings of Speech Prosody, pp 776–780.

Bidelman GM, Krishnan A (2010) Effects of reverberation on brainstem representation of speech in musicians and non-musicians. Brain Res 1355:112–125.

Bidelman GM, Gandour JT, Krishnan A (2011) Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. Brain Cogn 77:1–10.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Bregman MR, Creel SC, Creel S (2012), Learning to recognize unfamiliar voices: The role of language familiarity and music experience. Proceedings of the 34th Annual Cognitive Science Society DOI 144–149.

Burgaleta M, Sanjuán A, Ventura-Campos N, Sebastian-Galles N, Ávila C (2016) Bilingualism at the core of the brain. Structural differences between bilinguals and monolinguals revealed by subcortical shape analysis. NeuroImage 125:437–445.

Byers-Heinlein K, Burns TC, Werker JF (2010) The roots of bilingualism in newborns. Psychol Sci 21:343–348.

Campbell T, Kerlin JR, Bishop CW, Miller LM (2012) Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. Ear Hear 33:144.

Chandrasekaran B, Kraus N, Wong PC (2012) Human inferior colliculus activity relates to individual differences in spoken language learning. J Neurophysiol 107:1325–1336.

Chimento TC, Schreiner CE (1990) Selectively eliminating cochlear microphonic contamination from the frequency-following response. Electroencephalogr Clin Neurophysiol 75:88–96.

Coffey EB, Herholz SC, Chepesiuk AM, Baillet S, Zatorre RJ (2016) Cortical contributions to the auditory frequency-following response revealed by MEG. Nat Commun 7:11070.

Corriveau K, Pasquini E, Goswami U (2007) Basic auditory processing skills and specific language impairment: a new look at an old hypothesis. J Speech Lang Hear Res 50:647–666.

Costa A, Sebastián-Gallés N (2014) How does the bilingual experience sculpt the brain? Nat Rev Neurosci 15:336–345.

Darwin CJ, Brungart DS, Simpson BD (2003) Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. J Acoust Soc Am 114:2913–2922.

Dawson C, Aalto D, Šimko J, Putkinen V, Tervaniemi M, Vainio M (2016) Quantity language speakers show enhanced subcortical processing. Biol Psychol 118:169–175.

Dolson M (1994) The pitch of speech as a function of linguistic community. Music Perception 11:321–331.

Engineer ND, Percaccio CR, Pandya PK, Moucha R, Rathbun DL, Kilgard MP (2004) Environmental enrichment improves response strength, threshold, selectivity, and latency of auditory cortex neurons. J Neurophysiol 92:73–82.

Fujioka T, Ross B, Kakigi R, Pantev C, Trainor LJ (2006) One year of musical training affects development of auditory cortical-evoked fields in young children. Brain 129:2593–2608.

Galbraith GC, Arbagey PW, Branski R, Comerci N, Rector PM (1995) Intelligible speech encoded in the human brain stem frequency-following response. NeuroReport 6:2363–2367.

Gardi J, Merzenich M, McKean C (1979) Origins of the scalp recorded frequency-following response in the cat. Audiology 18:358–381.

Gervain J (2015) Plasticity in early language acquisition: the effects of prenatal and early childhood experience. Curr Opin Neurobiol 35:13–20.

Hairston WD, Letowski TR, McDowell K (2013) Task-related suppression of the brainstem frequency following response. PLoS One 8:e55215.

Hakvoort B, van der Leij A, Maurits N, Maassen B, van Zuijen TL (2015) Basic auditory processing is related to familial risk, not to reading fluency: An ERP study. Cortex 63:90–103.

Hamers JF, Blanc M (2000) Bilinguality and bilingualism. Cambridge University Press.

Dv Hapsburg, Peña ED (2002) Understanding bilingualism and its impact on speech audiometry. J Speech Lang Hear Res 45:202–213.

Hartsuiker RJ (2015) Why it is pointless to ask under which specific circumstances the bilingual advantage occurs? Cortex 73:336–337.

Hepper PG, Shahidullah BS (1994) Development of fetal hearing. Arch Dis Child 71:F81–F87.

Herdman AT, Lins O, Van Roon P, Stapells DR, Scherg M, Picton TW (2002) Intracerebral sources of human auditory steady-state responses. Brain Topogr 15:69–86.

Hoormann J, Falkenstein M, Hohnsbein J, Blanke L (1992) The human frequency-following response (FFR): Normal variability and relation to the click-evoked brainstem response. Hear Res 59:179–188.

Intartaglia B, White-Schwoch T, Meunier C, Roman S, Kraus N, Schön D (2016) Native language shapes automatic neural processing of speech. Neuropsychologia 89:57–65.

Jeng FC, Hu J, Dickman B, Montgomery-Reagan K, Tong M, Wu G, Lin CD (2011) Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. Ear Hear 32:699–707.

Johnson JS, Newport EL (1989) Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. Cogn Psychol 21:60–99.

Johnson KL, Nicol T, Zecker SG, Bradlow AR, Skoe E, Kraus N (2008) Brainstem encoding of voiced consonant-vowel stop syllables. Clin Neurophysiol 119:2623–2635.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev 84:541–577.

Kral A, Sharma A (2012) Developmental neuroplasticity after cochlear implantation. Trends Neurosci 35:111–122.

Kraus N, Nicol T (2014) The cognitive auditory system: the role of learning in shaping the biology of the auditory system. In: Popper NA, Fay RR, editors. Perspectives on auditory research. New York, New York, NY: Springer. p. 299–319.

Krishnan A, Xu Y, Gandour J, Cariani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. Cogn Brain Res 25:161–168.

Krishnan A (2006) Human frequency following response. In: Auditory evoked potentials: Basic principles and clinical applications, (Burkard RF, Don M, Eggermont J, eds), Lippincott, Williams & Williams.

Krishnan A, Gandour JT, Bidelman GM, Swaminathan J (2009) Experience-dependent neural representation of dynamic pitch in the brainstem. NeuroReport 20:408–413.

Krishnan A, Gandour JT (2009) The role of the auditory brainstem in processing linguistically-relevant pitch patterns. Brain Lang 110:135–148.

Krishnan A, Gandour JT, Bidelman GM (2010) The effects of tone language experience on pitch processing in the brainstem. J Neurolinguistics 23:81–95.

Krishnan A, Gandour JT, Ananthakrishnan S, Bidelman GM, Smalt CJ (2011) Linguistic status of timbre influences pitch encoding in the brainstem. NeuroReport 22:801–803.

Krizman J, Marian V, Shook A, Skoe E, Kraus N (2012) Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. Proc Natl Acad Sci U S A 109:7877–7881.

Krizman J, Skoe E, Marian V, Kraus N (2014) Bilingualism increases neural response consistency and attentional control: evidence for sensory and cognitive coupling. Brain Lang 128:34–40.

Krizman J, Slater J, Skoe E, Marian V, Kraus N (2015a) Neural processing of speech in children is influenced by extent of bilingual experience. Neurosci Lett 585:48–53.

Krizman J, Tierney A, Fitzroy AB, Skoe E, Amar J, Kraus N (2015b) Continued maturation of auditory brainstem function during adolescence: a longitudinal approach. Clin Neurophysiol 126:2348–2355.

Krizman J, Bradlow AR, Lam SS-Y, Kraus N (2016a) How bilinguals listen in noise: linguistic and non-linguistic factors. Bilingualism: Language and Cognition. 10.1017/S13667289160004441-10.

Krizman J, Skoe E, Kraus N (2016b) Bilingual enhancements have no socioeconomic boundaries. Dev Sci 19:881–891.

Kuipers J-R, Thierry G (2010) Event-related brain potentials reveal the time-course of language change detection in early bilinguals. NeuroImage 50:1633–1642.

Kuwada S, Anderson JS, Batra R, Fitzpatrick DC, Teissier N, D'Angelo WR (2002) Sources of the scalp-recorded amplitude-modulation following response. J Am Acad Audiol 13:188–204.

Kyle Danielson D, Seidl A, Onishi KH, Alamian G, Cristia A (2014) The acoustic properties of bilingual infant-directed speech. J Acoust Soc Am 135:EL95–EL101.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 60:1799–1822.

Langner G (1992) Periodicity coding in the auditory system. Hear Res 60:115–142.

Li X, Jeng FC (2011) Noise tolerance in human frequency-following responses to voice pitch. J Acoust Soc Am 129:EL21–EL26.

Liu LF, Palmer AR, Wallace MN (2006) Phase-locked responses to pure tones in the inferior colliculus. J Neurophysiol 95:1926–1935.

Llanos F, Dmitrieva O, Shultz A, Francis AL (2013) Auditory enhancement and second language experience in spanish and english weighting of secondary voicing cues. J Acoust Soc Am 134:2213–2224.

Maddieson I (1984) Patterns of sounds. New York: Cambridge University Press, Cambridge [Cambridgeshire].

Majewski W, Hollien H, Zalewski J (1972) Speaking fundamental frequency of polish adult males. Phonetica 25:119–125.

Marmel F, Linley D, Carlyon RP, Gockel HE, Hopkins K, Plack CJ (2013) Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently. J Assoc Res Otolaryngol 14:757–766.

Mayo LH, Florentine M, Buus S (1997) Age of second-language acquisition and perception of speech in noise. J Speech Lang Hear Res 40:686–693.

Mennen I, Schaeffler F, Docherty G (2012) Cross-language differences in fundamental frequency range: a comparison of English and Germana). J Acoust Soc Am 131:2249–2260.

Middlebrooks JC (2008) Auditory cortex phase locking to amplitude-modulated cochlear implant pulse trains. J Neurophysiol 100:76–91.

Montagni I, Peru A (2011) Effects of language and music learning on pitch and duration perception: an experimental investigation. Int J Psychol Bev Sci 1:33–40.

Moore DR (2002) Auditory development and the role of experience. Br Med Bull 63:171–181.

Moore JK, Perazzo LM, Braun A (1995) Time course of axonal myelination in the human brainstem auditory pathway. Hear Res 87:21–31.

Moushegian G, Rupert AL, Stillman RD (1973) Laboratory note. Scalp-recorded early responses in man to frequencies in the speech range. Electroencephalogr Clin Neurophysiol 35:665–667.

Musacchia G, Sams M, Skoe E, 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.

Parbery-Clark A, Skoe E, Kraus N (2009) Musical experience limits the degradative effects of background noise on the neural processing of sound. J Neurosci 29:14100–14107.

Pelli DG (1997) The videotoolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442.

Ressel V, Pallier C, Ventura-Campos N, Díaz B, Roessler A, Ávila C, Sebastián-Gallés N (2012) An effect of bilingualism on the auditory cortex. J Neurosci 32:16597–16601.

Sanes DH, Bao S (2009) Tuning up the developing auditory CNS. Curr Opin Neurobiol 19:188–199.

Sanes DH, Woolley SM (2011) A behavioral framework to guide research on central auditory development and plasticity. Neuron 72:912–929.

Shi LF (2012) Contribution of linguistic variables to bilingual listeners' perception of degraded English sentences. J Speech Lang Hear Res 55:219–234.

Skoe E, Kraus N (2010) Auditory brain stem response to complex sounds: a tutorial. Ear Hear 31:302–324.

Skoe E, Nicol T, Kraus N (2011) Cross-phaseogram: objective neural index of speech sound differentiation. J Neurosci Methods 196:308–317.

Skoe E, Kraus N (2012) A little goes a long way: how the adult brain is shaped by musical training in childhood. J Neurosci 32:11507–11510.

Skoe E, Krizman J, Spitzer E, Kraus N (2013) The auditory brainstem is a barometer of rapid auditory learning. Neuroscience 243:104–114.

Skoe E, Kraus N (2013) Musical training heightens auditory brainstem function during sensitive periods in development. Front Psychol 4:622.

Skoe E, Kraus N (2014) Auditory reserve and the legacy of auditory experience. Brain Sci 4:575–593.

Skoe E, Chandrasekaran B (2014) The layering of auditory experiences in driving experience-dependent subcortical plasticity. Hear Res 311:36–48.

Skoe E, Krizman J, Anderson S, Kraus N (2015) Stability and plasticity of auditory brainstem function across the lifespan. Cereb Cortex 25:1415–1426.

Skoe E, Brody L, Theodore RM (2017) Reading ability reflects individual differences in auditory brainstem function, even into adulthood. Brain Lang 164:25–31.

Song JH, Skoe E, Wong PC, Kraus N (2008) Plasticity in the adult human auditory brainstem following short-term linguistic training. J Cogn Neurosci 20:1892–1902.

Song JH, Skoe E, Banai K, Kraus N (2010) Perception of speech in noise: neural correlates. J Cogn Neurosci 23:2268–2279.

Song JH, Skoe E, Banai K, Kraus N (2012) Training to improve hearing speech in noise: Biological mechanisms. Cereb Cortex 22:1180–1190.

Steinschneider M, Arezzo J, Vaughan HG (1980) Phase-locked cortical responses to a human speech sound and low-frequency tones in the monkey. Brain Res 198:75–84.

Tichko P, Skoe E (2017) Frequency-dependent fine structure in the frequency-following response: the byproduct of multiple generators. Hear Res. http://dx.doi.org/10.1016/j.heares.2017.01.014.

Verschooten E, Robles L, Joris PX (2015) Assessment of the limits of neural phase-locking using mass potentials. J Neurosci 35:2255–2268.

Vihla M, Kiviniemi K, Salmelin R (2002) Auditory cortical activation in Finnish and Swedish speaking Finns: A magnetoencephalographic study. Neurosci Lett 322:141–144.

Webb AR, Heller HT, Benson CB, Lahav A (2015) Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. Proc Natl Acad Sci U S A 112:3152–3157.

Weinberger NM, Kitzes LM, Goodman DA (1970) Some characteristics of the "auditory neurophonic". Experientia 26:46–48.

Weiss MW, Bidelman GM (2015) Listening to the brainstem: musicianship enhances intelligibility of subcortical representations for speech. J Neurosci 35:1687–1691.

Whalen DH, Levitt AG (1995) The universality of intrinsic f0 of vowels. J Phonet 23:349–366.

Wong PC, Warrier CM, Penhune VB, Roy AK, Sadehh A, Parrish TB, Zatorre RJ (2008) Volume of left Heschl's gyrus and linguistic pitch learning. Cereb Cortex 18:828–836.

Xu J, Yu L, Cai R, Zhang J, Sun X (2009) Early auditory enrichment with music enhances auditory discrimination learning and alters nr2b protein expression in rat auditory cortex. Behav Brain Res 196:49–54.

Xu M, Homae F, R-i Hashimoto, Hagiwara H (2013) Acoustic cues for the recognition of self-voice and other-voice. Front Psychol 4:735.

Zhu X, Wang F, Hu H, Sun X, Kilgard MP, Merzenich MM, Zhou X (2014) Environmental acoustic enrichment promotes recovery from developmentally degraded auditory cortical processing. J Neurosci 34:5406–5415.