Neuroscience Letters 777 (2022) 136582

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Research article

Automatic sound encoding is sensitive to language familiarity: Evidence from English monolinguals and Spanish-English bilinguals

Erika Skoe [a,*], Adrián García-Sierra [a], Nairán Ramírez-Esparza [b], Shu Jiang [b]

[a] *Department of Speech, Language and Hearing Sciences, Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut, Storrs, CT, USA*
[b] *Department of Psychological Sciences, University of Connecticut, Storrs, CT, USA*

ARTICLE INFO

Keywords:
Bilingual
Auditory
Frequency-following response
Language familiarity

ABSTRACT

We investigated whether language familiarity has a modulatory effect on automatic sound encoding in the auditory brainstem by measuring frequency-following responses (FFRs) to repeating speech syllables that played in the background while monolingual English speakers and Spanish-English bilingual speakers watched cartoon videos in English and Spanish. For the English monolinguals, we found that the FFR signal quality was different between the two language conditions, with higher signal to noise ratios emerging for the Spanish compared to the English condition. For the Spanish-English bilinguals, the FFR signal quality was overall higher than the monolinguals, but there no effect of language condition on the FFR. Thus, both language familiarity of the environment and bilingual language experience, may modulate automatic sound encoding.

1. Introduction

The influence of language experience on human brain development and function is profound, extending across sensory and cognitive brain areas [1,2] and emerging prenatally [3]. While language is not an auditory specific phenomenon, language is first experienced auditorily for most humans. Here we investigated the effect of language experience on subcortical auditory processing from two angles, by looking at both the effect of bilingual language experience and the familiarity of the language environment.

Cortical brain areas have been the target of most studies of language; a considerably smaller literature exists for subcortical auditory areas. This subcortical literature has largely focused on the frequency-following response (FFR), a pre-attentive early latency automatic response (emerging 5–20 ms after stimulus onset) that follows (i.e., phase-locks to) the spectral and temporal components of the stimulus [4]. The FFR is used as an objective index of auditory function in clinical populations (e.g., hearing loss, language impairment) and as a tool to investigate experience-dependent enhancements in clinical and neurotypical populations (e.g., arising from musical training, short-term auditory training, language learning) [4,5]. In neurotypical populations, the stimulus capture is so robust that the FFR waveforms look like, and when sonified sound like, the evoking stimulus [4]. For speech, the FFR captures multiple acoustic elements, including the fundamental frequency (F0) of the voice, voice onset time (VOT), speech formants, and the amplitude envelope [4], and when measured from scalp electrodes, the dominant source of the FFR to speech stimuli is the inferior colliculus, a brainstem structure [6].

FFR studies have shown that automatic sound encoding is sensitive to the language background of the listener, including what their native language is and how many languages they speak [7,8]. For example, native-language tuning of the FFR has been demonstrated in adult tonal language speakers (e.g., Mandarin Chinese), who, relative to English monolinguals, show more faithful neural tracking of complex pitch contours, a dimension of speech that carries critical lexical information in tonal languages [7]. Language-dependent tuning of another speech cue, the VOT, has been investigated in Spanish monolinguals and English monolinguals [9]. VOT is an acoustic dimension that differentiates phonologically contrastive pairs (e.g., ba-pa, ga-ka). While VOT exists along a linear continuum, the perceptual mapping of this physical continuum to a specific speech sound category is both non-linear and context-dependent [10]. Depending on the language, the same acoustic token, with the same VOT, can be perceived differently [9,10]. This creates an interesting model for studying how different listener groups with different language backgrounds process the same physical stimulus. A recent study of monolinguals capitalized on this to show that native Spanish speakers and native English speakers have different FFR latency patterns for a speech token that is perceived as a different speech

---

* Corresponding author at: Dept of Speech, Language, and Hearing Sciences, University of Connecticut, 2 Alethia Drive, U-1085, Storrs, CT 06269, USA.
E-mail address: Erika.skoe@uconn.edu (E. Skoe).

https://doi.org/10.1016/j.neulet.2022.136582
Received 25 August 2021; Received in revised form 16 March 2022; Accepted 17 March 2022

sound between English and Spanish [9]. Further insights into the connection between automatic sound processing and language background have emerged from perceptual learning paradigms [11], manipulations of speech intelligibility (e.g., forward vs. reversed speech) [12], and by studying populations with language disorders [13] and bilingual populations [14–18].

Bilingual populations present the opportunity for understanding how the brain has wired itself in response to the complex sensory, perceptual, communicative, and social demands of learning two complex language systems. Specific to the FFR, studies have found that bilinguals compared to monolinguals have more robust responses (i.e., stronger neural responses to speech at the stimulus F0 and greater cross-trial neural synchrony) [14–18]. This bilingual enhancement effect has been demonstrated across a variety of stop-consonant syllables (ba, da, ga) for school-age and adolescent children and young adults [14–17]. In these groups, the bilingual enhancement effect emerged as stronger for the vowel component of stop-consonant stimuli compared to the initial format transition where these syllables are acoustically contrastive [14,17].

FFR studies typically use auditory stimuli that are highly reduced forms of language, often through a single syllable that plays repetitively while the listener sleeps or watches a video [4]. This approach reflects the challenges of recording early latency, low amplitude responses of subcortical origin. Such techniques, while standard, do not capture the multidimensionality of language, and the diversity of cues that listeners use to gauge whether the language in the environment is familiar, something humans, even neonates, do with relative ease [19]. In natural settings, cues to language familiarity come in various forms: speech acoustics, visual information including text, lip movements, and gestures, as well as socio-pragmatic information (e.g., talker familiarity), and cultural signifiers (e.g., food, symbols). Thus, in natural environments language familiarity may be determined by a single modality or cue (e.g., printed text) or set of cues.

Here we investigated how bilingual status and language familiarity interact to influence automatic encoding of speech sounds early in the auditory pathway using videos to establish different language conditions. Specifically, we tested adult Spanish-English bilinguals and monolingual English speakers while they watched captioned cartoons in Spanish and English. The Spanish and English videos were rich with linguistic and cultural information (both visual and auditory in nature)

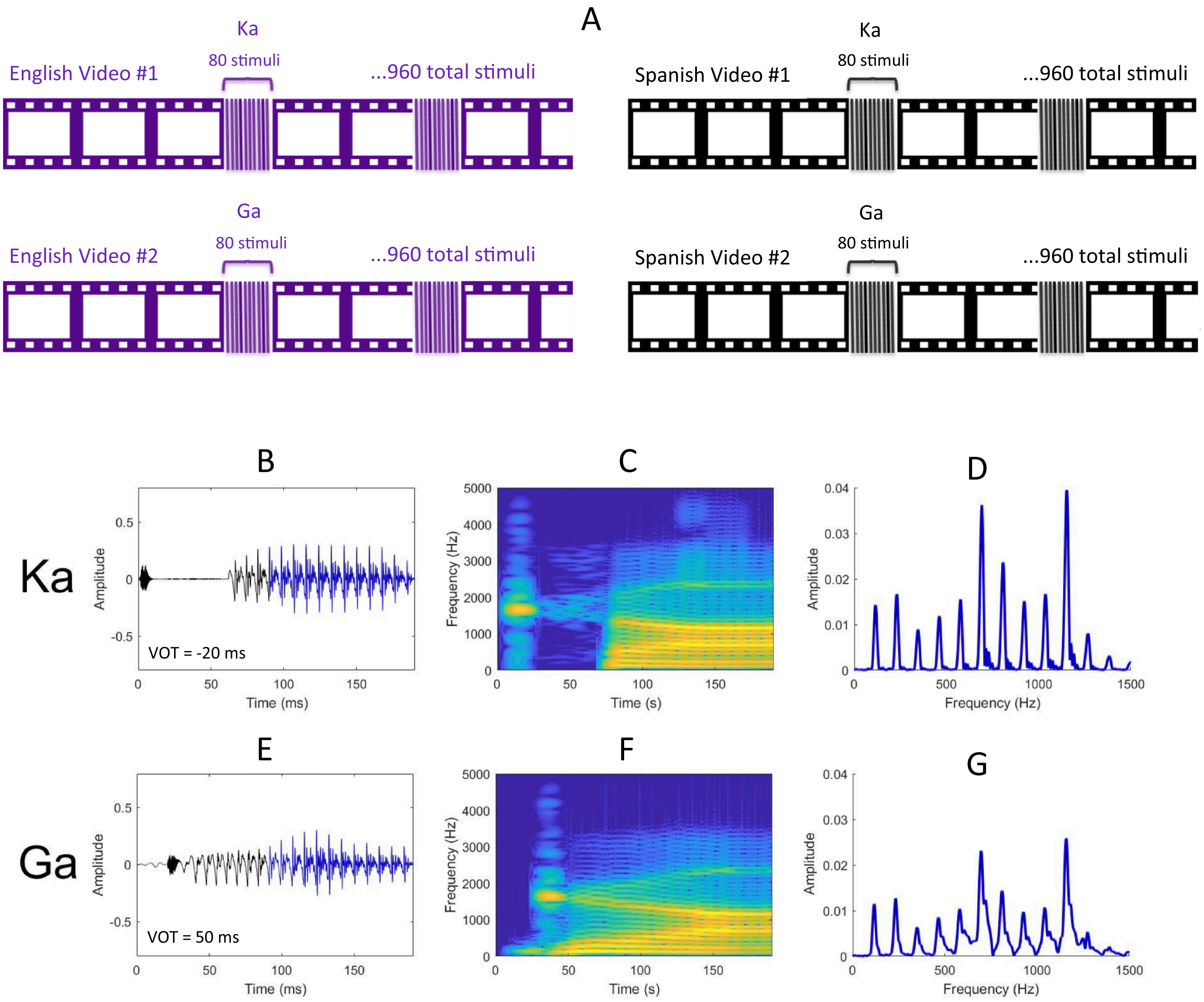


Fig. 1. (A) Experimental Design and (B-G) Stimulus Characteristics. (A) The FFR stimuli played in 1-minute blocks during the videos. The temporal and spectral characteristics of the "ka" and "ga" stimuli are illustrated using time domain waveforms (B, E), spectrograms (C, F), and spectra (D, G). The final 100 ms "a" period is plotted in blue in B, E, D, and G, to highlight the region of analysis. Spectra are zoomed in to show frequencies that fall within the filter used for FFR extraction. Note that the spectral energy distribution is the same for the two stimuli, although the amplitude of "ga" is lower during the vowel period.

that cued that the video was in English or Spanish. This information included the subtitles and the soundtrack but also visuals within the video itself (e.g., food) and the story narrative. While watching the videos, the FFR stimuli ("ga" and "ka") played intermittently in the background in 1-minute blocks (Fig. 1). During these blocks, the soundtrack was muted to avoid acoustic masking of the FFR stimulus [20]; however, the soundtrack was audible when the FFR was not present. Given that bilinguals are especially sensitive to visual cues in speech [21], we used cartoons, not live action videos, to prevent the activation of auditory areas from lip reading [22]. Bilinguals were predicted to have overall more robust FFRs to the vowel component of the stimuli than monolinguals regardless of the language condition [14,17]. When comparing the effects of language familiarity, we predicted that bilinguals would pattern differently than monolinguals given their different levels of exposure to Spanish contexts.

2. Materials and methods

2.1. Participants

Fifty-six young adults (18–23-year-old), all students at the University of Connecticut with normal hearing, participated in the study. Participants provided written informed consent in English. The FFR analysis is part of a larger study of the impact of language and culture on the brain, mind, and social behaviors. Since portions of the data collected in this larger study will be used to answer questions exclusively about culture, this informed the decision to focus on monolinguals of White-European backgrounds and bilinguals who identify as Latinx.

Participants were grouped according to whether they were monolingual English speakers (n = 31, 7 males) or Spanish-English bilinguals (n = 25, 8 males). The English monolinguals in our sample had no more than incidental exposure to Spanish, making it an unfamiliar language in the sense that they could not speak or read it and had limited Spanish exposure in their daily routines. Participants were classified as "bilingual" if they self-reported being a Spanish-English bilingual. Nineteen bilinguals reported living in the USA since birth and being exposed to both languages as children. The others reported immigrating to the USA as children (range < 1 year-15 years old). See Supplemental Methods for more information.

2.2. Experimental design and stimuli (Fig. 1)

The Spanish and English Conditions were tested on different days, with a counterbalanced order. Each session was ~ 2 h, including one hour for electrode cap placement. To establish the Language Condition, animated videos from the US and Latinx pop culture cannons were used. In total four unique videos, all featuring humans as the main characters were presented to each participant (two for the English Condition and two for the Spanish). Images showing text in the non-target language were blurred in the video. Each video was ~ 29 min (SD = 3.5). During data collection, participants were seated in a sound booth and the investigator monitored their EEG activity and their movements through an observation window to ensure that they stayed alert. Throughout the video, subtitles appeared, and participants were instructed to watch the video and ignore the FFR stimuli. The FFR stimulus played in 1-minute blocks (12-blocks total) (Fig. 1), during which the video soundtrack was temporarily muted but the subtitles stayed on.

For each Language Condition, FFRs were recorded to two stop consonant–vowel (CV) syllables ("ga" and "ka"). The stimuli were 190-ms long, had a constant F0 of 117 Hz, and they were presented binaurally. Each stimulus was presented during a separate video (960 stimuli per video). The stimuli differed in voice onset time (VOT, −20 for "ga" and + 50 for "ka") and they were selected from a VOT-speech continuum used previously [23]. Previous studies confirmed that the two stimuli are perceived as a "ga" and "ka", respectively, by Spanish and English monolingual speakers [10,23]. For details, see Supplemental Methods.

The "ga" and "ka" stimuli were selected because negative VOT stop consonants ("ga") occur less frequently in English than in Spanish and because Spanish and English speakers differ in how they phonetically categorize these sounds. This later point was critical for a separate component of the larger study, which focuses on neural indices of categorical perception. For the FFR analysis, we focused solely on FFR to the F0 of the vowel component of the stimuli, where the two stimuli are acoustically similar (Fig. 1 B-G, 90–190 ms). The vowel was selected because the bilingual effect has been shown to be stronger for the vowel component of CV stimuli compared to the initial transition period containing the VOT [14]. Additionally, given the relatively small number of trials per stimulus (960 compared to 4000–6000 presentations typical for FFRs to speech [4]), the fluctuating spectrotemporal characteristics of the transition period (0–100 ms) yielded weaker FFRs than the steady-state vowel.

Within each video, stimuli were presented using an oddball paradigm, with "ga" and "ka" serving as standard stimuli and playing with a probability of 0.80 (960 sounds). A deviant stimulus (VOT of + 15 ms) was presented with a probability of 0.20 (240 sounds). As previously reported, the perception of this deviant is expected to differ between language conditions for Spanish-English bilinguals [10,23] (perceived as "ka" in the Spanish Condition, but "ga" in the English Condition). Given the limited number of trials of the deviant, it did not produce sufficiently robust FFRs and was excluded from the FFR analysis. Only responses to the standard sounds are reported for the FFR analysis.

2.3. FFRs

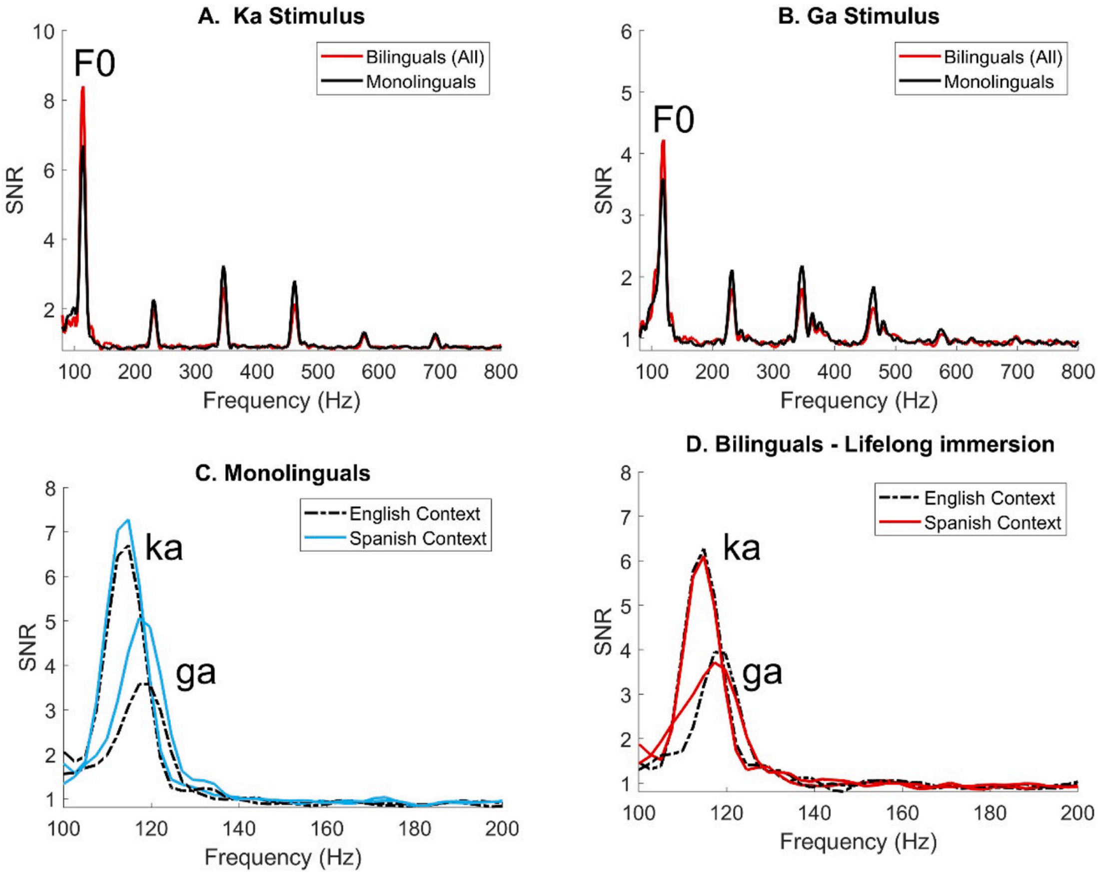
Scalp electrodes were applied using a 64-electrode headcap. See Supplemental Methods. Complex principal component analysis, cPCA, was performed on the recordings [24] to derive the FFR to the vowel (100–200 ms) from the first eigenvalue. cPCA allows for the FFR to be extracted from multichannel data to improve the signal to noise ratio (SNR). cPCA yielded a phase-locking value (PLV) for different frequency bins (in 2.44 Hz steps). PLVs were normalized by the noise floor to derive the SNR of the FFR, using the average PLV over the filter bandpass to define the FFR noise floor. For both stimuli, the response to the F0 emerged as a spectral peak (~105–125 Hz bandwidth) (Fig. 2). The analysis used the maximum SNR across this range; recordings, where the average SNR was < 1.5 over this range, were excluded from analysis due to insufficient quality.

2.4. Statistical analysis

Nine bilinguals and 13 monolinguals had partial datasets (missing one or more conditions), due to low SNRs and/or technical difficulties during data collection. Linear mixed-effects (LME) models were therefore selected for their ability to handle missing data. A linear mixed-effects model was used to test for main effects of Condition, Group, and the interaction between Group and Condition (See Supplemental Methods).

3. Results

The bilingual group had more robust FFRs (i.e., higher SNRs to the F0) overall compared to monolinguals when collapsing across stimuli and condition, with an average SNR group difference of 1.01 (main effect of Group, t(193) = 2.05, p = 0.04) (Fig. 2A, B). In addition, the two groups were found to be differently affected by Language Condition (interaction between Group and Condition (t(193) = -2.40, p = 0.02) (Fig. 2 C & D, Supplemental Table 2). Post-hoc analysis showed that the for the bilingual group, the FFR SNR to the F0 did not statistically differ between Conditions ($\bar{x}$ = 1.39, SD = 4.2, t(89) = -1.43, p = 0.15). By contrast, monolingual group had lower SNRs in the familiar English Condition compared to the Spanish Condition ($\bar{x}$ = −0.85, range −7.3 to 2.6, SD = 2.00; t(103) = 2.12, p = 0.03) (Fig. 2C). In the Spanish


A. Ka Stimulus

B. Ga Stimulus

C. Monolinguals

D. Bilinguals - Lifelong immersion

Fig. 2. Effect of Bilingual Language Experience and Language Familiarity on the Frequency Following Response (FFR). A-B: Bilinguals (red) have more robust FFRs to the fundamental frequency (F0) than monolinguals (black) across conditions and stimuli. This is illustrated for the "ka" (A) and "ga" (B) using the group average FFR spectra in the English Condition. Data from all bilinguals are plotted in panels A and B. C-D: Bilinguals and monolinguals also show a different effect of Language Condition. (C) The monolingual group shows a statistically significant difference between the English compared to Spanish Conditions, suggestive of a language familiarity effect. This is shown for vowels of both the "ga" and "ka" stimuli. (D) The bilingual group shows no statistical difference between conditions, which is consistent with both languages being familiar to them. For illustrative purposes, panel D shows the subset of bilinguals (n = 19) who were born in the USA and were immersed in a dual language environment since birth. See Results and Methods for a summary of the data from the remaining subset of bilingual participants who immigrated to the USA.

Condition, the SNR was nominally larger by 0.84 SNR for the monolingual group compared to the bilingual group, but the difference was not statistically significant ($t(93) = 0.94$, $p = 0.34$), potentially because FFR SNR is influenced not just by familiarity but also by bilingual status (i.e., greater language familiarity may decrease the SNR but this is offset by the boost from being a bilingual).

However, the effect of Language Condition on the SNR was not uniform across the bilingual group. For the bilingual group, the difference between the English and Spanish Conditions ranged from a decrease of 5.2 SNR to an increase of 14.25 SNR across participants. Two bilinguals showed large increases (English Condition > Spanish Condition), which drove the group mean into the positive range. Both learned English as their second language after immigrating to the USA. Fig. 2D shows data from the subset of the bilinguals who were born in the USA to Spanish-speaking parents ($N = 19$); in this subset of participants with lifelong bilingual language immersion, the two conditions are closely matched.

4. Discussion

We found that language familiarity, established using videos, influences automatic sound encoding of speech taking place early in the auditory pathway. For English monolinguals, we found that the signal quality of the FFR was different when watching a video in the highly familiar English condition vs. less familiar Spanish condition. For the bilingual group, for whom both language conditions were highly familiar, the FFR was statistically similar between the two conditions.

In addition to a language familiarity effect, we observed a main effect of group, with bilinguals having stronger FFRs overall (when collapsing across the two language conditions). This suggests that automatic sound encoding may remain generally enhanced in bilinguals than monolinguals, even in different language environments. This global enhancement of the stimulus F0 may allow bilinguals to continuously monitor their auditory environment to facilitate rapid perceptual and cognitive transitions when the language in the environment switches [14,25]. Consistent with this, an emerging literature suggests that bilingual speakers modulate their vocal F0 when speaking different languages, with the degree of language familiarity influencing this vocal modulation [26,27]. For bilinguals, the need to be sensitive to the language of the environment is routine, especially if they live in environments where both languages are spoken, and code switching is frequently required [25]. In addition to helping the bilingual select the language most appropriate for a given communication setting, cues that are specific to one language vs. the other have been shown to influence how bilinguals access words in their mental lexicon [28] and map acoustic information to different phonetic categories [10,23,29,30]. By contrast, English monolinguals in the USA do not have the same need to continuously monitor their auditory environment for language switches. For the monolinguals, the functional significance of the findings is therefore less clear. One interpretation is that in a language condition that is not highly familiar to them, monolinguals increase the gain on incoming acoustic stimulation to compensate for reduced familiarity of the language environment at later stages of language processing. Another interpretation that is that irrelevant auditory stimulation (i.e., the FFR stimuli) is suppressed for monolinguals in familiar language environments compared to unfamiliar or less familiar ones.

Consistent with the bilingual experience being a complex nonhomogeneous phenomenon [31], we found that the effect of language condition varied across the bilinguals. Indeed, two of the bilinguals showed strongly suppressed FFRs in the Spanish compared to English condition. Both immigrated in later childhood (ages 9 and 11, respectively), so although proficient in both languages, Spanish could be argued to be more familiar to them given early-life exposure to a Spanish monolingual culture and consequently less exposure to English. Interestingly, these two participants followed the pattern observed in monolinguals, but to the opposite language (lower SNRs in the Spanish (more familiar) language condition than in the English (less familiar) language condition). However, these two participants did not differ in any obvious way in their language learning history and ability from the others who immigrated to the USA. Still, the factors that dictate whether two language conditions are processed similar are likely more nuanced than can be captured with our language survey and ability measurements (See Supplemental Results). To better understand variation in the bilingual sample, future work should adopt more formal assessments of language dominance and proficiency.

Because the FFR stimulus was physically the same across the Spanish and English conditions, any differences observed between language conditions should not have been caused by the specifics of the stimuli. Thus, our findings provide strong evidence for a top-down, not bottom-up, effect of language familiarity on the FFR. Top-down effects on the FFR have previously been studied in various fashions, including intermodal, intramodal, and multimodal conditions [32–34]. We took a novel approach that uses a more naturalistic paradigm to engage top-down effects. While videos are a common approach to distract and calm the participant during the recording of the FFR — especially for paradigms involving children or when the recording session is long [4] — the influence of the video on the FFR has not been investigated. Our findings call for a more systematic look at the influence of a video.

For a brain that is bombarded with sensory information from multiple modalities, being able to focus on the most relevant sensory information in an environment, and suppress irrelevant stimulation, is computationally advantageous. It is therefore not surprising that attentional effects have been observed throughout the entire auditory system, from the cochlea to the cortex [35,36]. Yet, whether top-down attentional effects can be observed in early latency auditory evoked potentials such as the FFR remains controversial. Some studies suggest that attentional modulation does not occur [32,37], while others suggest that the FFR undergoes real-time modulation based on the attentional focus and load [33,38,39]. Studies supporting top-down modulation suggest that phase-locking is weaker (leading to lower SNRs) than an auditory-alone baseline when attention is focused on a task that directs the listener away from the FFR stimulus, and stronger than baseline when attention is focused on the FFR stimulus. Also, in bilinguals, but not monolinguals, FFR enhancements have been shown to correlate with better performance on standardized tests of selective attention in both the auditory and visual modalities [14].

This raises another possible interpretation for our findings. Perhaps language familiarity does not influence the FFR directly but instead, modulates automatic sound processing via processes involved in the suppression of unattended sound stimulation. In our experimental paradigm, the FFR stimulus played as an irrelevant background sound during the video, and participants were instructed to attend to the video and ignore the sound stimuli. (Indeed, our analysis of the P3a component of the auditory response (See Supplemental Results) suggests that participants were not actively attending to the sound stimuli). However, for the bilingual group, the two language conditions may have been more similar, in terms of how easy it was to attend to the videos and ignore the background sounds, leading to similar FFR results for the two conditions. Yet for the English bilinguals, it may have been easier to attend to the video and ignore the background sounds in the English relative to the Spanish condition. This could explain why the FFRs were less robust in the English compared to the Spanish Condition for the English monolinguals. However, it cannot explain why the FFR was overall more robust in the bilinguals compared to the monolinguals, unless, as argued above, the functional needs to remain sensitive to unattended background stimulation is greater for bilinguals.

Future studies should better control that all participants were doing the same task during the FFR recording (e.g., all reading the subtitles). With the current design, we cannot be certain about how much either group focused on the subtitles vs. the FFR stimulus vs. other language-relevant elements of the videos, or rule out that some participants looked alert weren't actually engaged in the video. Thus, it is currently unclear whether the language familiarity effects were induced by the written subtitles during the FFR stimulation or by other non-textual language cues during the FFR stimulation. However, continuously reading the subtitles may not be necessary to enter/stay in a language mode, as the videos were replete with cues about whether the videos were in a highly familiar or less familiar language (e.g., the video images and narrative, as well as soundtrack) and the language did not change within the video. It is also possible that the language familiarity effect could be due to a spillover effect of the video soundtrack that played before each FFR block. On a related point, we also recognize that many English monolinguals in the USA have had enough passive exposure to Spanish to distinguish Spanish from a language they have never heard before, even if they have limited to no ability to speak or understand Spanish. Isolating the different language cues within the videos, and better monitoring of what the participant is attending to would help to clarify the role of attention on our findings and the functional significance of the findings. It would also help explicate which environmental cues have a dominant influence on how language familiarity modulates the FFR. Future investigations should consider adopting methods to manipulate language familiarity that do not involve a concurrent visual stimulus, including an auditory-only condition (with no language condition cues) as a baseline for comparison, and expanding the analysis to other dimensions of the neural response (e.g., the response to the VOT and harmonics). Testing Spanish monolinguals and using a rare language that is entirely unfamiliar to both monolinguals and bilinguals (where incidental exposure would be zero), would help better delineate how language familiarity and bilingual experience collectively influence the FFR.

In summary, our results provide evidence that automatic sound processing early in the auditory pathway is shaped by two interacting phenomena: language familiarity and bilingual experience. Findings provide critical new insight into the real-time operations that act upon the FFR in a top-down fashion, and the powerful role that language experience might have on how sound is processed early in the auditory pathway.

CRediT authorship contribution statement

Erika Skoe: Conceptualization, Methodology, Software, Writing – original draft, Visualization, Funding acquisition. **Adrián García-Sierra:** Conceptualization, Methodology, Funding acquisition. **Nairán Ramírez-Esparza:** Conceptualization, Methodology, Funding acquisition. **Shu Jiang:** Conceptualization, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We wish to thank Kelly Linehan for her assistance with collecting pilot data, Karl Lerud for his assistance with the complex principal component code, and Ashley Parker and Luan Tonelli for their feedback on the manuscript. For data collection assistance we also thank Lexi Arcomano, Molly Barnett, Sydney Bates, Christine Cammisa, Kaleigh Constantine, Leiah Cutkomp, Tayla Duntz, Kristen Fagan, Crystal Flores, Lina Kane, Ashley Lombardi, Alondra Marmolejos, Amy O'Rourke, Sarah Polcardo, Calli Smith, Allison Tozzi, Eilis Welsh, and Noelle Wig.

Funding

This work was supported by a Research Excellence Program award to ES, AGS, and NRE from the Office of the Vice President for Research at the University of Connecticut, and startup funds awarded to ES and AGS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neulet.2022.136582.

References

[1] R.R. Romeo, J.A. Leonard, S.T. Robinson, M.R. West, A.P. Mackey, M.L. Rowe, J. D. Gabrieli, Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function, Psychol. Sci. 29 (2018) 700–710.

[2] L. Tu, J. Wang, J. Abutalebi, B. Jiang, X. Pan, M. Li, W. Gao, Y. Yang, B. Liang, Z. Lu, Language exposure induced neuroplasticity in the bilingual brain: A follow-up fMRI study, Cortex 64 (2015) 8–19.

[3] L. May, K. Byers-Heinlein, J. Gervain, J. Werker, Language and the Newborn Brain: Does Prenatal Language Experience Shape the Neonate Neural Response to Speech? Front. Psychol. 2 (2011).

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

[5] E.B. Coffey, T. Nicol, T. White-Schwoch, B. Chandrasekaran, J. Krizman, E. Skoe, R. J. Zatorre, N. Kraus, Evolving perspectives on the sources of the frequency-following response, Nat. Commun. 10 (2019) 1–10.

[6] G.M. Bidelman, Subcortical sources dominate the neuroelectric auditory frequency-following response to speech, NeuroImage 175 (2018) 56–69.

[7] A. Krishnan, J.T. Gandour, The role of the auditory brainstem in processing linguistically-relevant pitch patterns, Brain Lang. 110 (3) (2009) 135–148.

[8] V. Marian, T.Q. Lam, S. Hayakawa, S. Dhar, Top-down cognitive and linguistic influences on the suppression of spontaneous otoacoustic emissions, Front. Neurosci. 12 (2018) 378.

[9] T.C. Zhao, P.K. Kuhl, Linguistic effect on speech perception observed at the brainstem, Proc. Natl. Acad. Sci. 115 (35) (2018) 8716–8721.

[10] A. Garcia-Sierra, R.L. Diehl, C. Champlin, Testing the double phonemic boundary in bilinguals, Speech Commun. 51 (4) (2009) 369–378.

[11] R. Reetzke, Z. Xie, F. Llanos, B. Chandrasekaran, Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood, Curr. Biol. 28 (9) (2018) 1419–1427.e4.

[12] G.C. Galbraith, E.M. Amaya, J.M. Diaz de Rivera, N.M. Donan, M.T. Duong, J. N. Hsu, K. Tran, L.P. Tsang, Brain stem evoked response to forward and reversed speech in humans, NeuroReport 15 (13) (2004) 2057–2060.

[13] J. Hornickel, N. Kraus, Unstable representation of sound: a biological marker of dyslexia, The Journal of neuroscience : the official journal of the Society for, Neuroscience 33 (2013) 3500–3504.

[14] J. Krizman, V. Marian, A. Shook, E. Skoe, N. Kraus, Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages, Proc. Natl. Acad. Sci. U.S.A 109 (20) (2012) 7877–7881.

[15] J. Krizman, E. Skoe, N. Kraus, Bilingual enhancements have no socioeconomic boundaries, Dev. Sci. 19 (6) (2016) 881–891.

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

[17] E. Skoe, E. Burakiewicz, M. Figueiredo, M. Hardin, Basic neural processing of sound in adults is influenced by bilingual experience, Neuroscience 349 (2017) 278–290.

[18] D. Bsharat-Maalouf, H. Karawani, P. Prabhu, Bilinguals' speech perception in noise: Perceptual and neural associations, PLoS ONE 17 (2) (2022) e0264282.

[19] K.A. Yoshida, J.R. Iversen, A.D. Patel, R. Mazuka, H. Nito, J. Gervain, J.F. Werker, The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study, Cognition.

[20] N. Russo, T. Nicol, G. Musacchia, N. Kraus, Brainstem responses to speech syllables, Clin. Neurophysiol. 115 (9) (2004) 2021–2030.

[21] J. Maye, J.F. Werker, LouAnn Gerken, Infant sensitivity to distributional information can affect phonetic discrimination, Cognition 82 (3) (2002) B101–B111.

[22] G.A. Calvert, E.T. Bullmore, M.J. Brammer, R. Campbell, S.C. Williams, P. K. McGuire, P.W. Woodruff, S.D. Iversen, A.S. David, Activation of auditory cortex during silent lipreading, Science 276 (1997) 593–596.

[23] A. García-Sierra, N. Ramírez-Esparza, J. Silva-Pereyra, J. Siard, C.A. Champlin, Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study, Brain Lang. 121 (3) (2012) 194–205.

[24] H.M. Bharadwaj, B.G. Shinn-Cunningham, Rapid acquisition of auditory subcortical steady state responses using multichannel recordings, Clinical neurophysiology : official journal of the International Federation of, Clin. Neurophysiol. 125 (2014) 1878–1888.

[25] F. Grosjean, Life as a Bilingual: Knowing and Using Two or More Languages, Cambridge University Press, Cambridge, 2021.

[26] K. Järvinen, A.-M. Laukkanen, K. Izdebski, Voice Fundamental Frequency Changes as a Function of Foreign Languages Familiarity: An Emotional Effect?, Emotions in the Human Voice, Volume 1, Foundations 1 (2007) 203.

[27] S. Schwab, J.-P. Goldman, Do speakers show different F0 when they speak in different languages? The case of English French and German?, 2016.

[28] K. Timmer, J.G. Grundy, E. Bialystok, The influence of contextual cues on representations in the mental lexicon for bilinguals, Bilingualism: A framework for understanding the mental lexicon (2017) 123-142.

[29] A. García-Sierra, E. Schifano, G.M. Duncan, M.S. Fish, An analysis of the perception of stop consonants in bilinguals and monolinguals in different phonetic contexts: A range-based language cueing approach, Attention, Perception, & Psychophysics (2021).

[30] N. Wig, A. García-Sierra, Matching the Mismatch: The interaction between perceptual and conceptual cues in bilinguals' speech perception, Bilingualism: Language and Cognition (2020) 1-14.

[31] V. DeLuca, J. Rothman, E. Bialystok, C. Pliatsikas, Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function, Proc. Natl. Acad. Sci. 116 (15) (2019) 7565–7574.

[32] L. Varghese, H.M. Bharadwaj, B.G. Shinn-Cunningham, Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses, Brain Res. 1626 (2015) 146–164.

[33] W.D. Hairston, T.R. Letowski, K. McDowell, P.V. Nerurkar, Task-Related Suppression of the Brainstem Frequency following Response, PLoS One 8 (2) (2013) e55215.

[34] G. Musacchia, M. Sams, E. Skoe, N. Kraus, Musicians have enhanced subcortical auditory and audiovisual processing of speech and music, Proc. Natl. Acad. Sci. U.S.A 104 (40) (2007) 15894–15898.

[35] P.H. Delano, D. Elgueda, C.M. Hamame, L. Robles, Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas, The Journal of neuroscience : the official journal of the Society for, Neuroscience 27 (15) (2007) 4146–4153.

[36] J.B. Fritz, M. Elhilali, S.A. Shamma, Adaptive changes in cortical receptive fields induced by attention to complex sounds, J. Neurophysiol. 98 (4) (2007) 2337–2346.

[37] R. Al Osman, H. Al Osman, Inter-modality influence on the brainstem using an arithmetic exercise, J. Acoust. Soc. Am. 144 (1) (2018) EL26–EL32.

[38] G.C. Galbraith, P.W. Arbagey, R. Branski, N. Comerci, P.M. Rector, Intelligible speech encoded in the human brain stem frequency-following response, NeuroReport 6 (17) (1995) 2363–2367.

[39] Z. Xie, R. Reetzke, B. Chandrasekaran, Taking attention away from the auditory modality: Context-dependent effects on early sensory encoding of speech, Neuroscience 384 (2018) 64–75.