Bilingualism: Language and Cognition

cambridge.org/bil

Determinants of voice recognition in monolingual and bilingual listeners

Rachel M. Theodore[1,2] and Erin G. Flanagan[3]

[1]Department of Speech, Language, and Hearing Sciences; University of Connecticut; [2]Connecticut Institute for the Brain and Cognitive Sciences; University of Connecticut and [3]Department of Speech, Language, and Hearing Sciences; University of Connecticut

Research Article

Cite this article: Theodore RM, Flanagan EG (2020). Determinants of voice recognition in monolingual and bilingual listeners. *Bilingualism: Language and Cognition* **23**, 158–170. https://doi.org/10.1017/S1366728919000075

Received: 13 May 2018
Revised: 15 January 2019
Accepted: 17 January 2019
First published online: 13 February 2019

Key words:
speech perception; voice processing; individual differences; bilinguals

Author for correspondence:
Rachel M. Theodore,
E-mail: rachel.theodore@uconn.edu

Abstract

Recent findings demonstrate a bilingual advantage for voice processing in children, but the mechanism supporting this advantage is unknown. Here we examined whether a bilingual advantage for voice processing is observed in adults and, if so, if it reflects enhanced pitch perception or inhibitory control. Voice processing was assessed for monolingual and bilingual adults using an associative learning identification task and a discrimination task in English (a familiar language) and French (an unfamiliar language). Participants also completed pitch perception, flanker, and auditory Stroop tasks. Voice processing was improved for the familiar compared to the unfamiliar language and reflected individual differences in pitch perception (both tasks) and inhibitory control (identification task). However, no bilingual advantage was observed for either voice task, suggesting that the bilingual advantage for voice processing becomes attenuated during maturation, with performance in adulthood reflecting knowledge of linguistic structure in addition to general auditory and inhibitory control abilities.

Introduction

The acoustic speech signal simultaneously cues a talker's communicative intent and the talker's identity. That is, from the same acoustic stream, listeners have access to both WHO is speaking and WHAT the talker is saying. There is now a large body of literature documenting an exquisite interplay between these two aspects of speech acoustics. For example, listeners show improved comprehension for familiar compared to unfamiliar talkers, especially in degraded listening environments (e.g., Nygaard & Pisoni, 1998). Talker-specificity effects for language comprehension emerge early in the processing stream, during the stage in which the acoustic signal is mapped to representations for individual speech sounds (Clayards, Tanenhaus, Aslin & Jacobs, 2008; Drouin, Theodore & Myers, 2016; Theodore, Myers & Lomibao, 2015). In addition, listeners show better voice recognition for speakers of a native compared to a nonnative language (Goggin, Thompson, Strube & Simental, 1991). The ability to learn to identify novel talkers is facilitated when the talkers are speaking a known language (e.g., Orena, Theodore & Polka, 2015; Perrachione, Del Tufo & Gabrieli, 2011; Perrachione & Wong, 2007; Xie & Myers, 2015). The language familiarity benefit for voice recognition has also been elicited using discrimination tasks. For example, the perceived acoustic similarity for voices speaking a known language is greater than for voices speaking an unknown language (Fleming, Giordano, Caldara & Belin, 2014). Results from some AX discrimination paradigms have shown higher sensitivity for voices speaking a known language compared to an unknown language (Levi, 2018; Levi & Schwartz, 2013; Winters, Levi & Pisoni, 2008), though other studies have not shown such a language familiarity effect in discrimination tasks, which may reflect methodological differences such as the use of word-length versus sentence-length stimuli (Fecher & Johnson, 2018a; Wester, 2012).

Recent investigations of the native language benefit for voice recognition have sought to identify the specific aspects of linguistic experience that drive this voice processing benefit, with findings implicating a role for both higher-level linguistic knowledge, such as lexical structure, and lower-level linguistic knowledge, including knowledge of the sound structure (e.g., Bregman & Creel, 2014; Goggin et al., 1991; Johnson, Bruggeman & Cutler, 2018; Johnson, Westrek, Nazzi & Cutler, 2011; Orena et al., 2015; Perrachione et al., 2011). Converging evidence for this account comes from studies showing that the native language benefit for voice recognition emerges very early in the developmental process, even before vocabulary knowledge (Johnson et al., 2011; Fecher & Johnson, 2018b).

Other findings point to a role for global auditory processing abilities in facilitating voice processing, independent of language familiarity (Köster & Schiller, 1997; Xie & Myers, 2015). Xie and Myers (2015) examined the relationship between pitch perception and voice recognition by testing two populations that had experience-driven pitch expertise,

CAMBRIDGE UNIVERSITY PRESS

monolingual English musicians and native speakers of Mandarin, and comparing their performance to monolingual English speakers with no musical experience. The Mandarin listeners were late bilinguals who spoke English as a second language. Participants were tested on a voice learning task for English, Mandarin, and Spanish speakers; no listener was fluent in Spanish, though most of the English monolinguals had previous exposure to Spanish. The expected language familiarity effect was observed, and overall the English musicians and Mandarin bilinguals had higher accuracy compared to English non-musicians. In a follow-up experiment, both native English and native Mandarin listeners with and without musical training completed a talker learning task for English and Mandarin voices in addition to discrimination tasks that assessed local (i.e., absolute) and global (i.e., relative) pitch processing ability. The results replicated the first study to show that musicians performed better than non-musicians and that Mandarin listeners performed better than English listeners. A mediation analysis revealed that the effects of native language and musical training were mediated by individual differences in pitch processing; the path by which having musical training and speaking Mandarin influenced voice processing was through their relationship to heightened pitch perception, though this mediating relationship was observed only for nonnative voice recognition.

Relevant to the current investigation, the results of Xie and Myers (2015) provided no evidence of a global bilingual benefit for voice recognition. In contrast, recent findings have demonstrated a bilingual advantage for voice recognition in children. Levi (2018) used discrimination and associative learning identification tasks to examine voice processing in monolingual and bilingual children from two age cohorts, children ages 7–9 years and children ages 10–12 years. The test languages for the discrimination task were English voices and German voices; with only the English voices used in the identification task. In contrast to Xie and Myers (2015), the bilingual children in Levi (2018) represented a wide variety of language backgrounds, though no monolingual or bilingual participant had experience with the nonnative language (i.e., German). English was shared among all bilingual children, but the second language spanned a wide range with the constraints that it was not German or a tone language. When voice processing was examined in the talker discrimination task, bilingual children showed higher discrimination accuracy compared to monolingual children for both the English and the German voices, and all children showed higher discrimination accuracy for the familiar compared to the unfamiliar language. For the identification task, bilingual children outperformed their monolingual peers with respect to the amount of training required to meet learning criterion and overall accuracy. There was no interaction between age and language status in either task; the magnitude of the bilingual benefit was equivalent between the two age cohorts.

A mechanistic account of the bilingual advantage for voice recognition was not explicated in Levi (2018), though she hypothesized that this may reflect increased cognitive control (i.e., inhibition) or experience with accented speech among the bilingual children. An additional possibility is that it may be due to enhanced pitch processing in bilingual listeners. Recent findings from Skoe and colleagues (Skoe, Burakiewicz, Figueiredo & Hardin, 2017) demonstrated that the electrophysiological response to fundamental frequency information is more robust in bilingual compared to monolingual adults. They measured the frequency following response (FFR), which is a phase locked response to sound that occurs in the central auditory nervous system, for monolingual English speakers and a diverse group of bilingual speakers (the second language varied across bilingual speakers, as did age of acquisition for the second language). Compared to monolinguals, bilinguals exhibited increased amplitude of the FFR, even after controlling for musical experience and excluding bilingual speakers of tone languages from the analysis. This finding raises the possibility that pitch perception may be enhanced in bilinguals, regardless of which two languages they speak. With respect to the findings of Levi (2018), the results of Skoe et al. (2017) point to a specific mechanism that may converge the bilingual advantage for voice recognition, that being heightened sensitivity to pitch information. This account is consistent with the role of pitch perception in mediating voice recognition as identified in Xie and Myers (2015), though as previously noted, a global bilingual advantage was not present for voice recognition in their study. Levi notes a key difference between the monolingual experience with the nonnative language in previous studies. In Levi (2018), neither the monolingual nor bilingual children had substantial experience with the nonnative language used in the voice processing tasks (i.e., German), but most of the monolingual listeners in Xie and Myers (2015) had some experience with the nonnative language (i.e., Spanish), which may have masked any bilingual advantage present in the bilingual Mandarin listeners.

The literature reviewed thus far converges to provide strong evidence that linguistic knowledge influences voice recognition, resulting in a language familiarity benefit for voice recognition. Pitch perception as a general auditory ability also influences voice recognition, with its role heightened for nonnative compared to native voices (Xie & Myers, 2015). Though bilinguals show enhanced encoding of fundamental frequency in the central auditory nervous system compared to monolinguals (Skoe et al., 2017), the evidence of that sensitivity leading a global bilingual advantage for voice recognition is mixed. Indeed, the inconsistent evidence of a bilingual advantage for voice recognition is not unique given the broader literature on bilingual advantages for cognitive processing (e.g., Paap & Greenberg, 2013). The goal of the current work is to provide an additional test of the bilingual advantage for talker processing and, in doing so, test mechanistic explanations of such an advantage if it is observed. Two research questions were examined in the current work: Do bilingual adults show a bilingual advantage for talker processing? If so, does the bilingual advantage reflect enhanced pitch perception and/or inhibitory control?

In experiment 1, we assessed talker processing for English monolinguals and English bilinguals using an associative learning identification task. Talker processing was assessed for English (a familiar language) and French (an unfamiliar language) voices. As in Levi (2018) and Skoe et al. (2017), bilingual participants represented diverse language backgrounds. In addition to the identification task, participants completed a pitch perception task (Xie & Myers, 2015), and auditory Stroop (Sommers & Danielson, 1999) and flanker (Eriksen & Eriksen, 1974) tasks to assess inhibition in the auditory and visual domains, respectively. Participants in experiment 2 completed the same individual differences tasks (i.e., pitch perception and inhibition), but talker processing was assessed using a discrimination task. We used two different ways of measuring talker processing following Levi (2018). As reviewed above, studies examining the language familiarity effect for voice recognition have reported mixed results regarding its emergence using discrimination tasks (e.g., Fecher

& Johnson, 2018a). One key distinction between the identification and discrimination tasks is a learning component in that the identification task requires listeners to learn to associate a voice with an identifier (e.g., a cartoon avatar). The bilingual advantage observed in Levi (2018) was observed in both tasks, and thus we had no a priori prediction that a bilingual advantage would be task-dependent in the current study.

We selected the pitch processing task as a potential predictor of talker processing in light of the findings of Xie and Myers (2015); in addition, assessing performance on this task between the monolinguals and bilinguals will allow us to examine whether the bilingual advantage of pitch processing as measured by the FFR (Skoe et al., 2017) is also observed in a behavioral task that can exploit sensitivity to pitch information. We selected the auditory Stroop and flanker tasks as measures of cognitive inhibition for two reasons. First, as outlined in Levi (2018), the tasks used to examine talker processing may be considered ones of inhibitory control given that performing the task (i.e., identifying voice) may require inhibiting irrelevant information (i.e., the semantic content of the stimulus). Second, though the evidence is not without debate (e.g., Bialystok & Grundy, 2018; Paap, Myuz, Anders, Bockelman, Mikulinsky & Sawi, 2017; Paap & Greenberg, 2013), some studies have reported a bilingual advantage in cognitive inhibition measures (e.g., Ye, Mo & Wu, 2017). Thus, the inclusion of the inhibition measures allows us to test whether a bilingual advantage in voice processing (if observed) reflects differences in cognitive control, in addition to testing whether voice recognition ability is related to performance on these inhibition measures more generally. Two measures for assessing cognitive control (i.e., flanker task with visual stimuli, auditory Stroop task with auditory stimuli) were selected in order to examine whether any observed effects of inhibition on voice processing were limited to the specific modality in which it was assessed.

The bilingual advantage for voice recognition that was observed in Levi (2018) will be replicated if performance on the talker identification and discrimination tasks is improved for bilinguals compared to monolinguals for both test languages (English and French). If this advantage is due to enhanced pitch perception, then we would expect to observe better performance on the pitch perception task for bilinguals compared to monolinguals in addition to a reliable relationship between pitch perception and performance on the talker tasks. Similar patterns for the auditory Stroop and flanker tasks would suggest that the bilingual advantage for voice processing reflects improved inhibitory control. We first present results for the talker identification and discrimination tasks, which examined (1) performance between monolingual and bilingual participants and (2) the degree to which performance on the individual difference measures predicted voice processing. We then present results that compared performance between the monolingual and bilingual participants for the three individual difference measures (i.e., pitch perception, flanker, and auditory Stroop tasks), pooling participants across the two experiments.

Experiment 1

Method: Participants

Forty adults (15 males, 25 females) between the ages of 18 and 28 years (mean = 21, SD = 3) were recruited from the University of Connecticut community for participation in the experiment. Three additional participants were tested but excluded due to high error rates on the flanker task (n = 2) or acquiring the second language past the age of 5 (n = 1). No participant had a history of speech, language, or hearing disorders according to self-report. All participants passed a pure tone hearing screen on the day of testing administered at 20 dB for octave frequencies between 500 and 4000 Hz. Participants received either monetary compensation or partial course credit for participation. All testing procedures and informed consent acquisition followed protocols approved by the University of Connecticut Institutional Review Board.

Twenty participants were monolingual speakers of English and 20 participants were bilingual speakers of English and various second languages. The second languages included Arabic (n = 1), Cantonese (n = 2), German (n = 3), Gujarati (n = 1), Hindi (n = 2), Japanese (n = 1), Nepali (n = 1), Polish (n = 1), Portuguese (n = 1), Russian (n = 2), Spanish (n = 3), Tagalog (n = 1), and Urdu (n = 1). All of the bilingual participants reported acquiring both languages prior to age 5, and thus are considered to be native speakers of English and their second language. None of the monolingual participants reported knowing a language other than English.

Experience with French was assessed using a questionnaire that asked listeners to report current exposure to French, past exposure to French, and level of proficiency in speaking, understanding, and reading French. Proficiency was assessed using an 11-point scale that ranged from 0 to 10 as follows: 0 = none, 1 = very low, 2 = low, 3 = fair, 4 = slightly less than adequate, 5 = adequate, 6 = slightly more than adequate, 7 = good, 8 = very good, 9 = excellent, 10 = perfect. No participant reported current exposure to French. Sixteen monolinguals and 14 bilinguals reported no past exposure to French. For the four monolinguals who did report past exposure to French, this ranged from instruction in elementary school (n = 2), high school (n = 1), and college (n = 1). For the six bilinguals who reported past exposure to French, this ranged from instruction in middle school (n = 3), high school (n = 2), and college (n = 1). With respect to proficiency in French, two monolinguals indicated 1 (very low) on the 11-point scale, with all other monolingual participants indicating 0 (none). For the bilingual participants, three participants indicated 2 (low) and four participants indicated 1 (very low), with all other bilinguals indicating 0 (none) as their level of proficiency in French.

According to self-report, six monolinguals and nine bilinguals indicated no past musical experience and no current engagement in musical activities. For the 14 monolinguals with past musical experience, current engagement in musical activities ranged from daily (n = 1), weekly (n = 4), rarely (n = 3), and never (n = 6). For the 11 bilinguals with past musical experience, current engagement in musical activities ranged from daily (n = 1), weekly (n = 2), rarely (n = 6), and never (n = 2).

Method: Stimuli

Talker identification

The stimuli for the talker identification task were used in Valji (2004) and Orena et al. (2015). They consisted of auditory recordings of 12 English sentences produced by four female native English speakers and 12 French sentences produced by four female native French speakers, for a total of 96 auditory tokens. Each sentence contained between 15 and 21 syllables. The English sentences were those used in Nazzi, Jusczyk and Johnson (2000) and the French sentences were created to match

the English sentences in number of syllables (Valji, 2004). All sentences were equated for root-mean-square amplitude using the Praat software (Boersma & Weenink, 2012). For each sentence, the "quantify Source" script from GSU Praat Tools (Owren, 2008) was used to calculate sentence duration, mean fundamental frequency (F_0), and the standard deviation of F_0 (which reflects F_0 variation within each sentence). Table 1 shows the mean and standard deviation of these parameters for each of the eight talkers. Pretesting of the stimuli (reported in Orena et al., 2015) confirmed that the two sets of sentences were equally easy to learn for native speakers of each language: that is, the number of trials to reach training criterion (85% correct) was equivalent between native English listeners tested on the English sentences and native French listeners tested on the French sentences. Visual stimuli consisted of a unique cartoon avatar (i.e., colored shape with schematized eyes, nose, and mouth) for each voice.

Table 1. Mean duration (seconds), fundamental frequency (F_0, in Hz) and standard deviation (SD) of F_0 (in Hz) for the English and French talkers. Values in parentheses indicate standard deviation.

Language	Talker	Duration (s)	F_0 (Hz)	SD F_0 (Hz)
English	En01	3.036 (0.331)	198 (11)	54 (16)
	En02	2.960 (0.482)	208 (9)	42 (6)
	En03	3.389 (0.504)	207 (6)	45 (11)
	En04	2.989 (0.480)	159 (12)	35 (11)
French	Fr01	3.596 (0.219)	250 (8)	47 (9)
	Fr02	3.561 (0.308)	259 (12)	54 (8)
	Fr03	3.390 (0.175)	219 (9)	58 (11)
	Fr04	2.964 (0.311)	224 (14)	50 (17)

Pitch perception

The stimuli for the pitch perception task were provided by Xie and Myers (2015). The task consisted of two blocks, one to assess absolute pitch perception and one to assess relative pitch perception (referred to as the local and global blocks, respectively, in Xie & Myers, 2015). Stimuli for the absolute block consisted of tone sequences presented in pairs. Each sequence consisted of five tones. There were 20 "same" trials where the two members of the pair consisted of identical tone sequences and 20 "different" trials where the two members of the pair consisted of different tone sequences. Sequences of a given pair were separated by 1000 ms of silence. For the different trials, the tone sequences differed by only one tone in the sequence. Stimuli for the relative block also consisted of 40 pairs of five-tone sequences separated by 1000 ms of silence, half of which formed same trials and half of which formed different trials. For same trials, the two members of the pair shared the same tone contour but differed in absolute frequency, similar to hearing the same melody but presented in a different key. For different trials, the two members of the pair differed in both absolute frequency and tone contour.

Flanker

Stimuli for the flanker test (Eriksen & Eriksen, 1974) consisted of six linear arrays of a medial character, (>) or (<), flanked by two characters on both sides. Two arrangements represented congruent trials, (< < < < <) and (> > > > >), two represented incongruent trials, (< < > < <) and (> > < > >), and two represented neutral trials, (+ + > + +) and (+ + < + +). On each trial, the linear array was presented in the center of the computer screen and was approximately 0.6 cm in height and 3.8 cm in width.

Auditory Stroop

Stimuli for the auditory Stroop task are those used in Johns (2016) and were created following the methods outlined in Sommers and Danielson (1999) and Sommers and Huff (2003). Stimuli were 72 auditory tokens consisting of single word utterances produced by two male talkers and two female talkers. Each talker produced the words *male*, *female*, and *person* six times. Accordingly, 24 tokens showed semantic congruence with talker gender (e.g., male talker producing *male*, female talker producing *female*), 24 tokens showed incongruence (e.g., male talker producing *female*, female talker producing *male*), and 24 tokens were neutral with respect to the talker's gender and semantic content (i.e., male talker producing *person*, female talker producing *person*).

Method: Procedure

After providing informed consent, participants completed the talker identification, pitch perception, flanker, and auditory Stroop tasks in this fixed order. Testing was completed in a sound-attenuated booth. Auditory stimuli were presented over headphones (Sony MDR-7506) and visual stimuli were presented on a computer monitor. Experimental presentation and response collection were controlled by the SuperLab software (version 4.5) running on a Mac operating system (OS X). Procedural details unique to each task are described below. Participants were given a brief break in between each task and the entire procedure lasted approximately one hour.

Talker identification

The talker identification task was completed in two blocks, one for English voices and one for French voices, with order of the two languages counterbalanced within the listener groups. For each language, participants completed a familiarization phase, a training phase, and a test phase. The familiarization phase consisted of eight trials, two for each talker. On each familiarization trial, an auditory sentence was presented with the corresponding cartoon avatar. Participants were directed to learn which avatar was paired with each voice; no responses were collected during familiarization. During training, a unique randomization of the 48 sentences (4 talkers X 12 sentences) was presented. On each trial, participants were asked to identify the voice by selecting the appropriate cartoon avatar from a fixed visual array. Visual feedback was provided following each response in the form of "YES" for correct responses and "NO" for incorrect responses. The interstimulus interval (ISI) was 2000 ms, measured from the offset of the visual feedback to the onset of the next auditory stimulus. During test, a unique randomization of the 48 sentences was presented for each participant. All responses were collected via a button box. The procedure was identical to that used during training save that no feedback was provided at test.

Flanker

During the flanker task, 10 repetitions of the six linear arrangements were presented in a unique randomized order for each participant, resulting in 20 incongruent trials, 20 congruent trials, and 20 neutral trials. Each trial began with a fixation point consisting of a red circle presented in the center of the screen. The fixation point remained on the screen for 1000 ms and was immediately followed by the linear arrangement. Participants were

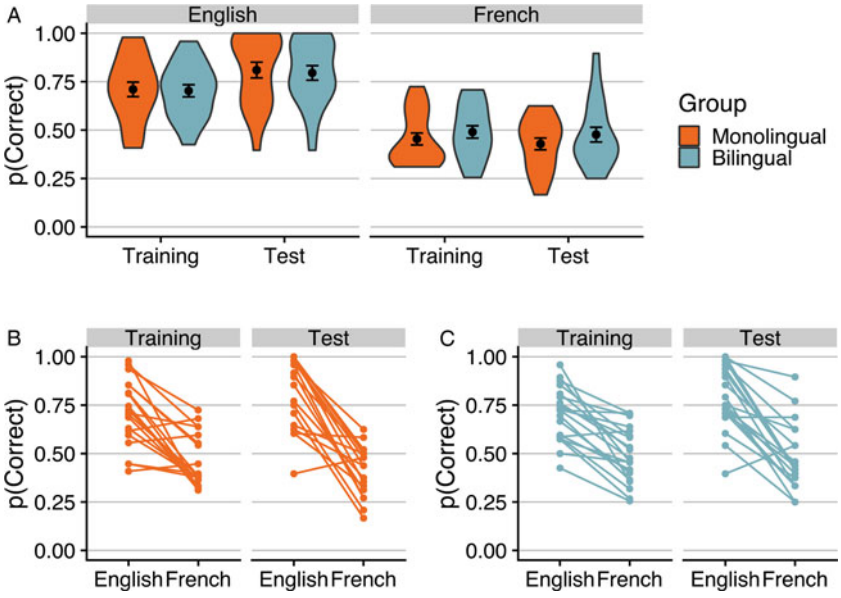

Fig. 1. Talker identification accuracy (proportion correct) for experiment 1. Panel A shows violin plots (with mean and standard error) for the monolingual and bilingual participants during training and test for the English and French voices. Panels B and C show performance for individual participants in the monolingual and bilingual groups, respectively.

directed to identify the direction of the medial arrow by pressing the appropriate arrow on the keyboard and were instructed to respond as quickly as possible without sacrificing accuracy. The ISI was 2000 ms, timed from the participant's response to the onset of the next fixation point.

Auditory Stroop

During the auditory Stroop task, a unique randomization of the 72 auditory tokens was presented for each participant. On each trial, participants were instructed to identify the gender of the talker's voice by pressing an appropriately labeled button on the response box. Each trial began with a 100 ms warning tone (1000 Hz) that was followed by 500 ms of silence prior to the presentation of the auditory token. Participants were directed to respond as quickly as possible without sacrificing accuracy. The ISI was 2000 ms, timed from the participant's response to the onset of the next warning tone.

Pitch perception

The pitch perception task consisted of two blocks, the absolute block and the relative block. In each block, a unique randomization of the 40 pairs of stimuli appropriate for each block was presented for each participant. All participants completed the absolute block followed by the relative block. On each trial, participants were instructed to press an appropriately labeled button to indicate whether the two members of the pair were the same tone sequence or were different tone sequences. Four practice trials (that included feedback) were presented prior to each block in order to ensure that the participant understood the instructions for completing the absolute and relative pitch perception tasks; feedback was not provided during the pitch perception task proper. The interstimulus interval was 2000 ms, timed from the participant's response to the onset of the next stimulus.

Results

Talker identification

All data and analysis scripts can be retrieved from https://osf.io/9dhqk/. Figure 1, panel A shows mean percent correct identification for the training and test blocks for the monolingual and bilingual listeners separately for the English and French voices. Figure 1, panels B and C show individual participants' performance so that performance for the within-subjects manipulation of language can be viewed for each participant. One-sample t-tests (reported in Table 2) confirmed that mean performance for the eight cells shown in Figure 1, panel A were all significantly above chance level. Visual inspection of Figure 1, panel A reveals a robust language familiarity effect, but no robust differences between the monolingual and bilingual groups are visually observed.

To analyze these patterns statistically, individual trial responses (0 = incorrect, 1 = correct) were fit to a generalized linear mixed-effects model (GLMM) using the glmer() function with the binomial response family from the lme4 package in R; all reported test statistics and p-values represent calculations from the glmer() function. Trials for which no response was given were excluded from the analysis (112 of 7680 trials, representing 1.5% of the data). The fixed effects were contrast-coded and included group (monolingual = −0.5, bilingual = 0.5), language (English = −0.5, French = 0.5), block (training = −0.5, test = 0.5), and all interactions. The model included random intercepts for subjects and items, and random slopes for language and block by subject. The full model results can be viewed in Table 3. The model revealed a significant effect of language ($\beta = -1.696$, SE = 0.219, $z = -7.737$, $p < 0.001$), a significant effect of block ($\beta = 0.314$, SE = 0.064, $z = 4.899$, $p < 0.001$), and a significant interaction between language and block ($\beta = -0.805$, SE = 0.112, $z = -7.157$, $p < 0.001$). No other main effect or interaction was reliable ($p \geq 0.326$ in all cases).

To explicate the nature of the interaction, two additional mixed-effects models were constructed, one for the English voices and one for the French voices. The fixed and random effects structure was identical to the full model save that language was removed as a fixed effect. For the English voices, the model showed a significant effect of block ($\beta = 1.018$, SE = 0.171, $z = 5.953$, $p < 0.001$), and neither the effect of group ($\beta = -0.204$, SE = 0.406, $z = -0.503$, $p = 0.615$) nor the block by group interaction ($\beta = -0.222$, SE = 0.318, $z = -0.700$, $p = 0.484$) was reliable.

Table 2. Mean, standard error of the mean (SE), t and p for one-sample t-tests that examined whether performance for the eight cells shown in Figure 1, panel A (experiment 1) and Figure 2, panel A (experiment 2) were significantly above chance performance. Chance was defined as proportion correct equivalent to 0.25 in experiment 1 and 0.50 in experiment 2. Degrees of freedom for the t-tests is 19 for experiment 1 and 17 for experiment 2.

	Experiment 1				Experiment 2			
Group	*Mean*	*SE*	*t*	*p*	*Mean*	*SE*	*t*	*p*
Monolinguals								
English – Training	0.711	0.038	12.324	<0.001	0.969	0.005	101.83	<0.001
English – Test	0.810	0.041	13.791	<0.001	0.978	0.005	87.701	<0.001
French – Training	0.455	0.031	6.687	<0.001	0.885	0.027	14.408	<0.001
French – Test	0.429	0.031	5.859	<0.001	0.885	0.021	18.731	<0.001
Bilinguals								
English – Training	0.703	0.031	14.497	<0.001	0.943	0.012	35.552	<0.001
English – Test	0.796	0.047	14.663	<0.001	0.966	0.008	61.711	<0.001
French – Training	0.491	0.032	7.589	<0.001	0.870	0.012	31.888	<0.001
French – Test	0.477	0.038	5.944	<0.001	0.885	0.017	22.823	<0.001

Table 3. Test statistics for the generalized linear mixed-effects models of accuracy as predicted by Group, Language, Block, and their interactions for the talker processing tasks in experiment 1 (talker identification) and experiment 2 (talker discrimination). As described in the main text, fixed effects were contrast coded (−0.5 vs. 0.5). Reference levels (in parentheses) indicate the level associated with the −0.5 contrast.

	Experiment 1				Experiment 2			
	ß	*SE*	*z*	*p*	*ß*	*SE*	*z*	*p*
Intercept	0.689	0.128	5.368	<0.001	3.244	0.234	13.863	<0.001
Group (Monolingual)	0.019	0.221	0.084	0.933	−0.375	0.225	−1.662	0.097
Language (English)	−1.696	0.219	−7.737	<0.001	−1.581	0.432	−3.660	<0.001
Block (Training)	0.314	0.064	4.899	<0.001	0.296	0.119	2.487	0.013
Group X Language	0.346	0.352	0.983	0.326	0.401	0.288	1.394	0.163
Group X Block	−0.019	0.126	−0.150	0.881	0.196	0.226	0.868	0.386
Language X Block	−0.805	0.112	−7.157	<0.001	−0.419	0.227	−1.842	0.065
Group X Language X Block	0.156	0.220	0.711	0.477	−0.046	0.448	−0.103	0.918
Observations	7568				6907			
Subjects	40				36			
Items	96				32			

For the French voices, there was no effect of group ($\beta = 0.189$, SE = 0.180, $z = 1.052$, $p = 0.293$), block ($\beta = -0.087$, SE = 0.081, $z = -1.084$, $p = 0.279$), nor an interaction between the two ($\beta = 0.062$, SE = 0.161, $z = 0.386$, $p = 0.700$). These analyses indicate that both groups of listeners showed better voice recognition for the English compared to the French voices, and that for the English voices, identification accuracy improved at test compared to training. There is no evidence indicating that identification accuracy differed between the monolingual and bilingual participants.

Predictors of talker identification

As shown in Figure 1, there was wide individual variability in performance for the talker identification task. The analyses reported above found no evidence to indicate that this variability can be attributed to the group manipulation. Because we did not observe a bilingual advantage for voice processing, the individual difference measures cannot be used as intended, and for that reason the following analysis should be considered exploratory in nature. In order to examine whether talker identification was linked to pitch perception or cognitive control, GLMMs were constructed with performance on the three individual difference measures serving as predictors of talker identification accuracy.

Performance on the pitch perception task was calculated as outlined in Xie and Myers (2015). Trials in which no response was provided (8 of 3200 trials) were removed from the analysis. Sensitivity (d') was calculated for each participant separately for the absolute and relative blocks using hit (H) and false alarm (FA) rates [d' = z(H) – z(FA)], applying a correction for participants with perfect accuracy to avoid an infinite d' (hit rate of 1.00 was corrected to 0.99; false alarm rate of 0.00 was corrected to 0.01). As in Xie and Myers (2015), performance on the pitch perception task was then quantified as the average d' across the absolute and relative blocks, which were highly correlated ($r = 0.559$, $p < 0.001$).

Performance for the congruent and incongruent trials of the flanker and auditory Stroop tasks was quantified separately for each participant as follows. First, missed and incorrect trials were removed from the analysis, representing 1.0% of the flanker data and 2.8% of the auditory Stroop data. Second, trials for which log RT deviated by more than 2.5 SDs of each participant's mean log RT were removed, representing 3.1% of the flanker data and 1.7% of the auditory Stroop data. Finally, an inhibition score was calculated as the difference in log RT between the incongruent and congruent trials; with this metric, higher values indicate decreased inhibitory control.

Separate GLMMs were performed, one for the English voices and one for the French voices, in light of previous results showing that pitch perception influences talker processing for a nonnative language, but not the native language (Xie & Myers, 2015). The structure of both models was identical. The dependent measure was accuracy at the trial level (0 = incorrect, 1 = correct). The fixed effects were block, pitch perception (average d'), flanker inhibition score (logRT incongruent – logRT congruent), auditory Stroop inhibition score (log RT incongruent – logRT congruent), and the interactions between block and each individual difference measure. Block was contrast coded (training = −0.5, test = 0.5). The three individual difference measures were each entered as a continuous variable, scaled and centered around the mean. The model also included random intercepts by subject and by item.

The results of the two models are shown in Table 4. For the English voices, neither pitch perception nor inhibition on the auditory Stroop task predicted accuracy on the talker identification task ($p \geq 0.126$ in both cases). However, the model showed a main effect of flanker inhibition score ($\beta = -0.384$, SE = 0.180, $z = -2.133$, $p = 0.033$), and an interaction between flanker score and block ($\beta = -0.247$, SE = 0.094, $z = -2.621$, $p = 0.009$). Results from follow-up models performed separately for each block showed that better cognitive control (i.e., a smaller flanker inhibition score) was associated with increased talker identification accuracy during test ($\beta = -0.577$, SE = 0.257, $z = -2.242$, $p = 0.025$), but not during training ($\beta = -0.236$, SE = 0.149, $z = -1.587$, $p = 0.112$). For the French voices, higher sensitivity on the pitch perception task was associated with increased talker identification accuracy ($\beta = 0.247$, SE = 0.082, $z = 3.021$, $p = 0.003$), as was better cognitive control for the flanker task ($\beta = -0.180$, SE = 0.083, $z = -2.170$, $p = 0.030$). No other main effects or interactions were observed ($p \geq 0.147$ in all cases).

The results of experiment 1 showed heightened talker processing for the native compared to the nonnative language, but did not reveal a bilingual advantage for either language. For the nonnative language, individual differences in talker identification accuracy were predicted by pitch perception, replicating previous findings (Xie & Myers, 2015). For both languages, enhanced cognitive control as measured by the inhibition score on the flanker task was associated with more accurate talker identification, though this relationship was limited to the test phase for the English voices. There was no relationship between auditory Stroop performance and talker identification in either language. In experiment 2 we provide an additional test of the bilingual advantage for voice recognition, using a discrimination task instead of an identification task. This experiment also provides an additional test of the relationships between pitch perception and cognitive control that were observed in experiment 1 in order to determine whether they will also be observed when talker processing is assessed in terms of discrimination.

Experiment 2

Method

Participants

Thirty-six adults (8 males, 28 females) between the ages of 18 and 31 years (mean = 21, SD = 3) who did not participate in experiment 1 were recruited from the University of Connecticut for participation in the experiment. Seven additional participants were tested but excluded from the study due to the button box becoming disconnected during the talker identification task (n = 1), experiment error (the participant was run twice in the English task instead of one for each of the English and French tasks, n = 1), acquiring the second language after the age of 5 (n = 4), or reporting proficiency in French above the level of 2 on the 11-point scale described for experiment 1 (n = 1). No participant had a history of speech, language, or hearing disorders according to self-report. All participants passed a pure tone hearing screen on the day of participation administered at 20 dB for octave frequencies between 500 and 4000 Hz. Participants received either monetary compensation or partial course credit for their participation. All testing procedures and informed consent acquisition followed protocols approved by the University of Connecticut Institutional Review Board.

Of the 36 adults, 18 participants were monolingual speakers of English and 18 participants were bilingual speakers of English and various second languages. The second languages included Bengali (n = 2), Bosnian (n = 1), Dutch (n = 1), German (n = 1), Hindi (n = 1), Kannada (n = 1), Mandarin (n = 1), Spanish (n = 5), Tagalog (n = 1), Taishanese (n = 1), Telugu (n = 1), and Ukrainian (n = 2). All bilinguals reported acquiring both languages prior to age 5; no monolingual reported proficiency in any language other than English.

Experience and proficiency with French was assessed using the questionnaire described for experiment 1. No participant reported current exposure to French. Fifteen monolinguals and 12 bilinguals reported no past exposure to French. The three monolinguals who did report past exposure to French indicated instruction in middle school. For the six bilinguals who reported past exposure to French, this ranged from instruction in elementary school (n = 1), middle school (n = 2), high school (n = 1), and college (n = 2). With respect to proficiency in French, two monolinguals indicated a 2 (low) on the 11-point scale, one indicated a 1 (very low), and all other monolinguals indicated 0 (none). For the bilingual participants, four participants indicated a 1 (very low) with all other bilinguals indicating a 0 (none) as their level of proficiency in French.

Two monolinguals and five bilinguals indicated no past musical experience and no current engagement in musical activities. For the 16 monolinguals with past musical experience, current engagement in musical activities ranged from daily (n = 1), weekly (n = 1), monthly (n = 2), rarely (n = 3), and never (n = 9). For the 13 bilinguals with past musical experience, current engagement in musical activities ranged from daily (n = 1), weekly (n = 1), monthly (n = 2), rarely (n = 4), and never (n = 5).

Stimuli

Stimuli for Experiment 2 were identical to those for Experiment 1. However, the auditory sentences used for the talker identification task in Experiment 1 were arranged into pairs for the talker discrimination task. For each language (i.e., English and French), 48 pairs of stimuli were created using the 48 auditory sentences described for Experiment 1 (4 talkers X 12 auditory sentences

Table 4. Test statistics for the generalized linear mixed-effects models of English and French talker identification accuracy as predicted by performance on the individual difference measures in experiment 1. Pitch perception was quantified by average d' for the absolute and relative pitch processing tasks. Inhibition on the flanker and auditory Stroop tasks was quantified as the difference in log reaction time for incongruent and congruent trials. As described in the main text, the individual difference measures were entered as continuous variables, each scaled and centered around the mean. Block was contrasted coded (training = −0.5, test = 0.5).

	English voices				French voices			
	ß	*SE*	*z*	*p*	ß	*SE*	*z*	*p*
Intercept	1.604	0.224	7.168	<0.001	−0.156	0.089	−1.757	0.079
Block (Training)	0.708	0.087	8.095	<0.001	−0.087	0.068	−1.275	0.202
Pitch perception	0.268	0.175	1.531	0.126	0.247	0.082	3.021	0.003
Flanker	−0.384	0.180	−2.133	0.033	−0.180	0.083	−2.170	0.030
Auditory Stroop	−0.066	0.172	−0.383	0.701	−0.117	0.081	−1.450	0.147
Block X Pitch perception	0.077	0.090	0.855	0.393	−0.068	0.070	−0.962	0.336
Block X Flanker	−0.247	0.094	−2.621	0.009	−0.086	0.072	1.196	0.232
Block X Auditory Stroop	0.094	0.080	1.174	0.241	−0.004	0.070	−0.055	0.956
Observations	3800				3768			
Subjects	40				40			
Items	48				48			

for each language). For each talker, six "same" pairs were created that held talker voice constant. Also for each talker, six "different" pairs were created such that the talker voice differed across the members of the pair; each talker was paired with the other three talkers twice, with order of the two voices reversed (i.e., first vs. second member of the pair) for the respective talker pairing. Lexical content always varied between the two members of a given pair, and each of the 48 sentences was presented exactly twice. Thus, there was no systematic relationship between the linguistic content of an utterance and a given speaker. This procedure resulted in 24 same pairs and 24 different pairs for the talker discrimination task; specific pair members for each language can be retrieved at https://osf.io/9dhqk/.

Procedure

The procedure followed that outlined for Experiment 1 except that listeners completed a talker discrimination task instead of an identification task. The stimuli for the voice discrimination task consisted of the 48 pairs of stimuli described above. Each of the language blocks (English vs. French) began with a familiarization phase, as described for experiment 1, which was followed by training and test phases. During training, a unique randomization of the 48 pairs was presented for each participant; on each trial, listeners were instructed to identify whether the two members of the pair were spoken by the same speaker or by different speakers. Feedback was provided during the training phase as outlined for Experiment 1. During test, a unique randomization of the 48 stimulus pairs was presented for each participant. Participants were again asked to indicate whether the two members of the pair were spoken by the same voice or by different voices, but no feedback was provided.

Results

Talker discrimination

All data and analysis scripts can be retrieved https://osf.io/9dhqk/. Figure 2, panel A shows mean proportion correct discrimination for the training and test blocks across the monolingual and bilingual listeners separately for the English and French voices. Figure 2, panels B and C show individual participants' performance. One-sample t-tests (shown in Table 2) confirmed that performance for each cell shown in Figure 2, panel A was significantly above chance. Visual inspection of Figure 2 shows a language familiarity effect, but no robust differences between the monolingual and bilingual groups are visually observed.

To analyze these patterns statistically, individual trial responses (0 = incorrect, 1 = correct) were fit to a generalized linear mixed-effects model using the glmer() function with the binomial response family from the lme4 package (Bates et al., 2014) in R; all test statistics and p-values represent the calculations from the glmer() function. Trials in which no response was given (6 of 6912 trials) were excluded. The fixed effects were contrast-coded and included group (monolingual = −0.5, bilingual = 0.5), language (English = −0.5, French = 0.5), block (training = −0.5, test = 0.5), and all interactions. The model included random intercepts for subjects and items, and random slopes for language by subject and block by subject. Item was coded as the talker pairing on each trial, preserving order (e.g., a pairing consisting of Talker 1 followed by Talker 2 was a different item than the pairing of Talker 2 followed by Talker 1). The full model results are shown in Table 3. The model revealed a significant effect of language ($\beta = -1.581$, SE = 0.432, $z = -3.660$, $p < 0.001$), with accuracy lower for the French compared to the English voices, and a significant effect of block ($\beta = 0.296$, SE = 0.119, $z = 2.487$, $p = 0.013$), indicating improved accuracy at test compared to training. No other main effect or interaction was reliable ($p \geq 0.065$ in all cases).

Predictors of talker discrimination

A set of GLMMs were performed to examine whether performance on the individual difference measures predicted talker discrimination accuracy, one for the English voices and one for the French voices. Performance on the individual difference measures was quantified as outlined for experiment 1. Pitch perception was

Table 5. Test statistics for the generalized linear mixed-effects models of English and French talker discrimination accuracy as predicted by performance on the individual difference measures in experiment 2. Pitch perception was quantified by average d' for the absolute and relative pitch processing tasks. Inhibition on the flanker and auditory Stroop tasks was quantified as the difference in log reaction time for incongruent and congruent trials. As described in the main text, the individual difference measures were entered as continuous variables, each scaled and centered around the mean. Block was contrasted coded (training = −0.5, test = 0.5).

	English voices				French voices			
	ß	*SE*	*z*	*p*	*ß*	*SE*	*z*	*p*
Intercept	4.040	0.330	12.239	<0.001	2.458	0.301	8.171	<0.001
Block (Training)	0.509	0.198	2.573	0.010	0.122	0.118	1.031	0.302
Pitch perception	0.289	0.140	2.063	0.039	0.284	0.117	2.429	0.015
Flanker	−0.199	0.124	−1.605	0.108	−0.205	0.112	−1.834	0.066
Auditory Stroop	−0.112	0.122	−0.923	0.356	−0.112	0.112	−0.994	0.320
Block X Pitch perception	0.025	0.210	0.118	0.906	−0.070	0.127	−0.533	0.580
Block X Flanker	−0.060	0.174	−0.347	0.729	−0.199	0.110	−1.816	0.069
Block X Auditory Stroop	0.245	0.173	1.415	0.157	−0.146	0.117	−1.247	0.212
Observations	3455				3452			
Subjects	36				36			
Items	16				16			

measured as the average d' for the absolute and relative blocks of the pitch perception task, which were highly correlated ($r = 0.666$, $p < 0.001$). Error rates for the flanker and auditory Stroop tasks were very low (0.7% and 2.0%, respectively). For the flanker task, 34 of 1430 correct responses were removed as RT outliers (i.e., trials for which log RT exceeded 2.5 SDs of the participant's mean log RT), representing 2.4% of the data. For the auditory Stroop task, 37 of 1694 correct responses were removed as RT outliers (representing 2.2% of the data). For each participant, an inhibition score for each of the flanker and auditory Stroop tasks was calculated as the difference in log RT for the incongruent and congruent trials.

The results of the two models are shown in Table 5. There was a significant relationship between pitch perception and talker discrimination such that higher sensitivity on the pitch perception task was associated with increased accuracy on the talker discrimination task for the English voices ($\beta = 0.289$, SE = 0.140, $z = 2.063$, $p = 0.039$) and the French voices ($\beta = 0.284$, SE = 0.117, $z = 2.429$, $p = 0.015$). No other main effects or interactions were observed for either the English or French voices ($p \geq 0.066$ in all cases).[1]

Comparisons between monolingual and bilingual listeners for individual difference measures

As stated in the introduction, performance on the pitch perception, flanker, and auditory Stroop tasks was measured in order to identify potential mechanisms in support of a bilingual advantage for voice processing. Because no bilingual advantage for voice processing was observed in the current work, these measures cannot be used as intended. Nonetheless, we pooled data across experiments in order to examine whether a bilingual advantage would be observed in any of the individual difference measures, explicitly noting that these are not planned analyses.

Performance for the monolingual and bilingual participants on the three individual difference measures is shown in Figure 3. The top row of Figure 3, panel A shows sensitivity on the pitch perception task in terms of average d' across the absolute and relative blocks; the bottom panel shows sensitivity in each block for each participant. There was no difference between the two groups regarding average d' in the pitch perception task [$t(74) = -0.502$, $p = 0.617$]. This pattern also held for sensitivity within the absolute [$t(74) = 0.137$, $p = 0.892$] and relative [$t(74) = -1.159$, $p = 0.250$] blocks.

The top row of Figure 3, panel B shows the inhibition score for the flanker task, measured in terms of the difference in RT between the incongruent and congruent trials, with the bottom panel showing RT for each trial type for each participant. Analysis of the flanker inhibition score (expressed in terms of log RT) showed a significant difference between the two groups [$t(74) = -3.000$, $p = 0.004$], with monolinguals (mean = 80 ms, SE = 11 ms) showing a smaller inhibition effect compared to bilinguals (mean = 139 ms, SE = 18 ms). Figure 3, panel C shows the inhibition score for the auditory Stroop task (top panel) and mean RT to incongruent and congruent trials for each participant (bottom panel). Analysis of the auditory Stroop inhibition score (in terms of log RT) showed no difference between monolingual and bilingual participants [$t(74) = 0.368$, $p = 0.714$].[2]

[1]Parallel models were constructed including Group (Monolingual vs. Bilingual) as an additional fixed effect for the analysis of the individual difference measures as predictors of talker identification (experiment 1) and talker discrimination (experiment 2). In all cases, the results converged with the models presented in the main text. These analyses can be viewed on the OSF repository associated with this manuscript.

[2]Performance for the flanker and auditory Stroop tasks was also analyzed using linear-mixed effects models of trial-level data in order to explicitly model subject-level variation. The results converged with the approach presented in the main text, which can be viewed on the OSF repository. In addition, visual inspection of the individual subject variation in Figure 3 reveals three participants who showed extreme deviation in their inhibition scores relative to other participants. The group comparisons were performed for the flanker and auditory Stroop tasks removing these participants, and the results patterned as those performed with all participants. For the flanker task, the inhibition score was smaller for the monolinguals (mean = 80 ms, SE = 11 ms) compared to the bilinguals [mean = 122 ms, SE = 13 ms; $t(71) = -2.678$, $p = 0.009$]. For the auditory Stroop task, there was no difference in the inhibition score between the two groups [$t(71) = -0.423$, $p = 0.674$].

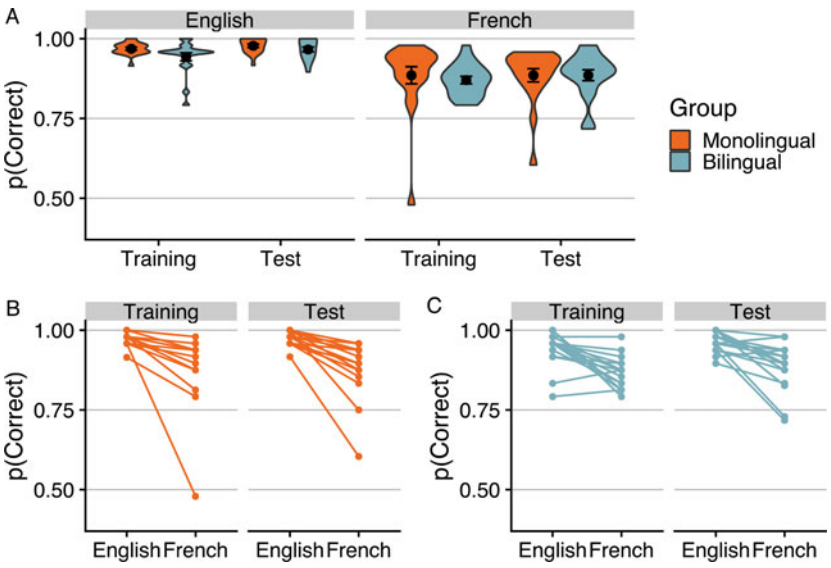

Fig. 2. Talker discrimination performance (proportion correct) for experiment 2. Panel A shows violin plots (with mean and standard error) for the monolingual and bilingual participants during training and test for the English and French voices. Panels B and C show performance for individual participants in the monolingual and bilingual groups, respectively.

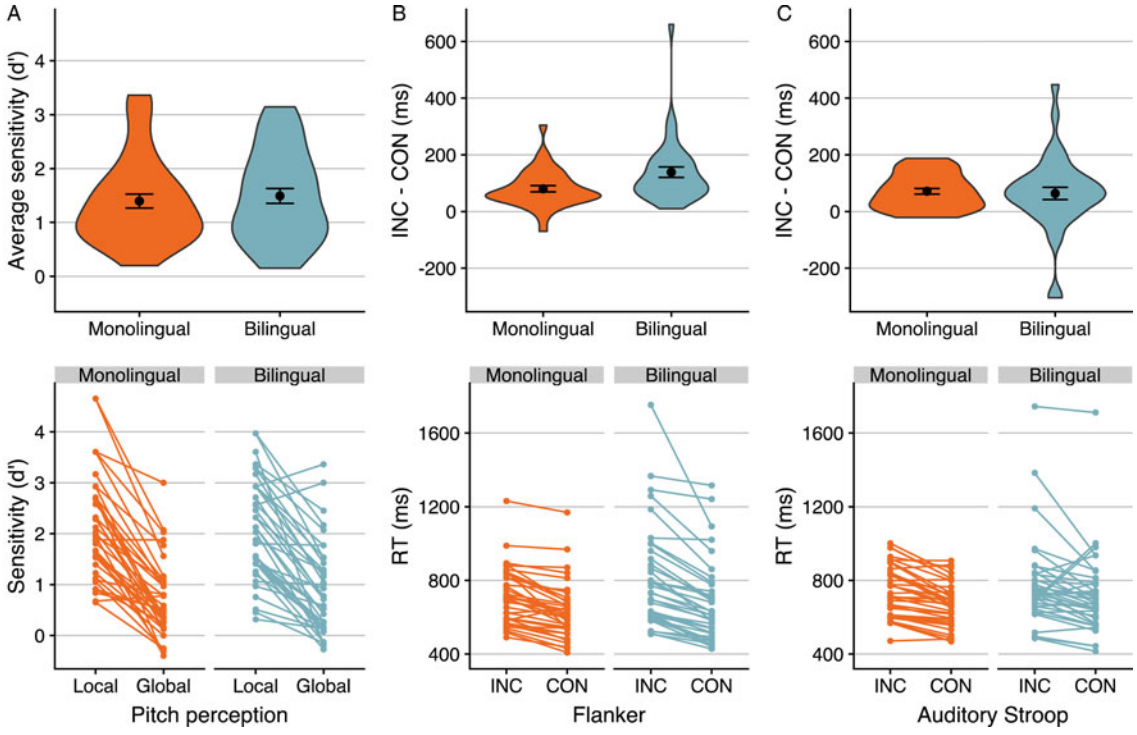

Fig. 3. Comparisons between monolingual and bilingual listeners for the pitch perception task (A), the flanker task (B), and the auditory Stroop task (C). Panel A shows performance in terms of average sensitivity (d') for the two blocks of the pitch perception task (top row) and sensitivity (d') in each block for each participant (bottom row). The top row of panels B and C show performance in terms of the inhibition score (INC – CON, in ms). The bottom row of panels B and C shows mean reaction time (ms) for each item type for each participant. In all panels, the top row shows violin plots (with mean and standard error). To ease interpretation, performance for the flanker and auditory Stroop tasks is shown in terms of reaction time; note that log-transformed reaction time was used for all statistical analyses.

Discussion

It is well established that voice recognition is heightened when listeners have familiarity with the linguistic structure of a talker's message (Goggin et al., 1991; Johnson et al., 2018b, 2011; Orena et al., 2015; Perrachione et al., 2011). Other investigations have implicated a role for language-independent mechanisms to support voice recognition, including pitch perception (Xie & Myers, 2015). Furthermore, a bilingual advantage for voice recognition has been observed in children for voices speaking a familiar language and for voices speaking an unfamiliar language (Levi, 2018). Here we examined whether a bilingual advantage for voice recognition would be observed in adults, and if so, whether it was linked to advantages in pitch perception or cognitive control. The results of the two experiments converged to provide no evidence of a bilingual advantage for voice processing, though

both groups showed improved voice processing for the familiar compared to an unfamiliar language. The lack of a bilingual advantage for voice processing observed here is consistent with results of Xie and Myers (2015), but contrasts with results from Levi (2018).

Here we consider four possibilities that may account for the discrepant results between the current study and Levi (2018). First, the discrepancy may reflect the different methodology of the two studies, including the length of the stimuli used to assess talker processing. The stimuli in Levi (2018) were individual words whereas the current study used sentence-length stimuli; stimuli in Xie and Myers (2015) were also of sentence length. Previous research has shown that talker identification is facilitated by stimuli with longer duration (e.g., Goggin et al., 1991). Accordingly, it may be the case that a bilingual advantage was masked in the current results due to an overall easier task than was used in Levi (2018). It could also be the case that the bilingual advantage reflects an enhancement in the use of cues to voice identification that are dominant in the limited temporal timescale of individual words. Second, the lack of a bilingual advantage in the current work may reflect the lack of novel stimuli presented at test. Levi (2018) tested generalization of learning to novel stimuli, whereas the current work tested stability of learning for trained stimuli. In some studies where performance at test for trained versus novel stimuli was directly examined, no robust differences in accuracy were observed for the two types of stimuli (e.g., Orena et al., 2015; Kadam, Orena, Theodore & Polka, 2016), but these studies did not examine how generalization may differ between monolinguals and bilinguals. Future work is needed in order to examine whether the bilingual advantage will emerge for generalization of learning when tested in adult listeners.[3]

A third explanation is that the adult monolinguals may have had previous exposure to other languages, as the monolingual children tested in Levi (2018) had no exposure to a nonnative language. Past research shows that CURRENT exposure to a nonnative language improves talker processing in that language (Orena et al., 2015) and that formal instruction in a nonnative language improves talker processing in THAT language (Sullivan & Schlichting, 2000). What is not yet known is whether exposure to any nonnative language promotes a talker processing benefit for either the NATIVE language or for OTHER nonnative languages. Future research that specifically examines voice processing between adults with and without past exposure to any nonnative language is needed to test this account.

A fourth possibility is that the disparate results of the current work and Levi (2018) may indicate that the bilingual advantage is present during childhood but not in adulthood, consistent with findings showing that bilingual advantages in other domains are more robust in children and older adults compared to college-aged populations (e.g., Bialystok, 2017). As for language acquisition in general, the emergence of voice recognition abilities occurs through a protracted period of development. Previous research has demonstrated increased voice processing abilities for adults compared to children (Fecher & Johnson, 2018a; Levi & Schwartz, 2013) and for older children compared to younger children (Levi, 2018; Levi & Schwartz, 2013). The developmental trajectory of the language familiarity effect for talker processing is complex in that it is observed in infancy but grows stronger during early childhood (Fecher & Johnson, 2018a; Johnson et al., 2011). Fecher and Johnson (2018a) found that 6-year-old children, but not 5-year-old children, exhibited a language familiarity effect, which they hypothesize reflects an increased awareness in the 6-year-olds of the linguistic sound structure that occurs concomitantly with the onset of literacy instruction. As elegantly outlined in Fecher and Johnson (2018a), a critical consideration for developmental accounts of the linguistic influences on voice recognition is to determine whether the strengthening of the language familiarity effect that occurs during childhood necessarily entails a trade-off in performance for native versus nonnative voices. That is, do the factors that support the development of voice recognition in the native language simultaneously diminish voice recognition in nonnative languages? If the native language benefit for voice recognition is indeed the consequence of "tuning in" to native language sound patterns at the expense of diminished sensitivity to nonnative sound patterns, then one possible explanation for the presence of a bilingual advantage in voice processing in children but not adults may be that the window of time before voice processing dominantly relies upon knowledge of the native language sound structure is extended in bilingual children.

As was observed in Xie and Myers (2015), better performance on the pitch perception task was linked to better voice processing for the nonnative voices for both the talker identification and discrimination tasks. For discrimination, this relationship was also observed for the English voices. Inhibition as assessed using the flanker task was a limited predictor of voice processing, with better cognitive control (i.e., lower inhibition scores) associated with increased talker identification at test for the English voices, and increased talker identification at both training and test for the French voices. Performance on the flanker task did not predict voice processing when measured using the talker discrimination task. Future research is needed in order to explicate the mechanisms by which inhibitory control influences talker identification but not talker discrimination. One possibility is that it may reflect the ability to ignore the irrelevant visual information provided in the avatars used for the talker identification task. Though there was no difference between monolingual and bilingual listeners in terms of talker processing, the monolingual participants did show enhanced cognitive control on the flanker task in terms of a smaller inhibition effect, which is perhaps not surprising given evidence that the bilingual advantage for cognitive control is minimally observed in college-aged populations (e.g., Bialystok, 2017) and other reports of a bilingual disadvantage for cognitive control in college-aged populations (e.g., Paap & Greenberg, 2013).

No reliable relationships were observed between voice processing and inhibition as assessed using the auditory Stroop task, which suggests that a model of auditory inhibition may not best characterize the voice recognition process. On the surface, the auditory Stroop task is very similar to the voice identification task used in the present work. In the auditory Stroop task, participants were asked to identify the gender of the talker while explicitly ignoring the semantic content, which in some cases was incongruent with the talker's gender. In the talker identification task, participants were asked to identify the talker, with semantic content being an irrelevant dimension for making this decision.

[3]A pilot study (n = 10 monolinguals, n = 10 bilinguals) was conducted following the procedures of the talker identification task used in experiment 1 except that novel sentences were included at test to assess generalization of learning. The results of this pilot test did not reveal any bilingual advantage for either the trained or novel items at test, though these results should be interpreted with caution given the small sample size. The raw data and analysis script for this pilot study are available in the OSF repository.

Despite the similarity between the two tasks, individuals who were better on one were not better on the other, which raises the possibility that performance on the two tasks reflects disassociated processes. Indeed, given the rich literature demonstrating that talkers systematically differ in how they implement individual speech sounds (Chodroff & Wilson, 2017; Hillenbrand, Getty, Clark & Wheeler, 1995; Newman, Clouse & Burnham, 2001; Theodore, Miller & DeSteno, 2009), attending to the linguistic stream – instead of inhibiting the linguistic content – provides access to a rich set of information that can facilitate identification and discrimination of voices beyond those cues available in the absence of talker-specific phonetic variation (e.g., vocal source characteristics such as fundamental frequency).

To conclude, the results of the current experiments contribute to the body of literature that demonstrates an exquisite interplay between the processing of talker and linguistic information. Indeed, the primary determinant of voice processing in the current work was the degree to which listeners were familiar with the linguistic content of the talker's message; both groups of listeners showed better voice processing for English compared to French voices, and being fluent in more than one language did not lead to improved processing for either language. Pitch perception was further identified as a determinant of voice processing, showing a facilitative effect for identification and discrimination of nonnative voices, and discrimination of native voices. Cognitive control as assessed via inhibition on the flanker and auditory Stroop tasks showed limited and null influences on voice processing ability, respectively, suggesting that language-specific linguistic knowledge and general auditory processing are the primary determinants among those evaluated here. In moving forward, theories that can account for the role of bilingual experience on voice processing will be advanced though further examination of how bilingual experience shapes voice processing across the developmental trajectory in order to specify the factors that underlie its influence during childhood but not in the mature perceiver.

Author ORCIDs. Rachel M. Theodore, 0000-0002-4601-8340

Acknowledgements. This study was funded by a seed grant from the Connecticut Institute for the Brain and Cognitive Sciences to R. Theodore. We express gratitude to Yinyin Gu for her assistance with experiment programming and data collection. We also thank Linda Polka, Xin Xie, and Alexis Johns for generously providing their stimuli for use in the current work.

References

Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H and Grothendieck G (2014) Package 'lme4.' *R Foundation for Statistical Computing, Vienna, 12.*

Bialystok E (2017) The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin* **143**(3), 233–262.

Bialystok E and Grundy JG (2018) Science does not disengage. *Cognition* **170**, 330–333.

Bradlow AR and Pisoni DB (1999) Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America* **106**(4), 2074–2085.

Bregman MR and Creel SC (2014) Gradient language dominance affects talker learning. *Cognition* **130**(1), 85–95.

Chodroff E and Wilson C (2017) Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* **61**, 30–47.

Clarke CM and Garrett MF (2004) Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America* **116**(6), 3647–3658.

Clayards M, Tanenhaus MK, Aslin RN and Jacobs RA (2008) Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* **108**(3), 804–809.

Drouin JR, Theodore RM and Myers EB (2016) Lexically guided perceptual tuning of internal phonetic category structure. *The Journal of the Acoustical Society of America* **140**(4), EL307–EL313. https://doi.org/10.1121/1.4964468

Eriksen BA and Eriksen CW (1974) Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics* **16** (1), 143–149. https://doi.org/10.3758/BF03203267

Fecher N and Johnson EK (2018a) Effects of language experience and task demands on talker recognition by children and adults. *The Journal of the Acoustical Society of America* **143**(4), 2409–2418.

Fecher N and Johnson EK (2018b) The native-language benefit for talker identification is robust in 7.5-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Fleming D, Giordano BL, Caldara R and Belin P (2014) A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences* **111**(38), 13795–13798.

Goggin JP, Thompson CP, Strube G and Simental LR (1991) The role of language familiarity in voice identification. *Memory & Cognition* **19**(5), 448–458.

Hillenbrand J, Getty LA, Clark MJ and Wheeler K (1995) Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America* **97**(5), 3099–3111.

Johns A (2016) *Sensory and cognitive influences on lexical competition in spoken word recognition in younger and older listeners* (Unpublished doctoral dissertation). University of Connecticut, Storrs, Connecticut.

Johnson EK, Bruggeman L and Cutler A (2018) Abstraction and the (misnamed) language familiarity effect. *Cognitive Science* **42**(2), 633–645. https://doi.org/10.1111/cogs.12520

Johnson EK, Westrek E, Nazzi T and Cutler A (2011) Infant ability to tell voices apart rests on language experience. *Developmental Science* **14**(5), 1002–1011.

Kadam MA, Orena AJ, Theodore RM and Polka L (2016) Reading ability influences native and non-native voice recognition, even for unimpaired readers. *The Journal of the Acoustical Society of America* **139**(1), EL6–EL12.

Köster O and Schiller NO (1997) Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics* **4**, 18–28.

Levi SV (2018) Another bilingual advantage? Perception of talker-voice information. *Bilingualism: Language and Cognition* **21**(3), 523–536.

Levi SV and Schwartz RG (2013) The development of language-specific and language-independent talker processing. *Journal of Speech, Language, and Hearing Research* **56**(3), 913–925.

Newman RS, Clouse SA and Burnham JL (2001) The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America* **109**(3), 1181–1196.

Nygaard LC and Pisoni DB (1998) Talker-specific learning in speech perception. *Attention, Perception, & Psychophysics* **60**(3), 355–376.

Orena AJ, Theodore RM and Polka L (2015) Language exposure facilitates talker learning prior to language comprehension, even in adults. *Cognition* **143**, 36–40.

Owren MJ (2008) GSU Praat Tools: Scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods* **40**(3), 822–829. https://doi.org/10.3758/BRM.40.3.822

Paap KR and Greenberg ZI (2013) There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology* **66**(2), 232–258.

Paap KR, Myuz HA, Anders RT, Bockelman MF, Mikulinsky R and Sawi OM (2017) No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology* **29**(2), 89–112.

Perrachione TK, Del Tufo SN and Gabrieli JD (2011) Human voice recognition depends on language ability. *Science* **333**(6042), 595–595.

Perrachione TK and Wong PC (2007) Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia* **45**(8), 1899–1910.

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

Sommers MS and Danielson SM (1999) Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging* **14**(3), 458.

Sommers MS and Huff LM (2003) The effects of age and dementia of the Alzheimer's type on phonological false memories. *Psychology and Aging* **18**(4), 791.

Sullivan KP and Schlichting F (2000) Speaker discrimination in a foreign language: First language environment, second language learners. *Forensic Linguistics* **7**(1), 95–111.

Theodore RM, Miller JL and DeSteno D (2009) Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America* **125**(6), 3974–3982. https://doi.org/10.1121/1.3106131

Theodore RM, Myers EB and Lomibao JA (2015) Talker-specific influences on phonetic category structure. *The Journal of the Acoustical Society of America* **138**(2), 1068–1078.

Valji A (2004) *Language preference in monolingual and bilingual infants* (Unpublished Master's thesis). McGill University, Montreal, Quebec.

Wester M (2012) Talker discrimination across languages. *Speech Communication* **54**(6), 781–790. https://doi.org/10.1016/j.specom.2012.01.006

Winters SJ, Levi SV and Pisoni DB (2008) Identification and discrimination of bilingual talkers across languages. *The Journal of the Acoustical Society of America* **123**(6), 4524–4538.

Xie X and Myers E (2015) The impact of musical training and tone language experience on talker identification. *The Journal of the Acoustical Society of America* **137**(1), 419–432. https://doi.org/10.1121/1.4904699

Ye Y, Mo L and Wu Q (2017) Mixed cultural context brings out bilingual advantage on executive function. *Bilingualism: Language and Cognition* **20**(4), 844–852. https://doi.org/10.1017/S1366728916000481