Cognition 130 (2014) 85–95

Contents lists available at ScienceDirect

Cognition

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

Gradient language dominance affects talker learning

Micah R. Bregman *, Sarah C. Creel

Department of Cognitive Science, UC San Diego, La Jolla, CA, United States

ARTICLE INFO

Article history:
Received 20 April 2012
Received in revised form 27 September 2013
Accepted 30 September 2013
Available online 5 November 2013

Keywords:
Speech perception
Talker recognition
Music perception
Voice identification
Bilingualism
Language dominance

ABSTRACT

Traditional conceptions of spoken language assume that speech recognition and talker identification are computed separately. Neuropsychological and neuroimaging studies imply some separation between the two faculties, but recent perceptual studies suggest better talker recognition in familiar languages than unfamiliar languages. A familiar-language benefit in talker recognition potentially implies strong ties between the two domains. However, little is known about the nature of this language familiarity effect. The current study investigated the relationship between speech and talker processing by assessing bilingual and monolingual listeners' ability to learn voices as a function of language familiarity and age of acquisition. Two effects emerged. First, bilinguals learned to recognize talkers in their first language (Korean) more rapidly than they learned to recognize talkers in their second language (English), while English-speaking participants showed the opposite pattern (learning English talkers faster than Korean talkers). Second, bilinguals' learning rate for talkers in their second language (English) correlated with age of English acquisition. Taken together, these results suggest that language background materially affects talker encoding, implying a tight relationship between speech and talker representations.

1. Introduction

Recent studies suggest a relationship between knowing a language and ability to identify talkers in that language (Goggin et al., 1991; Johnson et al., 2011; Perrachione et al., 2011; Sullivan and Schlichting, 2000; Winters et al., 2008). This radically departs from a viewpoint that speech processing operates over an abstract set of symbols. However, it is not clear what degree of language experience, or what specific *type* of language experience is relevant for talker recognition. On one hand, some studies suggest that limited exposure to a new language is sufficient to facilitate recognition of voices in that language (Sullivan and Schlichting, 2000, in adults; Johnson et al., 2011, in 7-month-olds). On the other hand, other studies suggest that the language-talker relationship is more complex. For instance, listeners are better at learning to identify voices in their native language than in a foreign language (Goggin et al.; Winters et al.). Further, Perrachione et al. recently showed that dyslexic listeners' degree of phonological impairment predicted difficulty in a talker learning task in a familiar (but not an unfamiliar) language. This suggests a link between subtle phonetic knowledge and talker identification. However, it leaves open the possibility that dyslexic listeners had broader deficits in auditory processing, rather than a linked deficit in phonological encoding and talker identification.

A different approach to investigating the link between language knowledge and talker recognition would be to assess normal listeners with extensive language experience but weaker phonetic knowledge—specifically, second-language listeners (Flege, 1988; Flege et al., 2006). If lengthy experience with a language permits excellent talker recognition in that language, then late but skilled learners should be good at talker recognition. However, if *subtle phonetic knowledge* acquired both over the long term and early in life is key, then early learners of a second language

* Corresponding author. Tel.: +1 213 210 8559.
E-mail address: mbregman@ucsd.edu (M.R. Bregman).

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.09.010

should be far more adept than late learners at talker recognition in that language.

The goal of the current study was to understand the relationship between language knowledge and listeners' abilities to encode talker characteristics. What role does one's language background play in learning to recognize voices? Below, we review evidence for and against connections between speech processing and talker recognition. We then describe a study designed to elucidate the nature of the relationship between these two abilities.

1.1. Evidence for interaction between speech processing and talker recognition

The speech signal contains not only linguistic information—*what is said*—but also information about the talker—*who says it*. Like vocal communication systems in other species (birds: Falls, 1982; primates: Cheney and Seyfarth, 1980), human speech contains acoustic cues that listeners use to recognize, for example, a talker's age, gender, race, emotional state, and their identity (Perrachione et al., 2010; Ramig and Ringel, 1983; Williams and Stevens, 1972).

A number of behavioral studies suggest that talker-specific acoustic cues are intertwined with speech recognition, with each affecting the other. First, talker variability affects speech processing. Listeners are better able to understand speech from familiar talkers than unfamiliar ones (Nygaard and Pisoni, 1998). Presenting words consistently from the same talker facilitates recognition of a previously presented word as familiar (Goldinger, 1996; see also Church and Schacter, 1994; Schacter and Church, 1992), and provides an extra cue for distinguishing phonologically similar words (Creel et al., 2008; Creel and Tumlin, 2011). Further, identification of speech sounds (Magnuson and Nusbaum, 2007; Nusbaum and Morin, 1992) and words (Mullennix et al. 1989; Nusbaum and Morin, 1992) in a sequence of elements is impaired when the talker changes from element to element. This suggests that variation in talker properties interferes with speech sound identification. One could interpret this to mean that listeners cannot selectively allocate attention to speech sound properties alone. Each of these lines of work suggests that talker information has effects on speech sound processing.

Additional studies suggest that the converse is also true—language knowledge affects talker recognition. Johnson et al. (2011) showed that 7-month-old infants detected a talker change in their native language (Dutch), but not in other languages. Sullivan and Schlichting (2000) looked at voice recognition among adults who had just begun studying a second language. They found an initial improvement in voice recognition after one semester of exposure, but multiple additional years of second-language study did not generate further improvement. However, this study did not include native-speaking controls, and the voice stimuli used were all intended to imitate the same voice. Additional studies have examined native listeners of varied abilities. Perrachione et al. (2011) found that individual dyslexic listeners' degree of impairment in phonological processing predicts their ability to recognize voices in their native language. Goggin et al. (1991) demonstrated that monolingual English speakers identified English-German bilinguals' voices better when those individuals spoke English than when they spoke German (see also Winters et al., 2008). On the other hand, Goggin et al. observed *no* significant difference in voice recognition abilities for English–Spanish bilinguals who were tested on English- vs. Spanish-speaking voices. They suggested that bilinguals might be equally able to recognize voices from either language since they have extensive knowledge of both. Finally, Perrachione et al. (2010) found that listeners even identified talkers better in their own *dialect* (General American vs. African-American English), suggesting that phonetic/phonological familiarity, alone or in conjunction with lexical familiarity, may underlie the language-familiarity effect.

Together, these studies suggest that differences in language processing are correlated with voice recognition. What remains unclear is *how much* (months? years?) and *what type* of language knowledge (lexical? phonological? phonetic?) is necessary for good talker recognition. Studies showing language-specific talker recognition benefits in infants (Johnson et al., 2011) and new language learners (Sullivan and Schlichting, 2000) suggest that relatively little exposure—months—is needed for language effects on talker recognition to emerge. However, if there is a strong relationship between speech-sound knowledge and talker recognition, then one would expect any differences in language exposure to affect talker recognition in that language.

1.2. Evidence for separation of talker and speech information

While the research just reviewed suggests intimate connections between speech processing and talker processing, other studies suggest that talker recognition and speech perception are computed by different cognitive processes and may be neurally dissociable. Much of the work suggesting dissociation comes from the neuroimaging and neuropsychology literatures. Neuroimaging results (e.g. Belin et al., 2004; Von Kriegstein et al., 2003; though see Perrachione et al., 2009) suggest that talker recognition is mediated by different brain structures (the right superior temporal sulcus) than those supporting speech-sound recognition in the left temporal lobe. (Van Lancker et al., 1989; see also Van Lancker et al., 1988) report that damage to the right temporal lobe is associated with difficulty recognizing famous voices, while other aspects of speech perception remain seemingly intact. Right-hemisphere damage leading to voice-recognition deficits is consistent with a functional dissociation between voice recognition and speech-sound processing. Interestingly, Van Lancker et al. also found that difficulty discriminating *unfamiliar* voices was associated with damage to either hemisphere, suggesting a more complex pattern.

Some behavioral evidence also suggests that language knowledge is not the sole factor in recognizing talkers. Specifically, listeners can identify time-reversed famous voices, indicating that they do not need identifiable verbal content to recognize at least some talkers (Van Lancker et al., 1985). Moreover, listeners with the same language background differ dramatically in their ability to recognize unfamiliar voices (Pollack et al., 1954) and in their judgments of talker

similarity (Kreiman et al., 1992). These studies suggest that voice recognition abilities may not be fully dependent on speech processing, and may vary across individuals.

1.3. Overview of the current study

We explore the nature of the language-talker interaction in monolingual and bilingual listeners. Prior studies of voice recognition in bilinguals suggest that bilinguals may perform similarly well at voice recognition in each language (Goggin et al., 1991) or that extensive exposure to a language is not necessary for talker recognition to become proficient (Johnson et al., 2011; Sullivan and Schlichting, 2000). However, these studies have not assessed *degrees* of bilingualism, or the age when each language was acquired. This is significant because developmental studies indicate that early exposure is crucial to developing sensitivity to a language's sound contrasts, and that this learning alters the neural processing of speech sounds (Kuhl and Rivera-Gaxiola, 2008; Werker and Tees, 1984). If talker recognition is strongly influenced by sound attributes learned early in life, then bilinguals who are early learners of a language may identify talkers as well as monolinguals do, while late learners may show worse performance. However, if the majority of talker recognition is carried by language-general sound properties, or sound properties that are easily learned by non-native speakers, then later learners who are reasonably proficient should show good talker recognition in a second language.

We conducted a study with Korean-English bilinguals and English speakers (who did not know Korean) to assess how linguistic and nonlinguistic factors contribute to talker recognition learning. First, we asked how bilinguals process talker information relative to monolinguals. Second, we assessed *gradient effects of bilingualism*: does degree of knowledge of each language predict talker recognition abilities in bilingual individuals? Finally, we explored whether other individual differences, such as music experience, contribute to voice recognition. We measured the ease of voice encoding in terms of *learning rate*, the speed with which individuals learn to accurately recognize unfamiliar voices. If listeners are better at encoding voices in their native language, then native-English-speaking participants should learn English-speaking voices faster than Korean-speaking voices, while native Korean-speaking L2 learners of English should learn Korean voices faster than English voices. Additionally, if there are gradient effects of bilingualism, then bilingual listeners with greater English knowledge should learn English voices faster than bilingual listeners with less English knowledge.

2. Methods

2.1. Participants

We tested 48 participants, 22 of whom were bilinguals who spoke Korean and English fluently. The remaining 26 participants had no background or experience with Korean and were native speakers of English (24 of whom had little background in a foreign language; 1 was bilingual in Spanish, 1 had early exposure to Thai but did not consider themselves bilingual). For convenience, we refer to native English speakers who do not know Korean in this study as "monolinguals," and Korean-English bilinguals as "bilinguals." All bilingual participants learned Korean as their first language, and learned English between 1 and 17 years of age (mean = 7.1 years). One Korean-English participant was natively bilingual, having learned both languages simultaneously. An additional bilingual participant did not reach criterion after 9 learning blocks in either language, indicating failure to understand the task, so this participant was excluded from all analyses. For bilingual participants, we defined "early" English learners via median split as those who reported learning English at age 5 or before, and "late" English learners as those reporting learning English after age 5. All participants were students at UC San Diego, where the language of instruction is English, meaning that all were fluent English speakers. They received course credit for participation.

2.2. Stimuli

We recorded 15 Korean sentences spoken by each of four female native Korean speakers, and 15 English sentences spoken by four female native American English speakers. English sentences were selected from the SPIN sentence set (Kalikow et al., 1977). All chosen English sentences were high predictability, and were statements rather than questions, e.g. "He caught the fish in his net." Direct translations of English sentences into Korean were often longer than the English versions in terms of syllable length. Thus, to better equate sentences in terms of the amount of spoken material, we asked a native Korean speaker to create similar Korean sentences that were simple, high predictability, and of similar syllabic length to the English sentences, e.g. "공책을 집에 놓고 왔다" ("Gongchek eul jibeh nohgo watda," "I left the notebook at home"). This approximately equated the amount of talker exposure in each language. Recordings were made in a sound-isolated recording booth. Each 16-bit, 44.1-kHz monaural recording was trimmed to begin at sentence onset and normalized to a mean intensity of 70 dB.

We divided the recorded stimuli into three sets. For each language, the first 5 sentences ($\times$4 talkers = 20 stimuli) were used for learning trials. Remaining sentences were reserved for the posttest phase, allowing us to assess whether listeners generalized talker recognition to new utterances, either as originally recorded (5 sentences), or with a shift in fundamental frequency (f0) to test generalization of learned voices to novel pitch ranges (5 sentences). Pitch-shifted stimuli were included to further explore a finding in European starlings (Bregman et al., 2012), in which recognition of their species-typical song was robust to changes in absolute pitch. Pitch shifting was performed by extracting the f0 contour using Praat software 5.1.20 (Boersma and Weenink, downloaded October 31, 2009), and then multiplying the pitch values by 1.2 and 1/1.2 (a 20% increase and a 17% decrease, equivalent log distances). We then resynthesized the sentences with the shifted f0 contours using overlap-add resynthesis (Moulines and Charpentier, 1990) as implemented in Praat.

2.3. Procedure

Each participant learned to recognize unfamiliar English and Korean voices in separate blocks. The order of stimulus language learned first was counterbalanced across participants. Learning and testing procedures were identical for each language, so we describe the learning and testing procedure for a single language. All stimuli were presented using the Matlab Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997) on the MacOS X operating system. Participants were individually tested in a sound isolated room, and audio was presented using Sennheiser HD 280 Pro headphones, which participants could adjust to a comfortable loudness level.

We designed this experiment to follow a set of training and testing procedures that have been widely applied in animal learning studies (e.g. Bregman et al., 2012; Gentner and Hulse, 1998; Hulse and Dorsky, 1979; Scharff et al., 1998) and human learning studies (e.g. Anderson, 1976; Gathercole and Baddeley, 1990; Reber, 1967). In these studies, participants are trained for a variable amount of time until they reach a criterion level of performance, and are then tested on novel exemplars to verify accurate generalization. Learning rate (time to reach criterion) is used as a measure of learning ability. Generalization, on the other hand, distinguishes whether the organism has acquired talker-related information during the learning process—conferring the ability to transfer talker recognition to novel stimuli—rather than memorizing the individual training stimuli as unrelated instances. This method allows us to investigate the relationship between language background and auditory processing and ability to encode novel voices.

Participants learned to associate each talker with one of four cartoon objects (Fig. 1a), which differed in both shape and color. We chose highly discriminable cartoon objects rather than faces to control for differences in face discriminability across participants of differing ethnicities (Bothwell et al., 1989). Prior to the first trial, participants read printed instructions (in English) telling them that the task was to learn to match voices to pictures. The experimenter emphasized and verbally confirmed that the particular phrases the talkers were saying were not relevant to the matching, only their voices. They then began the first block of learning trials.

Learning trials provided accuracy feedback. To initiate a trial, participants clicked a cross in the center of the screen. On each trial, two cartoon faces appeared to the left and right of the center cross. We used two visual stimuli rather than four to increase the speed of learning since previous studies in our lab indicated that increasing the number of alternatives increases learning time (e.g. Creel and Tumlin, 2011). Auditory stimulus presentation began simultaneously with picture display. During each learning trial, participants made a guess as to the "identity" of the talker by clicking one of the two cartoons with the computer mouse. A response could not be made until after the stimulus finished playing. After the participant made a selection, the correct cartoon remained on the screen, providing feedback, until they made a second confirmation click. Example learning and posttest trials are illustrated in Fig. 1b and c.

Learning took place in blocks of 60 trials, with order randomized within each block. Within a block, each talker and each sentence were heard equally often, and each pair of cartoons appeared together equally often. Learning continued until participants reached 85% correct—that is, they chose the target object on at least 51 of 60 trials in a single block (chance = 50%)—or reached a maximum of 9 learning blocks. After reaching criterion, participants immediately completed two posttest blocks without interruption or additional instruction. Individuals who completed 9 learning blocks without reaching criterion were not tested for that stimulus language.

Each posttest block contained 120 trials. During posttest blocks, no feedback was provided and the screen was blank after the participant's response. Each posttest block contained 60 trials that were identical to the learning trials (without feedback) to verify continued high performance on the learning stimuli, plus 60 trials containing 5 novel sentences produced by the 4 learned talkers, each of which was presented three times. The second posttest block contained 60 trials of the learning stimuli and 60 trials of novel sentences with modified f0. The language of talkers in the first learning set (Korean or English), the cartoon objects associated with each voice, and the positions of the two images on the screen on each trial were counterbalanced across participants.

2.4. Language background measures

In addition to completing the voice learning task, participants completed assessments to identify individual

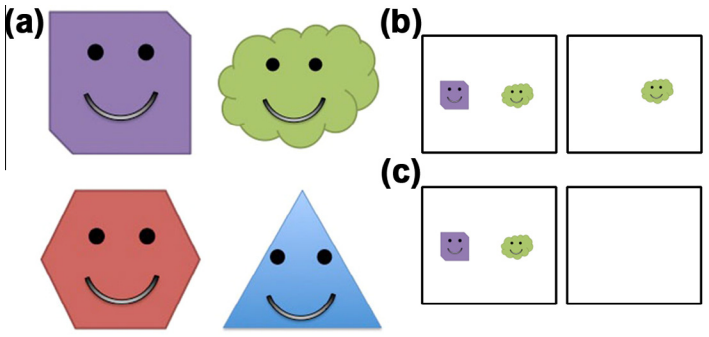

Fig. 1. (a) Four shape-based pseudo-faces that participants learned to associate with four voices. (b) Schematic of a single learning trial, with visual feedback provided after response. (c) Schematic of a single posttest trial, with no feedback after response.

differences in language background. All participants completed a language background questionnaire describing the age they were exposed to each of their language(s), as well as the amount of time spent speaking each language per week. For bilingual participants, we assessed the relative dominance of English and Korean in two ways. Each participant completed a Bilingual Dominance Scale (BDS) (Dunn and Fox Tree, 2009) and a picture naming task assessing lexical inventory (MiNT) in English and Korean (modified from Gollan et al., 2011 by removing words that are cognates in English and Korean). Their BDS was scored as described in Dunn and Fox Tree (2009), and recorded as Korean score minus English score. This score can range from −30 to +30. Thus, positive scores indicate Korean dominance and negative scores English dominance, with larger magnitudes indicating increasing dominance. According to the BDS, bilingual participants were balanced in their language dominance, with a mean difference between English and Korean of −0.22, and a range of −15 (English dominant) to 20 (Korean dominant). Note that no participants were near the ends of the range at −30 or +30. The MiNT (picture naming task) was scored as the number of correct uncued responses in Korean minus correct uncued responses in English, following Gollan et al., 2011 procedure. This score could theoretically range from −56 to +56 if a participant recognized all words in one language and none in the other. Again, a positive score indicates a higher vocabulary score in Korean. In our population of bilinguals the score ranged from −27 (English dominant) to 18 (Korean dominant) with a mean of −9.48. Thus, the MiNT suggests that our sample was slightly more English-dominant, at least in terms of vocabulary, than the BDS does.

Phonological working memory in English was estimated by measuring each participant's digit span. Digit span has been used as an index of phonological working memory in many experiments (Baddeley and Hitch, 1977). Participants heard a series of 16 audio recordings with a female voice reading random sequences of English digits at a rate of 1 digit per second. Two sequences for each length were presented, in order, from 2 to 9 digits. After each recording, participants verbally repeated the numbers they had heard. Scores were recorded as the number of sequences correctly repeated, with a maximum score of 16 (observed range = 7–15, mean = 10.7). Digit spans for bilinguals and monolinguals did not differ (Welch's $t(45.95) = 0.83$, $p = 0.41$).

2.4.1. Music background and auditory perception

To allow us to explore effects of music background, all participants completed a questionnaire describing their formal training and current performance activity. They also completed the pitch contour subtest from the Montreal Battery for the Evaluation of Amusia (MBEA) to measure differences in music perception ability (Peretz et al., 2003). During the MBEA test, participants heard 2 example melody-pair trials, followed by 31 test trials (15 same, 16 different). For each pair of melodies on a trial, they provided a same/different judgment. All melody pairs had the same melodic contour and there were no out-of-key notes, meaning that the "different" trials were fairly subtle changes. Each participant's score was recorded as the number of correct responses (observed range = 12–30, mean = 23.5). Finally, pitch discrimination thresholds (in Hz change discriminable with a 500 Hz standard) were captured using a web-based adaptive pitch threshold task (www.tonometric.com).

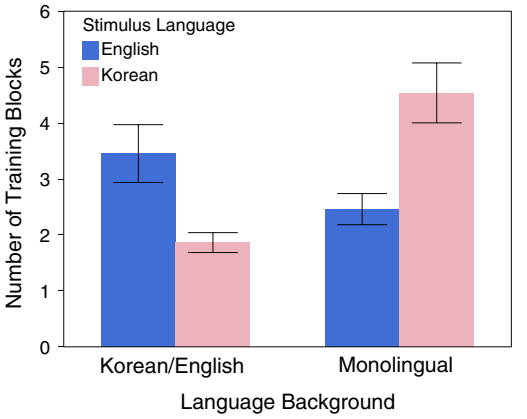

Fig. 2. Number of learning blocks to reach criterion of 85% on each stimulus language for Korean/English bilinguals ($n = 22$) and English-only speakers ($n = 26$). Each bar represents the mean number of learning blocks ± s.e.

3. Results

3.1. Basic language effects

3.1.1. Learning rate

Participants learned to recognize voices more quickly when they had knowledge of the language being spoken (Fig. 2). To measure talker learning ability, we measured the number of blocks required to reach a threshold of 85% correct in a single block.[1] A 3-way mixed model ANOVA on learning rate with factors of Participant Language Background (monolinguals, bilinguals; between-participants), Stimulus Language (English, Korean; within-participants) and Block Order (English first vs. Korean first; between-participants) revealed no significant main effects of participant language background ($F(1, 44) = 3.19$, $p = 0.08$),[2] stimulus language ($F(1, 44) = 0.44$, $p = 0.51$), or block order ($F(1, 44) = 1.09$, $p = 0.30$). However, there was a strong interaction between stimulus language and participant language background ($F(1, 44) = 24.02$, $p < 0.0001$). Bilingual participants learned the Korean talkers faster than the English talkers ($M = 1.9$ blocks vs. 3.5 blocks; paired t-test $t(21) = -3.03$, $p = 0.006$). Inversely, monolingual participants learned English voices faster than Korean voices ($M = 2.5$ blocks vs. 4.5 blocks; paired t-test $t(25) = 4.14$, $p = 0.0003$). No other interactions were statistically significant (all Fs < 0.08 and ps > 0.78).

[1] Eight participants did not reach 85% correct after 9 learning blocks (540 trials) in just one of the two talker-learning tasks, always in their non-dominant language. These participants were included in the analysis of learning rate, with a learning rate of 9 blocks. Note that this is an *under*estimate of the amount of time it would have taken them to learn, which could range from 10 blocks to infinity.

[2] Bilingual participants learned marginally faster overall.

Table 1
Correlation matrix for measures of bilingualism. Entries in bold are significantly correlated, $p < 0.01$.

	Learning rate – English (blocks)	Learning rate – Korean (blocks)	Age learned English	Bilingual dominance	MiNT	Digit span
Learning rate – English (blocks)	**1**	0.13	**0.62**	**0.55**	**0.53**	**−0.51**
Learning rate – Korean (blocks)		**1**	0.24	0.08	−0.14	−0.34
Age learned English			**1**	**0.89**	**0.72**	−0.52
Biling. dominance (Korean–English)				**1**	**0.76**	**−0.54**
MiNT (Korean–English)					**1**	−0.14
Digit span						**1**

3.1.2. First learning block accuracy

The above analysis did not allow us to assess item differences, because each item occurred an equal number of times for each participant. To assess item effects, and to confirm the above analysis based on a time point where all subjects had received identical exposure to stimuli, we performed an additional analysis on recognition accuracy during the first learning block using a 3-way ANOVA. Analyses were conducted both with participants as random factors ($F1$ and $t1$ statistics) and with items (sentences) as random factors ($F2$ and $t2$ statistics). The results of this analysis followed the same pattern described for learning rates: there was no main effect of participant language background ($F1(1, 44) = 0.34$, $p = 0.56$; $F2(1, 16) = 1.00$, $p = 0.33$), stimulus language ($F1(1, 44) = 0.47$, $p = 0.50$; $F2(1, 16) = 0.80$, $p = 0.38$) or block order ($F1(1, 44) = 1.03$, $p = 0.32$; $F2(1, 16) = 2.45$, $p = 0.14$). As in the learning rate data, there was a strong interaction between participant language background and stimulus language ($F1(1, 44) = 16.59$, $p = 0.0002$; $F2(1, 16) = 34.53$, $p < 0.0001$). Bilinguals showed higher accuracy for Korean talkers than English talkers ($t1(21) = 2.52$, $p = 0.02$, $t2(9) = 3.40$, $p = 0.01$), while monolinguals showed higher accuracy for English talkers than Korean talkers ($t1(25) = -2.94$, $p = 0.007$, $t2(9) = -4.05$, $p = 0.004$).

3.2. Language effects within bilinguals

3.2.1. Language measures

Next we considered measures of individual differences in language processing among bilinguals. First, we asked whether voice learning speed was captured by measures of English proficiency among bilinguals, such as the age English was learned. Table 1 shows the pair-wise correlations between these measures.

We next asked whether bilinguals showed different learning patterns as a function of age of acquisition. Among the group of bilinguals we studied, the MiNT (Gollan et al., 2011), BDS (Dunn and Fox Tree, 2009), and reported age of English acquisition were highly correlated with one another, so all measures of bilingual dominance behaved similarly to the age of acquisition effects (Table 1). We report correlations with age of acquisition due to theoretical interest. However, given the correlations between learning and the MINT and BDS, it is important to keep in mind that language dominance, rather than age of acquisition, may alternatively explain our results.

We computed an ANCOVA on the number of blocks to reach criterion for bilingual listeners, with age of English acquisition and stimulus language as predictors. This analysis revealed statistically significant main effects of acquisition age ($F(1, 43) = 13.76$, $p = 0.001$), and stimulus language ($F(1, 43) = 12.46$, $p = 0.001$). There was also a statistically significant interaction ($F(1, 43) = 8.03$, $p = 0.007$), suggesting a different relationship between age of English acquisition and learning time in each stimulus language. For Korean stimuli, the correlation between age of English acquisition and learning rate was small and not statistically significant ($r(20) = 0.24$, $p = 0.28$). However, for English stimuli, learning rate was positively correlated with age of English acquisition (Fig. 3a, $r(20) = 0.62$, $p = 0.002$). Note that this means that early English learners learned English voices as rapidly as monolinguals, and also learned Korean voices as rapidly as more Korean-dominant later English learners (Fig. 3b). This data pattern suggests that the later a listener acquires a language, the more difficulty they have encoding voices in that language.

One counter-explanation of this phenomenon is that, rather than age-of-acquisition-related phonetic knowledge affecting talker learning, it is instead listeners' difficulty encoding English phrases in working memory. That is, later learners of English are more taxed by retaining English phrases in working memory, impairing their ability to process talker differences. This seems unlikely, as our bilinguals were fairly English proficient (and according to the MiNT, more English-dominant than Korean-dominant on average; see Table 2). However, if this account were true, then one would expect that a measure of phonological working memory would explain the variability in learning rate. To test this, we computed a multiple regression with age of acquisition and English digit span as predictors of learning rate. This analysis showed that age of English acquisition remained a significant predictor of learning time for English voices even when digit span was controlled for ($F(1,21) = 5.83$, $p = 0.026$). Thus, phonological working memory differences cannot explain away the age of acquisition effect on voice learning.

3.3. Generalization posttest reflected extraction of voice properties

To verify that subjects were encoding voice properties and not simply memorizing the learning stimuli, we assessed their ability to generalize to novel sentences. We

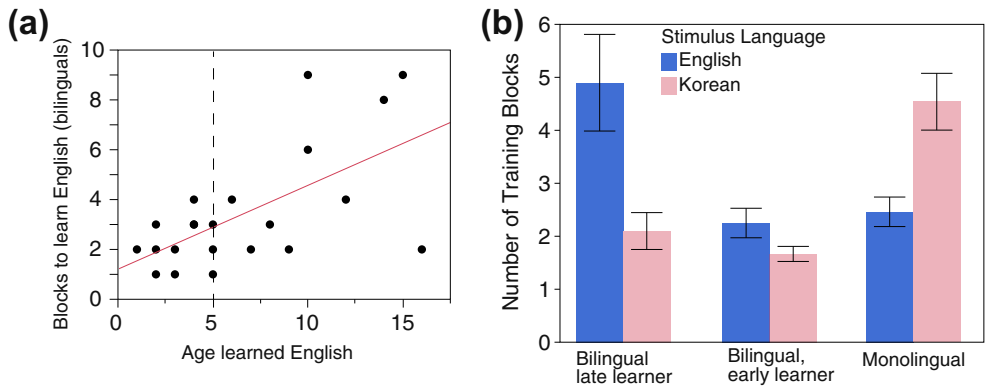

Fig. 3. (a) Each point represents a single bilingual participant's learning rate for English stimuli vs. the age they learned English. For all bilinguals, English was their second language. (b) Number of learning blocks to reach criterion of 85% on each stimulus language for bilinguals who learned English late (n = 10), early (n = 12) and monolingual speakers (n = 26). Each bar represents the mean number of learning blocks (60 trials/block) ± s.e.

Table 2
Bilinguals' performance, late vs. early learners. Recall that negative numbers on the BDS and MiNT imply English dominance, positive numbers imply Korean dominance.

	Early bilinguals	Late bilinguals
Learning rate – English (blocks)	2.25	4.90
Learning rate – Korean (blocks)	1.67	2.10
Mean age learned English	3.3	10.7
Biling. dominance (Korean–English)	−7.8	6.9
MiNT (Korean–English)	−15.6	−4.3
Digit span	11.5	9.6

also asked whether participants were better at generalizing to new utterances within their dominant language, even after achieving equally accurate recognition.

Four English monolinguals failed to reach 85% correct after 9 learning blocks for Korean stimuli, reaching a mean of 75.8% correct in the final learning block (range: 71.7–80.0%). However, they did reach the threshold for English talkers. Three bilingual participants did not reach criterion for English stimuli, reaching a mean of 71.1% correct (range: 61.7–78.3%; though they did reach criterion for Korean stimuli). These participants did not participate in posttest blocks on the language on which they did not reach criterion, and their results are not included in the generalization posttest results for either language.

For each of the 41 participants who reached criterion in both languages, we measured recognition performance during two posttest blocks. Recall that the first posttest block included the learning-phase stimuli, plus novel sentences. The second block included the learning-phase stimuli plus a different set of novel sentences whose pitch had been shifted.

Table 3 and Fig. 4 show mean performance for each of the three posttest stimulus types compared to performance achieved after learning. Participants continued to perform well on the learning-phase sentences (mean = 90.8% correct in first posttest block, 89.0% correct in second posttest block). More importantly, participants performed just as well on novel sentences in the first generalization block as they did on the learning-phase sentences. This verified that participants had extracted voice characteristics rather than overlearning the learning-phase stimuli.

We conducted an ANOVA on generalization accuracy. If listeners generalized poorly, there should be an overall effect of stimulus type, with less accurate recognition performance for novel sentences than for learning stimuli. If listeners generalized poorly only in their less-dominant language, there should be a 3-way interaction of participant language background, stimulus language, and stimulus type, with novel sentence performance less accurate only in their less dominant language.

Analysis revealed main effects of stimulus language ($F1(1, 39) = 16.98$, $p = 0.0002$; $F2(1, 24) = 7.07$, $p = 0.01$), and stimulus type (learning without feedback, novel, novel pitch shift; $F1(2, 78) = 164.81$, $p < 0.0001$; $F2(2, 24) = 61.31$, $p < 0.0001$). There was an interaction between participant language background and stimulus language ($F1(1, 39) = 90.31$, $p < 0.0001$; $F2(1, 24) = 86.27$, $p < 0.0001$), resulting from each participant showing modestly greater overall accuracy in their native language across both learning and test blocks. Because criterion accuracy was evaluated in a block of trials, individuals who learned more

Table 3
Mean performance for the learning stimuli and each of the stimulus types during testing.

	Learning blocks		Test blocks	
Group	Max learning perf.	Learning stimuli	Novel sentence	Novel sentence (shifted)
Bilingual – English voices	86.4% (6.5)	88.8% (4.7)	89.8% (8.4)	79.1% (6.4)
Bilingual – Korean voices	90.9% (3.3)	93.5% (4.4)	93.3% (7.6)	80.4% (3.7)
Monolingual – English voices	88.4% (3.9)	91.6% (6.7)	90.6% (7.3)	83.5% (6.0)
Monolingual –Korean voices	85.8% (5.1)	85.2% (5.4)	83.7% (8.0)	73.1% (5.2)

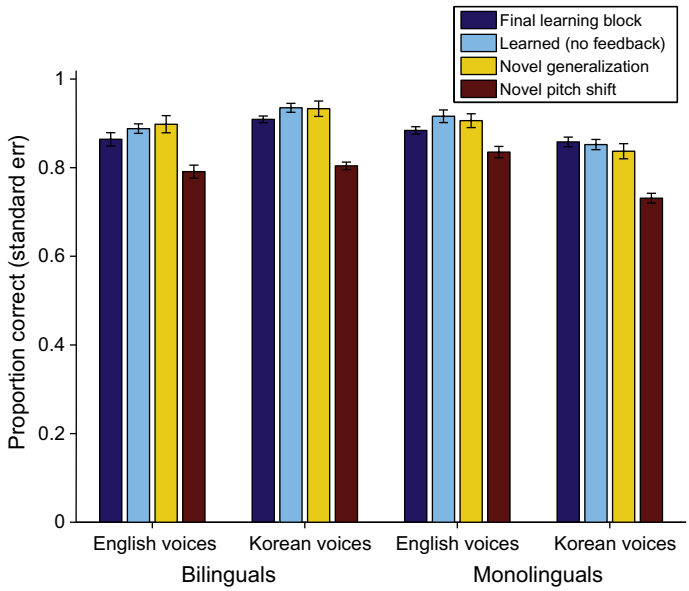

Fig. 4. Comparison of mean performance for learning and posttest stimuli. Vertical axis is mean proportion correct, and error bars represent standard error. Very little difference in performance was observed for learning and posttest stimuli in all stimulus groups, while pitch shifting disrupted performance for all listeners.

quickly achieved high accuracy across fewer blocks and hence had an overall higher accuracy rate during training. For instance, a native language participant might improve from 80% to 100% within a block, peaking at 100% accuracy; whereas a nonnative participant might improve more slowly, from 80% to 85% in one block and 85% to 90% in the next block, with a peak at 90%. This higher peak accuracy would then be expected to carry over to the generalization trials. There was a trend toward a main effect of language background ($F1(1, 39) = 3.02$, $p = 0.089$; $F2(1, 24) = 22.87$, $p < 0.0001$), with slightly higher accuracy for bilinguals overall. There was also an interaction between stimulus language and stimulus type, though the effect was significant by participants only, likely due to the small number of items per cell ($F1(2, 78) = 5.71$, $p = 0.0049$; $F2(2, 24) = 1.57$, $p = 0.23$). This resulted from a larger decrease in accuracy due to pitch shifting in the Korean stimuli than in the English stimuli (mean decrease in percent correct of 12.4% vs. 8.4%; $F1(1, 39) = 7.82$, $p = 0.0080$; $F2(1, 16) = 1.84$, $p = 0.18$),[3] but no decrease in accuracy from trained to novel stimuli ($F1(1, 39) = 0.66$, $p = 0.42$; $F2(1, 16) = 0.14$, $p = 0.71$). However, generalization accuracy did not differ as a function of language background (no interaction of participant language background, stimulus language, and stimulus type, $F1(2, 39) = 0.72$, $p = 0.49$; $F2(2, 24) = 0.10$, $p = 0.90$). That is, participants maintained terminal accuracy equally well regardless of language familiarity. No other effects approached significance. The observed generalization to novel sentences demonstrates that participants accurately learned to recognize the experimental voices, rather than rote acoustic representations of the learning tokens.

[3] There was no difference in performance between upward shifted and downward shifted stimuli (paired $t(40) = -1.27$, $p = 0.21$, $t2(9) = -0.71$, $p = 0.50$).

3.4. Individual differences in talker learning

3.4.1. A "talent" for talker learning?

One simple question is whether some listeners are better overall at learning to recognize voices, regardless of language: is there a "talent" for learning voices? This would be indicated by a correlation between participants' learning rate for Korean talkers with their learning rate for English talkers, without respect to language expertise. This correlation was near zero and not significant ($r(46) = 0.04$, $p = 0.77$), providing no support for a language-general talent for recognizing talkers.

3.4.2. Auditory processing differences

Previous research has suggested that musicians are better at phonological speech-processing tasks in second languages (Slevc and Miyake, 2006). Motivated by this finding, we performed an exploratory analysis investigating whether individual differences in auditory processing including pitch perception, music background, and music perception ability predict voice recognition ability. While some of these correlations approached statistical significance, when adjusted for multiple comparisons, only length of music training correlated significantly with average voice learning time ($r(46) = -0.40$, $p = 0.0045$). This relationship seemed to be driven by better learning in the *less*-familiar language ($r(46) = -0.40$, $p = 0.0055$), rather than the more-familiar language ($r(46) = -0.22$, $p = 0.13$), implying a potential advantage for musical experience in encoding highly-novel voices (or new linguistic information, which then yields better encoding of voices in that language). It may be that musicians are superior at auditory encoding generally. Another possibility is that they are leveraging an advantage in pitch processing—derived from music experience—to classify talkers by pitch characteristics,

rather than homing in on the talker characteristics that native speakers would use. This analysis should be treated as tentative, as participants were selected based on language background, not music background. Nonetheless, it suggests future directions for studies that more explicitly vary music experience while controlling for language background and cognitive differences.

4. Discussion

Our main goal was to characterize talker recognition as a function of native language background. We made two major discoveries. First, there was an overall effect of native language: native listeners performed better than second-language listeners (Korean–English bilinguals hearing English voices) or listeners unfamiliar with the language (English listeners hearing Korean voices). While previous studies showed that listeners learn to recognize unfamiliar voices better when they speak the language or dialect of the talker (Goggin et al., 1991; Perrachione et al., 2011; Winters et al., 2008), we demonstrate that listeners with lengthy exposure, but without native-speaker status, had increased difficulty recognizing voices in their second language. Our second major discovery was that within bilinguals, age of second-language acquisition impacted performance on voice learning. Within the studied group of bilinguals, earlier age of English acquisition predicted faster learning on the English (but not Korean) voices. That is, bilinguals who learned their second language (English) at an early age learned voices equally quickly in each language, while those who learned English later learned English voices slower than Korean voices.

A secondary goal was to explore the role that individual differences in auditory perception might play in voice recognition. While no clear differences emerged from auditory-acuity tasks (pitch-threshold, MBEA), length of music training appeared to confer a benefit to encoding voices in unfamiliar (or less familiar) languages.

To summarize, our study is the first demonstration of second-language and age of acquisition effects on talker recognition: not only is it easier to learn voices in a language you know, it is also easiest to learn voices in a language learned *early*. This pattern of data is consistent with shared or overlapping representations between language knowledge and talker knowledge, and inconsistent with accounts postulating strong neural/cognitive separation between these two faculties. It is also potentially consistent with neural/cognitive overlap between processing sound in *language* (to recognize speech and talkers) and processing sound in *music*, though the exploratory nature of our music-experience analyses suggests the need for further investigation.

4.1. The gradient role of language background in voice recognition

Previous studies in adults and infants have revealed that knowledge of a language improves ability to recognize voices in that language (Goggin et al., 1991; Johnson et al., 2011; Perrachione et al., 2011). We expanded upon this work by contrasting two language groups learning the same set of voices, and by looking at degrees of language dominance among bilinguals. Not only did we find a crossover interaction between listeners' native-language backgrounds and talkers' language, but we also found that early second-language acquisition facilitated talker learning without loss in performance on the first language. This acquisition effect—if viewed as such—is particularly interesting because it mimics acquisition of phonology: as age of acquisition increases, receptive and productive phonology are less native-like (Flege et al., 2006; Oh et al., 2011). Our study suggests that talker representations are also less native-like as age of acquisition increases. We consider why this might be the case below.

Our age-of-acquisition results suggest a more nuanced relationship between language familiarity and talker learning than previous studies. Recall that Goggin et al. (1991) observed no significant difference in how Spanish–English bilinguals responded to Spanish vs. English speaking voices, and that Sullivan and Schlichting (2000) observed rapid improvement in voice learning by inexperienced adult language learners, but little gain after multiple years of further study. Further, Johnson et al. (2011) found that even 7-month-olds, who have limited exposure to their language, detected native-language talker changes. These studies imply that exposure to a new language or phonological system permits talker recognition in that language. Our findings, though, suggest a *gradient* effect of language experience—listeners proficient enough in English to attend an English-speaking university have surprising difficulty learning to recognize English voices if they learned the language relatively late. This result is consistent with a large body of research suggesting that acquisition of phonological categories has a characteristic developmental trajectory, with early exposure important to developing robust representations of language-specific phonemes (e.g. Iverson et al., 2003; Werker and Tees, 1984).

We emphasize that caution is needed in interpreting the age-of-acquisition results. In our bilingual population, age of acquisition was highly correlated with amount of English exposure and likely to some extent with English proficiency: all of the early learners of English spent more time in the United States and had spoken English for longer than the late learners of Korean in school (although many continued to speak Korean at home). Indeed, age-of-acquisition is likely to be correlated with multiple independent physiological and cognitive factors that co-vary with age (Flege and Mackay, 2010).

Nonetheless, we argue that age of acquisition is the relevant variable here. Despite having relatively less exposure to Korean, those bilinguals who learned English early were still as good at learning Korean voices as their more Korean-dominant counterparts. That is, if language dominance alone predicted talker learning, then English-dominant bilinguals should show better learning for English voices. However, our earlier-learning bilinguals, who were English dominant by two measures (BDS score: −7.8, MiNT: −15.6; recall that negative scores suggest English dominance) appeared similarly adept at encoding voices in both languages. Our later-learning bilinguals looked more balanced (scores near 0; BDS score: 6.9, MiNT: −4.3), yet

were much faster to learn Korean voices than English voices. Thus, language dominance scores do not predict the exact data pattern of native-like voice learning by early learners, but less-adept voice learning by later learners. Only age of acquisition predicts this pattern. Whether our effects reflect age of acquisition or amount of exposure, though, they clearly demonstrate that long-term differences in language exposure result in large differences in talker encoding. We do recognize that though language dominance and proficiency are highly correlated, it is possible to be proficient in a language in which one is not dominant. Future work should investigate populations of bilinguals with different patterns of exposure to dissociate age of acquisition effects from language-dominance effects, and to examine the role of language dominance vs. proficiency in voice learning.

4.2. What knowledge do early learners have?

A question raised by the age-of-acquisition effect in talker recognition is: what acoustic–phonetic knowledge drives the effect? At least two accounts are plausible. One possibility is that talker variation is encoded with respect to speech sound representations (and vice versa). Thus, the better one's representations of speech sounds in a language, the better one can encode talker variation. This is consistent with evidence that adults, who encode words more accurately than young children (e.g. Ohde and Haley, 1997), also encode talkers more accurately (Creel and Jimenez, 2012; Mann et al., 1979).

A second account is that, just as languages make different distinctions among speech sounds, languages or cultures may make different distinctions among *talkers*. For example, speakers of Language A might have a higher characteristic fundamental frequency (f0) but vary in other characteristics, while speakers of Language B might vary in f0 but less so in other properties. In this case, Language B listeners might be biased to listen for talker differences in f0, putting them at a disadvantage in recognizing Language-A talkers. This is consistent with research (Johnson, 2005) showing cultural variation in gender differences in f0 and formant frequencies: language, and even accent within a language, seems to modulate voice properties that are often thought of as biologically determined. This is also consistent with our own finding that f0 and formant differences seemed more important for recognizing Korean voices than English voices (i.e., pitch-shifting the voices disrupted recognition more). This account implies that early learners, in addition to acquiring their native language's speech sounds, are also acquiring talker-varying characteristics unique to a particular culture—perhaps because, as we outlined in the Introduction, speech-sound and talker-identifying sound characteristics are tightly coupled. Future work is needed to dissociate these two accounts.

4.3. More exposure, but equivalent eventual recognition?

In contrast to differences in learning rate, we observed equally good generalization to new utterances regardless of language background. Does this mean that slower-learning listeners eventually formed native-like representations of the non-dominant-language talkers? This is possible, but we view it as unlikely. Listeners with less phonological knowledge could have been encoding talkers with a smaller set of cues than listeners with more phonological knowledge. This may have sufficed to distinguish within our small sets of talkers, but with a larger set—or with new "lure" talkers presented at posttest—less-expert listeners may have been at a greater disadvantage because they have fewer cues available to distinguish talkers.

4.4. Conclusions

We explored how language experience contributed to talker identification in two different languages. Listeners learned voices in their native language faster than voices in their second-language or in an unfamiliar language. Among bilinguals, earlier L2 acquisition predicted faster learning. Our work suggests a role for early language learning in talker identification. This is consistent with a tight linkage between language processing and talker identification, which presents an interesting puzzle to accounts of specialized neural mechanisms for speech recognition and talker identification.

Acknowledgments

MRB was supported by the Kavli Institute for Brain and Mind at UC San Diego, and SCC was supported by an NSF CAREER Award (BCS-1057080). We would like to acknowledge the help of undergraduate research assistants Shawn Cho and Hye Young Lee who were instrumental in developing Korean language stimuli and collecting data on participants' language backgrounds. We thank M. Hall for providing stimuli used to measure digit span.

References

Anderson, L. W. (1976). An empirical investigation of individual differences in time to learn. *Journal of Educational Psychology, 68*(2), 226–233. http://dx.doi.org/10.1037//0022-0663.68.2.226.

Baddeley, A. D., & Hitch, G. J. (1977). Working memory. *The Psychology of Learning and Motivation: Advances in Research and Theory*, 47.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129–135. http://dx.doi.org/10.1016/j.tics.2004.01.008.

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin, 15*(1), 19–25. http://dx.doi.org/10.1177/0146167289151002.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2012). Stimulus-dependent flexibility in non-human auditory pitch processing. *Cognition, 122*(1), 51–60. http://dx.doi.org/10.1016/j.cognition.2011.08.008.

Cheney, D., & Seyfarth, R. M. (1980). Vocal recognition in free-ranging vervet monkeys. *Animal Behaviour, 28*, 362–367.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 20*(3), 521–533.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition, 106*(2), 633–664. http://dx.doi.org/10.1016/j.cognition.2007.03.013.

Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology, 113*, 487–509.

Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language, 65*(3), 264–285. http://dx.doi.org/10.1016/j.jml.2011.06.005.

Dunn, A. L., & Fox Tree, J. E. (2009). A quick gradient bilingual dominance scale. *Bilingualism: Language and Cognition, 12*(03), 273. http://dx.doi.org/10.1017/S1366728909990113.

Falls, J. B. (1982). Individual recognition by sound in birds. In D. E. Kroodsma & H. E. Miller (Eds.), *Acoustic Communication in Birds* (pp. 237–278). New York: Academic Press.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America, 84*(1), 70–79.

Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics, 34*(2), 153–175. http://dx.doi.org/10.1016/j.wocn.2005.05.001.

Flege, J. E., & Mackay, I. R. A. (2011). "Age" effects on second language acquisition. In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010* (pp. 65–82). Bern, Switzerland: Peter Lang.

Gathercole, S., & Baddeley, A. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 4*.

Gentner, T. Q., & Hulse, S. H. (1998). Perceptual mechanisms for individual vocal recognition in European starlings, Sturnus vulgaris. *Animal Behaviour, 56*, 579–594.

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition, 19*(5), 448–458.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 22*(5), 1166–1183.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2011). Self-ratings of spoken language dominance: A multilingual naming test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition, 1–22*. http://dx.doi.org/10.1017/S1366728911000332.

Hulse, S. H., & Dorsky, N. P. (1979). Serial pattern learning by rats: Transfer of a formally defined stimulus relationship and the significance of nonreinforcement. *Animal Learning & Behavior, 7*(2), 211–220.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., & Diesch, E. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, 47–57. http://dx.doi.org/10.1016/S0.

Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 363–389). Oxford: Blackwell.

Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science, 14*, 1002–1011. http://dx.doi.org/10.1111/j.1467-7687.2011.01052.x.

Kalikow, D., Stevens, K. N., & Elliott, L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61*(5), 1337.

Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research, 35*(3), 512–520.

Kuhl, P. K., & Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annual Review of Neuroscience, 31*, 511–534. http://dx.doi.org/10.1146/annurev.neuro.30.051606.094321.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33*(2), 391–409. http://dx.doi.org/10.1037/0096-1523.33.2.391.

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology, 27*, 153–165.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* (Vol. 9(5–6), pp. 453–467). Elsevier. doi:10.1016/0167-6393(90)90021-Z.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*(1), 365–378.

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, speech production, and linguistic structure*. Washington: IOS Press.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*(3), 355–376.

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics, 39*(2), 156–157. http://dx.doi.org/10.1016/j.wocn.2011.01.002.

Ohde, R. N., & Haley, K. L. (1997). Stop-consonant and vowel perception in 3- and 4-year-old children. *The Journal of the Acoustical Society of America, 102*(6), 3711–3722. <http://www.ncbi.nlm.nih.gov/pubmed/9407663>.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: The montreal battery of evaluation of Amusia. *Annals of the New York Academy of Science, 999*, 58–75.

Perrachione, T. K., Chiao, J. Y., & Wong, P. C. M. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition* (Vol. 114 (1), pp. 42–55). Elsevier B.V.. doi:10.1016/j.cognition.2009.08.012.

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science (New York, N.Y.), 333*(6042), 595. http://dx.doi.org/10.1126/science.1207327.

Perrachione, T. K., Pierrehumbert, J. B., & Wong, P. C. M. (2009). Differential neural contributions to native- and foreign-language talker identification. *Journal of Experimental Psychology. Human Perception and Performance, 35*(6), 1950–1960. http://dx.doi.org/10.1037/a0015869.

Pollack, I., Pickett, J., & Sumby, W. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America, 26*(3), 403–406.

Ramig, L. A., & Ringel, R. L. (1983). Effects of physiological aging on selected acoustic characteristics of voice. *Journal of speech and hearing research, 26*(1), 22–30.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 18*(5), 915–930.

Scharff, C., Nottebohm, F., & Cynx, J. (1998). Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway. Journal of Neurobiology, 36(1), 81–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9658340.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second language proficiency: Does musical ability matter? *Psychological Science, 17*(8), 675–681.

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

Van Lancker, D., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex, 24*(2), 195–209. http://dx.doi.org/10.1016/S0010-9452(88)80029-7.

Van Lancker, D. R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology, 11*(5), 665–674. http://dx.doi.org/10.1080/01688638908400923.

Van Lancker, D. R., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters Part I: Recognition of backward voices. *Journal of Phonetics, 13*, 19–38.

Von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research, 17*(1), 48–55.

Werker, J. F., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63. http://dx.doi.org/10.1016/S0163-6383(84)80022-3.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America, 52*(4), 1238–1250.

Winters, S. J., Levi, S. V., & Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *The Journal of the Acoustical Society of America, 123*(6), 4524–4538. http://dx.doi.org/10.1121/1.2913046.