General Language Ability Predicts Talker Identification

Xin Xie (Xin.Xie@Uconn.Edu)
Department of Psychology, University of Connecticut
406 Babbidge Road, Unit 1020
Storrs, CT 06269 USA

Emily B. Myers (Emily.Myers@Uconn.Edu)
Department of Speech, Language, and Hearing Sciences, University of Connecticut
850 Bolton Road, Unit 1085
Storrs, CT 06269 USA

Abstract

Individuals can use both linguistic and non-linguistic features of the speech signal to identify talkers. For instance, listeners have more difficulty identifying talkers in unfamiliar languages compared to a native language (language familiarity effect), implying that language-specific knowledge aids talker identification. In the present study, the source of the language familiarity benefit on talker identification was investigated as listeners identified talkers in their native language as well as non-native languages. Experiment 1 was designed to explore the influence of L2 proficiency on talker identification across languages. Experiment 2 further investigated individual differences in L1 phonetic perception and their contribution to talker identification by comparing English listeners' performance across different language conditions that varied in the availability of linguistic cues. Results imply that familiarity with a specific language (L1 or L2) did not explain individual variation in language familiarity effect. Rather, in addition to the native language benefit, talker identification may be supported by general sensitivity to sound structures in language, modulated by the availability of higher-level linguistic information.

Keywords: talker identification; language proficiency; speech perception; bilingualism; individual differences

Introduction

Listeners differ in their ability to identify or recognize human voices, but the source of these underlying individual differences is poorly understood. Recently, research has revealed an important role of linguistic knowledge in voice perception. First, there is strong evidence that speaker-related acoustic-phonetic properties (e.g., formants in vowels, voice onset time of consonants) can inform listeners of talker gender or identity (e.g., Remez et al., 1997). Second, the language background of listeners qualitatively affects talker identification performance, contributing to the *language familiarity effect* (LFE) in talker identification. This effect establishes that listeners have more difficulty identifying talkers in unfamiliar languages compared to their native language (Perrachione et al., 2009; Goggin et al., 1991). This finding raises two important questions about talker identification: first, *how much* prior linguistic knowledge is required to promote the LFE? Second, what *type* of linguistic knowledge drives the LFE?

With respect to the first question, a number of cross-linguistic studies investigated the influence of second language (L2) knowledge on talker identification by comparing participants who had *qualitatively* different language experience: naïve listeners who had no familiarity at all, listeners with some knowledge of the target language and native listeners. However, conflicting results were obtained: on the one hand, Spanish- and Chinese-native speakers who spoke German as a L2 had significantly poorer performance than native-German speakers in identifying speakers in German (Köster & Schiller, 1997), suggesting a decisive role of native language; on the other hand, native-English speakers who learned German as a L2 reached native-like performance in the same task (Köster, Schiller, & Künzel, 1995). It is possible that the similarity between languages contribute to the larger transfer of language knowledge. Notably, English is typologically closer to German than are either Spanish or Mandarin; English and German have more overlap in phonology, among other linguistic structures. Alternatively, the discrepancy between the studies might arise from individual differences in L2 proficiency, which may itself arise from differences in age of acquisition (AoA). Bregman & Creel (2014) compared monolingual English speakers and English-Korean bilinguals in their ability to learn English voices. In this study, the AoA of L2 predicted the speed of voice learning in the L2. While late English learners (L1 Korean) were significantly slower in learning English voices, early English learners approached native-like performance on this measure, suggesting a *gradient* effect of language background on talker learning.

In Experiment 1, we tested a more specific hypothesis that this gradient effect is driven by individual L2 proficiency. To this end, we tested a homogeneous group of Mandarin speakers who started English acquisition around the same age[1] but achieved different L2 proficiency. In particular, we examined talker identification in participants' native language and two non-native languages varying in their similarity to the native phonology, as an attempt to eliminate the confounding effects of language similarity in previous studies (e.g., Köster & Schiller, 1997; Schiller et

[1] All of them were late L2 learners using Bregman & Creel's criteria (AoA > 5).

al., 1995). If high L2 proficiency of late bilinguals predicts more native-like voice perception in the L2 (but not in other languages), then it provides a more *quantitative* explanation of the LFE. Such a finding will complement work of Bregman & Creel (2014) by suggesting a more plastic functional integration between speech processing and talker identity perception that continues to be influential after the critical period.

While language knowledge apparently facilitates talker identification, much less is known with respect to the second question: which *type* of language knowledge (other than broad terms such as "linguistic knowledge/proficiency" or "language-dependent indexical cues"; e.g., Winters et al., 2008) has a direct effect on LFE in talker identification. Individuals exhibit vast differences in their mastery of languages (native or non-native) on various levels of processing, from acoustic-phonetic level to the lexical level to the sentence level. Recent studies have implied that knowledge of the sound structure of one's native language leads to the LFE (Perrachione et al., 2011; Fleming et al., 2014). Perrachione et al. (2011) showed that in dyslexic adults, difficulty in talker identification in their native language was correlated with measures of phonological impairment. Other research has implied a link between subtle phonetic knowledge and talker perception (e.g., Bregman & Creel, 2014). It remains unclear whether individual differences in phonetic perception are linked to those in voice learning among typical adults and can uniquely predict voice learning performance in a particular language. The *second* goal of the current study is to explicitly measure individuals' ability to encode subtle phonetic detail and link it to performance in talker identification. If subtle phonetic knowledge is critical in voice learning, then even among native listeners, performance in native phonetic perception should predict performance in native talker identification and account for the native language benefit.

We adapted the talker identification task from Perrachione et al. (2009) and added different linguistic measures in a cross-language context. A sentence-in-noise transcription task (Experiment 1) and phonetic categorization tasks (Experiment 2) were used to assess listeners' familiarity with English. These tasks helped to verify whether any observed language familiarity effect could be explained by listeners' familiarity with a specific language, by comparing listeners' talker identification performance across language conditions. Experiment 1 focused on L2 proficiency of late L2 learners and Experiment 2 on phonetic encoding ability of native L1 speakers. Together, the experiments were designed to further elucidate the mechanism underlying the observed relation between speech processing and voice perception.

Experiment 1

In Experiment 1, we examined the relationship between listeners' L2 proficiency and talker identification performance, by assessing English and Mandarin listeners' ability to identify talkers across three language conditions: Mandarin, Spanish and English. The selection of these languages allowed control over potential effect originating from language similarities (cf. Köster et al., 1995 and Köster & Schiller, 1997). All Mandarin listeners were late L2 learners of English. A sentence-in-noise transcription task provided a direct measure of Mandarin listeners' L2 proficiency in English. Given that musical ability enhances talker identification (Xie & Myers, 2015) and L2 phonological processing (Slevc & Miyake, 2006), we included musical experience as a covariate in the analysis.

Methods[2]

Participants Two groups of listeners (44 native-English listeners and 39 native-Mandarin listeners who speak English as L2) participated in the study. All English listeners were monolingual speakers who were naïve to Mandarin, although some listeners learned basic Spanish in school. All Mandarin listeners were naïve to Spanish. Mandarin listeners were late bilinguals who learned English in classroom setting in Mainland China; their average age of acquisition was 10.33 (SD = 2.73) years old, and their age of arrival in the U.S. was 22.00 (SD = 4.23) years old. English and Mandarin groups did not differ in terms of years of musical training received prior to test time (English: M = 2.89, SD = 4.52; Mandarin: M = 2.44, SD = 4.08, $t(81) = .474$, $p = .64$). All participants were Uconn students with no known hearing or visual disorders.

Materials All recordings were made in a sound-proof room and digitally sampled at 22.05 kHz. Stimuli for the sentence transcription task consisted of three sentence lists adapted from the Revised Bamford-Kowal-Bench (BKB-R) sentence lists (Bench & Bamford, 1979). Each list consisted of 16 simple English declarative sentences with 3 or 4 keywords each, resulting in a total of 50 keywords per list. A male native speaker of American English recorded the sentence set. The recordings were then embedded in white noise at a +5 dB signal-to-noise ratio and normalized for RMS amplitude to 70 dB SPL. Stimuli for the talker identification task consisted of recordings of 10 sentences in each of the three languages: Mandarin, Spanish and English, recorded by five male native speakers of that language. Five sentences in each language were arbitrarily designated as training sentences, and the remaining five as test sentences.

Procedure Participants were seated in a sound-attenuated booth in front of a computer monitor and listened to stimuli delivered by headphones. Each participant completed the talker identification task followed by the sentence transcription task. The talker identification task was blocked

[2] A subset of the data (36 English and 25 Mandarin listeners) was reported in Xie & Myers (2015). An additional group of participants were recruited and combined with previous subjects for analysis in the current study. The talker identification task was reported in detail in Xie & Myers (2015). The sentence transcription task was analyzed and reported here for the first time.

by language condition (Mandarin, Spanish and English), with the order of language condition counterbalanced across participants. Each condition consisted of a familiarization phase, a practice phase and a generalization phase. During the familiarization phase, listeners heard all five training sentences, each spoken by five speakers twice. Talker identity information was provided with a generic label (e.g., Talker 1). During the practice phase, listeners identified the talkers speaking five training sentences with feedback. During the generalization phase, listeners identified talkers when five novel sentences were spoken. No feedback was provided during this phase.

During the sentence transcription task, participants transcribed each sentence in standard English orthography. All three BKB-R lists were presented with the order of lists counter-balanced across participants. The order of sentence presentation was randomized within each list; each sentence was played only once. Upon completion of listening tasks, participants filled out a survey on their language and musical background.

Results and Discussion

Talker Identification Task The dependent measure was the percentage of correct identifications of talkers during the generalization phase (Table 1). We compared the groups' accuracies using 2 between-subject (listener group: English or Mandarin) × 3 within-subject (language condition: Mandarin, Spanish, or English voices) ANCOVA with years of musical training as a covariate. Musical experience had a significant main effect ($F(1,80) = 9.823$, $p = .002$) but did not interact with other factors. Of our primary interest, the listener group × language condition interaction ($F(2,160) = 67.487$, $p < .001$) was significant, reflecting the language familiarity effect. Pairwise t-tests revealed that for both listener groups, the effect was due to significantly better performance identifying talkers speaking their native language versus other languages (*p*s < .001); no difference was observed between the two unfamiliar languages (*p*s > .10). Thus, the result indicated that with some knowledge of English, Mandarin participants did not perform better than they did with Spanish, an entirely unfamiliar language.

Table 1: Mean accuracy of talker identification results as a function of listeners' native language (listener group) and talker language condition.
Standard deviations are listed in parentheses.

Language condition	Listener group	
	English listeners	Mandarin listeners
Mandarin	0.49 (0.12)	0.74 (0.14)
Spanish	0.52 (0.15)	0.53 (0.12)
English	0.70 (0.18)	0.55 (0.13)

Sentence Transcription Task The sentence transcription score (the percentage of keywords correctly recognized) averaged across three lists was calculated. English listeners' performance exhibited a ceiling effect (all above .99). As late bilinguals, Mandarin listeners' accuracy ranged from .49 to .97 (M = .85, SD = .10). The score was used as a measure of participants' L2 proficiency in English.

Focusing on Mandarin listeners, we asked whether late bilinguals' talker identification differed as a function of their L2 proficiency; and whether the language influence predicted their talker identification performance in each language condition (Table 2). Surprisingly, sentence transcription scores significantly correlated with talker identification accuracy in all language conditions, not only in the English talker condition. We also computed the magnitude of the language familiarity effect (LFE), i.e., difference in performance on Mandarin versus on English (LFE1) or Spanish (LFE2). The L2 proficiency measure was not correlated with the size of either LFE (LFE1: $r = .11$, $p = 0.52$; LFE2: $r = -.05$, $p = 0.76$). Furthermore, talker identification accuracies across different language conditions were highly correlated[3]. These correlations became non-significant after we further controlled for listeners' L2 proficiency (all *p*s > .10). That is to say, L2 proficiency as measured by the sentence transcription score shared common variance with the variation in talker identification across language conditions.

Table 2: The matrix of partial correlations (controlling for years of musical training) between Mandarin listeners' L2 proficiency and talker identification performance.
**p* < .05, ***p* < .01.

Language condition	Mandarin	Spanish	English
Mandarin	1		
Spanish	0.35*	1	
English	0.36*	0.36*	1
L2 proficiency	0.45**	0.59**	0.34*

Thus, although a LFE was robustly observed in both listener groups, the results from Mandarin listeners were contrary to our prediction in two ways. First, Mandarin listeners' individual L2 proficiency did not predict the magnitude of LFE. On the group level, there was no difference observed between the two non-native language conditions (English and Spanish) either. Second, the evidence pointed to a *general* talker learning ability related to second language proficiency. Regarding the first result, it is possible that some minimum proficiency threshold with a language (in this case, English) must be achieved before any LFE can be observed. In our study, the late bilinguals did not reach the native range in terms of sentence transcription accuracy (see also Bregman & Creel, 2014). Future research may test early bilinguals varying in their L2 proficiency to validate or falsify this "threshold" hypothesis.

[3] The same pattern was found among English participants. Their talker identification accuracy in English correlated with performance in Mandarin ($r = 0.62$, $p < .001$) and Spanish ($r = 0.31$, $p < .05$), after controlling for individual musical experience.

Regarding the second result, it may be that some individuals have superior cognitive or auditory ability such that they are the better performers across the listening tasks for *nonlinguistic* reasons. Alternatively, correlations among voice perception across languages may have a *linguistic* root. Recent research has given special attention to the knowledge of speech sound structures (Perrachione et al., 2011; Fleming et al., 2014). However, the actual source of LFE is underspecified and may arise from acoustic-phonetic, phonological, or lexical levels of processing. For example, abstract phonological knowledge, as tested in Perrachione et al. (2011), could be at play. Alternatively, it is equally possible that listeners' ability to track fine-grained phonetic detail and link it with talker-specific variation is associated with LFE (see Theodore & Miller, 2010). We intended to provide a more rigorous test of cognitive or perceptual factors subserving the LFE in Experiment 2.

Experiment 2

In Experiment 2, we intended to refine previous hypothesis that familiarity with one's native phonology underlies the LFE (e.g., Fleming et al. 2014), by testing individual differences in acoustic-phonetic analysis of speech and their relation to talker identification in monolingual English listeners. Mastery of subtle phonetic knowledge was assessed by two phonetic categorization tasks.

Studies have shown that higher-level linguistic cues affect acoustic-phonetic processing. However, the role of lexical information in talker perception is much less investigated. Fleming et al. (2014) recently found evidence of LFE even when sentences were time-reversed (lexical information was not available) and concluded that comprehension is not necessary in eliciting the LFE. However, results of Goggin et al. (1991) indicated that talker identification is better for comprehensible sentences than for incomprehensible sentences. These findings together suggest phonetic or phonological level processing of speech may be at the root of LFE in talker identification, but top-down lexical information may further facilitate the use of sound patterns. To test this hypothesis, we manipulated the availability of higher-level linguistic cues across two native language conditions in Experiment 2. In one condition, lexical-semantic cues were eliminated by rearranging words or mixing syllables to create nonsense, "Jabberwocky" sentences. The comparison between this Jabberwocky English (JE) condition and normal English condition (compared to an unfamiliar language condition) helped us to examine the extent to which voice perception relies on the presence of meaningful linguistic content, holding phonology constant. If individual sensitivity to acoustic-phonetic detail is underlying the LFE, we predict correlations between the phonetic measures and talker identification in both native language conditions (normal English and Jabberwocky English). If lexical information provides an additional benefit (by facilitating the use of acoustic-phonetic detail), we predict performance in the normal condition should be better than that in the JE condition. Lastly, as Experiment 1, we intended to control individual variability in pitch processing ability as it relates to individual differences in talker perception (Xie & Myers, 2015). In Experiment 2, instead of controlling for musical experience (an indirect measure of pitch processing ability), we used pitch tasks to provide a more direct measure of pitch processing ability, in order to parse out any influence arising from this *nonlinguistic* auditory processing ability.

Methods

Participants 63 monolingual English participants from the Uconn community were included in the following analysis[4].

Materials The stimuli for the talker identification task in the English and Mandarin conditions were the same as in Experiment 1. Stimuli in the Jabberwocky English (JE) condition consisted of recordings of phonologically scrambled versions of the original 10 sentences in the English condition. We rearranged syllables to make nonsense sentences such as 'More in a tri- campic lingting turress angra the forture' (mixed syllables from 'Try angling the camera for a more interesting picture'). Five[5] native American-English speakers (all males) recorded the stimuli. A local and a global pitch perception task were used to assess listeners' sensitivity to changes in pitch height and pitch contours, respectively. In each trial, listeners reported whether two pure tone sequences were same or not, based on the criteria of the specific task. Stimuli used in the pitch perception tasks were reported in Xie & Myers (2015) in detail. Phonetic perception tasks consisted of a vowel categorization task and a consonant categorization task. For the vowel categorization task, tokens of a female American-English speaker producing the vowel /ɛ/ and /æ/ were recorded. Resynthesis of two natural productions spoken by this speaker provided the endpoints of the continuum and seven equally-spaced synthesized vowels along the /ɛ/-/æ/ continuum were created using PRAAT, following Sebastián-Gallés & Baus (2005). For the consonant categorization task, tokens of a male American-English speaker producing stop consonant /da/ and /ta/ were recorded. Nine synthesized syllables along the /da/-/ta/ continuum were created by varying the voice onset time of the consonant from 0ms to 80ms, in 10ms steps.

[4] Participants had to achieve above-chance performance in the English condition in talker identification and in both pitch tasks. For the phonetic perception task, response rate had to exceed 75% across all continuum steps and accuracy of responses must be below 25% at one end of the continuum, and above 75% at the other end. A total of 13 participants were excluded. The pitch task and talker identification task (Mandarin and English) were reported in Xie & Myers (2015). We now report the phonetic tasks and talker identification in the JE condition for the first time.

[5] They were different speakers from those in the normal English condition. All speakers practiced reading the sentences until they could read the sentences fluently as if they were real English sentences.

Procedure Participants first completed the talker identification task, followed by the pitch perception tasks and the phonetic categorization tasks, with the order of the latter two types of tasks counterbalanced across participants.

In the vowel categorization task, participants were told to press one button when hearing a vowel like the one in '*bed*', and another for the vowel in '*bad*'. The task began with seven practice trials, one for each stimulus, presented in randomized order. The experimental session included 12 blocks, with the continuum steps randomized within each block. A similar procedure was used in the consonant categorization task.

Results and Discussion

Table 3 lists the descriptive statistics for all the measures. Accuracy in the English condition was significantly higher than that in the JE condition ($t(62) = 4.33$, $p < .001$), and both were significantly higher than that in the Mandarin condition (*p*s < .001), suggesting that the magnitude of LFE was larger when lexical information was present in the native language.

For pitch perception tasks, the average sensitivity score (log-transformed d′) across the two pitch tasks was computed to reflect listeners' sensitivity to pitch (see Xie & Myers, 2015 for details). For the vowel categorization task, to obtain a performance score for each individual, we calculated the vowel categorization score by subtracting the average log odds of steps 2 and 3 from the average log odds of steps 5 and 6 (see Sebastián-Gallés & Baus, 2005). We interpret the score as reflecting the slope of the categorization curve. The higher the score, the better separation between /ɛ/ and /æ/. We similarly computed the consonant categorization score by subtracting the average log odds of steps 3 and 4 (VOT = 20ms and 30ms) from the average log odds of steps 6 and 7 (VOT = 50ms and 60ms). All measures were normally distributed as assessed by Kolmogorov–Smirnov Z tests (*p*s > .05).

Table 3: Descriptive statistics of all measures.

	Measure	Mean	SD	Min	Max
Talker identification	Mandarin	0.48	0.15	0.17	0.85
	JE	0.62	0.19	0.15	0.97
	English	0.71	0.13	0.41	0.93
Auditory	Pitch	-0.01	0.22	-0.49	0.42
Phonetic	Vowel	6.30	2.23	0.10	9.19
	Consonant	7.49	1.77	1.45	9.19

The relationship between measures of individual differences in voice perception, pitch processing and native phonetic processing was examined by pair-wise correlations (Table 4). As reported in Xie & Myers (2015), there was a significant positive correlation between pitch sensitivity and talker identification accuracy that was specific to the unfamiliar language (Mandarin). The correlation was marginally significant in the Jabberwocky English condition ($p = .07$) and non-significant in the English condition ($p = .32$). If variation in talker identification performance arose from auditory-generic processes—if individuals with better talker identification simply had better pitch processing skills and exploited those skills to the same degree regardless of language conditions, then talker identification scores in general should correlate positively with pitch sensitivity. This was not the case. In addition, pitch sensitivity was not correlated with any of the phonetic categorization results. Thus, it was not the case that some participants were simply "more adaptable" than others.

Table 4: The correlation matrix for measures on talker identification, pitch sensitivity and phonetic categorization. *$p < .05$, **$p < .01$.

	M	JE	E	Pitch	V	C
Mandarin	**1**					
JE	0.52**	**1**				
English	0.39**	0.49**	**1**			
Pitch	0.32**	0.23	0.13	**1**		
Vowel	0.52**	0.41**	0.31**	0.16	**1**	
Consonant	0.38**	0.26*	0.29*	0.18	0.5**	**1**

The second key finding was that both vowel and consonant categorization scores (independent of pitch processing skills) positively correlated with talker identification accuracy across all language conditions (Table 4). In addition, none of the pitch or phonetic measures predicted the magnitude of language familiarity effects (English vs. Mandarin or JE vs. Mandarin). Thus, individual variation in native phonetic perception, as measured by the categorization scores (vowels and consonants), did not uniquely explain the language familiarity effect *per se*, but instead seems to be related to talker identification ability across language conditions. This finding was surprising, but it mirrored our finding in Experiment 1. These results are theoretically informative because they suggest that *general* speech processing abilities, rather than phonological knowledge of a s*pecific* language may aid in talker identification.

General Discussion

Previous studies established that native language experience (Perrachione et al., 2009) and early bilingualism (Bregman & Creel, 2014) enhance voice learning. We designed this study to test the hypothesis that such language familiarity effect is driven by individuals' knowledge of a *specific* language, and in particular, by individuals' perception of subtle phonetic detail in that language. We examined whether language-related abilities in L1 (among monolinguals) and/or L2 (among late bilinguals) predict talker identification accuracy in that language. We report two key findings.

The first finding was that talker identification is language-dependent, but in a less specific way than previously hypothesized. Although the native language familiarity effect was very robust, contrary to our prediction, the linguistic knowledge of a *specific* language, either nonnative (Experiment 1) or native (Experiment 2), did not explain how well a listener can identify speakers of that particular language, compared to the baseline talker identification accuracy in an entirely unfamiliar language. Instead, performance assessing language abilities (either native or nonnative) correlated with talker identification across all language conditions. Meanwhile, the language measures were independent of individual variation in nonlinguistic pitch perception abilities, suggesting that general auditory processing skills cannot explain away the close relation between performance in the language-related listening tasks and talker identification tasks. The results suggested that a *language-general* aptitude may exert a major influence on talker identification, regardless of the language being spoken. Note that we cannot exclude the possibility that correlations in performance on talker identification and other linguistic/non-linguistic tasks were mediated by other untested *nonlingusitic* cognitive factors. Future research should aim to disentangle these possibilities.

The second finding is that listeners readily exploit lexical information in the native speech in talker identification, as English listeners were more accurate identifying talkers in normal English than Jabberwocky English, and more accurate in Jabberwocky English than Mandarin. Taken together with the first finding, we suggest that the *language-general* aptitude may reflect processing ability of acoustic-phonetic detail in two ways. First, unlike the abstract phonological knowledge of a *specific* language, it could potentially function in a language-*independent* way. Second, top-down cues from the lexicon may strengthen acoustic-phonetic cues associated with the talker, and therefore facilitate talker identification in one's native language.

Conclusion

Our findings expand upon previous studies to demonstrate that a language-related capacity is underlying the observed individual variation in talker identification skills. This capacity is not specific to any particular language system and is used in talker identification across language conditions. Moreover, its functioning may be facilitated by lexical access. We thus suggest that sensitivity to acoustic-phonetic detail is a good candidate for this language-general capacity. Future studies should further investigate this phenomenon to provide a satisfactory theoretical framework of talker identification. Ultimately, the disparate avenues of research in talker identification and speech perception need to be more united to elucidate the cognitive mechanism of how humans recognize one another's voice.

Acknowledgments

We are grateful to Carol Fowler for her advice throughout the course of this project. We acknowledge support from "Fund for Innovation Fund in Science Education" at University of Connecticut. This work was supported by NIH R03 DC009495 (E. Myers, PI) and by NIH P30 DC010751 (D. Lillo-Martin, PI). The content is the responsibility of the authors and does not necessarily represent official views of the NIH or NIDCD.

References

Bench, J., & Bamford, J. (Eds.). (1979). *Speech-hearing tests and the spoken language of hearing-impaired children*. London: Academic Press.

Bregman, M. R., and Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition*, *130*, 85-95.

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

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

Köster, O., & Schiller, N. O. (1997). Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics*, *4*, 18–28.

Köster, O., Schiller, N. O., & Künzel, H. J. (1995). The influence of native-language background on speaker recognition. *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 4, pp. 306-309).

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. (2011). Human voice recognition depends on language ability. *Science*, *333*, 595-595.

Perrachione, T. K., Pierrehumbert, J. B., & Wong, P. (2009). Differential neural contributions to native-and foreign-language talker identification. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1950.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 651–666.

Sebastián-Gallés, N., & Baus, C. (2005). On the relationship between perception and production in L2 categories. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 279–292). New York: Erlbaum.

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

Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *Journal of the Acoustical Society of America*, *128*, 2090-2099.

Winters, S. J., Levi, S. V., & Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *Journal of the Acoustical Society of America*, *123*, 4524.

Xie, X., & Myers, E. (2015). The impact of musical training and tone language experience on talker identification. *The Journal of the Acoustical Society of America*, *137*, 419-432.