

Cite as:

Perrachione, T.K. (2018). "Recognizing speakers across languages," in S. Frühholz & P. Belin (Eds.), *The Oxford Handbook of Voice Perception*, Oxford: Oxford University Press.

Title:

Recognizing speakers across languages

Author:

Tyler K. Perrachione
Department of Speech, Language, and Hearing Sciences
College of Health and Rehabilitation Sciences: Sargent College
Boston University
635 Commonwealth Ave.
Boston, MA 02215

to appear in:

The Oxford Handbook of Voice Perception
Sascha Frühholz and Pascal Belin, Editors.

Contact:

Email: tkp@bu.edu
Phone: 617.358.7410

Keywords:

talker identification; voice recognition; language familiarity; native language; foreign language; speech perception; expertise; linguistic competence; phonetic variability

Abstract

Listeners identify voices more accurately in their native language than an unknown, foreign language, in a phenomenon known as the *language-familiarity effect* in talker identification. This effect has been reliably observed for a wide range of different language pairings and using a variety of different methodologies, including voice line-ups, talker identification training, and talker discrimination. What do listeners know about their native language that helps them recognize voices more accurately? Do listeners gain access to this knowledge when they learn a second language? Is linguistic competence necessary, or can mere exposure to a foreign language help listeners identify voices more accurately? In this chapter, I review the more than three decades of research on the language-familiarity effect in talker identification with an emphasis on how attention to this phenomenon can inform not only better psychological and neural models of memory for voices, but also better models of speech processing.

Introduction

Some talkers are more or less memorable because of distinctive acoustic properties of their voice: a particularly low pitch, a notable voice quality, or an unusual mismatch between fundamental frequency and formant dispersion. However, some talkers are more or less memorable not because of anything inherent to their voice, but because of something in the mind of the listener: whether the listener and speaker share a common language. Over the last three decades, one of the most reliable observations in studies of voice perception and talker identification has been that listeners are more accurate at identifying voices in their native language compared to a second or foreign language. This phenomenon, called the *language-familiarity effect* in talker identification, has been reported in numerous studies using diverse methodologies and a wide range of language pairings. The language-familiarity effect has been the subject of increasing scientific interest in the past decade, not only because of its importance for developing robust models of voice perception, but also because of the ways this phenomenon can provide new insights into models of speech perception, auditory plasticity, and even developmental disorders of language and communication.

My interest in the language-familiarity effect began in a café in Paris one rainy March afternoon, where I was struck by how much the voice of my waitress sounded like the recorded voice of the announcer at the train station, and even like the voices of some new friends I had met the day before. Surely speakers of French did not sound more alike one another than speakers of English? Yet this was my distinct impression. I returned from my trip to the laboratory of Patrick Wong with a pair of what I thought at the time were relatively simple questions: Does our native language affect our ability to recognize voices speaking other languages, and, if so, why? It turns out that the answer to the first question is a resounding “yes,” as many researchers had noted before (Goggin, Thompson, Strube, & Simental, 1991; Hollien, Majewski, & Doherty, 1982; Köster & Schiller, 1997; Sullivan & Schlichting, 2000; Thompson, 1987) and that finding the answer to the second question is not nearly as straightforward as I had hoped.

In this chapter, I review the extant and emerging research on the language-familiarity effect in talker identification, with a particular interest in addressing the question of what it is exactly that a listener knows about a language that helps them more accurately recognize voices speaking that language. It is worth starting out by noting that there is no *a priori* reason to assume that competence in a language should contribute to the ability to identify voices; indeed, the vast majority of studies of voice distinctiveness have revealed that obviously non-linguistic acoustic features such as pitch, voice quality, and vocal tract length by themselves provide robust dissociation of individual voices (Bachorowski & Owren, 1999; Baumann & Belin, 2008; Carrell, 1984; Latinus & Belin, 2011b; Latinus, McAleer, Bestelmeyer, & Belin, 2013; Lavner, Rosenhouse, & Gath, 2001; Remez, Fellowes, & Nagel, 2007; Schweinberger, Kawahara, Simpson, Skuk, & Zäske, 2014). And yet, studies of the effect of language on memory for voices routinely show a substantial improvement in talker identification when listening to a native versus foreign language, regardless of whether the speech is isolated words, short sentences, or longer samples.

It is not surprising that, in any task as behaviorally important as the social obligation to quickly and accurately recognize other individuals, the brain would seek to maximize the availability of potential sources of information. Indeed, there is an enormous amount of inter-talker variability in the phonetic information in speech, which remains relatively consistent for a given talker (Hillenbrand, Getty, Clark, & Wheeler, 1995; Theodore, Miller, & DeSteno, 2009) and to which listeners are decidedly sensitive during speech perception (Theodore & Miller, 2010). It stands to reason that listeners would also be able to use consistent inter-talker variation in speech phonetics not only to facilitate speech perception, but also to recognize individual talkers (Francis & Driscoll, 2006; Remez, Fellowes, & Rubin, 1997). Languages obviously differ in their phoneme inventories and thereby distributions of phonetic features. Correspondingly, the phonetic dimensions along which variation will meaningfully convey phonemic vs. idiolectic identity will be different across languages, and listeners' attention or

inattention to the relevant dimensions will help or hinder their ability to detect the individuating phonetic features of different talkers' speech and vocal identity.

However, at present, a number of key questions about the role of language processing in talker identification remain only poorly understood. Foremost among these is the question of what information, exactly, a listener gains access to by having competence in a language. Although the answer may trivially seem to be “everything,” a more specific understanding of the language-familiarity effect hinges on how much and what kinds of linguistic competence are required for improved talker identification abilities. Is passive exposure to the statistical distributions of phonetic features in native or foreign-language speech sufficient? Do listeners require access to higher-level linguistic processing, such as memory for words or even speech comprehension, to gain the full range of linguistic benefits that support enhanced talker identification? Or is the contribution of linguistic processing to talker identification more like language learning itself, in that it depends not only on exposure, but also on socially-relevant exposure in order to make its fullest contribution (Kuhl, Tsao, & Liu, 2003)?

In the following sections, I explore what we know about the language-familiarity effect in talker identification. I first briefly survey the extensive literature showing the reverse side of this phenomenon, that talker-specific variability affects speech processing. Then I describe the various studies that have investigated how listeners are able to recognize voices speaking a foreign language. I consider whether and how foreign-language learning or exposure affect the ability to identify voices speaking an unfamiliar language, as well as how early language exposure appears to establish a nascent language-familiarity effect in infants and children. Finally, I review the very limited evidence about the brain bases of the language-familiarity effect and describe how these numerous lines of evidence are beginning to converge on alternative psychological models of voice processing that account for the role of language abilities and experiences in talker identification.

Integration of indexical and phonological processing in speech perception

Variability in speech acoustics due to differences across talkers incur a cognitive cost during speech perception. Listeners are faster to make decisions about the content of speech when listening to a single consistent talker compared to multiple different talkers (Green, Tomiak, & Kuhl, 1997; Magnuson & Nusbaum, 2007; Mullennix & Pisoni, 1990). Listeners have better memory for words when they hear them spoken again by the same talker than when they are spoken by a new talker (Bradlow, Nygaard, & Pisoni, 1999; Palmeri, Goldinger, & Pisoni, 1993). Likewise, in a phenomenon known as the familiar-talker advantage, prior exposure to a talker's voice improves listeners' ability to perceive speech from that voice compared to an unfamiliar talker, particularly in adverse listening conditions like noise (Johnsrude et al., 2013; Newman & Evers, 2007; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994), and listeners' expectations about talker identity influence speech processing in real time (Creel, Aslin, & Tanenhaus, 2008; K. Johnson, 1990). Listeners' ability to extract the idiosyncratic but consistent source, filter, and dynamic phonetic features of a talker's voice to improve speech perception (Kleinschmidt & Jaeger, 2015) raises the possibility that our mental representations of speech and voice are indeed closely integrated (Kuhl, 2011) such that any information about one aids substantially in perceiving and remembering the other.

The language-familiarity effect in talker identification

Just as familiarity with a talker improves speech recognition, so too does familiarity with speech improve talker recognition. In the language-familiarity effect, listeners are more accurate at distinguishing voices when listening in their native language than when listening in a foreign language (**Figure 1**). This effect obtains across a host of different experimental design considerations: how many voices are included, the languages spoken by talkers and listeners, whether listeners are asked to identify or discriminate voices, how much exposure listeners have to the target voices, how long between exposure and test, and whether the speech content is the same at exposure and test. In a time of

increasing concern over reproducibility in psychological research (Open Science Collaboration, 2015), the language-familiarity effect remains one of the most robust and highly replicable phenomena in the psychology of voice processing. The reliability of this effect notwithstanding, a comprehensive model of voice recognition that parsimoniously integrates the potential cognitive, perceptual, and mnemonic bases for this effect has remained elusive.

Most contemporary researchers refer to the work of Judith Goggin and Charles Thompson (Goggin et al., 1991; Thompson, 1987) as the first to investigate how listeners are able to identify voices speaking another language; however, the observation of the language-familiarity effect in talker identification appears to have been originally described in a paper by Harry Hollien and colleagues (Hollien et al., 1982), and reported many years earlier in an abstract at the Acoustical Society of America (Hollien, Majewski, & Hollien, 1974). In that study, Polish-speaking listeners were significantly less accurate at identifying English-speaking voices than were English-speaking listeners. Although it has passed mostly unnoticed in the small field of language and voice, the Hollien study actually presaged many of the observations and approaches that have become standard today: Whereas the intervening years would see most research on the role of language in talker identification conducted using voice line-ups, the Hollien study explicitly trained listeners to identify talkers – the predominant method for studying the language-familiarity effect today. Additionally, Hollien and colleagues not only observed a similar magnitude for the effect of language familiarity to what we find today, they also noted that phonetic manipulations have a much smaller effect on talker identification by foreign-language listeners – an observation in accord with contemporary views of which features of speech are used by listeners to recognize native- versus foreign-language voices.

Identifying speakers of other languages in voice line-ups

Without a doubt, though, the experiments of Goggin and colleagues (1991), following the earlier report of Thompson (1987), represent one of the most comprehensive investigations of the role of language in talker identification. In a series of four experiments utilizing voice line-ups with paragraph-length recordings, Goggin and colleagues found that monolingual English listeners identified talkers' voices better when they were speaking English than German, but that monolingual German listeners showed the opposite effect from the same voices¹. English monolinguals were also more accurate at identifying voices when they spoke English than Spanish – replicating Thompson (1987) – but English-Spanish bilinguals were equally accurate regardless of which of those languages was being spoken.

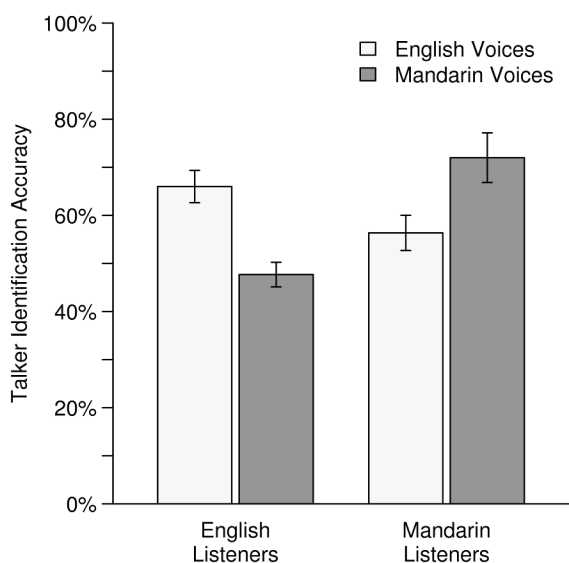


Figure 1: The language-familiarity effect in talker identification. People identify voices more accurately when listening to their native language than an unfamiliar foreign language. These data (redrawn from Perrachione & Wong, 2007) show that native English-speaking listeners are better at identifying English-speaking voices, whereas native Mandarin-speaking listeners are better at identifying Mandarin-speaking voices. This language-familiarity effect in talker identification has been observed for a wide range of language pairings and task designs.

Table 1.

The language-familiarity effect in voice line-ups

Study	Native Language	Foreign Language	Effect Size (<i>d</i>) of LFE	Participants
(Thompson, 1987)	English	Spanish	*	Control
(Goggin et al., 1991)	English	German	*	Control
		Spanish		
	German	English	*	Control
(Köster & Schiller, 1997)	Spanish	German	*†	
	Chinese			
(Sullivan & Schlichting, 2000)	English	Swedish	*†	Control
(Sullivan & Kügler, 2001)	English	Swedish	*†	Control
(Philippon et al., 2007)	English	French	*	Control
(Johnson et al., 2011)	English	Dutch	0.876	Control
		Japanese	0.827	
		Italian	0.125	

If studies reported multiple experiments, only the first experiment or only those testing the basic effect of language on talker identification are included here. "Control" participant groups were those selected in experiments that did target a specific population (e.g., musicians) or did not manipulate a between-group difference besides native language background. Effect sizes are calculated as Cohen's *d* from summary statistics reported in the papers or obtained from the authors.

*The type of data collected does not allow for calculation of the Cohen's *d* effect size statistic.

†Listeners were only tested in the foreign language.

Goggin and colleagues' (1991) fourth experiment was particularly clever: they examined talker identification accuracy as the sampled speech was made increasingly unlike English. They first recorded their talkers reading a paragraph in English; second, they read a paragraph consisting of the same words, but with their order randomized to produce nonsense; third, they read a paragraph consisting of the same syllables, but with their order again randomized to produce nonsense pseudo-English; fourth, they time-reversed the natural English recordings to produce incomprehensible backwards speech. Listeners' ability to accurately identify the target voice in the line-up fell with each manipulation, such that the less the target speech resembled English, the more poorly voices were identified.

The sum of these results were interpreted as suggesting that memories for voices were encoded via "schemata" that consisted of "norms for all aspects of a language, including its syntax, lexicon, and phonology", which were "learned through exposure to voices in a local area" (Goggin et al., 1991). This interpretation, although less formal, is perhaps not so unlike contemporary models of talker identification that posit prototype-based coding (Latinus & Belin, 2011a; Lavner et al., 2001). Furthermore, while no contemporary models – even those that are most assertive about a role for high-level linguistic processing as a basis for the language-familiarity effect (see below) – would suggest that syntactic processing plays a role in talker identification, there is evidence that listeners do learn talkers' idiosyncratic preferences for certain syntactic structures (Kamide, 2012).

Subsequent studies testing listeners' ability to identify foreign-language voices in voice line-ups (**Table 1**) tended not to compare performance against a native language condition, but between listeners of different proficiency levels (Köster & Schiller, 1997; Sullivan & Kügler, 2001; Sullivan & Schlichting, 2000). These studies are considered in greater detail in the section on the effects of foreign-language learning below. However, two more recent studies did test listeners in multiple language conditions and found converging results. When tested on line-ups of French-speaking voices, English-speaking listeners were more likely to select the incorrect target voice (more false alarms) than when tested on line-ups of English speaking voices (Philippon, Cherryman, Bull, & Vrij, 2007). Likewise, English-speaking listeners were less accurate at identifying Japanese- or Dutch-speaking voices in a line-up than they were English-speaking voices (E. K. Johnson, Westrek, Nazzi, & Cutler, 2011); however, interestingly, the same listeners were not impaired in their ability to identify Italian-speaking voices.

Table 2.

The language-familiarity effect in studies of talker discrimination

Study	Native Language	Foreign Language	Effect Size (d) of LFE	Participants
(Winters et al., 2008)	English	German	*	Control
(E. K. Johnson et al., 2011)	Dutch	Japanese Italian	*	Infants
(Wester, 2012)	English	German Finnish Mandarin	*	Control
(Levi & Schwartz, 2013)	English	German	1.60	Adults
	English	German	0.33	Children
(Fleming et al., 2014)	English	Mandarin	0.306	Control
	Mandarin	English	0.153	Control

Studies and effect sizes reported as in **Table 1**.

*Effect sizes could not be calculated because the necessary descriptive summary statistics were not available.

Discriminating speakers of other languages

A relatively less-frequently used technique to explore the language-familiarity effect is the discrimination paradigm, in which listeners hear pairs of voices and decide whether they are the same or different. There are a variety of reasons why discrimination paradigms may be less preferred for studying the role of language in voice processing. For one, listeners' voice discrimination abilities tend to be very good, such that ceiling effects may obfuscate differences between conditions. Additionally, in discrimination paradigms, there is reason to believe that listeners may be attending to different features of the voices, including particularly placing greater emphasis on low-level acoustic differences between pairs of stimuli, which may not accurately represent the psychological processes that are contributing to more ecological voice recognition behaviors (Perrachione, Stepp, Hillman, & Wong, 2014; Van Lancker & Kreiman, 1987).

Nonetheless, several experiments have shown an influence of listeners' native language on their ability to discriminate pairs of talkers (**Table 2**). Native English-speaking listeners are more accurate at discriminating pairs of voices speaking English than when the same voices speak in German (Winters, Levi, & Pisoni, 2008). Interestingly, performance falls even further when listeners are required to make the discrimination decision across the two languages. Discrimination performance also improves with explicit prior training on the voices, but there is no interaction between the language spoken during training and that during test, further suggesting that the acoustic features used in voice discrimination are based more on stimulus-specific acoustic features than those that contribute to memory for voices. Similar results have also been obtained for English-speaking listeners discriminating not only English- and German-, but also Finnish- and Mandarin-speaking voices (Wester, 2012), including especially reduced performance for across-language voice discrimination. More recently, Levi and Schwartz (2013) have reported a substantially larger difference between English-speaking listeners' ability to discriminate voices when listening to English speech compared to German.

In an interesting departure from the usual binary response of discrimination paradigms, Fleming and colleagues (2014) conducted a study in which native English- and native Mandarin-speaking listeners rated the subjective similarity of native-, foreign-, and across-language voice pairs. Critically, all speech samples in this experiment were time-reversed, rendering them incomprehensible to listeners regardless of which language was being spoken. Both native English- and Mandarin-speaking listeners rated pairs of voices in their native language as sounding more distinct than pairs of voices in the foreign language, despite never being able to comprehend the speech. These results accord well with the confidence ratings from studies using voice line-ups, in which listeners are more confident in their ability to detect target voices speaking their native language than a foreign one (Goggin et al., 1991;

Table 3.

The language-familiarity effect in studies that train talker identification

Study	Native Language	Foreign Language	Effect Size (<i>d</i>) of LFE	Participants
(Hollien et al., 1982)	Polish	English	*	Control
(Perrachione & Wong, 2007)	English	Mandarin	1.585	Control
	Mandarin	English	0.970	Control
(Winters et al., 2008)	English	German	*	Control
(Perrachione et al., 2009)	English	Mandarin	0.902	Control
	Mandarin	English	1.827	Control
(Perrachione et al., 2011)	English	Mandarin	1.153	Typical readers
	English	Mandarin	-0.091	Readers with dyslexia
(Bregman & Creel, 2014)	English	Korean	0.922	Monolinguals
	Korean	English	0.449	Bilinguals
(Orena et al., 2015)	English	French	*	High and Low L2 contact groups
(Xie & Myers, 2015a)	English	Mandarin, Spanish	1.474 0.976	Musicians
	English	Mandarin, Spanish	1.802 1.671	Non-Musicians
	Mandarin	English, Spanish	1.587 1.949	Control
(Zarate et al., 2015)	English	Mandarin, German	*	Control
(Kadam et al., 2016)	English	French	1.743	Average and advanced readers

Studies and effect sizes reported as in **Table 1**.

*Effect sizes could not be calculated because the necessary descriptive summary statistics were not available.

Thompson, 1987; cf. Philippon et al., 2007). Interestingly, listeners also judged across-language voice pairs as sounding significantly more distinct than either within-language pairing – a result curiously at odds with listeners' poor performance discriminating across-language voice pairs (Wester, 2012; Winters et al., 2008). However, the effect size of language-familiarity in similarity judgments of time-reversed speech is very small compared to studies in which listeners explicitly identify voices.

Training listeners to identify speakers of other languages

The most common method for studying the effect of language on voice recognition – and the one that produces the most consistent and largest effects of language – is to explicitly train listeners to identify talkers speaking in a known and unknown language (**Table 3**). Training explicit talker identification, in which listeners learn to associate different voices with a corresponding label, such as a name, number, photo, or avatar, has the advantage of being more similar to ecological voice recognition behaviors, as well as incorporating the full range of factors that may contribute to differences in talker identification abilities, including perception, learning, and memory of voices.

The earliest experiment in which listeners were trained to explicitly identify a slate of talkers found that Polish-speaking monolinguals identified English-speaking voices less accurately than did native English speakers (Hollien et al., 1982). More recently, native speakers of English have been found to be more accurate at explicitly identifying voices speaking English than those speaking Mandarin, whereas native Mandarin speakers are more accurate for Mandarin-speaking voices than English-speaking ones (**Figure 1**) (Perrachione & Wong, 2007). This latter study further demonstrated that the language-familiarity effect is robust to training when listeners have no knowledge of the foreign language: Monolingual English speakers improve in their ability to recognize both English and Mandarin voices across six days of explicit talker identification training, but are always better at

identifying voices speaking English than those speaking Mandarin. Native Mandarin speakers who have some competence in English, however, show a different pattern: Although they are initially more accurate identifying Mandarin-speaking voices, after six days of explicit training on both Mandarin- and English-speaking voices, the effect of language disappears for these bilingual listeners. These results have been interpreted to mean that some sort of linguistic competence, not mere exposure, is critical for the enhanced talker identification accuracy associated with the language-familiarity effect.

Numerous design considerations come into play in developing talker identification training paradigms: How many voices will be trained? How much training will listeners receive? How should voices be labeled? Interestingly, although some researchers prefer one alternative over another, none of these design considerations appears to meaningfully affect the observation of the language-familiarity effect in talker identification studies. Experiments training four (Bregman & Creel, 2014; Kadam, Orena, Theodore, & Polka, 2016; Orena, Theodore, & Polka, 2015) or five voices (Perrachione, Del Tufo, & Gabrieli, 2011; Perrachione, Pierrehumbert, & Wong, 2009; Perrachione & Wong, 2007; Xie & Myers, 2015a; Zarate, Tian, Woods, & Poeppel, 2015) do not produce wildly different estimates of the magnitude of the language-familiarity effect. Similar effects of language-familiarity are also seen in studies that give all listeners a fixed amount of training (Perrachione et al., 2011, 2009; Perrachione & Wong, 2007; Xie & Myers, 2015a; Zarate et al., 2015) and those in which listeners are trained to a particular level of performance before a generalization test (Bregman & Creel, 2014; Kadam et al., 2016; Orena et al., 2015). Finally, there does not appear to be any difference in listeners' ability to learn voices – or to learn voices better in their native language – when the trained voices are paired with names (Perrachione & Wong, 2007), numbers (Perrachione et al., 2009; Xie & Myers, 2015a), or cartoon avatars (Bregman & Creel, 2014; Kadam et al., 2016; Orena et al., 2015; Perrachione et al., 2011; Zarate et al., 2015).

The numerous studies of talker identification that have confirmed a reliable presence and magnitude of the language-familiarity effect have also pushed forward our understanding of the sources of and factors affecting this phenomenon. Several have investigated how second-language learning and exposure impact the language-familiarity effect (Bregman & Creel, 2014; Orena et al., 2015) and will be addressed in greater detail below. Others have investigated how individual differences in cognitive and perceptual abilities affect native- and foreign-language talker identification. For instance, better native-language phonological skills appear to endow listeners with superior foreign- (but not native-) language talker identification abilities (Kadam et al., 2016). Moreover, individuals with superior pitch perception abilities – whether because they are musicians or speakers of a tone language like Mandarin – also have enhanced foreign- (but not native-) language talker identification abilities (Xie & Myers, 2015a). Still other studies have investigated the contribution of stimulus factors to the language-familiarity effect: Zarate and colleagues (2015) found that the language-familiarity effect was greater for native English speakers when identifying voices speaking Mandarin – a language with a very different phonology compared to English – and smaller when identifying voices speaking German, with its more similar phonology (however, cf. E. K. Johnson et al., 2011; Köster & Schiller, 1997; and Xie & Myers, 2015).

How foreign-language learning affects foreign-language talker identification

To what extent is the language-familiarity effect fixed, the result of some “critical period” in the early development of voice perception, or to what extent is it plastic to experience with new voices in adulthood? And, if the contribution of language skills to talker identification is plastic, what kinds of exposure or expertise are necessary to improve foreign-language talker identification abilities? Although many researchers have taken advantage of natural experiments or designed carefully constructed laboratory studies to address these questions, our poor understanding of how foreign-language talker identification can be improved with experience or training reaffirms how little we understand about the cognitive processes that underlie this effect in the first place.

There is some evidence that the kinds of foreign-language competence one acquires in the usual course of second-language study can improve talker identification abilities in the second language. Native Chinese-speaking students studying in the United States were found to be able to recognize English-speaking voices in a voice line-up with accuracy equal to that of native English-speaking students (Goldstein, Knight, Bailis, & Conover, 1981). Native speakers of Chinese or Spanish who had completed foreign-language study in German were also more accurate at identifying German-speaking voices than were Chinese or Spanish speakers without any prior German knowledge (Köster & Schiller, 1997). Likewise, native English speakers who had studied Swedish were more accurate than English speakers with no knowledge of Swedish at identifying Swedish-speaking voices in a voice line-up, but advanced students of Swedish did not outperform novices (Sullivan & Kügler, 2001; Sullivan & Schlichting, 2000).

However, more recent studies using explicit talker identification training may temper our enthusiasm for automatic gains in foreign-language talker identification abilities following foreign language study. In similarly obtained samples of native Mandarin-speaking students who have gained satisfactory proficiency in English to study abroad at American universities, Mandarin L1 listeners nonetheless were found to have significantly poorer identification of English-speaking voices compared to either Mandarin-speaking voices or performance by English-speaking listeners (Perrachione et al., 2009; Perrachione & Wong, 2007; Xie & Myers, 2015a). The persistence of the language-familiarity effect in these listeners can, however, be diminished and even eliminated with further explicit training on foreign-language talker identification (Perrachione & Wong, 2007). There is also some preliminary evidence to suggest that foreign-language learners who seek more immersive second language experiences may overcome the language-familiarity effect to a greater degree than those who still predominately use their native language while abroad (Dougherty & Perrachione, 2016).

Instead of second language skills acquired in adulthood, Bregman and Creel (2014) investigated whether earlier exposure in childhood to a second language was associated with improved second-language talker identification skills. They found that, for adult Korean L1, English L2 bilinguals, the younger the age of their English exposure, the faster they were able to learn to identify English-speaking voices and the smaller their relative language-familiarity effect for Korean-speaking voices. These results suggest that earlier acquisition of (or potentially greater lifelong experience with) a second language can improve individuals' ability to recognize voices speaking in that language.

Others have gone further to suggest that lifelong exposure to a foreign language need not involve any actual competence in that language to improve talker identification abilities: Monolingual English-speaking adults from Montreal are faster and more accurate at learning French-speaking voices than are monolingual English-speaking adults from Connecticut (Orena et al., 2015). These results suggest that merely being in an environment in which one regularly hears a foreign language, even without being able to speak it themselves, can improve talker identification abilities in that language – a result consistent with Goggin and colleagues' (1991) idea of voice schemata based on the local environment. However, these results also raise the question of how much passive exposure one can have in a language without gaining some degree of linguistic knowledge, and, furthermore, what that knowledge might be and how it might contribute to improved talker identification skills. Johnson and colleagues' (2011) reported a similar observation, that adult listeners' mere familiarity with a language – even in the alleged absence of any particular linguistic competence – may reduce the magnitude of the native-language advantage. However, in contrast, even explicit and extended training on foreign language talker identification in a controlled laboratory setting failed to diminish the language-familiarity effect for native English speakers attempting to identify Mandarin-speaking voices (Perrachione & Wong, 2007). Briefly, this small collection of mostly equivocal results have left unanswered many important questions about how listeners learn to deploy second-language skills and knowledge in the service of talker identification, and even what these skills and knowledge consist of.

The role of language in the development of voice recognition abilities

Voice recognition abilities emerge early in development (Blasi et al., 2011; Grossmann, Oberecker, Koch, & Friederici, 2010; Kisilevsky et al., 2003; Mehler, Bertoncini, Barrière, & Jassik-Gerschenfeld, 1978), paralleling early development of language-specific processing and representations (Kuhl, 2004; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Stager & Werker, 1997). It is therefore interesting to consider what role early language experience has on the development of voice recognition abilities, and what this tells us about the cognitive or mnemonic bases of the language-familiarity effect.

In an ambitious study, and the only one of its kind to date, infants as young as 7 months were found to be more sensitive to a change in talker when listening to speech in their (emerging) native language than when listening to speech in an unfamiliar foreign language or time-reversed speech (E. K. Johnson et al., 2011). This remarkable result suggests that there is a language-familiarity effect in voice perception even in infants who putatively know few words (cf. Bergelson & Swingley, 2012) and who are still learning the phonemic and phonotactic distribution of sounds in their native language. Given the limited linguistic knowledge or experience of these infants, it is a fascinating question to speculate on what distinctive features attract their attention to changing voices in their native language, but not in a foreign language. A lack of attention to foreign-language speech alone does not account for these observations, since infants in that study were just as likely to notice a change from one foreign language to another.

The role that language abilities play in voice recognition are unfortunately no better understood as children get older. In the only study to date of the role of language in voice processing by children, young English-speaking children age 7-9 years are significantly more accurate at discriminating voices when listening to English speech than German speech (Levi & Schwartz, 2013). Adults in the study also showed this language-familiarity effect in talker discrimination. However, older English-speaking children age 10-12 years showed no difference in their ability to discriminate German- or English-speaking voices. Why and how do foreign-language voice discrimination abilities develop across childhood, only to fall off again in adults? More work is clearly needed to understand the contribution of language abilities to talker identification across development.

Interestingly, the same study found that neither younger nor older children with specific language impairment (SLI) showed a language-familiarity effect in talker discrimination, and that children with SLI did not underperform in talker discrimination compared to children with typical language abilities. This observation stands in partial contrast to a study of the language-familiarity effect in language-impaired adults. Specifically, adults with dyslexia – a phonologically based disorder of reading development – also do not exhibit a language-familiarity effect in talker identification (Perrachione et al., 2011). This is because adults with dyslexia do not appear to gain the typical advantage for talker identification in a native language compared to a foreign one, even though they are not impaired in foreign-language talker identification compared to adults with typical reading abilities. The parallel between the cognitive processes that underlie reading, such as phonological awareness, and improved ability in talker identification skills has also been observed by others (Kadam et al., 2016).

Together, these results suggest that the early development of linguistic skills unfolds in parallel with the development of voice recognition abilities, and that early exposure to voices speaking a particular language yields listeners who are more sensitized to talker-specific differences in that language (Bregman & Creel, 2014; E. K. Johnson et al., 2011). However, the existence of the language-familiarity effect is poorly attested in children, mostly because the only study in this age range used a talker discrimination paradigm, which, as discussed above, is known to be less sensitive to the role of language in voice processing due to ceiling effects and differences in task demands. More research into how the talker identification abilities of children are shaped by language experience is clearly necessary.

Neural integration of speech and voice processing

Although there is a growing understanding of the neural bases of voice processing (Belin, Zatorre, & Ahad, 2002; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Latinus et al., 2013; Pernet et al., 2015; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003; von Kriegstein & Giraud, 2004), as well as the dynamic integration of representations of voice and speech information (Chandrasekaran, Chan, & Wong, 2011; Formisano, De Martino, Bonte, & Goebel, 2008; Kaganovich, Francis, & Melara, 2006; Kreitewolf, Gaudrain, & von Kriegstein, 2014; Perrachione et al., 2016; Sjerps, Mitterer, & McQueen, 2011; von Kriegstein, Smith, Patterson, Kiebel, & Griffiths, 2010; Wong, Nusbaum, & Small, 2004; Zhang et al., 2016), there has been disappointingly little research into the brain bases of the language-familiarity effect in talker identification. However, given the above results, we can postulate how neural systems for speech and voice perception might align to facilitate native-language talker identification.

Although voice recognition is generally associated with auditory areas predominately in the right hemisphere (Belin, Bestelmeyer, Latinus, & Watson, 2011; Belin et al., 2000; von Kriegstein & Giraud, 2004), there is reason to hypothesize that linguistically-derived representations of talker identity may reflect increased bilateral integration between the right and left hemispheres. Evidence from patients with brain injuries suggest that, while the right hemisphere is important for recognizing familiar voices, the left hemisphere plays a role in distinguishing new and unfamiliar voices (Van Lancker & Kreiman, 1987), similar to the important role of language in learning new voice identities. Likewise, although there is an overall left-ear / right-hemisphere bias for talker identification, there appears to be increased recruitment of the left hemisphere specifically in native- compared to foreign-language talker identification tasks (Perrachione et al., 2009), consistent with neuroimaging evidence for increased left-right integration during processing talker-specific information in speech (Kreitewolf et al., 2014; von Kriegstein et al., 2010). Neural representations of speech and voice information also overlap in auditory areas of both hemispheres (Formisano et al., 2008), suggesting that integration of both speech content and talker identity is likely to occur bilaterally. Furthermore, individuals with dyslexia are impaired in their ability to identify native- but not foreign-language voices (Perrachione et al., 2011), and the brains of these individuals show attenuated neural adaptation to the repetition of native-language voices in both the left and right hemispheres compared to controls, but only in the left hemisphere for the repetition of speech content (Perrachione et al., 2016).

Despite these inferences and tangential data, there has been no targeted investigation of the brain bases of native- versus foreign-language talker identification to date. The acquisition paradigms and statistical analysis tools are now in place to ascertain how neural systems integrate speech and voice information in a more sophisticated way than was previously possible using classical cognitive subtractions (Fristen, 1997), and the time is right to make good on decades-old speculation about the brain bases of the language-familiarity effect (Perrachione & Wong, 2007).

Theories and models of the language-familiarity effect

Based on the converging evidence from voice line-ups, talker discrimination, talker identification training, foreign language learning, development, and neuroscience, do we now have sufficient evidence to build a psychological model of voice processing that accurately accounts for the contribution of language to talker identification? Although a dominant model has not emerged, we are closer today than ever before. Prior explanations for why voices were more identifiable in a familiar language relied on broadly defined “schemata” based on the sum of one's experience with voices (Goggin et al., 1991). While this interpretation has held up remarkably well to much of the new evidence collected over the subsequent twenty-five years, it nonetheless falls short of explanatory adequacy in several ways. For one, it borders on tautological to assert that the reason voices are identified more accurately in a native language is because one is more familiar with such voices. What features in particular are more familiar, and what exactly does one know about the language that makes perception of, or memory for, those features more accessible? Additionally, there are some data for which a schema-based model cannot

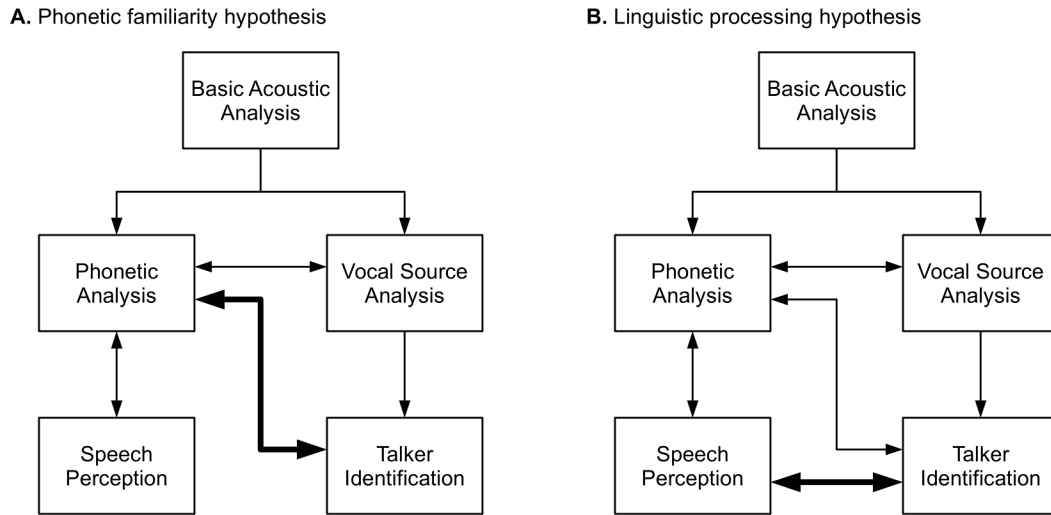


Figure 2: Psychological models of the language-familiarity effect posit different roles for linguistic processing and representations. Although the phenomenon of superior talker identification abilities in one's native language has been extensively documented, the psychological bases for this effect remain a matter of active research. **(A)** Familiarity with the phonetics and phonotactics of one's native language is a likely source for the language-familiarity effect. A growing body of evidence demonstrates that increasing familiarity with the phonetic features, phoneme inventory, and phonotactic distributions of a language plays a role in accurate talker identification. **(B)** Other evidence suggests that passive or statistical familiarity with the sound patterns of a language does not explain the full extent of the native-language advantage in talker identification. Beyond phonetics, higher-level linguistic processes, such as recognizing and remembering words and phrases may also play a role in enhancing listeners' ability to identify voices speaking a familiar language.

satisfactorily account: It is so far equivocal whether exposure without competence supports foreign-language talker identification (Dougherty & Perrachione, 2016; E. K. Johnson et al., 2011; Orena et al., 2015; Perrachione & Wong, 2007), even though a schema-based model predicts unequivocally that only exposure should matter. Additionally, the developmental trajectory of talker identification seems to be more complicated than an exposure-only model would predict (Bregman & Creel, 2014; E. K. Johnson et al., 2011; Levi & Schwartz, 2013); likewise, an exposure-only model that does not distinguish the relevant underlying features cannot account for impaired native-, but not foreign-language talker identification in disorders like dyslexia (Perrachione et al., 2011).

Contemporary researchers have, either implicitly or explicitly, begun to converge on two different descriptive models that go further in explaining how particular features of familiar and native-language speech may contribute to talker identification (**Figure 2**). The first of these – which I will call the *phonetic familiarity hypothesis* – is the idea that talker identification abilities take advantage of listeners' increased familiarity with the statistical distributions of phonetic features in their native language, including how variations in these features meaningfully reveal phonemic versus idiolectal identity (Fleming et al., 2014; E. K. Johnson et al., 2011; Orena et al., 2015; Zarate et al., 2015). A highly congruent model – which I will call the *linguistic processing hypothesis* – incorporates the role of familiar phonetics from the first model, but goes further in suggesting that a major source of improved talker identification accuracy in one's native language results from higher-level linguistic processing, such as recognition of words and comprehension of speech content, that depends on linguistic competence (Bregman & Creel, 2014; Perrachione, Dougherty, McLaughlin, & Lember, 2015; Perrachione & Wong, 2007). The principal questions in adjudicating between or synthesizing these models are: What linguistic factors contribute to talker identification, and how much do the various factors contribute independently versus depend on one another?

A phonetic familiarity model

The idea that the language-familiarity effect in talker identification arises due to increased familiarity with the phonological system of one's native language is supported by a number of converging lines of evidence. First, a language-familiarity effect has been found in infants at 7 months (E. K. Johnson et al., 2011), well before they have had the opportunity to establish higher-level linguistic representations such as words, but at an age when native-language phonological skills are also beginning to emerge (Kuhl, 2004). Second, listeners judge time-reversed native language voices to sound more distinct than time-reversed foreign language voices (Fleming et al., 2014), suggesting that familiar phonetic patterns, such as the vowel space, may contribute to voice processing even in the absence of comprehensible speech. Third, a larger magnitude language-familiarity effect has been reported between more phonologically dissimilar languages compared to more phonologically similar ones (Zarate et al., 2015), suggesting that the degree to which foreign-language speech can be mapped onto native-language phonological representations may facilitate talker identification. Fourth, the magnitude of the language-familiarity effect has been reported to diminish with increased exposure to a foreign language, even allegedly without any linguistic competence in that language (Orena et al., 2015). Together, these lines of research provide important and convincing evidence of a role for familiarity with language-specific phonetic patterns in enhancing voice processing.

However, there are alternative interpretations of or inconsistencies with the above results: Although infants exhibit a language-familiarity effect that presumably cannot be based on higher-level linguistic processing, it need not be the case (nor is it even likely) that language-familiarity effects in infants and adults arise from identical mechanisms. Additionally, although listeners rate time-reversed native-language voice pairs as sounding more distinct, we have seen above that the results of paradigms in which listeners discriminate voices do not always map cleanly onto the results of tasks in which they identify talkers. Indeed, more recent evidence has suggested that when asked to learn and identify voices from time reversed speech, listeners exhibit no language-familiarity effect whatsoever (Perrachione et al., 2015). Evidence for enhanced talker identification due to phonological similarity between languages is likewise equivocal, with some studies finding a reduced language-familiarity effect when two languages are more typologically similar (Zarate et al., 2015), whereas others that looked for such an effect do not find one (E. K. Johnson et al., 2011; Köster & Schiller, 1997; Xie & Myers, 2015a). Likewise, the observation that passive exposure to a language reduces the language familiarity effect has been reported by some studies (E. K. Johnson et al., 2011; Orena et al., 2015), but other studies have shown that the language-familiarity effect can be equally large whether listeners have prior experience with the foreign language or not (Perrachione & Wong, 2007; Xie & Myers, 2015a).

A linguistic processing model

The second descriptive model inherits all the connections of the phonetic familiarity model, but adds an important connection between speech processing (including especially word recognition and comprehension) and talker identification. There are a number of observations which do not seem to be well-accounted for by the phonetic familiarity model and which seem to suggest that processing the higher-level linguistic content of speech further facilitates native-language talker identification. First, even explicit and extended training on foreign language talker identification does not reduce the magnitude of the language-familiarity effect for listeners with no linguistic competence in the foreign language, but such training does improve foreign-language talker identification for emerging bilinguals (Perrachione & Wong, 2007). Additionally, talker identification improves for meaningful speech compared to meaningless but phonologically-balanced nonsense speech (Goggin et al., 1991; Perrachione et al., 2015; Xie & Myers, 2015b), further indicating that meaningful and familiar linguistic units like words facilitate talker identification. Finally, listeners are more accurate identifying voices when they can compare and remember how different talkers say words, an effect that obtains only in a native language, not a foreign one (McLaughlin, Dougherty, Lember, & Perrachione, 2015), which also

suggests that part of the representation in memory for voices is memory for how they sound saying particular words. Together, these results suggest that accurate memory for voices takes advantage of high-level linguistic processes and representations that come with linguistic competence, not just passive familiarity with sound structure.

There are limitations on these data as well that nonetheless underscore a core importance for familiarity with phonological structure. In particular, the decrement in talker identification performance between meaningful speech and meaningless but phonologically well-formed speech is much smaller than the corresponding decrement between meaningful and foreign-language speech (Goggin et al., 1991; Perrachione et al., 2015; Xie & Myers, 2015b), an observation that provides some insight into the relative importance of these features to accurate native-language talker identification.

Additionally, there is also some evidence that listeners cannot fully take advantage of familiar words in the absence of familiar phonetics. There are numerous studies demonstrating a decrement in talker identification accuracy when listening to one's native language, but in an unfamiliar regional or social accent (Doty, 1998; Goggin et al., 1991; Kerstholt, Jansen, Van Amelsvoort, & Broeders, 2006; Perrachione, Chiao, & Wong, 2010; Stevenage, Clarke, & McNeill, 2012). Goldstein and colleagues (1981) even found that English-speaking listeners were just as accurate recognizing voices in a voice line-up when they spoke Spanish as when the same voices spoke heavily Spanish-accented English, although most other studies have found smaller magnitude effects in an unfamiliar accent than an unfamiliar language (Goggin et al., 1991; Perrachione et al., 2010). Why the availability of a higher-level speech processing route might depend on the familiarity of the underlying phonetics is also an interesting question. Perhaps the contributions of phonetics and higher-order linguistic representations are indeed hierarchical. Alternatively, perhaps cognitive resources that would otherwise be available for talker identification are being usurped in the service of speech perception, mediating any performance gain from familiarity with the lexical content of speech. Future work is necessary to adjudicate whether and how the putative contribution of linguistic processing depends on the presence of familiar phonetics.

Unresolved questions about recognizing speakers across languages

- Is improved talker identification accuracy in a native language due to enhanced ability to *perceive* the relevant features that distinguish an individual talker, to *learn* those features, or to *remember* them when one encounters a voice again?
- Under the phonetic familiarity hypothesis, is the information learned from passive exposure to foreign-language sounds (Orena et al., 2015) the same kind of information about these sounds that is gained through developing linguistic competence (Fleming et al., 2014; E. K. Johnson et al., 2011)?
- Can listeners develop passive phonological familiarity with a foreign language (Orena et al., 2015) without actually gaining some linguistic competence (particularly lexical representations) (E. K. Johnson et al., 2011; Perrachione & Wong, 2007)?
- Under the linguistic processing hypothesis, can higher-level linguistic processing give rise to the language-familiarity effect even for phonologically unfamiliar (i.e., heavily accented) speech?
- How much do the typological or phonological similarities of the talker and listener's languages matter to the magnitude of the language-familiarity effect in talker identification (cf. Bent & Bradlow, 2003)?
- What role do familiar prosodic patterns in a native language play in talker identification?

Summary & Conclusion

One of the most reliable observations in voice perception research is that listeners identify talkers more accurately when they can understand the language being spoken compared to when they cannot. This language-familiarity effect arises from lifelong experiences listening to and recognizing talkers speaking in one's native language, and appears to depend primarily on familiarity with the phonological system of

a language, but also takes advantage of higher level linguistic representations, particularly memory for words. Superior native-language voice processing skills emerge early and are refined throughout development. Although this voice processing advantage is most prominent for one's native language, extensive and substantive experience with speech in a second language can improve talker identification abilities in that language, as well. The language-familiarity effect in talker identification reveals the closely integrated psychological, and therefore neural, representations of talker identity and speech content, providing further insight into how the mind and brain extract core linguistic and social information from the single, convolved, communicative signal of speech.

Acknowledgments

This work was supported in part by NIH R03 DC014045.

References

- Bachorowski, J.-A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *The Journal of the Acoustical Society of America*, 106(2), 1054–1063. <https://doi.org/10.1121/1.427115>
- Baumann, O., & Belin, P. (2008). Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychological Research PRPF*, 74(1), 110–120. <https://doi.org/10.1007/s00426-008-0185-z>
- Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding Voice Perception. *British Journal of Psychology*, 102(4), 711–725. <https://doi.org/10.1111/j.2044-8295.2011.02041.x>
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, 13(1), 17–26.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767), 309–312.
- Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, 114(3), 1600–1610. <https://doi.org/10.1121/1.1603234>
- Bergelson, E. & Swingle, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253–3258.
- Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., ... Murphy, D. G. M. (2011). Early Specialization for Voice and Emotion Processing in the Infant Brain. *Current Biology*, 21(14), 1220–1224. <https://doi.org/10.1016/j.cub.2011.06.009>
- Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, 61(2), 206–219.
- Bregman, M. R., & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition*, 130(1), 85–95.
- Carrell, T. D. (1984). *Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification*. UMI.
- Chandrasekaran, B., Chan, A. H. D., & Wong, P. C. M. (2011). Neural processing of what and who information in speech. *Journal of Cognitive Neuroscience*, 23(10), 2690–2700. <https://doi.org/10.1162/jocn.2011.21631>
- 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. <https://doi.org/10.1016/j.cognition.2007.03.013>
- Doty, N. D. (1998). The influence of nationality on the accuracy of face and voice recognition. *The American Journal of Psychology*, 111(2), 191–214.
- Dougherty, S. C., & Perrachione, T. K. (2016). The language-familiarity effect in talker identification by highly proficient bilinguals depends on second-language immersion. Presented at the 171st meeting of the Acoustical Society of America, Salt Lake City, UT.
- 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(38), 13795–13798.
- Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science (New York, N.Y.)*, 322(5903), 970–973. <https://doi.org/10.1126/science.1164318>
- Francis, A. L., & Driscoll, C. (2006). Training to use voice onset time as a cue to talker identification induces a left-ear/right-hemisphere processing advantage. *Brain and Language*, 98(3), 310–318. <https://doi.org/10.1016/j.bandl.2006.06.002>
- Fristen, K. J. (1997). Imaging cognitive anatomy. *Trends in Cognitive Sciences*, 1(1), 21–27. [https://doi.org/10.1016/S1364-6613\(97\)01001-2](https://doi.org/10.1016/S1364-6613(97)01001-2)
- 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.
- Goldstein, A. G., Knight, P., Bailis, K., & Conover, J. (1981). Recognition memory for accented and unaccented voices. *Bulletin of the Psychonomic Society*, 17(5), 217–220. <https://doi.org/10.3758/BF03333718>
- Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics*, 59(5), 675–692.
- Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The Developmental Origins of Voice Processing in the Human Brain. *Neuron*, 65(6), 852–858. <https://doi.org/10.1016/j.neuron.2010.03.001>
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97(5 Pt 1), 3099–3111.
- Hollien, H., Majewski, W., & Doherty, E. T. (1982). Perceptual identification of voices under normal, stress and disguise speaking conditions. *Journal of Phonetics*, 10(2), 139–148.

- Hollien, H., Majewski, W., & Hollien, P. A. (1974). Perceptual identification of voices under normal, stress, and disguised speaking conditions. *The Journal of the Acoustical Society of America*, 56(S1), S53–S53. <https://doi.org/10.1121/1.1914230>
- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, 14(5), 1002–1011.
- Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *The Journal of the Acoustical Society of America*, 88(2), 642–654.
- Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science*, 24(10), 1995–2004. <https://doi.org/10.1177/0956797613482467>
- Kadam, M. A., Orena, A. J., Theodore, R. M., & Polka, L. (2016). Reading ability influences native and non-native voice recognition, even for unimpaired readers. *The Journal of the Acoustical Society of America*, 139(1), EL6–EL12. <https://doi.org/10.1121/1.4937488>
- Kaganovich, N., Francis, A. L., & Melara, R. D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain Research*, 1114(1), 161–172. <https://doi.org/10.1016/j.brainres.2006.07.049>
- Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, 124(1), 66–71. <https://doi.org/10.1016/j.cognition.2012.03.001>
- Kerstholt, J. H., Jansen, N. J. M., Van Amelsvoort, A. G., & Broeders, A. P. A. (2006). Earwitnesses: effects of accent, retention and telephone. *Applied Cognitive Psychology*, 20(2), 187–197. <https://doi.org/10.1002/acp.1175>
- Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., ... Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, 14(3), 220–224.
- Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. <https://doi.org/10.1037/a0038695>
- Köster, O., & Schiller, N. O. (1997). Different influences of the native language of a listener on speaker recognition. *International Journal of Speech Language and the Law*, 4(1), 18–28.
- Kreitewolf, J., Gaudrain, E., & von Kriegstein, K. (2014). A neural mechanism for recognizing speech spoken by different speakers. *NeuroImage*, 91, 375–385. <https://doi.org/10.1016/j.neuroimage.2014.01.005>
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831–843. <https://doi.org/10.1038/nrn1533>
- Kuhl, P. K. (2011). Who's Talking? *Science*, 333(6042), 529–530. <https://doi.org/10.1126/science.1210277>
- Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15), 9096–9101. <https://doi.org/10.1073/pnas.1532872100>
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606–608.
- Latinus, M., & Belin, P. (2011a). Anti-Voice Adaptation Suggests Prototype-Based Coding of Voice Identity. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00175>
- Latinus, M., & Belin, P. (2011b). Human voice perception. *Current Biology*, 21(4), R143–R145. <https://doi.org/10.1016/j.cub.2010.12.033>
- Latinus, M., McAleer, P., Bestelmeyer, P. E. G., & Belin, P. (2013). Norm-Based Coding of Voice Identity in Human Auditory Cortex. *Current Biology*, 23(12), 1075–1080. <https://doi.org/10.1016/j.cub.2013.04.055>
- Lavner, Y., Rosenhouse, J., & Gath, I. (2001). The Prototype Model in Speaker Identification by Human Listeners. *International Journal of Speech Technology*, 4(1), 63–74. <https://doi.org/10.1023/A:1009656816383>
- Levi, S. V., & Schwartz, R. G. (2013). The development of language-specific and language-independent talker processing. *Journal of Speech, Language, and Hearing Research*, 56(3), 913–920. [https://doi.org/10.1044/1092-4388\(2012/12-0095\)](https://doi.org/10.1044/1092-4388(2012/12-0095))
- 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. <https://doi.org/10.1037/0096-1523.33.2.391>
- McLaughlin, D. E., Dougherty, S. C., Lember, R. A., & Perrachione, T. K. (2015). Episodic memory for words enhances the language familiarity effect in talker identification. Presented at the Proceedings of the International Congress of Phonetic Sciences XVIII, Glasgow, Scotland.
- Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, 7(5), 491 – 497. <https://doi.org/10.1068/p070491>
- Mullenix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47(4), 379–390.
- Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics*, 35(1), 85–103.
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3), 355–376.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5(1), 42–46.
- Open Science Collaboration. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Orena, A. J., Theodore, R. M., & Polka, L. (2015). Language exposure facilitates talker learning prior to language comprehension, even in adults. *Cognition*, 143, 36–40.
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2), 309.
- Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E. G., ... Belin, P. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage*, 119, 164–174. <https://doi.org/10.1016/j.neuroimage.2015.06.050>
- Perrachione, T. K., Chiao, J. Y., & Wong, P. C. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition*, 114(1), 42–55.
- Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science*, 333(6042), 595–595. <https://doi.org/10.1126/science.1207327>
- Perrachione, T. K., Del Tufo, S. N., Winter, R., Murtagh, J., Cyr, A., Chang, P., ... Gabrieli, J. D. E. (2016). Dysfunction of Rapid Neural Adaptation in Dyslexia. *Neuron*, 92(6), 1383–1397. <https://doi.org/10.1016/j.neuron.2016.11.020>
- Perrachione, T. K., Dougherty, S. C., McLaughlin, D. E., & Lember, R. A. (2015). The effects of speech perception and speech comprehension on talker identification. Presented at the Proceedings of the International Congress of Phonetic Sciences XVIII, Glasgow, Scotland.
- 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(6), 1950.
- Perrachione, T. K., Stepp, C. E., Hillman, R. E., & Wong, P. C. (2014). Talker identification across source mechanisms: Experiments with laryngeal and electrolarynx speech. *Journal of Speech, Language, and Hearing Research*, 57(5), 1651–1665.
- Perrachione, T. K., & Wong, P. C. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*, 45(8), 1899–1910.
- Philippon, A. C., Cherryman, J., Bull, R., & Vrij, A. (2007). Earwitness identification performance: The effect of language, target, deliberate strategies and indirect measures. *Applied Cognitive Psychology*, 21(4), 539–550.
- Remez, R. E., Fellowes, J. M., & Nagel, D. S. (2007). On the perception of similarity among talkers. *The Journal of the Acoustical Society of America*, 122(6), 3688–3696. <https://doi.org/10.1121/1.2799903>
- 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(3), 651.
- Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zäske, R. (2014). Speaker perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(1), 15–25. <https://doi.org/10.1002/wcs.1261>
- Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Listening to different speakers: on the time-course of perceptual compensation for vocal-tract characteristics. *Neuropsychologia*, 49(14), 3831–3846. <https://doi.org/10.1016/j.neuropsychologia.2011.09.044>
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388(6640), 381–382. <https://doi.org/10.1038/41102>
- Stevenage, S. V., Clarke, G., & McNeill, A. (2012). The “other-accent” effect in voice recognition. *Journal of Cognitive Psychology*, 24(6), 647–653. <https://doi.org/10.1080/20445911.2012.675321>
- Sullivan, K. P. H., & Kügler, F. (2001). Was the knowledge of the second language or the age difference the determining factor? *Forensic Linguistics*, 8(2), 1–8. <https://doi.org/10.1558/sll.2001.8.2.1>
- Sullivan, K. P. H., & Schlichting, F. (2000). Speaker discrimination in a foreign language: first language environment, second language learners. *Forensic Linguistics*, 7(1), 95–111.
- Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *The Journal of the Acoustical Society of America*, 128(4), 2090–2099. <https://doi.org/10.1121/1.3467771>
- Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America*, 125(6), 3974–3982. <https://doi.org/10.1121/1.3106131>
- Thompson, C. P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, 1(2), 121–131.
- Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia*, 25(5), 829–834.
- 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.
- von Kriegstein, K., & Giraud, A.-L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, 22(2), 948–955. <https://doi.org/10.1016/j.neuroimage.2004.02.020>
- von Kriegstein, K., Smith, D. R. R., Patterson, R. D., Kiebel, S. J., & Griffiths, T. D. (2010). How the human brain recognizes speech in the context of changing speakers. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 30(2), 629–638. <https://doi.org/10.1523/JNEUROSCI.2742-09.2010>
- Wester, M. (2012). Talker discrimination across languages. *Speech Communication*, 54(6), 781–790.
- 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.
- Wong, P. C. M., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience*, 16(7), 1173–1184. <https://doi.org/10.1162/0898929041920522>
- Xie, X., & Myers, E. (2015a). The impact of musical training and tone language experience on talker identification. *The Journal of the Acoustical Society of America*, 137(1), 419–432.
- Xie, X., & Myers, E. B. (2015b). General language ability predicts talker identification. Presented at the Proceedings of the 37th Annual Meeting of the Cognitive Science Society, Austin, TX.
- Zarate, J. M., Tian, X., Woods, K. J. P., & Poeppel, D. (2015). Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Scientific Reports*, 5, 11475. <https://doi.org/10.1038/srep11475>
- Zhang, C., Pugh, K. R., Mencl, W. E., Molfese, P. J., Frost, S. J., Magnuson, J. S., ... Wang, W. S.-Y. (2016). Functionally integrated neural processing of linguistic and talker information: An event-related fMRI and ERP study. *NeuroImage*, 124(Pt A), 536–549. <https://doi.org/10.1016/j.neuroimage.2015.08.064>