Psychonomic Bulletin & Review (2025) 32:941–950
https://doi.org/10.3758/s13423-024-02598-x

BRIEF REPORT

Foreign language talker identification does not generalize to new talkers

Jayden J. Lee[1] · **Jessica A. A. Tin**[1] · **Tyler K. Perrachione**[1]

Accepted: 26 September 2024 / Published online: 23 October 2024

Abstract

Listeners identify talkers less accurately in a foreign language than in their native language, but it remains unclear whether this *language-familiarity effect* arises because listeners (1) simply lack experience identifying foreign-language talkers or (2) gain access to additional talker-specific information during concurrent linguistic processing of talkers' speech. Here, we tested whether sustained practice identifying talkers of an unfamiliar, foreign language could lead to generalizable improvement in learning to identify new talkers speaking that language, even if listeners remained unable to understand the talkers' speech. English-speaking adults with no prior experience with Mandarin practiced learning to identify Mandarin-speaking talkers over four consecutive days and were tested on their ability to generalize their Mandarin talker-identification abilities to new Mandarin-speaking talkers on the fourth day. In a "same-voices" training condition, listeners learned to identify the same talkers for the first 3 days and new talkers on the fourth day; in a "different-voices" condition, listeners learned to identify a different set of voices on each day including the fourth day. Listeners in the same-voices condition showed daily improvement in talker identification across the first 3 days but returned to baseline when trying to learn new talkers on the fourth day, whereas listeners in the different-voices condition showed no improvement across the 4 days. After 4 days, neither group demonstrated generalized improvement in learning new Mandarin-speaking talkers versus their baseline performance. These results suggest that, in the absence of specific linguistic knowledge, listeners are unable to develop generalizable foreign-language talker-identification abilities.

Keywords Speech perception · Talker identification · Voice recognition · Language/memory interactions

Introduction

Comprehending speech and identifying its speaker are fundamentally intervolved processes (Creel & Bregman, 2011; Scott, 2019). Listeners identify talkers more accurately when hearing their native language than an unfamiliar, foreign language – a phenomenon called the *language-familiarity effect* in talker identification. First documented half a century ago (Hollien et al., 1974, 1982), this effect has been widely replicated and is reliable across languages and experimental paradigms (Levi, 2019; Perrachione, 2019). Recent work on voice processing has explored factors affecting native- versus foreign-language talker identification (e.g., Bregman & Creel, 2014; Fecher & Johnson, 2018a; McLaughlin et al., 2019; Perrachione et al., 2011; Xie & Myers, 2015a, 2015b), but there is little theoretical consensus for *why* or *how* listeners are better able to learn to identify talkers in their native language compared to one they do not know.

The language-familiarity effect was initially explored in cognitive psychology studies showing that native English listeners recognized English-speaking talkers better than Spanish-speaking ones (Thompson, 1987), and English and German listeners both recognized talkers more accurately in their native language than the other language (Goggin et al., 1991). These early accounts postulated that listeners develop prototype-based schemata for voices over their lifetimes (what might contemporarily be called a listener's "voice space," e.g., Latinus & Belin, 2011a), such that new voices are distinguished based on how they deviate from the prototype (Latinus & Belin, 2011b). This *exposure-based schematic hypothesis* asserts that voice prototypes reflect listeners' cumulative lifetime exposure to talkers, which is

✉ Tyler K. Perrachione
tkp@bu.edu

[1] Department of Speech, Language, and Hearing Sciences, Boston University, 635 Commonwealth Ave, Boston, MA 02215, USA

typically dominated by experiences in their native language. Listeners' abilities to identify voices may rely on multiple schemata, such as for male or female voices, or for particular accents, and greater experience with a class of voices yields the ability to detect more subtle variations (i.e., better distinguish individuals) within a schema (Goggin et al., 1991, p449).

Under an exposure-based hypothesis, newly encountered talkers' memorability depends on *how* they are distinguished from the prototype: More prototypical voices are distinguished based on the degree to which they differ from the prototype on familiar dimensions of the voice space. However, less prototypical voices (e.g., those speaking a foreign language) are identified less accurately because listeners primarily encode them as "a foreign-language speaker," which maximizes their distinctiveness from the prototype but simultaneously obscures their differences from other speakers of that language. However, as listeners gain further exposure to a new class of voices, such as those speaking a particular foreign language, they should be able to develop a new schema that captures the relevant dimensions on which those voices are distinct from others of the same class and, with more extensive exposure, further refine this schema and thus better learn to identify new individuals of that class (Goggin et al., 1991, p457).

An alternative explanation for the language-familiarity effect comes from studies in psycholinguistics and laboratory phonology, which find that talker identification improves when listeners can access more and higher-level native-language linguistic representations: Listeners identify talkers least accurately in a foreign language, more accurately from nonsense speech that matches the phonological structure of their native language, and most accurately from speech that contains familiar native-language words (Goggin et al., 1991; Perrachione et al., 2015; Xie & Myers, 2015a; Zarate et al., 2015). Hearing talkers say the same speech content also improves talker identification in a familiar language, but not an unfamiliar one (McLaughlin et al., 2015; Perrachione & Wong, 2007), suggesting that linguistic structures like words provide a lens through which listeners can focus on the between-talker phonetic variation that plays a unique role in identifying talkers in a familiar language. Thus, talker identification likely depends on both language-independent and language-dependent information sources (Wester, 2012; Winters et al., 2008), with talker-specific variation in low-level acoustic cues such as average voice pitch and vocal tract length available to listeners in both familiar and unfamiliar languages equally (Fecher & Johnson, 2018a, 2018b; Perrachione et al., 2019; Winters et al., 2008), but talker-specific variation in higher-level linguistic cues, such as idiosyncratic phonetic variation, available primarily or exclusively in a familiar language (Perrachione et al., 2015). Because the low-level acoustic properties of talkers' voices are structured similarly – both within and between talkers – regardless of what language they are speaking (Johnson & Babel, 2023; Lee & Kreiman, 2020; Lee et al., 2022), any language-based advantage in talker identification likely arises from processes or representations that are unavailable to listeners hearing an unfamiliar language. These findings advance a *competence-based linguistic hypothesis*, which posits that superior native-language talker identification abilities result not from the mere lack of experience hearing foreign-language voices, but from access to additional talker-specific information that can only be obtained during simultaneous linguistic processing of speech, such as recognizing language-specific phonetic-phonological variation or encoding higher-level linguistic structures such as words (McLaughlin et al., 2019; Perrachione, 2019; Perrachione et al., 2011, 2015).

Evidence from prior studies examining how variation in language experience influences the language-familiarity effect is equivocal with respect to these two hypotheses, as these studies have relied on natural experiments that conflate exposure to, and competence in, the less-familiar language. For example, English listeners from Connecticut show a greater difference in their ability to learn to identify English- versus French-speaking talkers than do native-English listeners from Montreal (Orena et al., 2015). Hearing French is common in Montreal, and so the *exposure-based account* explains the smaller English versus French language-familiarity effect in Montreal as a result of these listeners' having more experience with French voices and thus a more refined schema that can better encode the individually distinctive features of a newly encountered French-speaking voice. However, despite reportedly not understanding French, the Montreal participants in fact had received "formal classroom instruction in French" (Orena et al., 2015, p37), challenging the idea that French was an "unfamiliar language" for these listeners in the same way it was unfamiliar for the listeners from Connecticut. Consequently, the *competence-based account* therefore posits instead that the smaller English versus French language-familiarity effect among those Montreal English-speakers likely resulted from the additional and exclusive information they could encode for the French talkers during rudimentary linguistic processing of their speech, as small amounts of linguistic information can improve talker identification even in the absence of complete comprehension (Perrachione et al., 2015; Xie & Myers, 2015a; Zarate et al., 2015).

Similarly, the size of the language-familiarity effect in Korean-L1, English-L2 bilinguals depends on when they learned English as children (Bregman & Creel, 2014), and adult Mandarin-English bilinguals have smaller language-familiarity effects than English monolinguals (Perrachione & Wong, 2007). Under an exposure-based account, these listeners' improved talker identification in their non-native

language reflects their greater experience hearing voices speaking that language; whereas under a competence-based account, this improvement results from these listeners' ability to simultaneously process the linguistic content of those talkers' speech. Even for adults, learning a second language can improve the ability to identify talkers speaking in that language (Sullivan & Kugler, 2001). However, second-language learners gain both knowledge of the language and experience hearing different people speak it: More immersive second-language exposure improves listeners' ability to learn to identify talkers of that language (Dougherty & Perrachione, 2016) but likely also improves their language skills. Ultimately, the convolved effects of exposure and linguistic competence in bilinguals and second-language learners have previously been impossible to disentangle, especially in natural experiments and with small sample sizes (Hartshorne et al., 2018). However, prior results do hint that improving foreign-language talker identification as an adult may nonetheless require concurrent linguistic processing: while English monolinguals and Mandarin learners of English as a second language both improve in identifying speakers of their non-native language with training, only the bilinguals close the language-familiarity gap (Perrachione & Wong, 2007).

The inability to disentangle the effects of exposure versus competence in prior studies leaves it unclear whether listeners can refine a schema for identifying talkers in a particular foreign language that will improve their ability to identify new talkers in that language, even in the absence of the ability to understand their speech. Although listeners can improve foreign-language talker identification after extensive practice (Drozdova et al., 2017; Perrachione & Wong, 2007; cf. McLaughlin et al., 2019), this improvement has only ever been shown for the specific voices heard during training. It remains unknown whether foreign-language talker identification training can yield generalizable talker identification abilities (as predicted by the exposure-based hypothesis) or whether the improvements apply only to the talkers used in training (as predicted by the competence-based hypothesis).

In this study, we tested whether listeners can develop generalizable talker identification skills in an unfamiliar foreign language while lacking any linguistic competence in that language. We trained listeners with no prior experience with Mandarin Chinese to identify Mandarin-speaking talkers over a 4-day training paradigm. Participants were divided between two training conditions that manipulated whether they practiced identifying the same slate of talkers every day (*same-voices condition*) or whether new talkers were introduced on each training day (*different-voices condition*). On the fourth day, we tested whether either condition led to generalized improvement in listeners' Mandarin talker-identification abilities compared to baseline performance on the first day. Importantly, the exposure-based hypothesis and the competence-based hypothesis make distinct predictions regarding how listeners should learn and generalize foreign language talker identification abilities resulting from these training conditions, allowing us to adjudicate between these hypotheses based on the pattern of results.

If the language-familiarity effect depends solely on exposure (Goggin et al., 1991; Orena et al., 2015) and not linguistic processing, then training to identify talkers in an unfamiliar language should allow listeners to learn the relevant features that differentiate those voices, incorporate this knowledge into a schema for recognizing talkers in that language, and deploy this knowledge more effectively when learning new voices speaking that language. Importantly, this improvement should occur even though listeners have not learned to speak or understand the foreign language (e.g., they still lack the ability to recognize words or discriminate unfamiliar phonological contrasts). According to the exposure-based hypothesis, we should see day-over-day improvement in talker identification accuracy in both the same- and different-voices conditions (Fig. 1A). While the learning rate may be greater in the same-voices condition (since listeners retain all their knowledge about the voices from the previous day), listeners in the different-voices condition should nonetheless exhibit consistent, albeit slower, learning as they begin to develop a schema that affords more generalizable foreign-language talker-identification skills (Goggin et al., 1991; Lavan et al., 2019). On the fourth day, when listeners in both groups hear a new slate of voices, generalization performance should be best after different-voices training, as having heard a greater diversity of voices should have led to a more refined schema (Goggin et al., 1991) and thus yielded more extensible learning of Mandarin-speaking voices. Importantly though, both groups' training-related exposure to Mandarin-speaking talkers should improve their ability to learn these new voices compared to baseline learning on Day 1.

On the other hand, if the language-familiarity effect has a basis in concurrent linguistic processing (Perrachione et al., 2011, 2019), then the predictions are quite different (Fig. 1B): Day-over-day talker-identification improvement is expected only for participants in the same-voices condition as they continue to learn to identify the language-independent voice qualities that distinguish these talkers (Johnson & Babel, 2023; Winters et al., 2008). However, their generalization abilities to new voices on Day 4 will not differ from their baseline performance on Day 1 because – lacking access to language-specific representations – their ability to learn to identify these new voices will again be limited to learning to distinguish them based on the degree to which these voices differ from each other on low-level acoustic features (Lee et al., 2019; Perrachione & Wong, 2007; Perrachione et al., 2019). Similarly, there will be no day-over-day

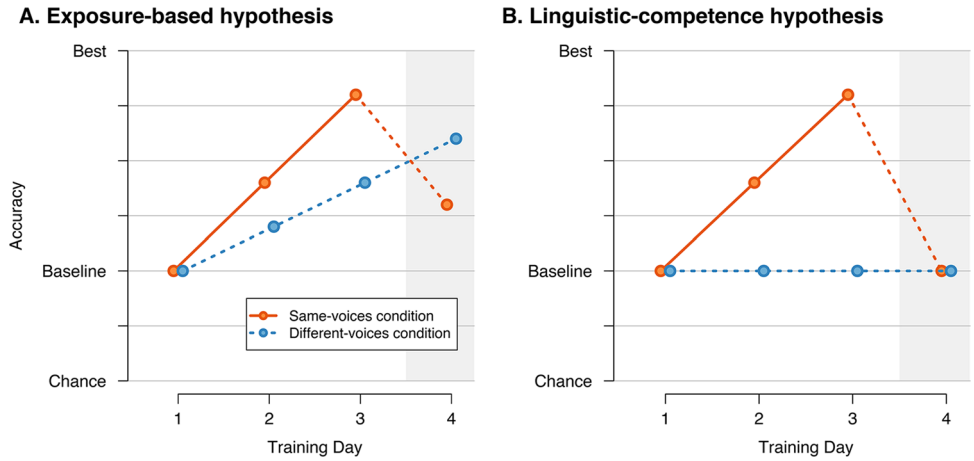

Fig. 1 Contrasting predictions of two theoretical accounts of the language-familiarity effect. These panels show listeners' predicted talker identification accuracy when trained on the same-voices condition (orange lines) or the different-voices condition (blue lines) for Days 1–3, and their ability to generalize to foreign-language talker identification abilities to new voices on Day 4 (shaded background). Solid lines indicate that the same voices were heard on subsequent days; dashed lines indicate that new voices were heard vs. the previous day. **(A)** Under an *exposure-based* account, listeners can improve in their ability to identify talkers in a foreign language as they gain more experience listening to different speakers of that language, despite no change in their inability to process the linguistic content of the speech. (**B**) In contrast, under a *linguistic competence-based* account, listeners can learn to more accurately identify a specific slate of foreign-language talkers as they become increasingly familiar with those talkers' language-independent vocal features; however, this knowledge will be specific to the talkers in that particular set (Lee et al., 2019) and, lacking the ability to simultaneously process the linguistic content of the speech, listeners will not be able to generalize their talker-identification skills to new talkers in the foreign language

improvement (nor evidence of generalization vs. baseline) for listeners in the different-voices condition, whose learning must essentially start afresh with new talkers each day.

Methods

Participants

English-speaking adults ($N = 96$; 73 female, 23 male; age 18–27 years, mean = 19.5 years) who reported no knowledge of Mandarin Chinese nor any history of speech, hearing, or language disorder participated in this study. Two additional subjects completed the experiment, but their data were excluded from analysis due to counterbalancing errors.

Stimuli

Forty phonetically balanced Mandarin Chinese sentences from the *Mandarin Speech Perception Test* (Fu et al., 2011) were recorded by 20 female adult native speakers of Mandarin, all of whom were born and raised in mainland provinces of China and who were, at the time of recording, residing in the USA for undergraduate or graduate study. All recordings were made in a sound-attenuated chamber with a Shure SM58 microphone and Roland Quad Capture sound card sampling at 44.1 kHz and 16 bits; all stimuli were RMS amplitude normalized to 65 dB SPL via Praat (Boersma, 2001). All sentences were seven syllables long and 1.08–2.26 s in duration (mean = 1.52 ± 0.20 s); recording durations varied systematically by talker and sentence content, with systematic speech rate differences across talkers as the largest source of variance. Participants heard ten different sentences on each of the 4 days of training, with these 4 sets of ten sentences counterbalanced across talkers and training condition and phonetically balanced between sets (Fu et al., 2011). The 20 talkers were also divided into four groups of five talkers that were balanced for ease of identification based on pilot testing (McLaughlin et al., 2019). All auditory stimuli were presented binaurally over Sennheiser HD 380 Pro headphones; accompanying written instructions and cartoon avatars representing each voice were presented visually on a computer monitor.

Procedure

All participants underwent four consecutive days of training and test (~30 min per day), during which they learned to identify previously unfamiliar Chinese-speaking talkers. Participants were randomly assigned to one of two training regimens that manipulated the number of unique Chinese-speaking talkers they heard during training. For the *same-voices condition*, participants ($n = 48$) practiced identifying the same slate of five voices during each of the first 3 days of training, followed by a novel set of five voices on the fourth and final day of the experiment (Fig. 2A). (The specific sets

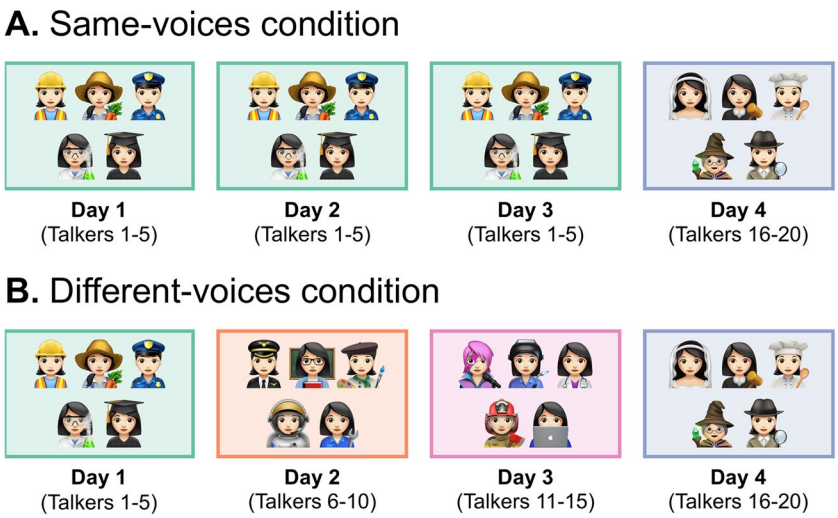

Fig. 2 Paradigm design. Participants were assigned to either **(A)** the *same-voices condition,* where they learned to recognize the same five talkers on Days 1–3 and a new slate of five talkers on Day 4, or (**B**) the *different-voices condition*, where they learned five different talkers on Days 1–4. Face avatars denote unique talkers. (**C**) During the daily training sessions, participants indicated on each trial who they heard talking by selecting the corresponding unique avatar from the display; corrective feedback was given during training trials. The daily test followed the same structure, but no feedback was given

of voices used in Days 1–3 and Day 4 were counterbalanced across listeners from among the four available sets; the sets were determined to be roughly equally identifiable by both naive English and Mandarin listeners through extensive piloting and from the results of prior talker identification studies by our group.) For the *different-voices condition*, participants ($n = 48$) practiced identifying a new set of five voices during each of the four training days (with the order of talkers counterbalanced across listeners), amounting to a total of 20 voices heard over the course of 4 days (Fig. 2B).

Participants learned to identify the five talkers on each day through a training phase consisting of 150 randomized trials (5 talkers × 10 sentences × 3 repetitions). During training, listeners heard a sentence spoken in Mandarin by one of the talkers, and they then indicated the identity of the current talker by selecting the corresponding visual avatar (Fig. 2C). The avatars were highly visually and semantically distinct, and the unique avatar assigned to each voice was consistent across all participants. Corrective feedback was given immediately after each trial of the training phase, indicating if the response was correct or what the correct response should have been. At the end of each session, participants completed a 50-trial test phase (5 talkers × 10 sentences × 1 repetition). The task during test was identical to the training phase, but no feedback was given. The ten sentences used as stimuli on each day were counterbalanced across participants from among the 40 sentences recorded, such that each listener heard ten different sentences on each day. This ensured that listeners were learning abstract representations of the talkers' voices rather than episodic memories for particular stimuli (Lee & Perrachione, 2022; McLaughlin et al., 2019).

Statistical analysis

Data analysis was conducted in *R* (v4.0.3), and the data and analysis code are available online as part of this project's open access dataset (see below). Participants' accuracy on each trial of the test phase was submitted to a generalized linear mixed-effects model for binomial data implemented in the packages *lme4* (v1.1.26) and *lmerTest* (v3.1.3). This model tested for differences in how the two training conditions affected learning on each day of training versus participants' Day 1 baseline. Categorical fixed factors were *condition* (sum-coded contrast: same- vs. different-voices conditions), *training day* (treatment-coded contrast: Day 1 vs. Days 2, 3, or 4), and their interactions. The random factors were by-participant slopes (for the within-subject fixed factor *training day*), correlated by-participant intercepts, by-talker slopes (for the between-subject fixed factor *condition*), and correlated by-talker intercepts (to account for possible differences in the memorability of individual voices). Significance of fixed factors was determined via contrasts on model terms with the criterion of $\alpha = 0.05$ based on the Satterthwaite approximation of the degrees of freedom.

Next, we conducted post hoc contrasts using *emmeans* (v1.6.2) to examine patterns of learning in each of the training conditions. These selected pairwise comparisons assessed different hypotheses: First, we examined how accuracy on each day of training differed from that condition's Day 1 baseline. Second, we compared day-over-day changes in accuracy in each condition. We also compared performance between conditions on Day 4. Holm-Bonferroni

p-value adjustments were used to correct for multiple comparisons.

Results

Participants' test accuracy across all 4 days of training (within-subjects) and the two conditions (between-subjects) is shown in Table 1.

Differences between training conditions

We first examined how learning on each training day differed from Day 1 baseline, and how this was affected differently by the training conditions (Table 2). This analysis revealed a theoretically coherent pattern of significant differences (and lack of differences): Participants' accuracy on Day 1 did not differ between training conditions, suggesting that the groups' talker-identification abilities were matched at baseline. Overall, talker identification accuracy did not improve on Day 2 compared to Day 1, but was significantly better by Day 3. On Day 4 (when a new slate of voices was introduced for learners in the same-voices condition), talker identification accuracy again did not differ from baseline.

However, the pattern of significant interaction effects clarified how the two training conditions had different effects on talker identification learning (Fig. 3A): A significant *condition* × *day* interaction effect revealed that the same-voices group saw greater improvement in accuracy on Day 2 than the different-voices group – a pattern that persisted on Day 3 (vs. Day 1). However, there was no *condition* × *day* interaction between Day 4 and Day 1. Coupled with the lack of an overall Day 4 versus Day 1 difference, this indicates that neither training condition improved listeners' generalizable talker identification abilities at Day 4 compared to baseline.

In a post hoc pairwise test of Day 4 performance (when both groups were learning talkers they had not heard before), there was no effect of *condition* on talker identification accuracy ($\beta = 0.046$, *s.e.* = 0.069, $z = 0.667$, $p = 0.505$).

Improvement but not generalization from same-voices training

Compared to the Day 1 baseline, talker identification accuracy for participants undergoing same-voices training was significantly better on Day 2 and on Day 3 (Fig. 3B, Table 3). However, when a new slate of talkers was introduced on Day 4, accuracy was not different from Day 1. Compared to the previous day, talker identification accuracy for same-voices learners improved significantly on Day 2 and again on Day 3. However, on Day 4, these learners' accuracy was significantly worse than the previous day.

No improvement from different-voices training

Compared to the Day 1 baseline, talker identification accuracy for participants undergoing different-voices training surprisingly declined on Day 2, though not significantly (Fig. 2C, Table 3). However, there were no differences in participants' accuracy on Day 3 or Day 4 compared to baseline. Comparing day-to-day changes in performance, the modest drop in performance from Day 1 to Day 2 was followed by significant improvement from Day 2 to Day 3, as different-voices learners' performance returned to baseline levels. However, these learners saw no further improvement from Day 3 to Day 4.

Table 1 Talker identification accuracy (%, mean ± s.d. across participants) by condition and training day

Condition	Day 1	Day 2	Day 3	Day 4
Same-voices	40.8 ± 14.1	47.8 ± 17.2	51.9 ± 21.1	43.8 ± 15.0
Different-voices	43.7 ± 14.0	40.3 ± 12.7	45.3 ± 11.7	45.8 ± 14.1

Table 2 Comparison of the training conditions

Contrast	β	*s.e*	*z*	*p*	
Intercept	-0.340	0.134	-2.535	< 0.012	*
Condition (Same vs. Different at Day 1)	-0.067	0.065	-1.032	0.302	
Day 2 vs. Day 1 (both conditions)	0.084	0.058	1.432	0.152	
Day 3 vs. Day 1 (both conditions)	0.294	0.072	4.107	< 0.00005	***
Day 4 vs. Day 1 (both conditions)	0.116	0.069	1.680	0.093	
Condition (Same vs. Different) × Day (2 vs. 1)	0.239	0.058	4.094	< 0.00005	***
Condition (Same vs. Different) × Day (3 vs. 1)	0.221	0.072	3.094	0.002	**
Condition (Same vs. Different) × Day (4 vs. 1)	0.021	0.069	0.300	0.764	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

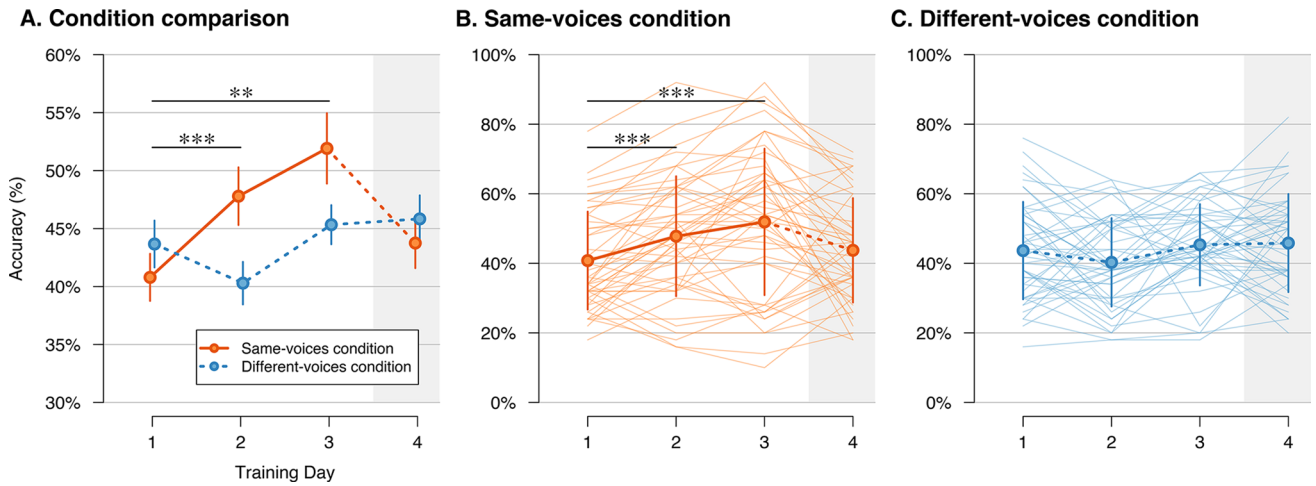

Fig. 3 Talker identification accuracy by condition and training day. **(A)** Mean accuracy in each training condition across days, indicating significant *condition* × *day* (vs. baseline) interactions. Solid lines indicate the same talkers were used as stimuli across days; dashed lines indicate new talkers were heard vs. the previous day. Day 4 is shown with a shaded background to indicate that all listeners heard new talkers on this day. Error bars in this panel indicate standard error of the mean. Note the restricted y-axis range used in this panel. **(B)** Condition mean (bold line) and individual participants' accuracy (light lines) in the *same-voices condition* across training days. Participants were significantly more accurate on Days 2 and 3 than Day 1, but returned to baseline accuracy when new talkers were introduced on Day 4. (**C**) Group mean (bold line) and individual participants' (light lines) accuracy in the *different-voices condition*. Accuracy never exceeded baseline performance on Day 1. Error bars in B, C show the standard deviation across participants. ** $p < 0.01$, *** $p < 0.001$

Table 3 Day-to-day learning effects by condition

	Same-voices condition				Different-voices condition			
Pairwise test	*estimate*	*s.e*	*z*	*P*	*estimate*	*s.e*	*z*	*p*
Change vs. Day 1 baseline								
Day 2 vs. 1	0.322	0.083	3.888	< 0.0003***	-0.155	0.082	-1.892	0.234
Day 3 vs. 1	0.515	0.102	5.056	< 0.0001***	0.072	0.100	0.721	0.982
Day 4 vs. 1	0.136	0.098	1.394	0.163	0.095	0.097	0.098	0.982
Change vs. previous day								
Day 2 vs. 1	0.322	0.083	3.888	< 0.0003***	-0.155	0.082	-1.892	0.234
Day 3 vs. 2	0.192	0.085	2.263	< 0.05*	0.228	0.084	2.719	< 0.05*
Day 4 vs. 3	-0.379	0.095	-3.998	< 0.0003***	0.023	0.093	0.241	0.982

Estimates are log odds ratios. Holm-Bonferroni corrected: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Discussion

When listeners repeatedly practice identifying the same foreign-language talkers, their ability to identify those specific talkers improves. However, this training does not cause listeners to develop talker-identification skills that generalize to better learning of new talkers of that language. When the same-voices trained listeners had to learn to identify a new slate of talkers on the fourth training day, their performance was no better than at baseline. Likewise, listeners who were trained on a new slate of talkers every day showed no day-over-day improvement or overall gain in accuracy. Ultimately, we found that extensive practice identifying Mandarin-speaking talkers without any linguistic knowledge of Mandarin did not lead to generalizable improvements in Mandarin talker-identification ability. These results support a model of talker identification that is not based on simply accumulating exposure to talkers speaking in one language or another (e.g., Goggin et al., 1991; Orena et al., 2015), but rather a model in which talker identification is improved by using the multiple additional sources of talker-specific information that are only available when processing the linguistic content of speech (Perrachione, 2019).

Listeners showed day-over-day improvement in the same-voices condition, but did not generalize this knowledge to new talkers. These listeners were likely learning to recognize systematic variation in the low-level acoustic features that was specific to their slate of talkers, such as differences in mean f_0, voice quality, and vocal tract length (Lee et al., 2019; Perrachione et al., 2019). Because these features are consistent within a particular talker regardless of the language they are speaking, they should have been available to listeners regardless of whether they understood the speaker or not (Johnson & Babel, 2023; Lee & Kreiman, 2020; Winters et al., 2008). However, learning how low-level acoustic features distinguish a particular set of voices may not afford generalization to new slates of voices for two reasons: First, the patterns of distinctive variation among low-level features differ depending on the selection of talkers in a set (Lee et al., 2019). Second, the low-level features that distinguish voices are much more similar across languages than they are different (Johnson & Babel, 2023; Lee & Kreiman, 2020; Perrachione et al., 2019). Thus, learning to identify a small number of foreign-language talkers is unlikely to help listeners learn to discern these language-invariant features over and above what they had already learned after a lifetime of identifying voices in their native language. That is, while training on select Mandarin-speaking voices improved listeners' familiarity with the acoustic nuances distinguishing *those* voices (Lee et al., 2019), such training could not reveal any new features, dimensions, or relationships that would improve Mandarin-talker identification *in general*. Slate-specific, rather than language-general, learning is also evident in the trend for lower talker-identification accuracy on training Day 2 versus Day 1 among listeners undergoing different-voices training. Here, listeners may have tried to use the low-level acoustic relationships they learned for the Day 1 voices to distinguish the new voices on Day 2 – ultimately to their detriment.

Alternatively, though not explicitly claimed by exposure-based hypotheses (Goggin et al., 1991), it is possible that talker identification ability – like language generally (Hartshorne et al., 2018; Werker & Hensch, 2015) – becomes less plastic in adulthood. The language-familiarity effect might reflect talker-identification expertise established during a critical period in early development, when listeners' experiences are dominated by hearing speakers of their native language (Fecher & Johnson, 2018b, 2019, 2021; Johnson et al., 2011). Our participants' inability to develop generalizable talker identification skills in Mandarin may reflect the inflexibility of their foundational English-language voice schema. A critical period for talker prototypes may explain age-of-acquisition gradation in the language-familiarity effect (Bregman & Creel, 2014). However, a critical period for voice schema is not consistent with numerous reports that talker-identification skills in a foreign language also improve after learning that language in adulthood (Sullivan & Kugler, 2001). The language-familiarity effect is smaller among adult second-language learners than monolinguals (Xie & Myers, 2015b), and a week of laboratory training on talker identification in an adult-acquired second language can overcome the language-familiarity effect (Perrachione & Wong, 2007), as does greater immersion in a real-world second-language environment (Dougherty & Perrachione, 2016).

Alternately, exposure to Mandarin-speaking voices for four consecutive days of training may not have been enough to update the English-language voice-recognition schema that listeners had developed over their lifetime of experience. It may be that sustained exposure over many more days may yet allow listeners to develop properly generalizable Mandarin talker-identification abilities. However, we doubt that further foreign-language talker identification training in the absence of gaining any linguistic competence in Mandarin would eventually be able to yield the substantial gains in Mandarin talker-identification abilities necessary to overcome the language-familiarity effect. For one, robust learning and generalization has been found for many other auditory skills after only short-term laboratory training, such as foreign-language consonant (Earle & Myers, 2015), vowel (Brosseau-Lapré et al., 2011), or tone identification (Reetzke et al., 2018), musical interval comparison (Little et al., 2019), vocoded speech perception (Huyck et al., 2017), and even recognition memory for random white noise sequences (Agus et al., 2010). Additionally, bilingual talkers show improvement in second-language talker-identification skills after brief laboratory training (Perrachione & Wong, 2007), so it would be surprising that speakers who are unfamiliar with a language could not also develop such skills, unless this latter group faced some other limitation that constrained learning, such as inability to encode or represent information sources that become available only during concurrent linguistic processing of speech (Lee & Perrachione, 2022; Perrachione, 2019). Indeed, degree of second-language proficiency has been related to listeners' ability to overcome the language-familiarity effect (Bregman & Creel, 2014; Orena et al., 2015; Perrachione & Wong, 2007; Yu et al., 2023). However, it remains an unresolved question exactly what the language-dependent sources of information are, and how listeners access and represent them, in the course of identifying native-language talkers (Abu El Adas & Levi, 2022; Drozdova et al., 2017; Fleming et al., 2014; Goggin et al., 1991; McLaughlin et al., 2019; Perrachione et al., 2015; Perrachione et al., 2019; Xie et al., 2015a; Yu et al., 2023).

Ultimately, the present work clarifies how exposure versus linguistic knowledge underlie listeners' poorer talker identification abilities in an unfamiliar language. While prior interpretations of the language-familiarity effect suggested that it may be affected by the amount of second- or

foreign-language exposure (Bregman & Creel, 2014; Goggin et al., 1991; Orena et al., 2015; Sullivan & Kugler, 2001), no previous work had explicitly tested whether listeners wholly unfamiliar with a foreign language could learn generalizable talker-identification skills in that language. Here, we found no evidence of improved talker identification when controlling foreign-language skills. Listeners could learn to better distinguish a specific slate of foreign-language talkers with training, but could not generalize this improvement when learning new talkers. Likewise, training to identify new talkers every day yielded no improvements over baseline performance. Taken together, these results suggest that a general enhancement in talker-identification abilities, such as that seen in the language-familiarity effect, likely requires access to additional, language-dependent sources of talker-specific information, which only become available during concurrent linguistic processing of talkers' speech.

Acknowledgements We thank Deirdre McLaughlin, Soyoung Jung, Yinuo Liu, Kristina Furbeck, and Yaminah Carter for their assistance.

Funding This work was supported by the National Institute on Deafness and Other Communications Disorders (NIDCD) of the National Institutes of Health under awards T32DC013017 (to Christopher Moore) and R01DC004545 (to Gerald Kidd).

Data availability The raw data, stimuli, and stimulus delivery scripts from this project are available online via our institutional repository: https://open.bu.edu/handle/2144/16460

Code availability The statistical analysis code from this project is available online via our institutional repository: https://open.bu.edu/handle/2144/16460

Declarations

Ethics approval This study was approved and overseen by the Institutional Review Board at Boston University.

Consent to participate All participants provided informed, written consent prior to undertaking the experiment.

Consent for publication All participants provided informed, written consent to publish their deidentified data.

Competing interests The authors declare no conflicts of interest or competing interests.

References

Abu El Adas, S., & Levi, S. V. (2022). Phonotactic and lexical factors in talker discrimination and identification. *Attention, Perception, & Psychophysics, 84*, 1788–1804.

Agus, T. R., Simon, J. T., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron, 66*(4), 610–618.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*, 341–345.

Bregman, M. R., & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition, 130*(1), 85–95.

Brosseau-Lapré, F., Rvachew, S., Clayards, M., & Dickson, D. (2011). Stimulus variability and perceptual learning of nonnative vowel categories. *Applied Psycholinguistics, 34*(3), 419–441.

Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Language & Linguistics Compass, 5*, 190–204.

Dougherty, S. C., & Perrachione, T. K. (2016). The language-familiarity effect in talker identification by highly proficient bilinguals depends on second-language immersion. *Journal of the Acoustical Society of America, 139*, 2161.

Drozdova, P., van Hout, R., & Scharenborg, O. (2017). L2 voice recognition: The role of speaker-, listener-, and stimulus-related factors. *Journal of the Acoustical Society of America, 142*(5), 3058–3068.

Earle, F.S., & Myers, E.B. (2015). Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds. *Journal of the Acoustical Society of America, 137,* EL91.

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

Fecher, N., & Johnson, E.K. (2018b). The native-language benefit for talker identification is robust in 7.5-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(12), 1911–20.

Fecher, N., & Johnson, E.K. (2019). By 4.5 months, linguistic experience already affects infants' talker processing abilities. *Child Development, 90*(5), 1535–43.

Fecher, N., & Johnson, E. K. (2021). Developmental improvements in talker recognition are specific to the native language. *Journal of Experimental Child Psychology, 202*, 104991.

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.

Fu, Q.J., Zhu, M., & Wang, X. (2011). Development and validation of the Mandarin speech perception test. *Journal of the Acoustical Society of America, 129*, EL267–73.

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.

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition, 177*, 263–277.

Hollien, H., Majewski, W., & Hollien, P. A. (1974). Perceptual identification of voices under normal, stress, and disguised speaking conditions. *Journal of the Acoustical Society of America, 56*, S53. https://doi.org/10.1121/1.1914230

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.

Huyck, J. J., Smith, R. H., Hawkins, S., & Johnsrude, I. S. (2017). Generalization of perceptual learning of degraded speech across talkers. *Journal of Speech, Language, and Hearing Research, 60*(11), 3334–3341.

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. A., & Babel, M. (2023). The structure of acoustic voice variation in bilingual speech. *Journal of the Acoustical Society of America, 153*, 3221–3238.

Latinus, M., & Belin, P. (2011a). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology, 2*, 175.

Latinus, M., & Belin, P. (2011b). Human voice perception. *Current Biology, 21*, R143–R145.
Lavan, N., Knight, S., Hazan, V., & McGettigan, C. (2019). The effects of high variability training on voice identity learning. *Cognition, 193*, 104026.
Lee, J. J., & Perrachione, T. K. (2022). Implicit and explicit learning in talker identification. *Attention, Perception, & Psychophysics, 84*, 2002–2015.
Lee, Y., Keating, P., & Kreiman, J. (2019). Acoustic voice variation within and between speakers. *Journal of the Acoustical Society of America, 146*, 1568.
Lee, Y., & Kreiman, J. (2020). Language effects on acoustic voice variation within and between talkers. *Journal of the Acoustical Society of America, 148*, 2473.
Levi, S. V. (2019). Methodological considerations for interpreting the Language Familiarity Effect in talker processing. *Wires Cogn Sci, 10*, e1483.
Little, D. F., Cheng, H. H., & Wright, B. A. (2019). Inducing musical-interval learning by combining task practice with periods of stimulus exposure alone. *Attention, Perception, & Psychophysics, 81*, 344–357.
McLaughlin, D. E., Carter, Y. D., Cheng, C. C., & Perrachione, T. K. (2019). Hierarchical contributions of linguistic knowledge to talker identification: Phonological versus lexical familiarity. *Attention, Perception & Psychophysics, 81*, 1088–1107.
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. *18th International Congress of Phonetic Science*s (Glasgow, August 2015).
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.
Perrachione, T.K. (2019). Speaker recognition across languages. In S. Früholz & P. Belin (Eds.), *The Oxford handbook of voice perception*. Oxford: Oxford University Press. Available online: https://hdl.handle.net/2144/23877
Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science, 333*, 595.
Perrachione, T.K., Dougherty, S.C., McLaughlin, D.E., & Lember, R.A. (2015). The effects of speech perception and speech comprehension on talker identification. *18th International Congress of Phonetic Sciences* (Glasgow, August 2015).
Perrachione, T. K., Furbeck, K. T., & Thurston, E. J. (2019). Acoustic and linguistic factors affecting perceptual dissimilarity judgments of voices. *Journal of the Acoustical Society of America, 146*(5), 3384–3399.
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.
Reetzke, R., Xie, Z., Llanos, F., & Chandrasekaran, B. (2018). Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood. *Current Biology, 28*(9), 1419–1427.
Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. *Science, 366*, 58–62.
Sullivan, F., & Kügler, F. (2001). Was the knowledge of the second language or the age difference the determining factor? *International Journal of Speech Language and the Law, 8*, 1–8.
Thompson,. (1987). A language effect in voice identification. *Applied Cognitive Psychology, 1*(2), 121–131.
Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology, 66*, 173–196.
Wester, M. (2012). Talker discrimination across languages. *Speech Communication, 54*, 781–790.
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*, 4524–4538.
Xie, X., & Myers, E.B. (2015a). General language ability predicts talker identification. In Noelle, D.C., Dale, R., Warlaumont, A.S., Yoshimi, J., Matlock, T., Jennings, C.D., & Maglio, P.P. (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, (pp. 2697–2702). Austin, TX: Cognitive Science Society.
Xie, X., & Myers, E. B. (2015b). The impact of musical training and tone language experience on talker identification. *Journal of the Acoustical Society of America, 123*, 4524–4538.
Yu, K., Zhou, Y., Zhang, L., Li, L., Li, P., & Wang, R. (2023). How different types of linguistic information impact voice perception: Evidence from the language-familiarity effect. *Language & Speech, 66*, 1007–1029.
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.