Published in final edited form as:

Biling (Camb Engl). 2018 May ; 21(3): 523–536. doi:10.1017/S1366728917000153.

Another bilingual advantage? Perception of talker-voice information

SUSANNAHV. LEVI

New York University

Abstract

A bilingual advantage has been found in both cognitive and social tasks. In the current study, we examine whether there is a bilingual advantage in how children process information about who is talking (talker-voice information). Younger and older groups of monolingual and bilingual children completed the following talker-voice tasks with bilingual speakers: a discrimination task in English and German (an unfamiliar language), and a talker-voice learning task in which they learned to identify the voices of three unfamiliar speakers in English. Results revealed effects of age and bilingual status. Across the tasks, older children performed better than younger children and bilingual children performed better than monolingual children. Improved talker-voice processing by the bilingual children suggests that a bilingual advantage exists in a social aspect of speech perception, where the focus is not on processing the linguistic information in the signal, but instead on processing information about who is talking.

Keywords

speech perception; talker processing; bilinguals

Introduction

Speech is a complex acoustic signal that simultaneously carries information about what is being said (linguistic information) and about who is saying it (talker information). While it is possible to process these two dimensions separately, numerous studies have shown that they interact during speech perception (e.g., Cutler, Andics & Fang, 2011; Green, Tomiak & Kuhl, 1997; Mullennix & Pisoni, 1990), showing the importance of exploring how people process talker information for understanding the larger questions related to spoken language processing. In this study, we test whether being bilingual is an advantage when processing information about who is talking. In particular, we investigate the perception of talker-voice information (who is talking) in both a familiar language (English) and in an unfamiliar language (German) to see whether a bilingual advantage exists and whether it extends to unfamiliar languages. By including talker-voice perception in an unfamiliar language, we are able to rule out the possibility that a bilingual advantage in English, a familiar language, is

due to being exposed to a wider variety of speech samples in English (presumably both native speakers of English and foreign-accented English).

The majority of people in the world grow up in multilingual environments, whether that means that they speak more than one language or are simply exposed to speakers of other languages (Hamers & Blanc, 2000). In recent years, there has been an increase in interest in how bilingualism or exposure to multiple languages affects cognitive and social processing. The majority of studies of a bilingual advantage have focused on executive functioning and cognitive control, and a few studies have examined the social benefits of bilingualism.

In terms of executive function, several studies have found a BILINGUAL ADVANTAGE where bilinguals respond faster and are better at inhibiting irrelevant information or shifting to a new task. This advantage has been found in children (Bialystok, 1999; Bialystok & Martin, 2004; Carlson & Meltzoff, 2008), older adults (Bialystok, Craik, Klein & Viswanathan, 2004; Gold, Kim, Johnson, Kryscio & Smith, 2013), and young adults who are at their peak of executive function / cognitive control ability (Costa, Hernández & Sebastián-Gallés, 2008; Friesen, Latman, Calvo & Bialystok, 2015). Bilingualism may even act a buffer for other factors that are known to negatively affect executive function skills, such as lower socio-economic status (SES) (Carlson & Meltzoff, 2008) or normal aging (Bialystok et al., 2004; Gold et al., 2013). More recently, Incera and McLennan (2015) used a mouse tracker to examine online processing in a Stroop task (Stroop, 1935) and found that bilinguals take longer to initiate their response, but then move faster to their target, a pattern that the authors argue matches the performance of experts in other domains (e.g., baseball players who take longer to initiate a swing of the bat but then swing faster). The benefits of bilingualism on executive function have not gone without criticism, however. Several articles argue against the bilingual advantage (de Bruin, Treccani & Della Sala, 2015; Paap & Greenberg, 2013; Paap, Johnson & Sawi, 2014; Paap & Sawi, 2014; Paap, Sawi, Dalibar, Darrow & Johnson, 2014), but see (Ansaldò, Ghazi-Saidi & Adrover-Roig, 2015; Arredondo, Hu, Satterfield & Kovelman, 2015; Saidi & Ansaldò, 2015) for a rebuttal.

In addition to studies showing a bilingual advantage in cognitive control, studies have also found a bilingual advantage in social perception. Greenberg, Bellana, and Bialystok (2013) asked school-age children to shift their perspective to that of another person (cartoon owl) in a visual task. Children saw a cartoon owl looking at four blocks and were asked to identify what the blocks looked like from the perspective of the owl, who differed in orientation from the child by 90°, 180°, or 270°. Bilingual children were significantly better at selecting the perspective of the other person. In a similar study using a visual perspective-taking task, Fan, Liberman, Keysar, and Kinzler (2015) tested whether being exposed to other languages in the environment – but not actually being bilingual – was sufficient to improve performance. They found that even children who were not bilingual but who were exposed to other languages in their community were better than monolinguals at interpreting the visual perspective of another person. Furthermore, they found no difference in performance between bilinguals and those children who were exposed to other languages in the ambient environment, suggesting that exposure is sufficient to generate a social advantage in visual perception

Processing who is talking is an important social component of communication and begins to develop even before birth. Differentiating between the mother's voice and an unfamiliar female has been found in fetuses (Kisilevsky, Hains, Lee, Xie, Huang, Ye, Zhang & Wang, 2003), newborns (DeCasper & Fifer, 1980), and infants in their first year of life (Purhonen, Kilpeläinen-Lees, Valkonen-Korhonen, Karhu & Lehtonen, 2004, 2005). In the first year, infants also begin to differentiate between unfamiliar voices (DeCasper & Prescott, 1984; Johnson, Westrek, Nazzi & Cutler, 2011). The ability to identify and discriminate voices continues to develop through elementary school and adolescence (Bartholomeus, 1973; Creel & Jimenez, 2012; Levi & Schwartz, 2013; Mann, Diamond & Carey, 1979; Moher, Feigenson & Halberda, 2010; Spence, Rollins & Jerger, 2002).

The close ties between linguistic processing and talker processing have led to studies exploring what determines performance on a talker-voice processing task. Because a salient feature of talker identity is a speaker's fundamental frequency (vocal pitch), Xie and Myers (2015) tested talker identification in native-English listeners with and without musical training. They found that the musicians performed better than non-musicians in an unfamiliar language (Spanish), suggesting shared resources between linguistic and non-linguistic perception. In another study, Krizman, Marian, Shook, Skoe, and Kraus (2012) found that bilinguals have more robust pitch processing than monolinguals.

Several recent studies have examined how bilinguals and people exposed to a second language process talker information. Bregman and Creel (2014) tested how age of acquisition of a second language affects learning to identify talkers in both the first and second language. They trained listeners to criterion over several learning sessions. Listeners heard sentences produced by both English and Korean talkers. They found that listeners learn talkers faster in their L1, similar to previous studies that have found that listeners are better at processing talker information in their native language (Goggin, Thompson, Strube & Simental, 1991; Goldstein, Knight, Bailis & Conover, 1981; Hollien, Majewski & Doherty, 1982; Köster & Schiller, 1997; Levi & Schwartz, 2013; Perrachione, Del Tufo & Gabrieli, 2011; Perrachione, Dougherty, McLaughlin & Lember, 2015; Perrachione & Wong, 2007; Schiller & Köster, 1996; Thompson, 1987; Winters, Levi & Pisoni, 2008; Zarate, Tian, Woods & Poeppel, 2015), even when the stimuli are time-reversed (Fleming, Giordano, Caldara & Belin, 2014). In addition, Bregman and Creel (2014) found that bilinguals who had learned the L2 early were faster at learning talkers in the L2 than were bilinguals who had learned the L2 later in life. They argue that the better performance by early bilinguals suggests a close link between speech representations and talker representations.

In another study, Orena, Theodore, and Polka (2015) tested how adult listeners learned to identify talkers. They compared English monolinguals who were living in Montréal and thus had extensive exposure to French and those living in Connecticut who did not have frequent exposure with French. Even though both groups of adults did not speak French fluently, those who had extensive exposure to French were better at learning to identify voices in that language.

In Xie and Myers (2015), they not only examined the effect of musical training on the perception of talker information (see above), but also compared performance in adult listeners who were monolingual, native speakers of English with those who were Mandarin–English bilinguals. In contrast to Bregman and Creel (2014) and to Orena et al. (2015), Xie and Myers did not find a difference in talker processing between these two groups of listeners. We return to this point in the discussion.

In the current study, we expand on these data by testing whether a bilingual advantage in processing talker information is present in children. The current study is a strong test of the benefits of bilingualism because it looks for differences in both a language familiar to all participants (English) and a language that is unfamiliar to all participants (German). The talkers used here are German(L1)–English(L2) bilinguals and thus have produced English with a foreign accent. Bilingual speakers – rather than two different sets of monolingual speakers – were used to minimize differences in discriminability in the two different languages. If a bilingual advantage is present even in an unfamiliar language (German), it would indicate that the ability is not limited to experience listening to foreign-accented talkers in the dominant language. We also extend the previous findings by testing talker processing across different tasks: discrimination, learning, and identification of talkers. An additional extension in the current study is testing children versus adults. This allows us to examine the development of talker processing and also allows us to better control language exposure of the monolingual group. For example, in Xie and Myers (2015) a large portion of the monolingual English, college-attending listeners had studied Spanish, the other UNFAMILIAR language, and no mention of other languages studied by the college students used as listeners was mentioned. We return to this point in the discussion.

We expect bilinguals to perform better than monolinguals for a variety of reasons. First, bilinguals are more likely to have heard the same person speaking different languages so they have experience perceiving the same voice in multiple languages¹. Second, knowing who is talking is relevant for interpreting both linguistic information and also social information, as who is talking may affect how to react in a particular social setting, such as weighing how much trust to put into someone's statement. Given that previous studies have found a bilingual advantage in other tasks tapping social processing (see above), we expect that bilinguals will be better at processing information about who is talking. Third, even though the bilinguals in the current study do not have experience listening to the particular type of foreign accented speech used here (German L1 talkers), they are likely to have more experience listening to foreign-accented speech than do monolingual children. Finally, given that previous research has found that bilinguals have more robust pitch perception and that pitch perception is an important component of processing talker-voice information, we expect bilinguals to perform better than monolinguals.

¹We recognize that not all bilingual children have heard the same person in multiple languages, such as in the case where a grandparent speaks the home language but not the language of local community.

Methods

Participants

Forty-one children (22 monolinguals, 19 bilinguals) were selected from a dataset of 97 subjects who had participated in a larger study. All children were attending school in New York City with English as the primary language of instruction. The 97 children all started a multiple-session study (7–12 sessions) that included the experimental tasks described below, as well as other experimental tasks which will not be described. In addition to the experimental tasks, children were given subtests from the Clinical Evaluation of Language Fundamentals-4 (CELF) (Semel, Wiig & Secord, 2003) and were also given the Test Of Nonverbal Intelligence-3 (TONI) (Brown, Sherbenou & Johnsen, 1997). The CELF-4 subtests generate a composite “Core Language” score that is normed with a mean of 100 and a standard deviation of 15. The TONI is a test of nonverbal cognition using black and white line drawings and tests problem solving and pattern matching. It is also normed to 100 with a standard deviation of 15. Parents were asked to fill out a questionnaire that included five questions about the child’s language use ((i) What language(s) does your child speak?, (ii) What language does your child use most?, (iii) What language does your child use best?, (iv) What language does your child prefer?, (v) What language does your child understand best?), questions about the language(s) spoken by other family members living with the child, and questions about the languages spoken by other caregivers (e.g., nanny, daycare provider).

First, children from the larger subject pool were categorized as monolingual, bilingual, or other as follows: Monolingual children were those for whom the parent report indicated only “English” for the five language questions listed above and indicated that there was no caregiver or family member living with the child who spoke another language (n=52). Children were categorized as bilingual if there was a family member living with the child who spoke another language or if the parent report indicated an additional language for any of the language questions listed above (n=40). Thus, the bilingual group includes both children who speak/understand another language and children who are exposed to another language on a daily basis. This latter set of children who are merely exposed to another language is similar to the groups used in the social processing (Fan et al., 2015) and talker processing tasks (Orena et al., 2015) discussed above. The remaining five children were monolingual, but had an additional caregiver not living with the child who spoke another language, or had incomplete language background questionnaires. These five children were excluded at this point. None of the participants in the monolingual or bilingual groups had any knowledge of German (the second language used in the perception task).

The monolingual and bilingual groups were then divided into two groups for age, with a younger group that was nine years and younger and an older group that was 10 years and older. In addition, children were only included if they had completed the CELF-4 and TONI-3 and scored 85 or above on these tests (greater than one standard deviation below the mean). Children were eliminated if they failed the hearing screening or if the parent report indicated (i) a failed hearing screening, (ii) a diagnosis of ADHD, or (iii) a diagnosis of Autism/Asperger’s. These exclusion criteria ensured that children who were included in the

study had language and nonverbal skills in the typical range for children their age. These additional subject criteria resulted in the following four groups of children: Younger monolingual (n=12), Older monolingual (n=10), Younger bilingual (n=11), Older bilingual (n=8). The children in the bilingual groups were exposed to the following languages: Spanish (n=11), Russian (n=2), Hindi (n=1), Greek (n=1), Italian (n=1), Tagalog (n=1), Japanese (n=1), Fulani (n=1). Two of the Spanish speaking children also had exposure to an additional language (Catalan and French). None of these languages are tone languages.

In addition to age, language, and nonverbal IQ measures, we also collected information about socioeconomic status (SES). We used mother's education level as a proxy for SES (Ensminger & Fothergill, 2003), and divided education into an 8-point scale following Buac and Kaushanskaya (2014) as follows: 1, less than high school; 2, high school; 3, one year of college; 4, two years of college / associate's degree; 5, three years of college; 6, four years of college; 7, master's degree; 8, Ph.D., J.D., M.D. The mean SES for the monolingual group indicates an average three years of college and for the bilingual group indicates two-to-three years of college. Information about the participant groups is in Table 1.

Three separate two-way ANOVAs were run with Group (monolingual, bilingual) and Age (younger, older) as between-subjects factors for the Core Language composite of the CELF-4, the TONI-3 score (scaled for age), and SES. No main effects or interactions were significant. See appendix A for the full ANOVA results.

Stimuli

Three female bilingual German(L1)–English(L2) speakers were recorded producing 360 monosyllabic CVC words in both English and German (see Winters et al., 2008 for details of the recording methods). The three bilingual speakers used in the current study were highly intelligible and selected from a larger group of bilinguals (Levi, Winters & Pisoni, 2011) – see Table 2. Additionally, the three speakers were selected as having relatively different average fundamental frequency across productions. Fundamental frequency, whose perceptual correlate is vocal pitch, is a salient feature used to distinguish speakers (Van Dommelen, 1987). The current study is designed to use the same talkers in the English (familiar) and German (unfamiliar) conditions to control for the fact that some speakers sound more or less similar to each other and to limit paralinguistic (e.g., acoustic) differences across the talkers. Thus, in the current study, the talkers speak English with a slight foreign accent (Levi et al., 2011).

Procedure

Participants completed three experimental tasks: two separate self-paced talker discrimination experiments in English and German on the first day of the study and a five-day talker-voice learning task which was completed on later days of the experiment. All experiments were conducted on a Panasonic Toughbook CF-52 laptop with a touch screen running Windows XP. All experiments were created with E-Prime 2.0 Professional (Schneider, Eschman & Zuccolotto, 2007). Children sat at a desk or table facing the computer screen. Stimuli were presented binaurally over Sennheiser HD-280 circumaural

headphones. Children were tested in a quiet room either at school, in their home, or in the Department of Communicative Sciences and Disorders at New York University.

For the two talker discrimination tasks, children were randomly assigned to completing the English talker discrimination task first or second. The English talker discrimination task was completed first for 7/12 of the younger monolingual group, for 6/10 for the older monolingual group, for 5/11 for the younger bilingual group and for 3/8 of the older bilingual group. Children were informed whether they would be listening to English or a language other than English, but were not told what the other language would be. Participants heard pairs of distinct words separated by a 750 ms inter-stimulus interval. Participants were asked to determine whether the words were spoken by the same talker or by two different talkers. In all trials, different lexical items were used, requiring listeners to process the talker information at an abstract level (e.g., Talker 1 *book*, Talker 2 *room*; Talker 1 *cat*, Talker 1 *name*). Participants listened to one practice block with 12 trials followed by two experimental blocks of 24 trials. Half of the trials were same-talker pairs and half were different-talker pairs.

Children also completed a talker-voice learning task in English. For this task, children completed five days of talker-voice training in which they learned to identify the voices of three unfamiliar talkers (the same talkers used in the discrimination task), represented as cartoon-like characters on the computer screen. Each day of talker-voice learning consisted of two sessions with feedback (“training”) and one test session without feedback (“test”). Children were instructed that they would hear a single English word and have to decide which of three characters produced the word by tapping the screen (Figure 1). During each session, children first had a familiarization phase in which they heard the same four words produced by all three talkers twice. Only the image of the actual character/talker appeared on the screen during the familiarization phase. After familiarization, children completed 30 trials (10 words produced by all three talkers).

During this phase, children heard a word and had to select which character had spoken the word. In the two feedback sessions, after their response, children received two forms of feedback: first they were shown a smiley or frowny face to indicate their accuracy and then they heard the word again while the image of the correct character/talker appeared on the screen. An outline of a single trial is provided in Figure 2. Each day children completed the exact same feedback session twice and then completed the test session with different lexical items. Test sessions had the same format as the training sessions except no feedback was provided. Thus, over the course of talker-voice learning, children were presented with 104 distinct lexical items produced by all three talkers as follows: same 4 lexical items during familiarization, 10 lexical items × 5 days for the feedback sessions, and 10 lexical items × 5 days for the test sessions without feedback.

Results

Talker discrimination in English and German

To assess performance on the two talker discrimination tasks, a logit mixed-effects model (Baayen, 2008; Baayen, Davidson & Bates, 2008; Jaeger, 2008) was fit to the English and

German data using the *glmer()* function with the binomial response family (Bates, Maechler, Bolker & Walker, 2010) in R (<http://www.r-project.org/>). The built-in p-value calculations from the *glmer()* function were used. A model was fit to the data with fixed effects for Group (monolingual, bilingual), Language (English, German), and Age (in months which was scaled), along with random intercepts by Subjects and random slopes for Language by Subject.

The model revealed a significant effect of Group (estimate = $-.364$, standard error = $.165$, z -value = -2.20 , $p = .027$), of Age (estimate = $.322$, standard error = $.130$, z -value = 2.47 , $p = .013$), and of Language (estimate = $-.290$, standard error = $.113$, z -value = -2.54 , $p = .010$), illustrated in Figure 3. None of the interactions reached significance (all $p > .59$). These main effects indicate that overall, bilinguals performed better than monolinguals, that older children performed better than younger children, and that children performed better in the English condition than in the German condition. The full table of results can be found in appendix B.

Because a primary question is whether the bilingual group would perform better than the monolingual group in the German condition, where none of the children have prior exposure to German, we constructed separate models for the two language conditions with Group and Age as fixed effects and with Subject as a random effect. For the German AX task, the model revealed a significant effect of Group (estimate = $-.378$, standard error = $.179$, z -value = -2.11 , $p = .034$) and of Age (estimate = $.338$, standard error = $.142$, z -value = 2.68 , $p = .007$). Similar results were found for the English AX task, where the model revealed a significant effect of Group (estimate = $-.361$, standard error = $.162$, z -value = -2.22 , $p = .026$) and of Age (estimate = $.315$, standard error = $.128$, z -value = 2.46 , $p = .013$).

Talker-voice learning

Two different analyses were conducted on the talk-voice learning data. First, we examined the number of days it took for children to reach 70% correct talker-voice identification in Training sessions (averaging performance for the two feedback sessions completed per day) and in the Test sessions where listeners received no feedback. This criterion was used because 70% has been used in other studies of voice learning to differentiate good learners from poor learners (Levi et al., 2011; Nygaard & Pisoni, 1998). Because a single data point is used per subject (number of days to criterion), a simple two-way ANOVA with Group (monolingual, bilingual) and Age (younger, older) as between-subjects factors conducted on these data. The ANOVA for the number of days to criterion for training sessions with feedback revealed a significant main effect of Group ($F(1,37) = 5.16$, $p = .028$, $\eta_p^2 = .12$).

Neither the main effect of Age ($F(1,37) = 1.71$, $p = .198$) nor the interaction ($F(1,37) = 0.57$, $p = .45$) were significant. As seen in Figure 4, bilingual children took fewer days to reach the 70% criterion than monolingual children.

Similarly, the ANOVA for the number of days to reach criterion on the test condition (no feedback) revealed a main effect of Group ($F(1,37) = 4.34$, $p = .044$, $\eta_p^2 = .105$). Neither the main effect of Age ($F(1,37) = 1.22$, $p = .27$) nor the interaction ($F(1,37) = .03$, $p = .85$) were

significant. As in the training sessions, bilingual children took fewer days to reach the 70% criterion than the monolingual children (see Figure 5).

Second, separate models were fit to the Training data (with feedback) and the Test data (no feedback) with fixed effects for Group (monolingual, bilingual), and Age (in months which was scaled), Day (1–5), and the Group by Age interaction, along with random intercepts by Subjects and random slopes for Day by Subject. For the Training data, a fixed effect for Session (1, 2) was also included.

For the training data, the model revealed a significant effect of Day (estimate = .322, standard error = .035, z -value = 9.15, $p < .001$) and a nearly significant effect of Group (estimate = $-.325$, standard error = .166, z -value = -1.95 , $p = .051$), illustrated in Figure 6. None of the remaining main effects nor the interaction reached significance (all $p > .13$). The full table of results can be found in Appendix B. These main effects indicate that, overall, bilinguals performed better than monolinguals and that children performed better with additional days of training. The effect of Day is clearly visible in Figure 6, where talker identification accuracy gradually improves across the five days of training. The effect of Group is also visible, where the bilinguals (dashed lines) are more accurate than the monolinguals (solid lines).

For the Test data, the model revealed a significant effect of Day (estimate = .244, standard error = .045, z -value = 5.36, $p < .001$) and of Age (estimate = .329, standard error = .151, z -value = 2.17, $p = .029$). None of the remaining main effects or the interaction reached significance (all $p > .09$). The full table of results can be found in Appendix B. The effect of Day is clearly visible in Figure 7, where talker identification accuracy gradually improves across the five days of training. As has been found in previous studies, both adults (Winters et al., 2008) and children (Levi, 2014) perform significantly worse on talker identification when no feedback is provided, ultimately reducing the range of performance.

Discussion

Several findings emerged from the experiments presented here. First, the current study confirmed the findings of many previous studies showing that processing talker-voice information in the speech signal improves across development. Second, the results of the current study show that a bilingual advantage exists for talker-voice perception in a familiar language (English) and in an unfamiliar language (German). In the familiar language, bilingual children were better at discriminating voices, better at learning to identify the voices, and also faster at learning the voices (days to criterion). In the unfamiliar language, bilingual children were better at discriminating voices. Third, the current study revealed a language familiarity effect, where children discriminated voices better in the familiar language (English) than in an unfamiliar language (German).

Before delving into the mechanisms that underlie these results, it is important to point out some differences between the current study and other studies that have examined talker-voice processing. The current study tested two different experimental paradigms, an AX discrimination task and a talker-voice training task that took place over five days of training.

As will be shown below, most other studies have used a single task and results are not always consistent across different studies. Early studies of talker-voice processing used a voice line-up task where listeners are exposed to a voice and then have to pick it out from an array of voices (Goggin et al., 1991; Goldstein et al., 1981; Stevenage, Clarke & McNeill, 2012; Sullivan & Schlichting, 2000; Thompson, 1987). Other experimental paradigms include familiarization with a voice followed by trials in which the listener has to respond with whether they match the target voice (Köster & Schiller, 1997; Schiller & Köster, 1996), brief voice familiarization with multiple talkers followed by an identification task (Bregman & Creel, 2014; Orena et al., 2015; Perrachione, Chiao & Wong, 2010; Perrachione et al., 2011; Perrachione et al., 2015; Perrachione & Wong, 2007; Xie & Myers, 2015; Zarate et al., 2015), extensive training and identification of talkers (Nygaard & Pisoni, 1998; Winters et al., 2008), voice similarity rating where listeners hear two speech samples and rate the likelihood that they were produced by the same speaker (Fleming et al., 2014), and discrimination tasks (Levi & Schwartz, 2013; Winters et al., 2008).

Another difference between the current study and many previous studies is in the type of stimuli that were used. Most other studies of talker-voice perception use sentence-length utterances (Bregman & Creel, 2014; Fleming et al., 2014; Orena et al., 2015; Perrachione et al., 2010). The study by Zarate et al. (2015) presented several polysyllabic words on a single trial. In the experiments presented here, listeners were provided much less information from each speaker on a given trial because stimuli were only monosyllabic CVC words. Previous research has shown that the type of talker information available from sentence-length utterances and word-length utterances is not the same. In a task looking at transfer between talker-voice training and word recognition of familiar talkers, Nygaard and Pisoni (1998) demonstrated that learning a voice from word-length stimuli results in better word recognition in noise, but learning a voice from sentence-length utterances does not result in better word recognition in noise. They argue that the phonetic information available in the sentence length utterances is different from the word-length utterances. This is not surprising given that sentences have more complex intonation patterns than words and thus provide different information about the speaker. We now discuss each of the findings in turn.

Age

The effect of age found in the discrimination tasks and in the accuracy analysis of the Test data replicates many findings of talker-voice perception showing improvement across age (Bartholomeus, 1973; Creel & Jimenez, 2012; Levi & Schwartz, 2013; Mann et al., 1979; Moher et al., 2010; Spence et al., 2002). The lack of an effect of Age in the other analyses related to the five days of training may be due to the extensive training and exposure that participants got, essentially washing out group differences. As will be discussed below, the language familiarity effect has been shown to be eliminated in an extensive training paradigm similar to the one used here.

A bilingual advantage for talker-voice processing

The current study also showed a bilingual advantage, which we can separate into two different types based on familiarity with the language. In the first type, bilinguals outperformed monolinguals in talker-voice tasks that were presented in foreign-accented

English. Previous research has demonstrated that listeners are worse when processing accented speech, whether it is due to a foreign accent (Goggin et al., 1991; Goldstein et al., 1981; Thompson, 1987) or to dialect differences (Kerstholt, Jansen, Van Amelsvoort & Broeders, 2006; Perrachione et al., 2010). In the case of dialect differences, these latter two studies found asymmetrical differences for own vs other dialect where speakers of the non-standard variety showed less of a difference between dialects than speakers of the standard dialect. The authors attribute this asymmetry to differences in exposure: Speakers of nonstandard varieties have extensive exposure to the standard variety, but not the other way around. Thus, one possible explanation for the bilingual advantage in the English conditions is that the children in the bilingual group simply have more experience with (foreign) accented speech and thus perform better.

Another possible explanation for the bilingual advantage in the English tasks is that the bilinguals have better cognitive control (inhibition) and are able to focus on the task (processing who is talking), while suppressing irrelevant information, namely that the speech is accented. Yet another possibility is that bilinguals have better social perception and perceiving the voice of a talker is highly relevant in social situations. Future studies will be needed to adjudicate between these alternatives.

The second way that the bilingual advantage manifested itself here was with better performance in an unfamiliar language. In this case, it is not possible to explain the better performance of the bilinguals as resulting from experience because none of the children had experience with German. As in the bilingual advantage for the experiments with English stimuli, it is possible that the advantage in the German condition is due to either better cognitive control (inhibition), to better social processing, or to better pitch perception. In the German condition, children must suppress the somewhat distracting fact that they cannot understand what is being said and must attend only to the talker-voice information. Thus, it is possible that the better cognitive control of the bilinguals translates to better processing in the German condition. Alternatively, bilinguals may perform better in an unfamiliar language because they simply have more experience listening to talker information in multiple languages, even if this experience does not come from German. This could be related to the literature on high variability training which shows that performance improves when listeners learn from many different speakers or speaker types versus a single speaker or single speaker type (Bradlow & Bent, 2008; Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997; Clopper & Pisoni, 2004; Logan, Lively & Pisoni, 1991; Sidaras, Alexander & Nygaard, 2009). Finally, better performance in the German condition could be due to better pitch perception by bilinguals (Krizman et al., 2012).

It should be noted that a bilingual advantage for an UNFAMILIAR language was not found in Xie and Myers (2015). They tested Mandarin(L1)–English(L2) (non-musician) and L1 English (non-musician) listeners on talker identification in Spanish. For these two groups of listeners, they did not find better performance for the bilingual Mandarin–English listeners, who presumably had two characteristics that could have made them better: (1) being a speaker of a tone language and thus having better pitch perception and (2) being a bilingual and thus having better pitch perception. There is no way of knowing why a bilingual advantage was not found in their study, but several possibilities exist. First, 17/26 of the L1

English listeners had studied Spanish so Spanish was not truly an unfamiliar language. Even though these listeners reported that they were not fluent in Spanish, the results of Orena et al. (2015) with Montreal listeners suggested that exposure and not fluency is sufficient for performing better in an exposed language. Second, Xie and Myers (2015) used sentence-length utterances with training and identification, whereas the current study used monosyllabic words with a discrimination task. Future studies will need to probe whether training tasks can eliminate or reduce predicted group differences, as has been found in Winters et al. (2008), where the generally robust language familiarity effect was absent in the training-identification task, but was present in the discrimination task.

Language familiarity effect

Finally, the current study replicated the language familiarity effect which has been found for both a native language (Goggin et al., 1991; Perrachione et al., 2011; Perrachione et al., 2015; Perrachione & Wong, 2007; Thompson, 1987; Winters et al., 2008) and a second language (Köster & Schiller, 1997; Schiller & Köster, 1996; Sullivan & Schlichting, 2000). The source of the language familiarity effect has been explored in several recent studies and has been attributed either to comprehension – retrieving lexical and/or semantic information – or to phonological knowledge.

Fleming et al. (2014) argue that comprehension is not necessary to exhibit a language familiarity effect. In their study, listeners showed the effect in time-reversed speech which they interpret as evidence that the phonological information that is retained in the time-reversed samples is sufficient to generate better performance in a familiar language. Zarate et al. (2015) also argue for the importance of phonological knowledge. In their study, native English-speaking listeners were better at perceiving talker information in German – a phonologically similar language – than in Mandarin – a phonologically dissimilar language: both of which were languages unfamiliar to the listeners. Additional support for the importance of phonetic/phonological processing in the absence of comprehension comes from Orena et al. (2015) who found that listeners residing in Montréal, but who were not second language speakers of French, performed better than listeners residing in Connecticut who did not have frequent exposure to French. This latter finding suggests that the language familiarity effect is indeed that, familiarity with the language, and does not require access to lexical or semantic information.

In contrast, Perrachione et al. (2015) argue for the importance of comprehension, or access to lexical and semantic information. Like Fleming et al. (2014), they tested listeners in both forward and backward speech. Listeners who did demonstrate a language familiarity effect in the forward condition did not demonstrate this effect in the backward condition. In a second experiment, they found that listeners were more accurate in learning talkers from real sentences than from nonword sentences. The authors take these two studies as evidence that lexical and semantic information is an important component of the language familiarity effect. Additional evidence for the contribution of lexical/semantic information comes from Goggin et al. (1991) who found that listeners were worse in time-reversed speech – which contains some amount of phonological information but lacks lexical/semantic information – than in forward speech. Perrachione et al. (2015) argue for the importance of both phonetic/

phonological and lexical/semantic processing due to a stepwise improvement from Mandarin to English nonwords and then to English words. Zarate et al. (2015) actually argue that phonological information is more important than lexical or semantic information because they did not find a significant difference in performance between German, English nonwords, and English words. It should be noted that Zarate et al. (2015) did not perform similar follow-up analyses to explore these differences, as they did between German and Mandarin, where this latter difference only emerged when examining data from the first 40% of the experiment.

If we delve deeper into the processing of English and German – two phonologically similar and genetically related languages – it is apparent that not all studies have found the same results, suggesting that task effects might modulate performance. Twenty years ago, Köster and Schiller (1997) tested three groups of listeners, all of whom had no knowledge of German: English L1, Spanish L1, and Mandarin L1. Using a voice line-up paradigm, they found that English-speaking listeners outperformed Spanish-speaking listeners, a result that was expected given the phonological similarity of English and German. They also found that English-speaking listeners performed better in English than in German, suggesting that phonological similarity is not the only factor affecting performance. Surprisingly, they also found that Mandarin-speaking listeners outperformed English-speaking listeners when listening to German, even though Mandarin and German are phonologically dissimilar and genetically unrelated languages. Phonological similarity appears to play some role in perceiving talker information in an unfamiliar language (English > Spanish), but is clearly not the only factor that influences performance. The better performance by the Mandarin-speaking listeners is consistent with the hypotheses about talker-voice perception outlined in Xie and Myers (2015), who predict that speaking a tone language would result in better talker-voice perception in an unfamiliar language. Although Xie and Myers (2015) did not actually find better performance for Mandarin versus English listeners when listening to Spanish stimuli, the results from Köster and Schiller (1997) suggest that other properties of the native language, aside from phonological similarity, affect performance.

In a study using speech samples from the same database as those used in the current study, Winters et al. (2008) found that adults who were native speakers of English with no knowledge of German exhibited the language familiarity effect in a discrimination task, but this effect disappeared in an identification and training paradigm where listeners had extensive exposure to the talkers. The lack of a language familiarity effect in the training paradigm may be due to the extensive exposure, which may have washed out any differences across the two language conditions. Better performance for English versus German in a discrimination task was also found for both children and adults in Levi and Schwartz (2013) and in the current study.

That several studies have found poorer performance in German than in English by native speakers of English suggests that phonological similarity is not sufficient to override other factors. Better performance in English than German could be related to comprehension or to experience with the specific phonetics and phonology of English. The best way to test this would be with English nonwords, and again, studies have shown different results:

Perrachione et al. (2015) found better performance for English words than English nonwords, Zarate et al. (2015) did not.

The different results across the studies of German and English suggest that additional research is needed. From the results presented in the current study, it is not possible to determine whether differences across studies are due to task effects, properties of the stimuli (monosyllabic words, multiple polysyllabic words, sentences), age of the participants, or other factors. That differences are found across studies suggests that we should be careful when drawing conclusions from a single experiment.

Conclusion

Taken together, the current study shows that bilingual children – here defined loosely as those children who speak, read, or understand another language or who have a family member who speaks another language living in the household – have a perceptual advantage when processing information about a talker's voice. They are faster to learn the voices of unfamiliar talkers and also perform better overall. They are even better in an unfamiliar language, suggesting that the benefits found for the English stimuli are not attributable only to experience listening to foreign-accented speech. Processing information about a talker combines both social skills, where listeners must understand that people sound different, and perceptual acuity, where listeners must attend to fine-grained phonetic detail in the speech signal. Testing perceptual acuity cannot be done with typical tests of speech sounds because a listener's first and second languages and their fluency in those languages affect how they perceive nonnative contrasts (Best, 1995; Best & Tyler, 2007; Flege, 1995; Kuhl & Iverson, 1995; Strange, 2006). In contrast, the tasks used here are able to demonstrate differences in perceptual acuity for acoustic information in the speech signal by focusing on how monolinguals and bilinguals process talker-voice information. The current findings demonstrate a perceptual advantage for bilingual listeners. This finding contributes to the literature on the various types of advantages afforded to bilingual speakers, where they have been found to be better at tasks tapping cognitive control, as well as socially relevant tasks such as taking on the perspective of another person.

Acknowledgments

This work was supported by a grant from the NIH-NIDCD: 1R03DC009851-01A2. We would like to thank Jennifer Bruno, Emma Mack, Alexandra Muratore, Sydney Robert, and Margo Waltz for help with data collection, three anonymous reviewers for their extremely helpful comments, and the children and families for their participation.

References

- Ansaldo AI, Ghazi-Saidi L, Adrover-Roig D. Interference Control In Elderly Bilinguals: Appearances Can Be Misleading. *Journal of Clinical and Experimental Neuropsychology*. 2015; 37(5):455–470. DOI: 10.1080/13803395.2014.990359 [PubMed: 25641572]
- Arredondo MM, Hu X-S, Satterfield T, Kovelman I. Bilingualism alters children's frontal lobe functioning for attentional control. *Developmental Science*. 2015; n/a-n/a. doi: 10.1111/desc.12377
- Baayen, RH. *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge: Cambridge University Press; 2008.
- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008; 59:390–412.

- Bartholomeus B. Voice identification by nursery school children. *Canadian Journal of Psychology*. 1973; 27(4):464–472. [PubMed: 4766153]
- Bates, D., Maechler, M., Bolker, B., Walker, S. lme4: Linear mixed-effects models using Eigen and S4. 2010. Retrieved from <http://lme4.r-forge.r-project.org/>
- Best, CT. A direct realist view of cross-language speech perception. In: Strange, W., editor. *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press; 1995. p. 171-204.
- Best CT, Tyler MD. Nonnative and second-language speech perception: Commonalities and complementarities. *Language experience in second language speech learning: In honor of James Emil Flege*. 2007:13–34.
- Bialystok E. Cognitive complexity and attentional control in the bilingual mind. *Child Development*. 1999; 70(3):636–644.
- Bialystok E, Craik FIM, Klein R, Viswanathan M. Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*. 2004; 19(2):290–303. DOI: 10.1037/0882-7974.19.2.290 [PubMed: 15222822]
- Bialystok E, Martin MM. Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science*. 2004; 7(3):325–339. DOI: 10.1111/j.1467-7687.2004.00351.x [PubMed: 15595373]
- Bradlow AR, Bent T. Perceptual adaptation to non-native speech. *Cognition*. 2008; 106(2):707–729. [PubMed: 17532315]
- Bradlow AR, Pisoni DB, Akahane-Yamada R, Tohkura Yi. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*. 1997; 101(4):2299–2310. [PubMed: 9104031]
- Bregman MR, Creel SC. Gradient language dominance affects talker learning. *Cognition*. 2014; 130(1):85–95. <https://doi.org/10.1016/j.cognition.2013.09.010>. [PubMed: 24211437]
- Brown, L., Sherbenou, RJ., Johnsen, SK. TONI-3: Test of nonverbal intelligence (third edition). Austin, TX: Pro-Ed; 1997.
- Buac M, Kaushanskaya M. The relationship between linguistic and non-linguistic cognitive control skills in bilingual children from low socio-economic backgrounds. *Frontiers in Psychology*. 2014; 5:1–12. [PubMed: 24474945]
- Carlson SM, Meltzoff AN. Bilingual experience and executive functioning in young children. *Developmental Science*. 2008; 11(2):282–298. DOI: 10.1111/j.1467-7687.2008.00675.x [PubMed: 18333982]
- Clopper CG, Pisoni DB. Effects of talker variability on perceptual learning of dialects. *Language and Speech*. 2004; 47(3):207–239. [PubMed: 15697151]
- Costa A, Hernández M, Sebastián-Gallés N. Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*. 2008; 106(1):59–86. [PubMed: 17275801]
- Creel SC, Jimenez SR. Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*. 2012; 113:487–509. [PubMed: 22958962]
- Cutler, A., Andics, A., Fang, Z. Inter-dependent categorization of voices and segments. In: Lee, W-S., Zee, E., editors. *Proceedings of the Seventeenth International Congress of Phonetic Sciences*. Hong Kong: Department of Chinese, Translation, and Linguistics, City University of Hong Kong; 2011. p. 552-555.
- de Bruin A, Treccani B, Della Sala S. Cognitive Advantage in Bilingualism: An Example of Publication Bias? *Psychological Science*. 2015; 26(1):99–107. DOI: 10.1177/0956797614557866 [PubMed: 25475825]
- DeCasper AJ, Fifer WP. Of human bonding: Newborns prefer their mothers' voices. *Science*. 1980; 208:1174–1176. [PubMed: 7375928]
- DeCasper AJ, Prescott PA. Human newborn's perception of male voices: Preference, discrimination, and reinforcing value. *Developmental Psychology*. 1984; 17(5):481–491.
- Ensminger, ME., Fothergill, KE. A decade of measuring SES: What it tells us and where to go from here. In: Bornstein, MH., Bradley, RH., editors. *Socioeconomic status, Parenting, and Child Development*. Mahwah, New Jersey: Lawrence Erlbaum Associates; 2003. p. 13-27.

- Fan SP, Liberman Z, Keysar B, Kinzler KD. The Exposure Advantage: Early Exposure to a Multilingual Environment Promotes Effective Communication. *Psychological Science*. 2015; doi: 10.1177/0956797615574699
- Flege, JE. Second language speech learning: Theory, findings, and problems. In: Strange, W., editor. *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press; 1995. p. 233-277.
- Fleming D, Giordano BL, Caldara R, Belin P. A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences*. 2014; 111(38):13795–13798. DOI: 10.1073/pnas.1401383111
- Friesen DC, Latman V, Calvo A, Bialystok E. Attention during visual search: The benefit of bilingualism. *International Journal of Bilingualism*. 2015; 19(6):693–702. DOI: 10.1177/1367006914534331 [PubMed: 26640399]
- Goggin JP, Thompson CP, Strube G, Simental LR. The role of language familiarity in voice identification. *Memory & Cognition*. 1991; 19(5):448–458. [PubMed: 1956306]
- Gold BT, Kim C, Johnson NF, Kryscio RJ, Smith CD. Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *The Journal of Neuroscience*. 2013; 33(2):387–396. DOI: 10.1523/jneurosci.3837-12.2013 [PubMed: 23303919]
- Goldstein AG, Knight P, Bailis K, Conover J. Recognition memory for accented and unaccented voices. *Bulletin of the Psychonomic Society*. 1981; 17(5):217–220.
- Green KP, Tomiak GR, Kuhl PK. The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics*. 1997; 59(5):675–692. [PubMed: 9259636]
- Greenberg A, Bellana B, Bialystok E. Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. *Cognitive Development*. 2013; 28(1):41–50. <https://doi.org/10.1016/j.cogdev.2012.10.002>. [PubMed: 23486486]
- Hamers, JF., Blanc, M. *Bilinguality and bilingualism*. Cambridge University Press; 2000.
- Hollien H, Majewski W, Doherty ET. Perceptual identification of voices under normal, stress, and disguise speaking conditions. *Journal of Phonetics*. 1982; 10:139–148.
- Incera S, McLennan CT. Mouse tracking reveals that bilinguals behave like experts. *Bilingualism: Language and Cognition, FirstView*. 2015; :1–11. DOI: 10.1017/S1366728915000218
- Jaeger TF. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*. 2008; 59:434–446. [PubMed: 19884961]
- Johnson EK, Westrek E, Nazzi T, Cutler A. Infant ability to tell voices apart rests on language experience. *Developmental Science*. 2011; 14(5):1002–1011. [PubMed: 21884316]
- Kerstholt JH, Jansen NJM, Van Amelsvoort AG, Broeders APA. Earwitnesses: effects of accent, retention and telephone. *Applied Cognitive Psychology*. 2006; 20(2):187–197. DOI: 10.1002/acp.1175
- Kisilevsky BS, Hains SMJ, Lee K, Xie S, Huang H, Ye HH, Zhang K, Wang Z. Effects of experience on fetal voice recognition. *Psychological Science*. 2003; 14(3):220–224. [PubMed: 12741744]
- Köster O, Schiller NO. Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics*. 1997; 4:18–28.
- Krizman J, Marian V, Shook A, Skoe E, Kraus N. Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences*. 2012; 109(20):7877–7881. DOI: 10.1073/pnas.1201575109
- Kuhl, PK., Iverson, P. Linguistic experience and the “Perceptual Magnet Effect”. In: Strange, W., editor. *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press; 1995. p. 121-154.
- Levi SV. Individual differences in learning talker categories: The role of working memory. *Phonetica*. 2014; 71(3):201–226. [PubMed: 25721393]
- Levi SV, Schwartz RG. The development of language-specific and language-independent talker processing. *Journal of Speech, Language, and Hearing Research*. 2013; 56:913–920.
- Levi SV. Talker familiarity and spoken word recognition in school-age children. *Journal of Child Language*. 2015; 42(4):843–872. [PubMed: 25159173]

- Levi SV, Winters SJ, Pisoni DB. Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America*. 2007; 121:2327–2338. [PubMed: 17471745]
- Levi SV, Winters SJ, Pisoni DB. Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible? *Journal of the Acoustical Society of America*. 2011; 130(6): 4053–4062. [PubMed: 22225059]
- Logan JS, Lively SE, Pisoni DB. Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*. 1991; 89(2):874–886. [PubMed: 2016438]
- Mann VA, Diamond R, Carey S. Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*. 1979; 27:153–165. [PubMed: 458368]
- Moher M, Feigenson L, Halberda J. A one-to-one bias and fast mapping support preschoolers' learning about faces and voices. *Cognitive Science*. 2010; 34(5):719–751. [PubMed: 21564234]
- Mullennix JW, Pisoni DB. Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*. 1990; 47(4):379–390. [PubMed: 2345691]
- Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. *Perception & Psychophysics*. 1998; 60(3):355–376. [PubMed: 9599989]
- Orena AJ, Theodore RM, Polka L. Language exposure facilitates talker learning prior to language comprehension, even in adults. *Cognition*. 2015; 143:36–40. <https://doi.org/10.1016/j.cognition.2015.06.002>. [PubMed: 26113447]
- Paap KR, Greenberg ZI. There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*. 2013; 66(2):232–258. <https://doi.org/10.1016/j.cogpsych.2012.12.002>. [PubMed: 23370226]
- Paap KR, Johnson HA, Sawi O. Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*. 2014; 26(6):615–639. DOI: 10.1080/20445911.2014.944914
- Paap KR, Sawi O. Bilingual Advantages in Executive Functioning: Problems in Convergent Validity, Discriminant Validity, and the Identification of the Theoretical Constructs. *Frontiers in Psychology*. 2014; 5doi: 10.3389/fpsyg.2014.00962
- Paap KR, Sawi OM, Dalibar C, Darrow J, Johnson HA. The Brain Mechanisms Underlying the Cognitive Benefits of Bilingualism may be Extraordinarily Difficult to Discover. *AIMS Neuroscience*. 2014; 1(3):245–256. <https://doi.org/10.3934/Neuroscience.2014.3.245>.
- Perrachione TK, Chiao JY, Wong PCM. Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition*. 2010; 114(1):42–55. <https://doi.org/10.1016/j.cognition.2009.08.012>. [PubMed: 19782970]
- Perrachione TK, Del Tufo SN, Gabrieli JDE. Human voice recognition depends on language ability. *Science*. 2011; 333:595. [PubMed: 21798942]
- Perrachione, TK., Dougherty, SC., McLaughlin, DE., Lember, RA. Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow: University of Glasgow; 2015. The effects of speech perception and speech comprehension on talker identification.
- Perrachione TK, Wong PCM. Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*. 2007; 45:1899–1910. [PubMed: 17258240]
- Purhonen M, Kilpeläinen-Lees R, Valkonen-Korhonen M, Karhu J, Lehtonen J. Cerebral processing of mother's voice compared to unfamiliar voice in 4-month-old infants. *International Journal of Psychophysiology*. 2004; 52:257–266.
- Purhonen M, Kilpeläinen-Lees R, Valkonen-Korhonen M, Karhu J, Lehtonen J. Four-month-old infants process own mother's voice faster than unfamiliar voices -Electrical signs of sensitization in infant brain. *Cognitive Brain Research*. 2005; 24:627–633. [PubMed: 16099371]
- Saidi LG, Ansaldo AI. Can a Second Language Help You in More Ways Than One? *AIMS Neuroscience*. 2015; 2(1):52–57. <https://doi.org/10.3934/Neuroscience.2015.1.52>.
- Schiller NO, Köster O. Evaluation of a foreign speaker in forensic phonetics: a report. *Forensic Linguistics*. 1996; 3(1):176–185.
- Schneider, W., Eschman, A., Zuccolotto, A. E-Prime 2.0 Professional. Pittsburgh, PA: Psychology Software Tools, Inc; 2007.

- Semel, E., Wiig, E.H., Secord, W.A. Clinical evaluation of language fundamentals, fourth edition (CELF-4). Toronto, CA: The Psychological Corporation/A Harcourt Assessment Company; 2003.
- Sidasas SK, Alexander JED, Nygaard LC. Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America*. 2009; 125(5):3306–3316. [PubMed: 19425672]
- Spence MJ, Rollins PR, Jerger S. Children's recognition of cartoon voices. *Journal of Speech, Language and Hearing Research*. 2002; 45(1):214–222.
- Stevenage SV, Clarke G, McNeill A. The “other-accent” effect in voice recognition. *Journal of Cognitive Psychology*. 2012; 24(6):647–653. DOI: 10.1080/20445911.2012.675321
- Strange W. Second-language speech perception: The modification of automatic selective perceptual routines. *The Journal of the Acoustical Society of America*. 2006; 120:3137.
- Stroop JR. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*. 1935; 18:643–662.
- Sullivan KPH, Schlichting F. Speaker discrimination in a foreign language: first language environment, second language learners. *Forensic Linguistics*. 2000; 17(1):95–111.
- Thompson CP. A language effect in voice identification. *Applied Cognitive Psychology*. 1987; 1(2): 121–131.
- Van Dommelen WA. The Contribution of Speech Rhythm and Pitch to Speaker Recognition. *Language and Speech*. 1987; 30(4):325–338. DOI: 10.1177/002383098703000403
- Winters SJ, Levi SV, Pisoni DB. Identification and discrimination of bilingual talkers across languages. *Journal of the Acoustical Society of America*. 2008; 123(6):4524–4538. [PubMed: 18537401]
- Xie X, Myers E. The impact of musical training and tone language experience on talker identification. *Journal of the Acoustical Society of America*. 2015; 137(1):419–432. [PubMed: 25618071]
- Zarate JM, Tian X, Woods KJP, Poeppel D. Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Scientific Reports*. 2015; 5:11475.doi: 10.1038/srep11475 [PubMed: 26088739]

Appendix A. Detailed output of the ANOVAs for the demographic variables

	CELF Core Language	TONI	SES
Group	F(1,37) = 1.26 p = 0.26 $\eta_p^2 = 0.033$	F(1,37) = 0.28 p = 0.59 $\eta_p^2 = 0.007$	F(1,36) = 1.87 p = 0.17 $\eta_p^2 = 0.049$
Age	F(1,37) = 0.26 p = 0.60 $\eta_p^2 = 0.007$	F(1,37) = 0.06 p = 0.79 $\eta_p^2 = 0.001$	F(1,36) = 0.36 p = 0.54 $\eta_p^2 = 0.010$
Group x Age	F(1,37) = 0.05 p = 0.81 $\eta_p^2 = 0.001$	F(1,37) = 0.49 p = 0.48 $\eta_p^2 = 0.013$	F(1,36) = 0.19 p = 0.66 $\eta_p^2 = 0.005$

Appendix B. Results for talker discrimination

	Estimate	Standard Error	z-value	Probability
(Intercept)	1.170	.123	9.512	<.001

	Estimate	Standard Error	z-value	Probability
Group (monolingual)	-.364	.165	-2.206	.027
Age	.322	.130	2.477	.013
Language (German)	-.290	.113	-2.546	.010
Group x Age	.024	.168	.147	.883
Group x Language	-.010	.148	-.074	.941
Age x Language	.064	.121	.528	.597
Group x Age x Language	-.011	.153	-.074	.940
<i>Results for talker identification (Training, with feedback).</i>				
(Intercept)	.391	.151	2.59	.009
Group (monolingual)	-.325	.166	-1.95	.051
Age	.149	.137	1.08	.277
Day	.322	.035	9.15	<.001
Session	.065	.044	1.49	.134
Group x Age	.010	.173	.06	.952
<i>Results for talker identification (Testing, no feedback).</i>				
(Intercept)	.440	.166	2.65	.007
Group (monolingual)	-.334	.201	-1.65	.098
Age	.329	.151	2.17	.029
Day	.244	.045	5.36	<.001
Group x Age	-.015	.200	-.078	.938

Figure 1.
Response screen for Talker Learning. Figure from Levi (2015). Reprinted with permission.

Figure 2. Sample trial procedure during the five days of training. Figure from Levi (2015). Reprinted with permission.

Figure 3. Proportion correct in the English (black lines) and German (grey lines) discrimination tasks. Error bars represent standard error.

Figure 4. Average number of days to reach criterion (70% accuracy) in the training phase (with feedback). Error bars represent standard error.

Figure 5.
Average number of days to reach criterion (70% accuracy) in the test phase (no feedback).
Error bars represent standard error.

Figure 6. Proportion correct on talker identification during the five days of training (with feedback). Error bars represent standard error.

Figure 7. Proportion correct on talker identification during the five days of testing (no feedback). Error bars represent standard error.

Table 1

Demographic data for the four groups.

	N	Age		CELF		TONI		SES	
		Mean (SD in months)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Monolingual	Younger	7:11 (6)	7:1–8:7	101.4 (12.4)	87–121	108.7 (14.7)	92–140	5.6 (1.8)	2–8
	Older	10:11 (6)	10:3–11:9	104.4 (12.7)	90–123	106.7 (17.2)	89–144	5.5 (1.9)	2–7
Bilingual	Younger	8:2 (7)	6:8–8:11	106.9 (10.8)	93–124	108.3 (16.1)	90–125	4.9 (2.6)	1–8
	Older	10:9 (7)	10:2–11:10	108.0 (16.6)	88–133	113.7 (19.6)	95–138	4.1 (2.6)	1–8

Table 2

Speaker information.

Speaker	Intelligibility	Foreign-accent Rating	F0 (Hz) English	F0 (Hz) German
F3	76.2%	2.8	224	225
F8	83.3%	1.8	177	187
F11	77.8%	4.2	260	291

Note: Intelligibility and F0 information about the speakers used in the current study. “Intelligibility” refers to the average number of words correctly identified in the clear (Levi, Winters & Pisoni, 2011). For reference, average intelligibility for five native speakers in the same task ranged from 74.6% to 92.1%. “Foreign-accent Rating” is the average raw rating on a 7-point scale ranging from 0–6 (Levi, Winters & Pisoni, 2007). For reference, five native speakers in the same task ranged from 1.1–1.5. “F0” is the average fundamental frequency of the vowels for the 360 English and German words.