Published in final edited form as:

J Child Lang. 2015 July ; 42(4): 843–872. doi:10.1017/S0305000914000506.

Talker familiarity and spoken word recognition in school-age children*

Susannah V. Levi

New York University

Abstract

Research with adults has shown that spoken language processing is improved when listeners are familiar with talkers' voices, known as the familiar talker advantage. The current study explored whether this ability extends to school-age children, who are still acquiring language. Children were familiarized with the voices of three German–English bilingual talkers and were tested on the speech of six bilinguals, three of whom were familiar. Results revealed that children do show improved spoken language processing when they are familiar with the talkers, but this improvement was limited to highly familiar lexical items. This restriction of the familiar talker advantage is attributed to differences in the representation of highly familiar and less familiar lexical items. In addition, children did not exhibit accent-general learning; despite having been exposed to German-accented talkers during training, there was no improvement for novel German-accented talkers.

INTRODUCTION

To understand spoken language, listeners must be highly sensitive to acoustic-phonetic information in the speech signal. Within the first year of life, infants tune their perception, by increasing sensitivity to acoustic-phonetic detail that is relevant for their native language while decreasing sensitivity to detail that is not (Best, 1993; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; Werker & Tees, 1984). Unfortunately for listeners, the speech signal itself is highly variable and thus children must learn to map these variable inputs onto stable linguistic categories. One major source of variability in the speech signal arises from differences across talkers who vary along myriad dimensions such as gender, age, dialect, nativeness, and idiolect. Perceiving the linguistic content of spoken language therefore requires that listeners' perceptions be sufficiently flexible, allowing them to adapt to this variability in the signal.

*This work was supported by a grant from the NIH-NIDCD (1R03DC009851-01A2). I would like to thank Gabrielle Alfano, Josh Barocas, Jennifer Bruno, Stephanie Lee, Emma Mack, Alexandra Muratore, Sydney Robert, and Margo Waltz for help with data collection, Adam Buchwald and Richard Schwartz for comments on previous versions of this paper, and the children and families for their participation.

© Cambridge University Press 2014

Address for correspondence: Susannah Levi, New York University, Department of Communicative Sciences and Disorders, 665 Broadway, 9th floor, New York, NY 10012.

Considerable research has shown that adult speech perception exhibits a high degree of plasticity; adults tune their perceptions on-line based on characteristics of a talker. For example, processing foreign-accented speech is speeded after exposure to only a few sentences (Clarke & Garrett, 2004). This type of attunement to a talker's productions appears very rapidly, after approximately 120 ms of exposure, even before a listener's conscious awareness (Sjerps, Mitterer & McQueen, 2011). Creel, Aslin, and Tanenhaus (2008) found that adult listeners who learned nonword rhyming pairs spoken by different talkers used this information to facilitate lexical decision. This TALKER-CONTINGENT processing, where listeners process speech through the lens of different talkers, has been examined both through studies of short-term implicit learning of talker productions and also through studies of long-term explicit learning of talker information. These studies have found that both types of exposure can affect speech perception and spoken language processing.

In an early study of implicit attunement to a talker, Ladefoged and Broadbent (1957) found that synthesizing two different VOICES in a carrier phrase changed listeners' perceptions of ambiguous vowel targets following the carrier phase, indicating that listeners had modified their perceptual spaces based on the voice of the talker in the carrier phrase. More recently, studies have found that exposure to a talker-specific pronunciation (e.g. long versus short voice onset time, ambiguous fricatives) changes the location of phoneme boundaries along an acoustic continuum, and also generalizes to the perception of novel words (Allen & Miller, 2004; Eisner & McQueen, 2005; Kraljic & Samuel, 2007; McQueen, Cutler & Norris, 2006; Norris, McQueen & Cutler, 2003). These studies have reported short-term adaptation effects within the context of an experimental protocol and often showed changes in how listeners perceive ambiguous stimuli.

A related line of research has demonstrated that long-term explicit learning of talker information improves spoken language processing. In several studies, adult listeners learned to identify the voices of either native (Nygaard & Pisoni, 1998; Nygaard, Sommers & Pisoni, 1994; Yonan & Sommers, 2000) or non-native (Levi, Winters & Pisoni, 2011) talkers over several days of training. After talker-voice learning, listeners were asked to identify what a talker said. Listeners were more accurate at processing linguistic information spoken by familiarized talkers, than by novel, unknown talkers, but not when listeners were familiarized with talkers in a different language (Levi *et al.*, 2011). This FAMILIAR TALKER ADVANTAGE shows that stored knowledge about a talker mediates spoken language processing in adults. Furthermore, the familiar talker advantage applies to the perception of novel words, indicating that listeners learn how familiarized talkers articulate different speech sounds. Additional evidence that listeners learn talker-specific information at the level of individual sounds comes from a study where listeners are faster in a lexical decision task when they had been exposed to the same talker producing the target sounds, even when the target sounds were re-ordered (e.g. *cab-back*) (Jesse, McQueen & Page, 2007). Taken together, these studies of the familiar talker advantage show that stored information about a talker's productions helps listeners when they are asked to retrieve the linguistic content of a novel stimulus, resulting in more accurate spoken language processing.

Unlike the implicit talker-voice attunement studies, the explicit talker-learning studies actually found that only those listeners who learned to identify the talkers' voices

sufficiently well (known as ‘good learners’) showed a familiar talker advantage (Levi *et al.*, 2011; Nygaard & Pisoni, 1998). Thus, even though poor learners were exposed to the same amount of speech during training, they did not learn the voices of the talkers, and did not store the link between a particular talker and her talker-specific articulations. A difference between the implicit talker attunement studies discussed above and the explicit talker training studies is that the former examined listeners’ perceptions of only a few phonetic contrasts within the experiment which were often ambiguous or anomalous productions, while the latter examined how this talker-learning affected open-set spoken word recognition or sentence processing which involved the full range of speech sounds.

In addition to improving spoken language processing, familiarity with a talker also facilitates auditory stream segregation. Both adults (Newman & Evers, 2007) and infants (Barker & Newman, 2004) are better at tracking familiar talkers than unknown talkers when presented with background, distractor talkers. This finding is especially important for school-age children who need to attend to a familiar talker (e.g. a teacher or peer) in the presence of many background talkers in a school environment.

The primary goal of the present study is to test whether school-age children show the familiar talker advantage, as has been found for adults. Improving spoken language processing is especially important for this population because children are still acquiring language and their perceptual skills are poorer than adults (Bent, 2014). As such, their language processing skills are less honed than those of adults. If talker familiarity facilitates language processing for children, then some cognitive resources will be freed up and be available for other tasks. Indeed, for adults, the talker dimension interacts with language processing at higher cognitive levels such as list recall and lexical priming (Goldinger, Pisoni & Logan, 1991; Palmeri, Goldinger & Pisoni, 1993; Schacter & Church, 1992). Furthermore, the effect of talker information is most noticeable when tasks are more effortful (McLennan & Luce, 2005; Vitevitch & Donoso, 2011). Thus, in children, for whom spoken language processing is more challenging than it is for adults, the talker effects may have even more impact. Finally, examining the benefit of talker familiarity is especially important in children because they are more adversely affected by background noise than adults (see Nelson & Soli, 2000, for a review). Therefore, examining factors that could improve spoken language processing in natural, noisy conditions in children is highly important, as these are the environments in which most language is encountered.

School-age children are expected to show a familiar talker advantage for the following reasons. First, school-age children are expected to be able to learn the voices of novel talkers, as previous work has found that accurately processing talker information improves across childhood (Mann, Diamond & Carey, 1979). Bartholomeus (1973) found that preschool children aged four to five could identify their classmates’ voices above chance (approximately 60% accuracy with 18 voices to identify), although this was highly variable across children. Spence, Rollins, and Jerger (2002) also found that children of this age could recognize the voices of familiar cartoon characters. Preschool children are also able to learn the voices of novel talkers above chance (Creel & Jimenez, 2012; Moher, Feigenson & Halberda, 2010) and can generalize talker identification to novel words (Moher *et al.*, 2010). Furthermore, school-age children can discriminate the same talkers used in the current study

(Levi & Schwartz, 2013), suggesting that children will be able to also learn to identify these same talkers.

Second, children are expected to use stored information about a talker to improve spoken word recognition because previous studies have found that toddlers are able to implicitly adapt their perception to different talkers' voices (Schmale, Cristia & Seidl, 2012; Schmale & Seidl, 2009; van Heugten & Johnson, 2014; White & Aslin, 2011). In several studies, toddlers were briefly exposed to a speaker of a different dialect (Australian English: van Heugten & Johnson, 2014), to a foreign-accented speaker (Spanish-accented English: Schmale *et al.*, 2012; Schmale & Seidl, 2009), or to a speaker producing a low front vowel in words that commonly have a low back vowel (e.g. *block* [blak] produced as [blæk]), simulating vowel fronting, which occurs in some dialects of American English (White & Aslin, 2011). In all cases, toddlers were able to adapt to these accents, correctly mapping these types of novel speech onto lexical items. Toddlers in White and Aslin's study generalized their perceptions to novel words, suggesting that they learned a general shift in a phonemic category and used knowledge of the shift to influence processing of novel items. Similarly, van Heugten and Johnson (2014) also found that perception generalized to words that were not used during the talker familiarization phase. These studies show that toddlers implicitly learn these variant productions, quickly tune their perception to these variants, and use this to process spoken language on-line.

An important component of these previous studies with toddlers is the use of highly familiar words where children already have extensive exposure to the expected target in their native dialect. In fact, Van Heugten and Johnson (2014) only found adaptation to their Australian-accented talker when toddlers were highly familiar with the story used during the exposure phase. In one experiment, toddlers heard *The Very Hungry Caterpillar* (Carle, 1969) spoken by an Australian talker as part of a familiarization phase and then performed the experimental listening phase in which they heard words and nonwords spoken by the same Australian talker. This brief familiarization to the Australian talker did not result in looking time differences for words versus nonwords. However, in a follow-up experiment, parents read *The Very Hungry Caterpillar* once a day for two weeks and then toddlers completed the same experiment where they heard the Australian talker read the story and then completed the experimental listening phase. In this version, toddlers did listen longer to words than nonwords spoken by the Australian-accented talker. The crucial difference between the two experiments was that toddlers in the latter version had knowledge of the target words because they were highly familiar with the story and therefore knew the intended target in their native dialect. Toddlers were able to learn a new mapping from Australian-accented speech sounds to stored lexical representations because they were familiar with the lexical items in the story.

Because familiarity with lexical items seems to facilitate toddlers' ability to learn the mapping between a novel accent and existing lexical representations, an additional goal of the current study was to explore the role of lexical familiarity in the familiar talker advantage. Children in the current study were presented with an equal number of high- and low-familiarity words. This manipulation of lexical familiarity allows us to determine whether the familiar talker advantage is influenced by lexical familiarity, which could have

broad implications for novel word learning depending on whether low-familiarity words show the advantage.

Research on children's lexical representations suggests that there is a shift from holistic representations to more phonetically specified representations as children become more familiar with the lexical items (Fennell & Werker, 2003; Jusczyk, 1986, 1992; Stager & Werker, 1997; Swingley & Aslin, 2000; Walley, 1993; White & Morgan, 2008). With respect to lexical processing in general, it has been suggested that children first process words holistically and then progressively change to representations that are more phonetically specified (Jusczyk, 1986, 1992; Walley, 1993), and that this change in lexical representations extends into the school-age years (Metsala, 1997). Several studies have suggested that the change from holistic representations to more phonetically specified representations may not be due to age or vocabulary size per se, but rather to familiarity with different lexical items (Fennell & Werker, 2003; Stager & Werker, 1997; Swingley & Aslin, 2000; White & Morgan, 2008). These studies use a paradigm where listeners (toddlers) are presented with words produced with phonetic errors, such as *tog* for *dog*. When these phonetic mismatches are presented in unfamiliar nonwords (e.g. *bih* for *dih*), toddlers do not perceive the error (Stager & Werker, 1997), a surprising finding given their sensitivity to these contrasts in nonsense syllables within the first year of life (Werker & Tees, 1984). However, when these errors are presented in highly familiar words (e.g. *dog*, *doll*), toddlers do perceive the errors, suggesting that toddlers are sensitive to acoustic-phonetic information in highly familiar words and that representations for highly familiar words include phonetic detail (Fennell & Werker, 2003; White & Morgan, 2008). White and Morgan (2008) tested the degree of mismatch where errors varied in one, two, or three phonetic features (e.g. *keys* [kiz] becomes *teys* [tiz] (place of articulation), *deys* [diz] (place of articulation and voicing), or *zeys* [ziz] (place of articulation, voicing, and manner of articulation)) and found that toddlers showed more disruption in processing when the error varied by more phonetic features. The authors argued that this graded sensitivity further shows that at this young age toddlers are highly sensitive to acoustic-phonetic detail and that the representations for highly familiar words include this phonetic detail.

A recent study of adult sensitivity to phonetic detail suggests that differences in representation extend into adulthood. White, Yee, Blumstein, and Morgan (2013) examined adults' sensitivity to phonetic errors in a study using the same types of phonetic mismatch that had been done with toddlers. Because even low-frequency words are familiar to adults, the authors examined adults' sensitivity to phonetic errors in nonwords. Target nonwords were either presented many times (high-frequency nonword) or only a few times (low-frequency nonword) during the experiment, allowing the authors to simulate lexical familiarity and control exposure in a way that was more similar to what toddlers have experienced. The target nonwords were then presented with phonetic errors. The results revealed that adults' sensitivity to phonetic mismatches mimicked that of toddlers, where adults were only sensitive to acoustic-phonetic errors in the high-frequency nonwords, but were not sensitive to acoustic-phonetic errors in the low-frequency nonwords. The authors argued that the apparent difference in adult and child spoken language processing is not due to different mechanisms underlying processing, but instead is due to the sheer magnitude of

experience that adults have with spoken language. Thus, sensitivity to phonetic detail is not solely a function of lexicon size, but instead depends on degree of familiarity with a lexical item. Given that both toddlers' and adults' sensitivity to phonetic detail is tied to familiarity with a lexical item, it is possible that the familiar talker advantage will also be mediated by lexical familiarity.

One final goal of the current study was to test generalization of accent type to novel talkers. Six German-L1–English-L2 talkers were used for the spoken word recognition tasks but only three of these six were used during talker learning. Previous research on adult adaptation to non-native accented speech has found that hearing multiple talkers of a particular foreign accent allows listeners to generalize this knowledge to improve language processing for novel talkers with the same foreign accent (Bradlow & Bent, 2008; Sidaras, Alexander & Nygaard, 2009). In other words, adult listeners show accent-general learning. The current study will examine both the benefit of talker familiarity and the potential benefit of accent familiarity by comparing baseline spoken word recognition accuracy to post-training accuracy for both familiarized and unfamiliar talkers. Bilingual talkers were used in the current study because they produce English with an unfamiliar accent, allowing us to test whether children also show accent-general learning.

METHODS

Participants

Forty-one native English-speaking children aged 7;4–12;0 (mean: 9;5) participated in the study. None of these children had exposure to German or German-accented English, as indicated on the parent questionnaire. Language and non-verbal cognition were assessed in all children using the Clinical Evaluation of Language Fundamentals-4 (CELF) (Semel, Wiig & Secord, 2003) and the Test Of Nonverbal Intelligence-3 (TONI) (Brown, Sherbenou & Johnsen, 1997). Both tests are normed to 100 with a standard deviation of 15. To be included in the study, children had to score 85 or above on the Core Language composite of the CELF, which assesses both expressive and receptive language skills. The mean CELF-Core Language score for the children was 101 ($SD=10$), with a range of 85–123. The mean score on the TONI was 104 ($SD=15$), with a range of 84–138. One participant who scored an 84 on the TONI was also included in the data analysis. All children passed a pure tone hearing screening at 25 dB HL at 1000 Hz, 2000 Hz, and 4000 Hz either with a portable Earscan3 Screening Audiometer (ES3S) in their school or home or with a GSI 61 Clinical Audiometer in a sound attenuated IAC Booth in the Department of Communicative Sciences and Disorders at New York University, with the following exceptions: one child only passed 4000 Hz at 30 dB in the left ear; fourteen children did not complete a hearing screening because the portable audiometer was not available at the time of testing. Parent questionnaires indicated that five of these children passed a hearing screening at a doctor's office within one year prior to participating in the study and the remaining nine had not experienced frequent ear infections and the parent report did not indicate any known hearing problems.

Stimuli

Six female bilingual German-L1–English-L2 talkers were recorded producing 360 monosyllabic CVC words in both English and German. Recordings were made in a sound-attenuated IAC booth at the Speech Research Laboratory at Indiana University using a SHURE SM98 head-mounted unidirectional (cardioid) condenser microphone with a flat frequency response from 40 to 20,000 Hz. Productions were digitized into 16-bit stereo recordings via Tucker-Davis Technologies System II hardware at 22,050 Hz and saved directly to a PC. Talkers read each word as it was presented to them on a computer monitor in random order, blocked by language. All sound files were normalized to have a uniform root mean squared (RMS) amplitude. The six bilingual talkers used in the current study had similar intelligibility (see Table 1) and were selected from a larger group of bilingual talkers (Levi *et al.*, 2011). Additionally, the three speakers that were used during the talker identification training were selected to have relatively different average fundamental frequency across productions.

The six talkers produced very few deviations from typical native norms. One talker (F3) was *r*-less, producing a schwa off-glide when *r* was in the coda. Only 16 trials out of 3936 trials of the spoken word recognition task (41 participants $\times$ 48 trials $\times$ 2 times (baseline, post-training)) contained *r*-less targets from this talker. Of these 16 trials, 13 were perceived as containing the target coda [r], 2 were heard and reproduced by the children with the schwa off-glide and were given full credit for these responses, and one response was completely different from the target. The other production difference that occurred for these non-native talkers was voiceless coda fricatives produced as voiced (e.g. *moose* [muz]). Only 36/3936 trials contained a target voiceless fricative that was voiced. Scoring the voicing as correct or incorrect did not change the results.

The 360 English words were rated independently by three individuals (two speech-language pathologists and the author) as likely to be known to children aged five to eight (HIGH LEXICAL FAMILIARITY), likely to be unknown to these children (LOW LEXICAL FAMILIARITY), or possibly known by these children. Out of these words, 125 were rated as high familiarity and 109 were rated as low familiarity by all three raters. These familiarity ratings were compared with age of acquisition (AoA) ratings from Cortese and Khanna (2008), where participants rated words on a 1–7 scale. Lower numbers indicate that the words were acquired earlier. The mean AoA for the high familiarity words was 3.01 (acquired between ages four and six) and for the low familiarity words was 5.34 (acquired after ages eight to ten). Only three of the high-familiarity words were not found in Cortese and Khanna's database (*house*, *mouse*, *pill*). Seventeen words from the low-familiarity list were not found in the database; 15/17 have a frequency of 1 based on Kuera and Francis (1967) and two (*couth*, *goon*) are not included. The complete list of English words that were used in the current study can be found in the 'Appendix'.

Children were assigned to one of eight groups, representing a different random sampling of the high- and low-familiarity words. For each group, a distinct set of 116 (104+12 practice items) high-familiarity and 104 low-familiarity words were selected. Thirty-six high- (24 experimental + 12 practice) and 24 (experimental) low-familiarity words were used for

baseline and 36 high (24 novel experimental and the same 12 practice) and 24 low novel experimental words were used for post-training spoken word recognition. An additional 56 high- and 56 low-familiarity words were used in the talker training task. There was no overlap of lexical items for any of these three tasks.

Words were selected for each part of the procedure to be balanced for lexical familiarity. In addition to familiarity, all words were coded for phonotactic probability as the sum of biphone frequencies plus one (Storkel, 2013; Storkel & Hoover, 2010). The range of phonotactic probabilities was 1.0002–1.0236 with a mean of 1.0078. Based on this mean value, words with a phonotactic probability greater than 1.0078 were coded as having a high phonotactic probability and those with less were coded as having a low phonotactic probability. The mean phonotactic probability for both pre-training spoken word recognition and post-training spoken word recognition was 1.0076. Phonotactic probability was examined for the high and low lexical familiarity words. An independent-samples t-test revealed no difference in phonotactic probability for the high versus low lexical familiarity words ($t(246) = 1.12, p = .26$). For the statistical analyses of the experimental tasks, phonotactic probability is included as both a continuous factor and a binary factor. Because this analysis was conducted after word selection based on lexical familiarity, the words are not necessarily evenly distributed for high and low phonotactic probability for each participant.

For the spoken word recognition tasks, stimuli were mixed with signal-dependent noise based on Schroeder (1968) and Benkí (2003), and using MatLab code from Felty (2007). Adding this type of signal-dependent noise to the original sound file results in a signal where each segment is masked to the same degree, rather than adding a uniform level of noise across the entire stimulus. The first twelve younger children completed the spoken word recognition tasks with easier signal-to-noise ratios (SNRs) of either 10 dB or 15 dB because of concern that the task would be too difficult at the less favorable SNR. Some of these children performed quite well during the pre-training spoken word recognition; thus, to avoid the possibility of ceiling effects, the remaining twenty-nine children completed the spoken word recognition tasks with an SNR of +5 dB.

Procedure

All children performed the following three tasks: pre-training (baseline) spoken word recognition, talker training, and post-training spoken word recognition. 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.

Talker training—Children completed five days of talker training in which they learned to identify the voices of three unfamiliar talkers, represented as cartoon-like characters, from words presented in the clear. Each day of training consisted of two learning sessions (with

feedback) and one test session (without feedback). Children were instructed that they would hear a single word and have to decide which of three characters produced the word by tapping the screen (Figure 1). During the learning sessions, children first had a familiarization phase in which they heard the same two high- and two low-familiarity 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 thirty trials (the same five high- and five low-familiarity words produced by all three talkers). During this phase, children heard a word and had to select which character had spoken the word. 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 learning 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. On the sixth day, children completed a generalization task with no feedback. This generalization task consisted of forty-eight trials: twenty-four of the trials contained OLD WORDS which had been heard during Day 1 of training (the same four high- and four low-familiarity words $\times$ three talkers) and twenty-four of the trials contained novel words (the same four high- and four low-familiarity words $\times$ three talkers). These data from the generalization task will not be analyzed in the current study. Over the course of training, children were presented with fifty-six distinct high- and fifty-six distinct low-familiarity lexical items produced by all three talkers as follows: same two high-/low-familiarity items during familiarization, five high-/low-familiarity items $\times$ five days of training for the learning tasks with feedback, five high-/low-familiarity items $\times$ five days of training for the test tasks without feedback, and an additional four high-/low-familiarity novel lexical items during generalization.

Pre-training (baseline) spoken word recognition—Prior to training, children completed a spoken word recognition task in which they heard words mixed with signal-dependent noise and were asked to say the word they heard. Because some of the low-familiarity words were likely to be unknown to the children, they were instructed that if they heard a word they did not know, to respond with what the person said. The experiment consisted of one practice block with twelve high familiarity words intended to acclimate the children to the noise, and two experimental blocks with a total of forty-eight distinct lexical items (twenty-four high and twenty-four low familiarity). Half of the lexical items were produced by the three talkers who would be learned during the talker training portion (FAMILIARIZED TALKERS) and half were produced by three talkers who would not be learned (UNKNOWN TALKERS). None of the lexical items used during spoken word recognition were used during talker training. One trained researcher transcribed phonetically children's responses at the time of testing. In addition, children's responses were recorded with a SHURE SM58 dynamic microphone with a flat frequency response from 50 to 15,000 Hz onto either a Marantz PMD-670 or Zoom H4n Handy. A second researcher listened to the recording and transcribed the responses. Any discrepancies between the two transcribers (one of whom was the author) were resolved by a third transcriber.

Post-training spoken word recognition—The post-training spoken word recognition task used the same procedure as the baseline task. An additional set of novel words (twenty-four high and twenty-four low familiarity) were used during post-training spoken word recognition. The same twelve high-familiarity words from the practice block of baseline spoken word recognition were used in the practice block of post-training spoken word recognition.

Coding

Responses for the talker training were coded as correct or incorrect, resulting in a percent correct for each session of training. Responses on the spoken word recognition task were coded for whole word accuracy and also for phoneme accuracy, allowing children to receive partial credit for their responses. For the phoneme analysis, the incorrect responses were matched to the target to give children the most credit for the response. For example, response *nick* [nɪk] for target *neck* [nek] or response *age* [edʒ] for target *sage* [sedʒ] were both coded as 2 phonemes correct. Additionally, response *close* [kloʒ] for target *loathe* [loð] would also be coded as 2 phonemes correct, for the [l] and the [o]. Thus, for each target, children could receive 0–3 phonemes correct.

RESULTS

Talker training

The first analysis examined whether children improved in identifying the talkers across the five days of training and whether their performance was affected by age or the lexical familiarity of the target items. A linear mixed-effects model (Baayen, 2008; Baayen, Davidson & Bates, 2008; Jaeger, 2008) was fit to the data using the *lmer()* function (Bates, Maechler, Bolker & Walker, 2010) in R (<http://www.r-project.org/>). The model included fixed effects for Lexical Familiarity (high, low), Day (1–5), and Age (in months) with random intercepts by subject. Only the test sessions without feedback were included in the analysis. Statistical significance was assessed using likelihood ratio tests (Baayen, 2008, p. 276). Likelihood ratio tests evaluate the change in goodness-of-fit when terms are added to a linear model. They compare the log likelihood, a measure of goodness-of-fit, of a linear model with the term of interest, to the log likelihood of a model without that term. The difference in log likelihoods can be evaluated for statistical significance against the chi-square distribution with degrees of freedom based on the difference in the number of parameters in each model. Using this type of model comparison, significant effects on talker identification accuracy were found for Day ($\chi^2(1)=40.20$, $p<.001$) and for Age ($\chi^2(1)=4.84$, $p=.027$), but not for Lexical Familiarity. The effect of Day is clearly visible in Figure 3, where talker identification accuracy gradually improves across the 5 days of training. To illustrate the effect of age, mean accuracy for the seventeen youngest children (range: 7;4 – 9;1: mean: 8;4) and the seventeen oldest children (range: 9;4–12;0; mean: 10;8) are also presented in Figure 3, although the statistical analyses were conducted on all forty-one participants with age as a continuous factor.

Spoken word recognition

For the spoken word recognition portion, analyses explored whether children's performance improved between Baseline and Post-training and whether this was mediated by lexical familiarity. As this question relates to whether children showed any improvement, the analyses were conducted for both the larger set of forty-one children that included twelve children with easier SNRs and also for the subset of twenty-nine children who completed the task with a +5 SNR. For the familiarized talkers, a logit mixed-effects model was fit to the whole word data with Time (Baseline, Post-training), Lexical Familiarity (high, low), and their interaction as fixed effects, and with random intercepts by subject. This model with the interaction was compared to an identical model but without the interaction term to test whether removal of the interaction term significantly changes the model fit. Logit mixed-effects models, rather than linear mixed-effects models were used because the whole word accuracy data is coded as binary. Comparisons of these models revealed a significant difference (all 41 participants: $\chi^2(1)=6.38$, $p=.011$; subset of 29 participants: $\chi^2(1)=3.35$, $p=.066$), indicating a significant interaction between Time and Lexical Familiarity as seen in Figure 4a. Post-hoc analyses of this interaction revealed that performance improved for familiarized talkers between baseline and post-training only for words with high lexical familiarity (all 41 participants: $p<.001$; subset of 29 participants: $p<.001$), but not for words with low lexical familiarity (all 41 participants: $p=.225$; subset of 29 participants: $p=.191$).

Linear mixed-effects models of the familiarized talkers were also fit to the phoneme accuracy data with and without the Time $\times$ Lexical Familiarity interaction term.

Comparisons of these models also showed a significant difference (all 41 participants: $\chi^2(1)=5.67$, $p=.017$; subset of 29 participants: $\chi^2(1)=5.84$, $p=.015$), indicating a significant interaction between Time and Lexical Familiarity. As above, post-hoc analyses of this interaction revealed improvement for familiarized talkers only for words with high lexical familiarity (all 41 participants: $p<.001$; subset of 29 participants: $p<.001$), but not for words with low lexical familiarity (all 41 participants: $p=.265$; subset of 29 participants: $p=.599$).

Analogous analyses for the unknown talkers revealed no statistically significant interaction between Time and Lexical Familiarity for either the whole word data ($p=.348$) or for the phoneme data ($p=.272$), as shown in Figure 4b. To check that the lack of an interaction was not the result of improvement for both high- and low-familiarity words, model comparisons were conducted between a model with fixed effects for Time and for Lexical Familiarity compared to a model without a fixed effect for Time, thus testing whether removal of Time significantly changed the model fit. As expected, there was no significant change in model fit by removing Time for the whole word data ($p=.366$) or for the phoneme accuracy data ($p=.738$). These results indicate that children did not show any improvement between Baseline spoken word recognition and Post-training spoken word recognition for the unfamiliar talkers, as shown in Figure 4b.

Additional analyses were conducted to examine whether the effect of lexical familiarity was due to differences in phonotactic probability. Recall that no significant difference in phonotactic probability between the high- and low-familiarity words was found, thus it is unlikely that phonotactic probability is the source of the lexical familiarity effect. A logit

mixed-effects model on the data for familiarized talkers with Time (baseline versus post-training), Phonotactic Probability (high, low), and their interaction as fixed effects and with random effects for subjects confirmed this and revealed no significant interaction ($p=.701$), indicating that high and low photoactic probability words did not result in different patterns of improvement. There was also no significant interaction when phonotactic probability was coded as a continuous measure ($p=.247$). Similar results were obtained for the phoneme accuracy data. Linear mixed-effects models of the phoneme accuracy data revealed no significant interaction for phonotactic probability, regardless of whether phonotactic probability was coded as binary ($p=.814$) or continuous ($p=.410$). These findings indicate that phonotactic probability does not mediate the familiar talker advantage. Both words with high and with low phonotactic probability are equally likely to show a familiar talker advantage.

Additional factors affecting the familiar talker advantage

Several correlations were conducted to explore which children exhibit the greatest familiar talker advantage. The amount of benefit to spoken word recognition (SWR) – that is, the familiar talker advantage – was operationalized in two different ways. The first way serves to normalize across children. This measure of the familiar talker advantage is calculated as the percent of actual improvement relative to possible improvement, and is provided in equation (i). This method has been used to analyze the benefit of visual information on spoken word recognition (Sumby & Pollack, 1954). As such, a child who improves from 60% to 80% correct, has improved 50% of her possible range of improvement. Similarly, a child who improves from 90% to 95%, has also improved 50% of the possible improvement. This measure of assessing amount of improvement has the benefit of not artificially inflating the familiarity benefit for children at lower levels of performance during the baseline task.

(i) Familiar Talker Advantage 1 (adapted from Sumby & Pollack, 1954)

$$100 \times \frac{(\text{post} - \text{trainingSWR} - \text{baselineSWR})}{(100 - \text{baseline SWR})}$$

While calculating the familiar talker advantage in (i) is useful for normalizing across children, the importance of the improvement is partially obscured. That is, a child who improves from 60% to 80% has made an important increase in performance, while a child who improves from 90% to 95% has not. Furthermore, the equation in (i) is most useful for data near ceiling, but in the current whole word data the best baseline performance is only around 60%, therefore all children have a large range for improvement. For these reasons, the familiar talker advantage was operationalized in a second way, with the more traditional method of simply using difference scores (Post-training SWR minus Baseline SWR). To avoid the potential confound that children with an easier SNR start at an artificially higher baseline spoken word recognition accuracy (Whole word: 58.8 versus 41.3; Phonemes: 81.2 versus 73.4) and would be less likely to show a familiar talker advantage, Pearson's correlations below were all conducted only for those twenty-nine children who performed the spoken word recognition task with a +5 dB SNR.

First, correlations were conducted between the talker identification accuracy and the familiar talker advantage. This was examined because Levi *et al.* (2011) found that talker identification accuracy correlated significantly with the familiar talker advantage. In addition, Nygaard and Pisoni (1998) found that only those listeners who were GOOD LEARNERS (attained at least 70% accuracy for talker identification) showed a familiar talker advantage. Talker identification accuracy was operationalized as the average talker identification accuracy for the last two days of training. Regardless of whether the familiar talker advantage was calculated as in (i) or as a difference score, talker identification accuracy was not significantly correlated with the familiar talker advantage (all $r < 0.12$, all $p > .50$).

Second, correlations were conducted between age and the familiar talker advantage. As above, regardless of whether the familiar talker advantage was calculated as in (i) or as a difference score, age was not significantly correlated with the familiar talker advantage (all $r < 0.15$, all $p > .410$).

Third, correlations were conducted between children's baseline performance for the to-be-familiarized talkers and the familiar talker advantage. When the familiar talker advantage was operationalized as in (i), there is a significant negative correlation between children's baseline spoken word recognition scores for familiar talkers and their Familiar Talker Advantage for both the whole word data ($r = -0.45$, $p = .013$, Figure 5a) and for the phoneme data ($r = -0.67$, $p < .001$, Figure 5b). The significant correlation in Figure 5b is driven in part by the outliers, those four children with negative familiar talker advantage values of less than -25 . When these data points are removed, the negative correlation is weakened and only approaches significance ($r = -0.35$, $p = .084$). Correlations between baseline performance and the familiar talker advantage, as a child's difference scores, were also conducted. For the whole word data, there was a significant negative correlation between children's baseline spoken word recognition scores and their Familiar Talker Advantage ($r = -0.595$, $p = .006$), as shown in Figure 6a. There was also a significant negative correlation for the phoneme data ($r = -0.718$, $p < .001$), as shown in Figure 6b. All of these negative correlations indicate that the children with the poorest performance on the spoken word recognition task during baseline showed the most improvement as a result of talker familiarization.

DISCUSSION

The current study examined whether school-age children show a familiar talker advantage, as has been found for adults. Children learned to identify the voices of three German-English bilinguals over five days of training. Prior to training, children completed a baseline spoken word recognition task with six speakers, three of whom they would learn to identify during training. Following training, they completed a post-training spoken word recognition task with novel words spoken by the same six speakers. Three major findings emerged from the data. First, children do show a familiar talker advantage, whereby spoken word recognition improves once children are familiarized with talkers. Related to this, children with the poorest performance at baseline showed the most benefit as a result of talker familiarization. Second, this familiar talker advantage was limited to high-familiarity words. Third, children did not show accent-general learning, as no improvement was found for the

unfamiliarized, unknown talkers who were also German–English bilinguals. Each of these findings is discussed below in more detail.

Familiar talker advantage in school-age children

The main goal of the current study was to determine whether school-age children show a familiar talker advantage, namely that familiarization with a talker's voice results in improved spoken language processing. The presence of a familiar talker advantage in this population indicates that children are able to take stored information about how a talker articulates various speech sounds and use that information to improve spoken language processing. That is, children are able to perceive phonetic detail in the speech signal, store this information, and harness this information when listening to spoken language. It is well known that adults' perceptions are highly flexible (Samuel & Kraljic, 2009), allowing them to modify their perceptions for different dialects or accents (Bradlow & Bent, 2008; Kraljic, Brennan & Samuel, 2008; Sidaras *et al.*, 2009) and idiolectal productions (Allen & Miller, 2004; Eisner & McQueen, 2005; Kraljic & Samuel, 2007; McQueen *et al.*, 2006; Norris *et al.*, 2003). Thus, the current study shows that children's perceptual spaces also exhibit plasticity, as they are able to adapt and improve their perceptions of familiar talkers.

In addition, three factors were examined to determine which children exhibit the most benefit from talker familiarity. Surprisingly, there was no correlation between how well the children learned the talkers' voices (talker identification accuracy) and the familiar talker advantage or between age and the familiar talker advantage. The difference between this study and previous studies that have found a relationship between talker familiarity and the familiar talker advantage may be due to how well the children learned the talkers' voices. Previous studies with adults used five to ten talkers during talker training, while the current study only used three, making the task itself easier. The children with the lowest scores for talker learning, who also tended to be the youngest children, were 60 – 75% correct. Therefore, the children may all have reached the necessary threshold for familiarity with the talkers.

In contrast to talker identification accuracy, children's baseline performance did correlate significantly with how much of a familiar talker advantage they displayed. These correlations indicated that those children with the lowest performance at baseline received the most benefit for being familiarized with the talkers. These results indicate that talker familiarity is a useful way to improve spoken language processing in children, especially those who have the most difficulty. This particular finding has broad implications for both children with language impairments and for adults with language processing problems. If spoken language processing can be improved in these populations, then cognitive resources that might have otherwise been used to understand the speech are freed up to be used for other tasks such as interpreting syntactic structure or semantic information. Future research will need to examine whether children with language impairments do actually show the same ability to take advantage of stored talker information to improve perception.

Markham and Hazan (2004) reported a similar finding where on-line attunement to talkers was most beneficial to listeners with the lowest performance on a baseline task. In their study, younger children (seven- to eight-year-olds), older children (eleven- to twelve-year-

olds), and adult listeners completed a spoken word recognition task with two conditions. In one condition, listeners heard a single word and were asked what they heard. In the second condition, listeners heard the precursor phrase ‘and now please say’ followed by a list of three words in the same voice as the precursor phrase. The benefit of perceptual attunement was calculated as the difference in performance between the no-precursor single-word condition and the precursor multiple-word condition. Many of the listeners were near ceiling in the no-precursor condition so no overall benefit was found. However, an additional analysis of the bottom quartile of listeners in each group showed a significant increase in accuracy in the condition with the precursor phrase, leading the authors to conclude that brief exposure to a talker is most beneficial for listeners who have the most difficulty with spoken word recognition.

The role of lexical familiarity in the familiar talker advantage

An additional finding of the current study was that the familiar talker advantage is limited to highly familiar words. Due to the surprising nature of this finding, additional analyses were conducted for previously collected data on adults (Levi *et al.*, 2011). These previously collected data did not include a pre-training baseline measure of spoken word recognition skills, so the analyses described below compare performance for familiarized and unknown talkers in post-training spoken word recognition only. In the adult data, words were originally coded as having high, medium, or low lexical FREQUENCY, instead of lexical FAMILIARITY. There was a high rate of overlap between these two measures: 67% of the high-familiarity words were coded as high frequency (90/133) and 67% of low-familiarity words were coded as low-frequency words (79/117). The analyses below were conducted only on the high- and low-frequency words, as these were the most likely to show a difference if one existed. Additionally, the analyses were conducted only on the GOOD LEARNERS – those participants who attained at least 70% accuracy during talker training – as these were the participants who actually showed the benefit of talker familiarity. For the adult whole word data, logit mixed-effects model were fit to the data with and without the Talker Familiarity (familiarized versus unknown talkers) by Lexical Frequency (high, low) interaction term. In contrast to the child data, this model comparison for adult listeners revealed no significant difference in model fit ($\chi^2(1)=.08, p=.771$), indicating that the interaction is not significant. Linear mixed-effects models were fit to the phoneme accuracy data with and without the interaction term Talker Familiarity $\times$ Lexical Frequency to examine whether lexical familiarity modulates the familiar talker advantage. Comparisons of these models also showed no significant difference in fit between them ($\chi^2(1)=.03, p=.842$). As is visible in Figure 7, both high- and low-frequency words show a similar advantage for familiarized talkers.

This additional analysis of the familiar talker advantage for adult listeners reveals a difference in the scope of the familiar talker advantage in children and adults. The question that arises is: What underlies this difference? Previous studies of sensitivity to acoustic-phonetic information in toddlers and in adults provide a possible explanation for these seemingly conflicting findings that is based on the nature of lexical representations in these two populations.

As discussed in the ‘Introduction’, there is evidence that children first process words holistically, and then shift to lexical representations that are more phonetically specified (Jusczyk, 1986, 1992; Walley, 1993). The evidence from how both toddlers and adults process words (and nonwords) point to the importance of lexical representations in this shift in processing. Studies with toddlers found that they are sensitive to phonetic detail in highly familiar words, but not for nonwords (Fennell & Werker, 2003; Stager & Werker, 1997; Swingley & Aslin, 2000; White & Morgan, 2008). Similar conclusions about sensitivity to phonetic detail can be drawn from van Heugten and Johnson’s (2014) study of adaptation to Australian-accented English where toddlers only showed attunement to the accent when the accent was presented with known lexical items (i.e. when children were familiar with the target words of *The Very Hungry Caterpillar* story). White *et al.*’s (2013) study showed an analogous difference in adults by using two sets of nonwords. One set was presented many times while the other was presented only a few times in the experiment. Adults were sensitive to phonetic errors for the high-frequency nonwords, which were likely stored, at least temporarily, in the adults’ mental lexicons, but were not sensitive to phonetic errors for the low-frequency nonwords, which either had no lexical representations or very weak lexical representations.

In light of these findings, the limitation of the familiar talker advantage in school-age children to highly familiar words is not surprising. When these highly familiar words are produced by a familiar talker, children are sensitive to fine-phonetic detail and integrate information about how these talkers produce different phonetic contrasts into their spoken word recognition. In contrast, when children are confronted with low-familiarity words (or nonwords in some cases), they are less sensitive to acoustic-phonetic detail and do not integrate knowledge of acoustic-phonetic detail about familiar talkers into spoken language processing. It is also clear why the familiar talker advantage in adults is not limited to high-frequency words; low-frequency words were simply not unfamiliar enough. Future studies with adults are needed to determine whether the familiar talker advantage would be absent in nonwords.

The finding that the familiar talker advantage is limited to high-familiarity words in children raises an additional issue about children’s spoken word recognition in particular, and their spoken language processing more generally. This result implies that talker familiarity would not enhance novel word learning, as familiarity with a talker is only beneficial for already acquired, highly familiar words.

No accent-general retuning for school-age children

An additional finding of the current study is that children do not exhibit an accent-general, talker-independent benefit for German–English bilingual talkers. That is, there was no improvement between baseline and post-training spoken word recognition for the UNFAMILIAR talkers. This result for children is surprising in light of studies that have found an accent-general benefit from exposure to multiple talkers with the same foreign accent. For example, Bradlow and Bent (2008) found that adult listeners are able to generalize knowledge of Chinese-accented talkers to a novel Chinese-accented talker after being exposed to multiple (five) Chinese-accented talkers during training. Similarly, Sidaras *et al.* (2009) found that

exposure to six Spanish-accented talkers resulted in improvement for novel Spanish-accented talkers for adult listeners. In the current study, children were exposed to three German-accented talkers and were also exposed to 112 different lexical items produced by all three familiarized talkers during training (56 high familiarity, 56 low familiarity) and did not exhibit an accent-general benefit in spoken word recognition. While the number of talkers used in the current study is less than in the two other studies of foreign-accented speech, a study of the effects of high-variability training on learning dialects used three talkers as the high-variability condition and did find generalization (Clopper & Pisoni, 2004). Furthermore, the talkers used during training showed a wide range of degree of foreign accent (Levi, Winters & Pisoni, 2007), from very little accent (2.9 for F8) to high accent (5.4 for F11), providing listeners with a wide range of accentedness, similar to Sidaras *et al.* (2009), where the accent ratings ranged from 2.83 to 5.59 for one group of talkers and 2.68 to 6.17 for the second group of talkers. Thus, it was expected that the three familiarized talkers would provide sufficient variability to support generalization to novel German–English talkers. Despite this range in foreign accent ratings, these talkers did not exhibit many segmental deviations, as indicated in the ‘Stimuli’ section. The lack of generalization to other German-L1–English-L2 talkers may be due to the few deviations from native norms. In addition to research on adults, Schmale and Seidl (2009) found that toddlers are able to generalize across foreign-accented speech produced by two Spanish-accented talkers. Given these previous results, it is unclear why children in the current study did not show some benefit for having been exposed to German-accented English.

Differences in whether listeners exhibit accent-general adaptation to foreign accented speech may result from differences in the experimental design. Schmale and Seidl (2009) examined whether toddlers could perceive similarity across spoken word productions in the clear, Bradlow and Bent (2008) presented sentences in noise in both the exposure phase and in the testing phase, and Sidaras *et al.* (2009) examined single word recognition with exposure and test in the clear. In the current study, exposure was in the clear, but the relevant testing condition was in noise. Research on memory encoding and retrieval processes reveals that performance in a test phase depends on the similarity between how information was initially encoded and how it is retrieved at test (Morris, Bransford & Franks, 1977; Tulving & Thomson, 1973). Therefore, acoustic differences between encoding and test in the current study could have made accent-general adaptation more difficult.

Sidaras *et al.*’s (2009) first experiment with sentence-length stimuli did mirror the current study in that exposure stimuli were in the clear and test stimuli were mixed with noise. In this latter case, the task that presumably generated familiarity with the accent was quite different. In fact, in both Bradlow and Bent (2008) and Sidaras *et al.* (2009), the tasks that presumably generated familiarity with the accent were very different from the task used here. In their studies, listeners attended to the linguistic content of the utterances during exposure (i.e. transcribed what they heard), making the test procedure identical to that of the exposure procedure. While focusing attention to the talkers’ voices as was done in the current study does not prevent facilitation of spoken word recognition, as evidenced by the existence of a familiar talker advantage, the lack of an accent-general perceptual change

may be due to the structure of the exposure phase and whether listeners' attention is directed towards a talker's voice or to the content of their speech.

It remains to be seen whether the lack of accent-general learning in school-age children is due to less variability during exposure in the current study (only three talkers), to aspects of the experimental design (noise versus clear; attentional focus during exposure), or to something different in how children and adults generalize across accented speech.

CONCLUSION

The current study shows that children use stored information about a talker to improve spoken word recognition, similar to what has been found for adult listeners. Importantly, children not only store talker information – as seen by the high level of talker identification accuracy – but also harness this knowledge during spoken word recognition. The benefit of talker familiarity was most noticeable in children with poor baseline performance, suggesting that talker familiarity may be especially useful for listeners who have difficulty understanding spoken language. This benefit represents a possible method for not only improving spoken word recognition (especially in those children who have the most difficulty), but also for possibly freeing up cognitive resources that could then be used for performing other cognitively demanding tasks. Future research is needed to explore this possibility. In the current study, the benefit of talker familiarity was limited to highly familiar lexical items and may indicate that children do not tap into stored familiarity with the talker unless there is already a phonetically specified lexical representation.

REFERENCES

- Allen JS, Miller JL. Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*. 2004; 115:3171–3183. [PubMed: 15237841]
- 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.
- Barker BA, Newman RS. Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*. 2004; 94:B45–B53. [PubMed: 15582622]
- 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. Online: <<http://lme4.r-forge.r-project.org>>
- Benkí J. Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition. *Journal of the Acoustical Society of America*. 2003; 113(3):1689–1705. [PubMed: 12656401]
- Bent T. Children's perception of foreign-accented words. *Journal of Child Language*. 2014 (forthcoming).
- Best, CT. Emergence of language-specific constraints in perception of native and non-native speech: a window on early phonological development. In: de Boysson, B.; de Schonen, S.; Jusczyk, P.; McNeilage, P.; Morton, J., editors. *Developmental neurocognition: speech and face processing during the first year of life*. Dordrecht: Kluwer; 1993. p. 289-304.
- Bradlow AR, Bent T. Perceptual adaptation to non-native speech. *Cognition*. 2008; 106(2):707–729. [PubMed: 17532315]

- Brown, L.; Sherbenou, RJ.; Johnsen, SK. TONI-3: test of nonverbal intelligence. 3rd ed.. Austin, TX: Pro-Ed; 1997.
- Carle, E. The very hungry caterpillar. New York, NY: Philomel Books; 1969.
- Clarke CM, Garrett MF. Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*. 2004; 116(6):3647–3658. [PubMed: 15658715]
- Clopper CG, Pisoni DB. Effects of talker variability on perceptual learning of dialects. *Language and Speech*. 2004; 47(3):207–239. [PubMed: 15697151]
- Cortese MJ, Khanna MM. Age of acquisition ratings for 3,000 monosyllabic words. *Behavior Research Methods*. 2008; 40(3):791–794. [PubMed: 18697675]
- Creel SC, Aslin RN, Tanenhaus MK. Heeding the voice of experience: the role of talker variation in lexical access. *Cognition*. 2008; 106:633–664. [PubMed: 17507006]
- Creel SC, Jimenez SR. Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*. 2012; 113:487–509. [PubMed: 22958962]
- Eisner F, McQueen JM. The specificity of perceptual learning in speech processing. *Perception & Psychophysics*. 2005; 67:224–238. [PubMed: 15971687]
- Felty, RA. Unpublished PhD dissertation. University of Michigan; 2007. Context effects in spoken word recognition of English and German by native and non-native listeners.
- Fennell CT, Werker JF. Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*. 2003; 46(2/3):245–264. [PubMed: 14748446]
- Goldinger SD, Pisoni DB, Logan JS. On the nature of talker variability effects in recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1991; 17(1):152–162.
- 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]
- Jesse, A.; McQueen, JM.; Page, M. The locus of talker-specific effects in spoken-word recognition. In: Trouvain, J.; Barry, WJ., editors. *International Congress of Phonetic Sciences (ICPhS 200J)*. Dudweiler: Pirrot; 2007. p. 1921-1924.
- Jusczyk, PW. Toward a model of the development of speech perception. In: Perkell, JS.; Klatt, DH., editors. *Invariance and variability in speech processes*. Hillsdale, NJ: Erlbaum; 1986. p. 1-19.
- Jusczyk, PW. Developing phonological categories from the speech signal. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. *Phonological development: models, research, implications*. Parkton, MD: York Press; 1992. p. 17-64.
- Kraljic T, Brennan SE, Samuel AG. Accommodating variation: dialects, idiolects, and speech processing. *Cognition*. 2008; 107:54–81. [PubMed: 17803986]
- Kraljic T, Samuel AG. Perceptual adjustments to multiple talkers. *Journal of Memory and Language*. 2007; 56:1–15.
- Ku era, H.; Francis, WN. *Computational analysis of present-day American English*. Providence, RI: Brown University Press; 1967.
- Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*. 1992; 255:606–608. [PubMed: 1736364]
- Ladefoged P, Broadbent DE. Information conveyed by vowels. *Journal of the Acoustical Society of America*. 1957; 29:98–104.
- 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, 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]
- 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]

- Markham D, Hazan V. The effect of talker- and listener-related factors on intelligibility for a real-word, open-set perception task. *Journal of Speech, Language, and Hearing Research*. 2004; 47:725–737.
- McLennan CT, Luce PA. Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31(2): 306–321.
- McQueen JM, Cutler A, Norris D. Phonological abstraction in the mental lexicon. *Cognitive Science*. 2006; 30(6):1113–1126. [PubMed: 21702849]
- Metsala JL. An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & Cognition*. 1997; 25(1):47–56. [PubMed: 9046869]
- 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]
- Morris CD, Bransford JD, Franks JJ. Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*. 1977; 16:519–533.
- Nelson PB, Soli S. Acoustical barriers to learning: children at risk in every classroom. *Language, Speech, and Hearing Research Services in Schools*. 2000; 31:356–361.
- Newman RS, Evers S. The effect of talker familiarity on stream segregation. *Journal of Phonetics*. 2007; 35:85–103.
- Norris D, McQueen JM, Cutler A. Perceptual learning in speech. *Cognitive Psychology*. 2003; 47:204–238. [PubMed: 12948518]
- Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. *Perception & Psychophysics*. 1998; 60(3):355–376. [PubMed: 9599989]
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. *Psychological Science*. 1994; 5(1):42–46. [PubMed: 21526138]
- Palmeri TJ, Goldinger SD, Pisoni DB. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1993; 19(2):309–328.
- Samuel AG, Kraljic T. Perceptual learning for speech. *Attention, Perception & Psychophysics*. 2009; 71(6):1207–1218.
- Schacter DL, Church BA. Auditory priming: implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1992; 18(5):915–930.
- Schmale R, Cristia A, Seidl A. Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental Science*. 2012; 15(6):732–738. [PubMed: 23106727]
- Schmale R, Seidl A. Accommodating variability in voice and foreign accent: flexibility of early word representations. *Developmental Science*. 2009; 12(4):583–601. [PubMed: 19635085]
- Schneider, W.; Eschman, A.; Zuccolotto, A. E-Prime 2.0 Professional. Pittsburgh, PA: Psychology Software Tools, Inc; 2007.
- Schroeder MR. Reference signal for signal quality studies. *Journal of the Acoustical Society of America*. 1968; 44:1735–1736.
- Semel, E.; Wiig, EH.; Secord, WA. Clinical Evaluation of Language Fundamentals. 4th ed.. Toronto, CA: Psychological Corporation/A Harcourt Assessment Company; 2003. (CELF-4)
- Sidaras 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]
- Sjerps MJ, Mitterer H, McQueen JM. Listening to different speakers: on the time-course of perceptual compensation for vocal-tract characteristics. *Neuropsychologia*. 2011; 49:3831–3846. [PubMed: 22001313]
- Spence MJ, Rollins PR, Jerger S. Children's recognition of cartoon voices. *Journal of Speech, Language and Hearing Research*. 2002; 45(1):214–222.
- Stager CL, Werker JF. Infants listen for more phonetic detail in speech perception tasks than in word-learning tasks. *Nature*. 1997; 388:381–382. [PubMed: 9237755]

- Storkel HL. A corpus of consonant-vowel-consonant real words and nonwords: comparison of phonotactic probability, neighborhood density, and consonant age of acquisition. *Behavior Research Methods*. 2013; 45(4):1159–1167. [PubMed: 23307574]
- Storkel HL, Hoover JR. An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods*. 2010; 42(2):497–506. [PubMed: 20479181]
- Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*. 1954; 26(2):212–215.
- Swingle D, Aslin RN. Spoken word recognition and lexical representation in very young children. *Cognition*. 2000; 76:147–166. [PubMed: 10856741]
- Tulving E, Thomson DM. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*. 1973; 80(5):352–373.
- van Heugten M, Johnson EK. Learning to contend with accents in infancy: benefits of brief speaker exposure. *Journal of Experimental Psychology General*. 2014; 143(1):340–350. [PubMed: 23506084]
- Vitevitch MS, Donoso A. Processing of indexical information requires time: evidence from change deafness. *Quarterly Journal of Experimental Psychology*. 2011; 64(8):1484–1493.
- Walley AC. The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*. 1993; 13(3):286–350.
- Werker JF, Tees RC. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*. 1984; 7:49–63.
- White KS, Aslin RN. Adaptation to novel accents by toddlers. *Developmental Science*. 2011; 14(2): 372–384. [PubMed: 21479106]
- White KS, Morgan JL. Sub-segmental detail in early lexical representations. *Journal of Memory and Language*. 2008; 59:114–132.
- White KS, Yee E, Blumstein SE, Morgan JL. Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language*. 2013; 68:362–378. [PubMed: 24065868]
- Yonan CA, Sommers MS. The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging*. 2000; 15(1):88–99. [PubMed: 10755292]

APPENDIX

Words with high lexical familiarity

Bake	chop	hair	leash	peace	such
Beach	Coin	ham	leave	pet	tail
Bean	comb	hang	lick	pill	then
Bear	come	hatch	line	rash	these
Beef	cop	hate	look	rat	thin
Beer	couch	have	lose	ride	thumb
Bib	date	head	mad	room	tight
Big	dead	heel	map	rough	wag
Bike	dish	here	match	run	web
Bit	duck	him	meat	sale	whale
Both	fair	hit	miss	sauce	what
Bug	fat	hoop	moose	seed	where
Buzz	fill	hot	mop	shake	white
Cab	fish	house	mouse	shape	win
Cake	fit	hug	neck	share	wing
Cash	foot	jail	need	shave	wish
Cat	fun	jar	net	should	with
Cave	gas	kill	nice	shout	write
Cheek	goose	kite	night	shut	wrong
Cheese	gun	knife	page	son	year
Choke	gym	leaf	pan	soon	

Words with low lexical familiarity

Bane	fate	hick	mime	rife	vole
Bid	faze	hitch	mock	rile	wade
Bile	feign	hock	mode	rut	wane
Cad	fib	hone	mope	sage	wean
Chafe	foul	jeer	muss	sake	whim
Char	gab	knack	neap	sate	whip
Chum	gape	lass	nick	sheath	whiz
Cog	gauge	laud	node	shuck	wick
Cope	gauze	leach	noose	siege	wit
Core	gawk	ledge	notch	souse	womb
Couth	gig	loathe	peal	sup	wraith
Cuff	gin	loom	peer	tithe	wrath
Cull	gnash	lore	pith	tone	wreath
Dame	gnat	luff	poach	tout	zeal
Debt	goad	lush	puck	vague	
Dock	goon	mace	raid	vain	
Doff	gull	maim	reap	vat	
Dose	hail	mauve	reek	veil	
Dung	haze	mesh	retch	void

Julia

Lisa

Anne

Fig. 1.
Response screen for Talker Training.

Fig. 2.
Sample trial procedure during the five days of training.

Fig. 3. Talker identification accuracy during the five days of training for the seventeen youngest and seventeen oldest children.

Fig. 4. a (top) and 4b (bottom) Average percent correct for whole words and phonemes for Familiarized talkers (a) and Unknown talkers (b), separated by High versus Low Lexical Familiarity during baseline spoken word recognition (white bars) and Post-training spoken word recognition (gray bars). Error bars represent 95% confidence intervals around the mean ($n=41$).

Fig. 5. Scatter plots with each listener's Baseline spoken word recognition accuracy for familiarized talkers on the *x*-axis and the Familiar Talker Advantage (as in (i)) on the *y*-axis. (a) (top) shows these values for whole words correct, and (b) (bottom) for phonemes correct ($n=29$).

Fig. 6. Scatter plots with each listener's Baseline spoken word recognition accuracy for familiarized talkers on the x -axis and the degree of Familiar Talker Advantage (Post-training accuracy – Baseline accuracy) on the y -axis. (a) (top) shows these values for whole words correct, and (b) (bottom) for phonemes correct ($n=29$).

Fig. 7.

Average adult spoken word recognition accuracy for whole words correct and for phonemes correct, separated by High versus Low Lexical Frequency for unknown talkers (white bars) and familiarized talkers (gray bars). Error bars represent 95% confidence intervals around the mean ($n=20$).

TABLE 1

Information on familiarized and unknown talkers

Talker	Intelligibility	Fo	AoA	Years of English	Length of residence	Proficiency	Degree of foreign accent (raw)	Degree of foreign accent (z-scores)
Familiarized Talkers	F3	49.1%	222	10	1	5	4.1	0.22
	F8	54.7%	170	13	4	5	2.9	-0.27
	F11	41.2%	260	-	2	4.5	5.4	0.94
Unknown Talkers	F4	43.8%	186	13	3	4.5	3.5	0.02
	F9	48.4%	202	9	16	4	3.9	0.13
	F12	54.7%	186	7	26	5	-	-

NOTES: 'Intelligibility' refers to the average number of words correctly identified across three signal-to-noise ratios (+10, +5, 0 dB) and a clear listening condition (Levi, Winters & Pisoni, 2011). 'Fo' is the average fundamental frequency in hertz at the vowel midpoint for the 360 English words. 'AoA' is the age of acquisition of English. 'Years of English' refers to the number of years speakers have been learning/using English (age at test - age of acquisition). 'Length of residence' is how long they have lived in the US. 'Proficiency' is a self-reported measure of English proficiency (1=poor, 5=fluent). 'Degree of foreign accent' provides the mean foreign accent rating on a 1-7 Likert scale and also z-scores (Levi, Winters & Pisoni, 2007). Larger raw ratings and z-scores reflect a higher degree of foreign accent. Talker F12 was not included in the study of degree of foreign accent.