Research Article

Adults Show Initial Advantages Over Children in Learning Difficult Nonnative Speech Sounds

Pamela Fuhrmeister,[a] Brianna Schlemmer,[a] and Emily B. Myers[a]

Purpose: Children and early adolescents seem to have an advantage over adults in acquiring nonnative speech sounds, supported by evidence showing that earlier age of acquisition strongly predicts second language attainment. Although many factors influence children's ultimate success in language learning, it is unknown whether children rely on different, perhaps more efficient learning mechanisms than adults.
Method: The current study compared children (aged 10–16 years) and adults in their learning of a nonnative Hindi contrast. We tested the hypothesis that younger participants would show superior baseline discriminability or learning of the contrast, better memory for new sounds after a delay, or improved generalization to a new talker's voice. Measures of phonological and auditory skills were collected to determine whether individual variability in these skills predicts nonnative speech sound learning and whether these potential relationships differ between adults and children.
Results: Adults showed superior pretraining sensitivity to the contrast compared to children, and these pretraining discrimination scores predicted learning and retention. Even though adults seemed to have an initial advantage in learning, children improved after a period of off-line consolidation on the trained identification task and began to catch up to adults after an overnight delay. Additionally, perceptual skills that predicted speech sound learning differed between adults and children, suggesting they rely on different learning mechanisms.
Conclusions: These findings challenge the view that children are simply better speech sound learners than adults and suggest that their advantages may be due to different learning mechanisms or better retention of nonnative contrasts over the broader language learning trajectory.
Supplemental Material: https://doi.org/10.23641/asha.12735914

A central debate in second language research focuses on age-related constraints on acquisition, especially for acquiring the speech sounds in a second language. Although many experiential and environmental factors influence second language perception and production, the age at which a second language learner begins learning the target language is a robust predictor of ultimate attainment in the production of second language speech sounds (e.g., Flege et al., 1995, 1999; Granena & Long, 2013; Piske et al., 2001). Because of these strong age effects, many researchers have posited a critical or sensitive period for language acquisition (e.g., Johnson & Newport, 1989), and acquisition of the speech sounds of a language seems to be particularly susceptible to these effects (Granena & Long, 2013; see Werker & Hensch, 2015, for a review). Adults often exhibit persistent difficulties with perception and production of nonnative speech sounds, even after decades of language use (Flege et al., 1995) and following targeted perceptual training (e.g., Bradlow et al., 1999, 1997).

At the same time, this idea of hard developmental constraints on nonnative sound learning has not gone unchallenged. Several studies have tested this hypothesis and found no differences between adults and children or that adults were actually better at learning to perceive or produce second language speech contrasts (at least initially) in naturalistic (i.e., real world) or laboratory settings (Aoyama et al., 2004; Heeren & Schouten, 2010; Snow & Hoefnagel-Höhle, 1977, 1978; Wang & Kuhl, 2003). A perceptual learning study by Wang and Kuhl (2003) tested adults and children aged 6, 10, and 14 years and found that the older children were, the better they performed on their tasks measuring learning of a nonnative tonal contrast, with adults performing best. Heeren and Schouten (2010) tested

[a]Department of Speech, Language and Hearing Sciences, University of Connecticut, Storrs
Correspondence to Pamela Fuhrmeister: pamela.fuhrmeister@uconn.edu
Editor-in-Chief: Bharath Chandrasekaran
Editor: Chao-Yang Lee
Received November 25, 2019
Revision received March 30, 2020
Accepted May 26, 2020
https://doi.org/10.1044/2020_JSLHR-19-00358

Disclosure: The authors have declared that no competing interests existed at the time of publication.

the perceptual learning of 11- to 13-year-old children and found that they showed similar amounts and rates of learning as adult participants but that adults showed superior discrimination abilities.

These laboratory studies seem to suggest that adults may actually be better than children at learning nonnative speech sounds. However, these studies (e.g., Heeren & Schouten, 2010; Wang & Kuhl, 2003) have used phonetic contrasts (a geminate contrast and a tonal contrast, respectively) that were fairly dissimilar to participants' native language categories. Research suggests that the greater the similarity to native language categories, the more difficult learning will be, raising the possibility that perhaps the developmental advantage for nonnative learning will be more apparent in contrasts that are most difficult to learn (e.g., Best et al., 2001).

One contrast that has often been used to illustrate the difficulty of nonnative learning is the Hindi dental/retroflex contrast. The voiced dental and retroflex stop consonants in Hindi are especially challenging for native speakers of English to discriminate because they are perceptually similar to the voiced alveolar stop in English (Best et al., 2001). In addition, the voiced alveolar stop category in English encompasses allophonic variants that are produced with dental and retroflex points of articulation, as in the word "width" or "dry." This makes the perception of these sounds as independent categories especially challenging because native English speakers are accustomed to perceiving them as instances of the voiced alveolar stop category. By using this particular contrast in the current study, we test the hypothesis that children will show superior initial sensitivity or learning for sounds that are perceptually similar to a native language category as compared to adults or that sensitivity or learning decreases with increasing age. Even though some laboratory tests of nonnative learning have not favored younger learners, real-world differences in language attainment associated with younger age of arrival are well documented in the literature (Flege et al., 1995, 1999), and an important question is why more adults do not achieve native-like perception and production of second language speech sounds. To address this question, it is crucial to further elucidate the differences between the learning processes in adults and children.

There are several reasons why adults and children may approach the process of learning nonnative speech sounds differently. First, the brain may be less plastic in adulthood (see Werker & Hensch, 2015, for a review), which may impose constraints on how easily adults can learn new speech sounds. Second, extensive experience with the native language may inhibit the process of learning new sounds. Specifically, adults have a well-developed first language sound system, and this makes the process of learning additional sounds from another language more difficult because adults often perceive unfamiliar speech sounds through the filter of their native language speech categories (e.g., Best & Tyler, 2007). Indeed, most theories of nonnative speech sound learning attribute the difficulty of this process to the perceptual similarity of nonnative sounds to well-established existing native language speech sounds (e.g., the perceptual assimilation model, Best et al., 2001; Best & Tyler, 2007; the native language magnet model, Kuhl, 1994; Kuhl et al., 2007). Although these models differ in their details, each account is that listeners fail to perceive differences between nonnative speech sounds when they fall within the territory occupied by a native language speech category. This causes the listener to assimilate similar nonnative sounds to existing native language categories, and these nonnative sounds then get perceived as variants of the native language category, which makes it difficult to perceive them as distinct categories.

Children, on the other hand, do not have native language speech categories as fully developed or robust as those of adults. It is well established that infants perceptually reorganize their speech categories to reflect the ambient language by the end of the first year of life (e.g., Kuhl et al., 2006; Werker & Tees, 1984), but there is also evidence that some aspects of native language speech continue to develop throughout childhood and even adolescence (see Zevin, 2012, for a review). One way this has been demonstrated is by shallower categorization functions of native language speech sounds as compared to adults (Burnham et al., 1991; Hazan & Barrett, 2000). Hazan and Barrett (2000) found that children's native language categories continue to develop between the ages of 6 and 12 years as well as between the age of 12 years and adulthood, and more recent work has even found that children's native language speech representations continue to develop between the ages of 12 and 18 years (McMurray et al., 2018). Although some aspects of native language speech perception, such as weighting of certain acoustic cues, may be adultlike by the age of about 7 or 8 years (e.g., Nittrouer, 2004), weighting of other secondary acoustic cues is still not adultlike by the age of about 8.5 years (Idemaru & Holt, 2013). Empirical findings additionally suggest that children (aged 7–14 years) do not tend to assimilate nonnative speech sounds to native language categories as readily as adults do (Baker et al., 2008). On the whole, however, there is evidence that some aspects of native language speech perception are still developing through adolescence.

Taken together, these findings lend support to the notion that less developed native language speech categories are more malleable and will allow children to perceive nonnative speech sounds as distinct categories, rather than assimilate them to native language categories. This generates the hypothesis that children, by virtue of their less developed or malleable native language categories, will be more sensitive to subtle acoustic differences in (nonnative) speech stimuli and that this sensitivity may give them an advantage in acquiring the speech sound inventory in a second language.

Another reason why nonnative speech sound learning might look different in children compared to adults is that children may rely on different skills to accomplish this task. Specifically, a few studies have found relationships between auditory or phonological skills (in the native language) and various tasks involving nonnative speech among adult learners. For instance, Earle and Arthur (2017) tested adult learners on a sound blending task and a nonword repetition task from the Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 1999). In a sound blending task, participants hear a word presented one sound at a time (e.g., /k/ – /æ/ – /t/) and are asked to say the whole word. In a nonword repetition task, participants hear phonotactically legal nonwords and have to repeat them out loud. These two tasks, sound blending and nonword repetition, are measures of phonological skills that have been used in concert with other tests to diagnose reading and language disorders. Earle and Arthur found that sound blending positively predicted nonnative speech sound discrimination following training and a period of off-line consolidation. They also found that nonword repetition was related to performance on a posttraining (and postconsolidation) nonnative identification task.

Perrachione et al. (2011) similarly observed that higher scores on a sound blending task (from the Woodcock-Johnson III Tests of Cognitive Abilities [WJ-III COG]; Woodcock et al., 2001) predicted learning of a nonnative Mandarin tonal contrast. However, the most robust predictor of nonnative learning in this study was a pitch contour perception task, suggesting that auditory discrimination abilities rather than phonological skills best predict nonnative speech sound learning, at least for pitch-driven contrasts. MacKay et al. (2001) found that scores on a nonword repetition task of Italian nonwords predicted performance on a task involving English consonant identification in noise performed by native speakers of Italian. Results from these studies suggest that phonological skills in the native language, or auditory skills more generally, predict nonnative speech sound learning in adults; however, an open question is whether these relationships are present in child learners as well. For instance, it may be that adult learners approach nonnative phonetic learning through the lens of their native language, relying heavily on metalinguistic phonological skills, whereas younger learners are less affected by the structure of their native language. In the long run, relying on metalinguistic phonological skills may not be the optimal path for learning nonnative speech sounds.

One way to reconcile studies showing superior performance by adults in laboratory-based learning situations but better attainment by children in real-world language learning is to appeal to differences in consolidation or maintenance of learned phonetic information. A line of work from our laboratory shows that individuals vary substantially in their ability to retain learned phonetic information over an interval, and individual differences in ultimate attainment in phonetic learning may be linked to sleep-related consolidation processes (see Earle & Myers, 2014, for a review). It is possible that children may show poorer learning within the session but superior consolidation and retention of learned phonetic information.

The current study tests whether adult and child learners show comparable initial learning and retention (after a delay of 1 day) of a difficult nonnative phonetic contrast. For this study, we were particularly interested in sampling a wide age range of participants to span the ages that have been suggested to encompass the sensitive period for second language learning. Although some individuals who began learning a second language even as early as at 5 or 6 years old do not become native-like in their second language pronunciation (e.g., Granena & Long, 2013), we recruited participants with a minimum age of 10 years to ensure child participants could complete the same tasks as the adults (see Zevin, 2012, for a discussion on the difficulties of testing adults and young children with the same tasks). Additionally, a few studies have shown that some individuals who began learning a second language around the age of 10 years had achieved native-like production of the target language (Flege et al., 1995, 1999), and in these studies, individuals who began learning a language in adolescence typically have better speech production outcomes than even those who begin learning as young adults. We were also interested in testing adults older than the typical age range of laboratory studies examining the influence of age on nonnative speech sound learning (e.g., Heeren & Schouten, 2010; Wang & Kuhl, 2003), so we recruited participants between the ages of 10 and 60 years. This allowed us to test whether nonnative speech sound learning abilities change throughout adulthood as well.

The perceptual assimilation model (Best et al., 2001; Best & Tyler, 2007) and the native language magnet model (Kuhl, 1994) predict that nonnative speech sound learning is difficult because listeners assimilate perceptually similar nonnative speech sounds to existing categories. Because adolescent children's first language speech categories are still developing, we predict that they will show greater sensitivity to subtle acoustic differences in nonnative speech stimuli (i.e., they will be less likely to assimilate nonnative sounds to native categories), as indicated by more accurate discrimination and identification of the nonnative speech sounds as compared to adults.

Finally, we ask whether native language phonological skills or auditory skills (as measured by a pitch perception discrimination task) predict nonnative speech sound learning in adults and children. If we do not see the same relationships between phonological skills and nonnative speech sound learning in children that have been observed with adult learners, this would support a model of nonnative speech sound learning in which children and adults rely on different learning mechanisms and that those mechanisms change with age. However, if we see the same relationships between phonological or auditory skills in both adults and children, this would suggest that such skills are necessary for learning nonnative speech sounds and age is not the most critical predictor of learning.

Method

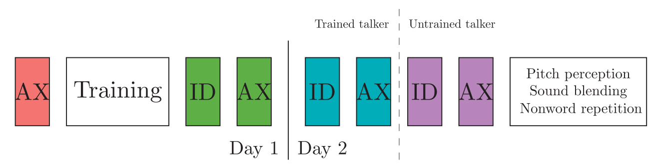
Participants

Forty-seven participants between the ages of 10 and 59 years were recruited from the University of Connecticut community and the UConn KIDS (University of Connecticut Kids in Developmental Science) research database. The sample included 23 children[1] (13 boys, 10 girls; age range: 10–16 years, *M* = 12.5, *SD* = 1.79) and 24 adults (three men, 21 women; age range: 18–59 years, *M* = 37.8, *SD* = 11.9). The experiment was advertised to native speakers of American English, and all participants passed a 20–dB HL hearing screening at 500, 1000, 2000, and 4000 Hz. Two child participants reported having learned some French in school (one of those participants had a caregiver who spoke French, but the participant was not fluent), one reported learning some Spanish in school, and one had been exposed to Spanish in a naturalistic context from birth to 10 months of life. One adult participant reported fluency in French, one reported fluency in American Sign Language, and one learned some Spanish in school. One child participant's data were removed from the analyses because the data file from one of the sessions was corrupted and did not save the data. One adult participant's data were removed from the analyses because the participant reported fluency in a language that contains the dental/retroflex contrast that was learned in this study. Data from the remaining 45 participants were processed for further analyses. Participants received $10 per hour for their participation and gave informed consent (with children giving written assent and obtaining parental permission) in accordance with the guidelines of the University of Connecticut Institutional Review Board.

Procedure

Participants participated in two sessions on consecutive days in order to test immediate learning and retention after an interval that allows for consolidation of the learned information (see Earle & Myers, 2014). Both visits occurred between 3:00 p.m. and 6:00 p.m. (see Figure 1) to produce an approximate 24-hr interval between training and test. In the first session, participants completed an AX discrimination pretest, an identification training to learn the Hindi contrast, and posttraining identification and AX discrimination assessments. In the second session, we assessed retention of participants' learning of the contrast with identification and AX discrimination tests. Participants then completed additional identification and AX discrimination tests with a novel speaker (the female speaker) to measure the generalizability of the trained contrast to an untrained talker's voice. Tests of generalization to a new talker were only administered on the second day, as exposure to phonological variability in the form of testing for generalization has been shown to diminish overnight improvement (Fuhrmeister & Myers, 2017). This assessment was followed by a pitch perception task to measure auditory discrimination skills and two standardized subtests of phonological skills: Nonword Repetition from the CTOPP (Wagner et al., 1999) and Sound Blending from the WJ-III COG (Woodcock et al., 2001).

Figure 1. Schematic of the procedure and tasks. AX = AX discrimination test; ID = identification test.



Training and Assessments

AX discrimination assessment. For the AX discrimination assessment, participants heard two presentations of the minimal pair nonwords (/ɖig/ or /d̪ig/) on each trial and were asked to indicate whether the initial sounds of each nonword were the same sound or different sounds. Participants completed 64 trials: On half of the trials, the sounds came from the same speech category, and for the other half of the trials, the sounds came from different categories. Same trials always included different instances of the same nonwords so that participants could not rely on low-level acoustic details of the stimuli to categorize the sounds.

Training. To learn the Hindi sounds, participants completed a two-alternative forced-choice identification training task, in which they learned that each of the minimal pair nonwords corresponded to a novel visual object. To familiarize participants with the pairings, each visual stimulus was shown on the screen while participants listened to five repetitions of the nonword that corresponded with the picture. For the training task, participants saw both visual stimuli and heard a nonword (/ɖig/ or /d̪ig/) and were asked to select the picture that correctly corresponded to the word they heard. Visual feedback ("Correct" or "Incorrect") was provided with each response, and training consisted of 200 trials with a 30-s break after the first 100 trials.

Identification assessment. The identification assessment contained 50 trials just like the training task, except no feedback was given.

Generalization assessments. To assess the generalizability of the learning that took place during the first session, the second session presented participants with a novel talker (a female native speaker of Hindi). A familiarization period similar to that of the initial identification training was given to participants where each novel visual stimulus was presented with five presentations of its corresponding auditory stimulus (either /ɖig/ or /d̪ig/). After familiarization, participants completed identification and AX discrimination assessments as described above.

Pitch perception. To measure auditory discrimination ability, participants completed an adaptive pitch perception discrimination task. The task was presented using MATLAB

[1]Three child participants reported a history of reading difficulties or dyslexia; however, all three scored well within the range of scores from our sample on the nonword repetition and sound blending tasks (raw scores of 9, 12, and 12 for nonword repetition, range: 6–16; raw scores of 25, 26, and 26 for sound blending, range: 18–33).

(The MathWorks, Inc.) and Psychtoolbox (Brainard & Vision, 1997). Participants completed a total of 50 trials. On each trial, participants heard two pure-tone pitches and were instructed to indicate whether the second tone presented was higher or lower than the first by pressing either the up or the down arrow key on a keyboard. The first tone served as a reference tone and was always 500 Hz. The pitch of the second tone was determined by participants' performance on the two-down, one-up adaptive staircase task, so that the difference in pitch of the two tones would decrease (become more difficult) after two consecutive correct responses. Each time a participant gave an incorrect response, the pitch difference would increase. Prior to the actual task, participants were given a short practice version of the task (five trials) to ensure they understood what to do. The pitch difference for the practice trials was larger (250 Hz, a 50% difference). For the actual task, the first two trials contained a pitch difference of 25 Hz (a 5% difference, or about a semitone), and depending on an individual's performance on the task, the pitches could contain a subsequent difference of 20 Hz (a 4% difference), 15 Hz (a 3% difference), 10 Hz (a 2% difference), 5 Hz (a 1% difference), 2.5 Hz (a 0.5% difference), or 1.25 Hz (a 0.25% difference).

Standardized measures. To measure native language phonological skills, the Nonword Repetition subtest from the CTOPP (Wagner et al., 1999) and the Sound Blending subtest from the WJ-III COG (Woodcock et al., 2001) were administered (order was counterbalanced) by the first and second authors and trained research assistants. In a nonword repetition task, participants hear a phonotactically legal nonword and are asked to repeat it out loud. For sound blending, participants hear a word presented one sound at a time and are instructed to say the whole word (e.g., /k/ – /æ/ – /t/ = "cat"). Administration of assessments followed protocols found in the examiner manuals, and assessments were audio-recorded and double-scored by the second author. Recordings of four participants' assessments (two children and two adults) were lost due to equipment or experimenter error and were therefore only scored once.

Stimuli

Auditory and visual stimuli for nonnative speech training and assessments were presented with OpenSesame 3.1.5 (Mathôt et al., 2012). Auditory stimuli were recorded by two native speakers of Hindi (one male talker, one female talker). Stimuli from the male talker were used in a previous study (Earle & Myers, 2015), and more details of stimulus creation can be found there. Stimuli produced by the female talker were recorded in a soundproof booth with a Roland R-05 digital voice recorder. Each talker produced five different exemplars of minimal pair nonwords /ɖig/ and /d̪ig/, and these were scaled to a mean amplitude of 70 dB SPL in Praat (Boersma & Weenink, 2018). Two different novel objects (Fribbles) were used as visual stimuli (stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University; http://www.tarrlab.org/). Pure-tone stimuli for the pitch perception task were created in Praat (Boersma & Weenink, 2018), with frequencies of 250, 475, 480, 485, 490, 495, 497.5, 498.75, 500, 501.25, 502.5, 505, 510, 515, 520, 525, and 750 Hz. All auditory stimuli were presented over headphones (Sony MDR-7506) at a comfortable listening level, adjustable by participants. Participants indicated responses by pressing the denoted keys on a keyboard.

Analysis Approach

Pitch perception task. To obtain a score for each participant's pitch discrimination ability, we took the mean of each participant's inflection points (a directional change) on the adaptive staircase. This value represents the average interval in hertz that the participant can reliably discriminate. Raw scores were scaled and centered and entered as fixed effects into the analyses, as described below.

Standardized measures. Raw scores for nonword repetition and sound blending tasks for each participant were scaled and centered and entered into analyses, as described below. Scores were not age-normed in order to capture maturational differences in these measurements.

Nonnative measures. To account for response bias, *d'* scores were calculated for discrimination (MacMillan & Creelman, 2005). Discrimination performance of the trained talker was analyzed using linear mixed-effects models (discrimination data of the untrained talker were analyzed with linear regression because there were no repeated measures), and mixed-effects logistic regression models were used to analyze identification data. All mixed-effects models were performed in R (R Development Core Team, 2008) using the lme4 package (Bates et al., 2015). The *p* values for linear mixed-effects models were estimated with the afex package using the Satterthwaite approximation (Singmann et al., 2019). For analyses of identification data, the random effects structure was determined by a backward-stepping procedure (Matuschek et al., 2017). In tasks of identification when no feedback is given, participants occasionally switch the category labels even when they can indeed differentiate the categories. To accommodate this, we performed a binomial test that determined that the probability of obtaining a score below 38% accuracy on tests of identification was less than chance ($p < .05$). For participants who scored below this threshold, data were recoded (i.e., 0 was recoded as 1, and 1 was recoded as 0), and this affected two participants: one child participant on both identification posttests of the trained talker and one adult participant on the identification posttest of the untrained talker. Raw data and analysis scripts can be found at https://osf.io/tfmg8.

Results

Relationships Between Age and Nonnative Learning

Pretraining Sensitivity

Does participant age predict pretraining discrimination ability? To test whether age predicted pretraining sensitivity to the contrast, we carried out a regression model with the *d'* score as the dependent variable and age as a predictor variable. The overall model fit was significant, $R^2 = .33$, $F(1, 43) = 21.35$, $p < .001$, and age was a significant predictor of baseline discrimination of the contrast, $\beta = .03$, $SE = 0.01$, $t = 4.62$, $p < .001$ (see Figure 2A). This indicates that baseline discrimination performance increased as age increased, which does not support the prediction that younger participants would show better pretraining discrimination.

Learning and Retention

Do adults and children identify and discriminate the sounds at above-chance levels after training? To test whether participants' performance on the identification and discrimination posttests was above chance, we ran an intercept-only model for each time point. All intercepts were significantly larger than zero, indicating that participants learned and retained above chance. We fit the same models with adult and child data separately to ensure that the above-chance performance was not being driven by one age group, and these models additionally revealed that children and adults learned and retained the phonetic contrast at above-chance levels for both identification and discrimination tasks. Details on the models and output can be found in Supplemental Material S1.

Does participant age predict identification accuracy? To test whether participant age predicted identification accuracy, we ran a mixed-effects logistic regression model. The model predicted a correct response on each trial (0 or 1), and fixed effects included time (immediate posttest and next-day posttest) and age. The random effects structure in the final model included by-subject random intercepts and slopes for time, and correlations of random effects were set to zero. The factor of time was deviation coded (immediate posttest = −.5, next-day posttest = .5). The model revealed a significant difference in the two time points, $\beta = .57$, $SE = 0.25$, $z = 2.28$, $p = .02$, such that performance was better at the next-day posttest than at the immediate posttest. Age was a significant predictor of identification performance, $\beta = .03$, $SE = 0.01$, $z = 3.10$, $p = .002$, such that increasing age predicted better identification accuracy; however, there was a significant interaction of age and time, $\beta = -.02$, $SE = 0.009$, $z = -1.99$, $p = .05$, which indicates that the relationship between age and identification performance was stronger at the immediate posttest than at the next-day posttest (see Figure 3A).[2] This may suggest that children were beginning to catch up to the adults after a period of off-line consolidation (see Figure 3A). We explore this further in the next analysis.

Does pretraining sensitivity to the contrast explain the relationship between age and identification performance? Because we found that age positively predicted pretraining discrimination sensitivity of the contrast, we wanted to test whether the effect of age in the previous model of identification data was driven entirely by pretraining sensitivity to the contrast. To test this, we fit another mixed-effects logistic regression model. This model predicted accuracy and included fixed factors of time (coded as in the previous model), age, and discrimination pretest scores. Random effects in the final model included by-participant random intercepts and slopes for time, and correlation parameters were set to zero. To get the model to converge, we used the glmerControl optimizer bobyqa and increased the iterations to 200,000. This model revealed only a significant interaction between age and discrimination pretest scores, $\beta = .03$, $SE = 0.01$, $z = 2.46$, $p = .01$, suggesting the relationship between pretraining discrimination and identification posttest performance changes as a function of age. To unpack this interaction, we split the factor age into a dichotomous variable with levels children (participants under the age of 18 years) and adults (participants 18 years

Figure 2. (A) Relationship between age and performance on the discrimination pretest. Age was a significant predictor of performance on the discrimination pretest. The shaded region represents 95% confidence intervals. (B) Discrimination performance on trained and untrained talkers at each time point. For visualization purposes only, children (under 18 years old) and adults (18 years old or older) are plotted separately. Error bars represent 95% confidence intervals. Please note the difference in scales on the *y*-axis.

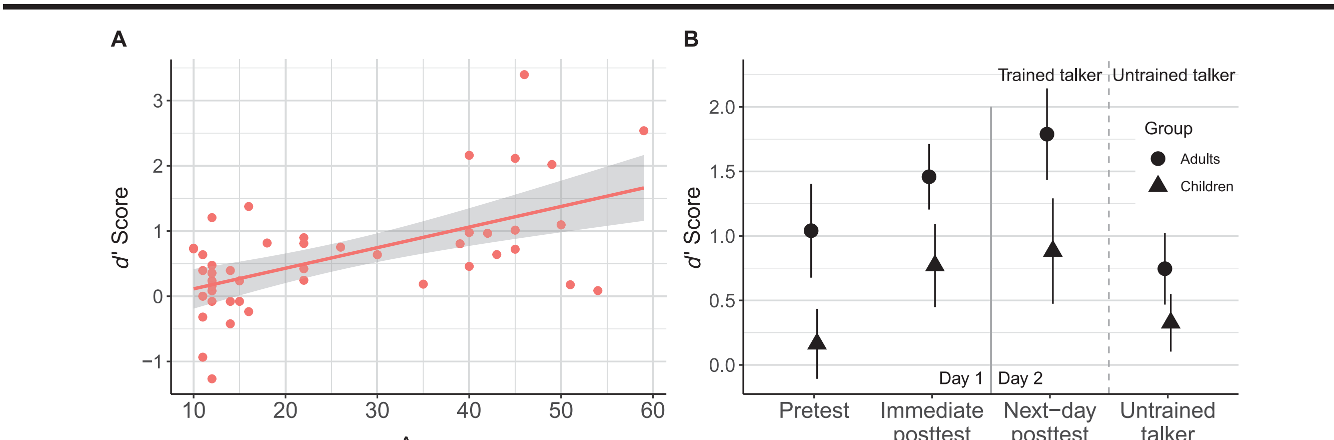
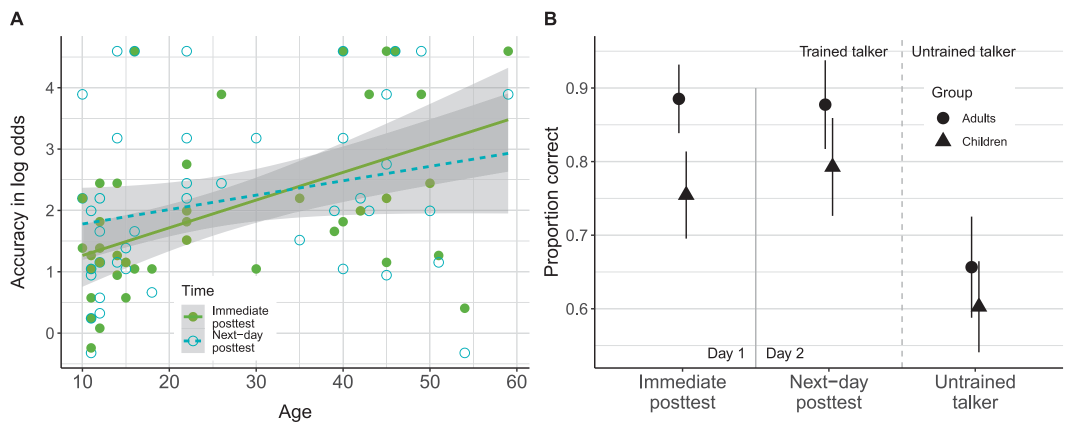


Figure 3. (A) Relationship between age and performance on the identification posttests. Age was a significant predictor of performance on both posttests, but the relationship was stronger on the first day. For purposes of visualization only, mean percent accuracy was converted into log odds, and any accuracy scores of 1 were changed to .99 to avoid infinite values. Children (under 18 years old) and adults (18 years old or older) are plotted separately. The shaded region represents 95% confidence intervals. (B) Identification performance for trained and untrained talkers at each time point. Children showed improvement from one day to the next on the trained talker, but adults did not. Error bars represent 95% confidence intervals.



old and over). We then fit two additional models: One model included data from the adult participants, and one model included data from the child participants. The model with the children's data predicted accuracy and included fixed factors of time (coded as before) and discrimination pretest. In the final model, by-participant random intercepts and slopes for time were included. In children, discrimination pretest was a significant predictor of identification performance, $\beta = .81$, $SE = 0.34$, $z = 2.36$, $p = .02$, and there was a significant main effect of time, $\beta = .38$, $SE = 0.17$, $z = 2.23$, $p = .03$, such that performance was better on the next-day posttest than at the immediate posttest. There was no interaction. The model of the adult data predicted identification accuracy and included fixed factors of time (coded as before) and discrimination pretest. The final model included by-participant random intercepts and slopes for time, and correlation parameters were set to zero. This model revealed a significant effect of the discrimination pretest, $\beta = 1.79$, $SE = 0.31$, $z = 5.69$, $p < .001$; no effect of time; and no interaction. Taken together, these results suggest that the discrimination pretest predicts posttraining identification performance for both adults and children; however, children show improvement after a delay, whereas adults simply maintain training-induced gains after a delay (see Figure 3B).

Does participant age predict discrimination of the sounds? In order to assess whether a participant's age predicted pre- or posttraining discrimination of the nonnative Hindi contrast, we carried out a linear mixed-effects model. The dependent variable in the model was the *d'* score. Fixed effects included time (pretest, immediate posttest, and next-day posttest) and participant age (as a continuous variable), and random effects included by-subject random intercepts. The factor of time was backward difference coded using the contr.sdif() function from the MASS package (Venables & Ripley, 2002) to test for differences between the pretest and the immediate posttest (learning) as well as differences between the immediate posttest and the next-day posttest (retention). This analysis revealed a difference between the pretest and the immediate posttest, $\beta = .76$, $SE = 0.22$, $t = 3.52$, $p < .001$, which suggests that participants improved their discrimination of the contrast after training. Age was a significant predictor, $\beta = .03$, $SE = 0.01$, $t = 4.53$, $p < .001$, suggesting that discrimination performance improved with increasing age. No interactions of age were found with either time contrast (see Figure 2B), suggesting that the initial advantage in discriminability of the contrasts conferred by age was maintained over learning and retention.

Does pretraining sensitivity to the contrast explain the relationship between age and discrimination performance? To test the possibility that pretraining sensitivity is what is driving the relationship between age and posttraining discrimination performance, we carried out an additional

[2]The way we set up the contrasts in this model does not tell us whether the interaction means that the factor age positively predicts performance on both days and it is simply a stronger association on the first day or whether it only predicts performance on the first day and not on the second day. To further explore this, we fit an additional exploratory model that instead nested the fixed factor age within time (see the Individual Differences section for an explanation of nested fixed factors). The final random effects structure of this model included by-participant random intercepts and slopes for time. As expected, the difference in the first two time points was found again, $\beta = .57$, $SE = 0.25$, $z = 3.10$, $p = .002$, and age positively predicted identification performance at the immediate posttest, $\beta = .04$, $SE = 0.01$, $z = 3.60$, $p < .001$, and the next-day posttest, $\beta = .02$, $SE = 0.01$, $z = 2.12$, $p = .03$. This indicates that age was a stronger predictor of identification performance at the immediate posttest than at the next-day posttest but that it significantly predicted performance at both time points.

mixed-effects model that predicted posttest *d'* scores. Fixed effects included time (deviation coded; immediate posttest = −.5, next-day posttest = .5), age, and discrimination pretest scores. By-participant random intercepts were also included. The discrimination pretest was a significant predictor of the discrimination posttest scores, $\beta = .67$, $SE = 0.28$, $t = 2.42$, $p = .02$, and no other main effects or interactions were found.

Generalization

Do adults and children identify and discriminate the sounds produced by an unfamiliar talker at above-chance levels after training? To test participants' ability to identify or discriminate the contrast produced by a new talker, we ran an intercept-only model for each task and then intercept-only models for adult and child participants on each task separately. All intercepts were significantly different from zero, indicating that participants' generalization performance was above chance. Details on the models and output can be found in Supplemental Material S1.

Does participant age predict identification accuracy of the sounds produced by a new talker? To assess whether age predicted a participant's ability to generalize their knowledge of the speech sound categories to a novel talker on the identification task, we fit a mixed-effects logistic regression model. This model predicted accuracy on each trial. Age was included as a fixed effect, and random intercepts for subject were included. Age did not significantly predict generalization to a novel talker on the identification task. To test whether the discrimination pretest predicted generalization in the identification task, another model was fit that was identical to the previous one except that the fixed effects included both age and the discrimination pretest scores as well as their interaction. Neither age nor the discrimination pretest scores predicted generalization performance on the identification task.

Does participant age predict discrimination of the sounds produced by a new talker? We additionally wanted to test whether participant age predicted how well learners could generalize their knowledge of the Hindi speech sounds to a new talker in the discrimination task. Because we only had data from one time point, we carried out a linear regression model with the *d'* scores as the dependent variable and age as a predictor. The overall fit of the model was significant, $R^2 = .22$, $F(1, 43) = 11.98$, $p = .001$. Age was a significant predictor of the *d'* score for the discrimination task with the untrained talker, $\beta = .02$, $SE = 0.01$, $t = 3.46$, $p = .001$, with *d'* scores increasing with increasing age. However, when we fit another model that included fixed effects of age and the discrimination pretest scores as well as their interactions, no significant predictors were found.

Individual Differences

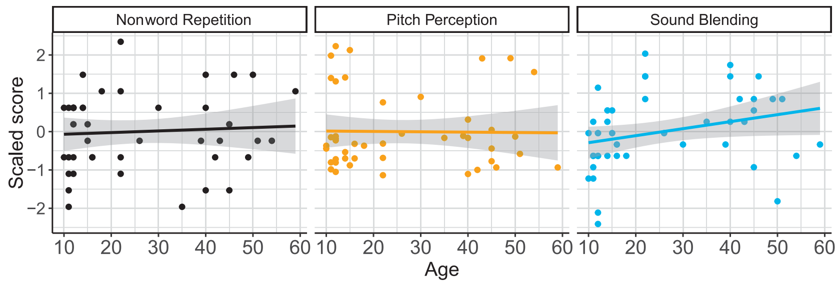
As we have shown, age and pretraining discrimination performance predict learning and retention of nonnative speech sounds. However, we see substantial individual variability in pretraining sensitivity, learning, and retention of the contrast even within age groups. Of interest is whether measures of individual differences in native language phonological skills (nonword repetition and sound blending) or auditory processing (pitch perception) predict performance on the nonnative measures. Tests of individual differences are necessarily exploratory, as (to our knowledge) there are no theories that make predictions about the sources or causes of individual differences in nonnative speech sound learning. Therefore, we tested measures that were identified from previous literature that have been found to predict nonnative speech sound learning (see the Introduction section). In addition, our sample size is likely too small to detect these relationships reliably, so we exercise caution in our interpretation of these analyses and suggest that future research should test these relationships with a larger sample size.

Does age predict performance on the individual differences measures? To test this question, we carried out three separate correlation tests that measured the relationships between age and sound blending, age and nonword repetition, and age and pitch perception, and *p* values were corrected for multiple comparisons using the Bonferroni–Holm method. Age did not significantly predict any of the individual differences measures: nonword repetition: $r = .07$, $t(43) = 0.44$, $p = 1.00$; sound blending: $r = .28$, $t(43) = 1.95$, $p = .17$; pitch perception: $r = -.01$, $t(43) = -0.10$, $p = 1.00$ (see Figure 4).

For the remaining analyses, we decided to dichotomize the age variable (children included participants under the age of 18 years, and adults included participants 18 years old and over) for two reasons. First, we have some evidence from previous studies that these phonological and auditory skills predict nonnative speech sound learning in adults, and we are interested in whether these relationships are also found in children or whether different skills predict baseline sensitivity to nonnative speech sounds in children. Second, we wanted to avoid potential four-way interactions with continuous variables. In this model, we took advantage of nested fixed factors. Nesting one fixed factor within another factor tests the simple effects of that factor at each level of the factor it is nested within without estimating the main effect of the nested factor (see Schad et al., 2020). In our case, we were interested in the relationship between individual differences measures (sound blending, nonword repetition, and pitch perception) and the posttest scores at each level of the age group factor (children and adults) and the time factor (pretest, immediate posttest, and next-day posttest; immediate posttest and next-day posttest only for identification).

Do individual differences measures predict discrimination ability differently in children or adults? To determine whether measures of sound blending, nonword repetition, and pitch perception predicted discrimination scores in adults or children, a linear mixed-effects model was performed, using *d'* scores as the dependent measure. Fixed effects of sound blending, nonword repetition, and pitch perception were nested within age group, and this was nested within time. Time was dummy coded with the pretest as

Figure 4. Relationship between age and individual differences measures. Age was not a significant predictor of any of the individual differences measures.



the reference level, and age group was deviation coded (adults = −.5, children = .5). Random by-subject intercepts were also included. Nonword repetition positively predicted discrimination performance on the next-day posttest for the child group, β = .88, *SE* = 0.33, t = 2.63, p = .01 (see Figure 5). Sound blending positively predicted discrimination performance in the adult group on the next-day posttest, β = .45, *SE* = 0.21, t = 2.09, p = .04, and pitch perception negatively predicted performance on the discrimination pretest, β = −.41, *SE* = 0.20, t = −2.01, p = .05; however, these effects barely reached significance by the $p < .05$ threshold with an already smaller-than-optimal sample size and should thus be interpreted with caution.

Do individual differences measures predict identification accuracy differently in children or adults? To test whether measures of sound blending, nonword repetition, and pitch perception predicted identification performance, we fit a mixed-effects logistic regression model that predicted accuracy. Fixed effects of sound blending, nonword repetition, and pitch perception were nested within age group, and this was nested within time. Time was deviation coded (immediate posttest = −.5, next-day posttest = .5), as was age group (adults = −.5, children = .5). The final model included by-subject random intercepts, and we used the glmerControl optimizer bobyqa and increased the iterations to 200,000 to facilitate model convergence. For the child group, nonword repetition positively predicted identification accuracy at the next-day posttest, β = 1.39, *SE* = 0.45, z = 3.12, p = .002 (see Figure 5). For the adult group, sound blending positively predicted identification accuracy at both the immediate posttest, β = .66, *SE* = 0.30, z = 2.17, p = .03, and the next-day posttest, β = .89, *SE* = 0.28, z = 3.12, p = .002 (see Figure 6).

Discussion

Children seem to have advantages over adults when it comes to their ultimate attainment in a second language, for learning certain grammatical constructions (e.g., Johnson & Newport, 1989) and especially for learning speech sounds (e.g., Flege et al., 1995, 1999). A possible explanation for this is that children's native language categories are not as well defined as adults' (i.e., children are less entrenched in their native language speech categories), and

Figure 5. Relationships between nonword repetition scores and (A) discrimination next-day posttest for child participants and (B) identification next-day posttest (log odds of accuracy) for child participants. For purposes of visualization only, mean percent accuracy was converted into log odds, and any accuracy scores of 1 were changed to .99 to avoid infinite values. Nonword repetition positively predicted discrimination and identification performance on the next-day posttests among child participants. The shaded regions represent 95% confidence intervals.

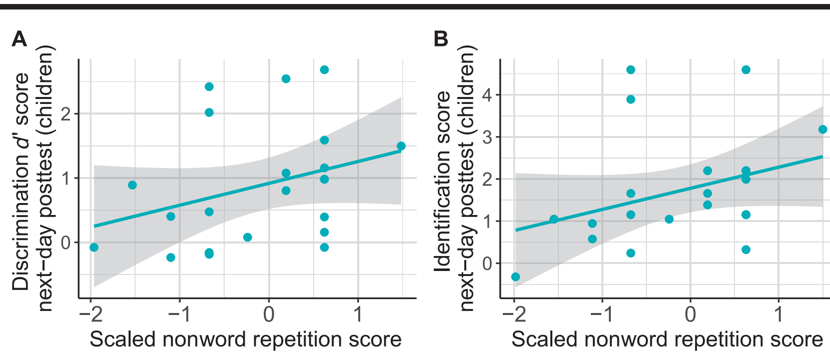
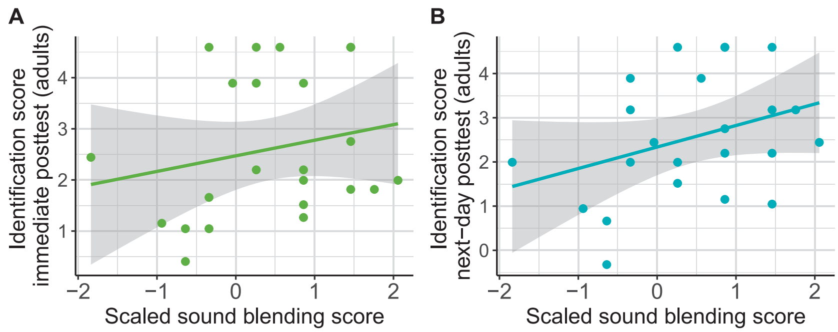


Figure 6. Relationships between sound blending scores and (A) immediate identification posttest scores (log odds of mean percent accuracy) and (B) next-day identification posttest scores for adult participants. Sound blending positively predicted identification performance on both posttests among adult participants. For purposes of visualization only, mean percent accuracy was converted into log odds, and any accuracy scores of 1 were changed to .99 to avoid infinite values. The shaded regions represent 95% confidence intervals.



therefore, their native language categories may be more malleable, giving them an advantage in detecting subtle acoustic details in nonnative speech stimuli.

In the current study, we tested whether age predicted pre- or posttraining discrimination and identification of a difficult nonnative phonetic contrast in Hindi. We indeed found a relationship between age and nonnative speech sound learning; however, it was in the opposite direction as would be predicted by the critical period hypothesis. First, we found that age positively predicted pretraining (baseline) discrimination of the contrast. This does not support the notion that children have an advantage in detecting fine-grained acoustic differences in the speech signal, and in fact, it suggests that perhaps adults have an advantage. One caveat, however, is that the results reported can only be interpreted for the age range we tested for children in the current study (ages 10–16 years), and it is possible that even younger children may show differences in initial perception or learning of a nonnative contrast, as younger children (usually about the age of 6 years or younger) have been shown to develop native-like production of nonnative speech sounds (e.g., Flege et al., 1999; Granena & Long, 2013).

Some recent work on the development of native language speech categories in children may offer an explanation for this counterintuitive finding, namely, that adults show superior initial perception of a nonnative contrast compared to children. We reasoned that listeners with more graded native language speech category representations may be less susceptible to perceptual assimilation effects, and because children have been found to have more graded native language category representations (e.g., Burnham et al., 1991; Hazan & Barrett, 2000), they may show superior naïve discrimination of a difficult nonnative contrast. A recent study, however, has suggested that children may actually have less graded or more categorical representations of speech sounds (McMurray et al., 2018). Similar to earlier work, this study found that younger children have shallower categorization slopes than older children on a behavioral categorization task. However, eye tracking data from this study suggest that perception of native language speech categories becomes more *graded* throughout adolescence; thus, children's shallower categorization functions may be a result of noisier representations. In fact, graded sensitivity to within-category differences may reflect a mature category representation, and this may be what allows older children and adults to show more precise and categorical responses to behavioral categorization tasks in the native language (McMurray et al., 2018, 2002). Assuming the adult participants in our study have more graded speech category representations, our results support the idea that sensitivity to within-category differences at least initially predicts performance on nonnative speech sound learning measures.

Another explanation is that working memory is needed to perform a discrimination task, and adults' working memory capacity facilitated their performance on this task. Specifically, when performing an AX discrimination task, a listener has to hold both sounds in memory and then decide whether they belong to the same speech category or different categories. Working memory capacity has been found to increase linearly from early childhood through adolescence (Gathercole et al., 2004), which could at least partially explain why age predicted discrimination performance. Notably, however, we found no relationship between age and nonword repetition, which has been used as a measure of *phonological* working memory (e.g., Coady & Evans, 2008). We would argue that phonological working memory would be more relevant for a phonetic discrimination task because listeners have to attend to subtle differences in sound, hold these in memory, and then make their same/different judgment. Ultimately, future studies will need to investigate the potential relationship between working memory and performance on (nonnative) phonetic discrimination tasks.

Yet another interpretation of these results is that older participants were more motivated to do well on the experimental tasks. If general motivation increases with age, we would expect that age would also predict performance on the individual differences measures that we collected (sound blending, nonword repetition, and pitch perception). Because we found no significant relationships between age and any of these three individual differences measures, we think age-related differences in motivation are an unlikely explanation for the relationship between age and nonnative tasks.

On balance, evidence points to age-related advantages in the initial discrimination of the nonnative contrasts, advantages that persist over learning and retention. Specifically, we found that age significantly predicted performance on the measures of posttraining discrimination. However, to better differentiate between learning and pretraining differences in sensitivity to the contrast, we conducted an additional analysis that included the discrimination pretest as a predictor. The effect of age went away when including discrimination pretest as a predictor, and the discrimination pretest scores significantly predicted discrimination posttest scores. This suggests that children and adults have similar learning and retention trajectories (at least over the age range that we tested, that is, early adolescence through adulthood). Importantly, what a learner comes to the table with (pretraining discrimination sensitivity) predicts learning and retention, although it is striking that adults seem to present with superior pretraining discrimination abilities.

These results present a puzzle, namely, that adults outperform children at every time point and on every task but that substantial real-world evidence shows children have better long-term outcomes in nonnative phonological learning. Hints of an explanation for this discrepancy can be found in the identification data. While age predicted accuracy in the identification task, the effect of age began to diminish by the next-day posttest (as indicated by the interaction of age and time). This finding was further supported by our second analysis, in which we included the discrimination pretest score as a predictor. There, we found that discrimination pretest scores predict posttraining identification performance for both children and adults, but children improved from one day to the next after accounting for pretraining sensitivity, whereas adults did not. This suggests that, while adults may have an initial advantage for learning speech sound categories, children may have an advantage in memory processes such as retention and consolidation, or they may experience less interference from native language speech sounds, which allows for better consolidation or retention of the (trained) identification task (see Fuhrmeister, 2019). In any case, despite their initial poorer performance, children seem to begin to catch up to adults after a period of off-line consolidation.

In order to account for age-related differences in nonnative learning, we explored the possibility that adults and children marshal different skills or use different strategies when approaching this problem. To test this, we asked whether measures of sound blending, nonword repetition, and pitch perception predicted performance on the nonnative speech sound learning tasks and whether these relationships differed between adults and children. The influence of these auditory and phonological skills is of interest since both the sound blending measure of the WJ-III COG and the nonword repetition score on the CTOPP have been characterized as indicators of native language phonological skills (sound blending; Sodoro et al., 2002) and phonological working memory (nonword repetition; Coady & Evans, 2008). Weaknesses in these skills (in combination with other subtests) are used to diagnose dyslexia and developmental language disorders. In our sample, there was a substantial range in scores in the typical to above-typical range.

We first found that, for children, nonword repetition positively predicted performance on the discrimination and identification posttests on the second day. Nonword repetition has been shown to be predictive of word learning, lexical knowledge, and phonological short-term memory (e.g., Coady & Evans, 2008). Children with stronger phonological short-term memory may have an advantage for holding nonnative sounds in memory temporarily in order to map them to the desired category. Interestingly, these relationships found between nonword repetition and nonnative speech sound learning in the child group emerged on the second day, or after a period of off-line consolidation. We speculate that phonological short-term memory is advantageous for children's ability to encode unfamiliar speech sounds more accurately, which in turn will lead to better consolidation of the newly formed memory traces.

For the adult group, sound blending positively predicted performance on both identification posttests (before and after a period of off-line consolidation). Sound blending is often used to measure phonological awareness, or the awareness that words are composed of individual sounds and the ability to manipulate those sounds (e.g., Sodoro et al., 2002), and it is likely strengthened by literacy skills (Morais et al., 1979). In this study, we asked participants to identify and discriminate Hindi sounds that were embedded in a lexical context. Therefore, better phonological awareness as measured by sound blending may have been beneficial for these particular tasks. The current study corroborates earlier findings in which sound blending ability predicts nonnative speech sound learning (Earle & Arthur, 2017; Perrachione et al., 2011) and suggests that sound blending may be a robust predictor of learning new speech sounds in adulthood. It is possible that phonological awareness, more generally, is what allows adults to form new speech sound categories successfully. For example, adults who have greater metalinguistic awareness of the sound structure and smaller units of sounds in their native language may be able to allocate the necessary attention to discriminate and identify unfamiliar sounds, especially when those sounds are embedded in a lexical context.

Overall, findings from the current study suggest that children do not necessarily have an initial advantage over adults for naïve perception or learning of nonnative speech sounds but that the cognitive or phonological skills that support this process may differ between adults and children. This study lends further support to the idea that children and adults approach the problem of nonnative speech sound learning differently. Nonetheless, these findings do not contradict a long line of work showing age-of-acquisition advantages for second language learning in general and for phonological learning specifically. Evidence is fairly clear that, on balance, those who begin to learn the sounds of a new language before adulthood go on to have superior abilities in nonnative speech perception and production (Flege et al., 1995, 1999; Piske et al., 2001). However, the current study calls into question some of the most sensible hypotheses for the root of this difference. That is, children show poorer naïve discrimination of nonnative contrasts compared to adults, suggesting that there is no initial perceptual advantage for detecting within-phonetic-category speech variants, nor is there an obvious advantage in learning these sounds or generalizing them to a new talker.

One hint of an explanation for better ultimate nonnative attainment in childhood comes from the fact that children show improvement after a delay on the (trained) identification task, whereas adults do not. Although future work will be needed to delineate a longer time course of nonnative learning, this finding raises the possibility that children may be better able to hold nonnative tokens in memory over the longer term. An advantage in memory consolidation might lead to cumulative gains in nonnative learning that could result in a significant advantage for learning nonnative speech sounds in childhood.

Acknowledgments

This material is based upon work supported in part by National Science Foundation (NSF) Division of Graduate Education Grant 1747486; NSF Division of Graduate Education Grant 1144399 through the Integrative Graduate Education and Research Traineeship program, awarded to the University of Connecticut; and NSF Division of Behavioral and Cognitive Sciences Grant 1554510, awarded to E. B. M. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

References

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics, 32*(2), 233–250. https://doi.org/10.1016/S0095-4470(03)00036-6

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child–adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech, 51*(4), 317–342. https://doi.org/10.1177/0023830908099068

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America, 109*(2), 775–794. https://doi.org/10.1121/1.1332378

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (Version 6.0.43) [Computer software]. http://www.praat.org/

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*(5), 977–985. https://doi.org/10.3758/BF03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299–2310. https://doi.org/10.1121/1.418276

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Burnham, D. K., Earnshaw, L. J., & Clark, J. E. (1991). Development of categorical identification of native and non-native bilabial stops: Infants, children and adults. *Journal of Child Language, 18*(2), 231–260. https://doi.org/10.1017/S0305000900011041

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders, 43*(1), 1–40. https://doi.org/10.1080/13682820601116485

Earle, F. S., & Arthur, D. T. (2017). Native phonological processing abilities predict post-consolidation nonnative contrast learning in adults. *The Journal of the Acoustical Society of America, 142*(6), EL525–EL531. https://doi.org/10.1121/1.5013141

Earle, F. S., & Myers, E. B. (2014). Building phonetic categories: An argument for the role of sleep. *Frontiers in Psychology, 5,* Article 1192. https://doi.org/10.3389/fpsyg.2014.01192

Earle, F. S., & Myers, E. B. (2015). Sleep and native language interference affect non-native speech sound learning. *Journal of Experimental Psychology: Human Perception and Performance, 41*(6), 1680–1695. https://doi.org/10.1037/xhp0000113

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97*(5), 3125–3134. https://doi.org/10.1121/1.413041

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78–104. https://doi.org/10.1006/jmla.1999.2638

Fuhrmeister, P. (2019). Interference in memory consolidation of non-native speech sounds. In S. Fuchs, J. Cleland, & A. Rochet-Capellan (Eds.), *Speech production and perception: Learning and memory* (pp. 207–243). Peter Lang.

Fuhrmeister, P., & Myers, E. B. (2017). Non-native phonetic learning is destabilized by exposure to phonological variability before and after training. *The Journal of the Acoustical Society of America, 142*(5), EL448–EL454. https://doi.org/10.1121/1.5009688

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*(3), 311–343. https://doi.org/10.1177/0267658312461497

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics, 28*(4), 377–396. https://doi.org/10.1006/jpho.2000.0121

Heeren, W. F. L., & Schouten, M. E. H. (2010). Perceptual development of the Finnish /t-tː/ distinction in Dutch 12-year-old children: A training study. *Journal of Phonetics, 38*(4), 594–603. https://doi.org/10.1016/j.wocn.2010.08.005

Idemaru, K., & Holt, L. L. (2013). The developmental trajectory of children's perception and production of English /r/-/l/. *The Journal of the Acoustical Society of America, 133*(6), 4232–4246. https://doi.org/10.1121/1.4802905

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology, 4*(6), 812–822. https://doi.org/10.1016/0959-4388(94)90128-7

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2007). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 979–1000. https://doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

MacKay, I. R. A., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica, 58*(1–2), 103–125. https://doi.org/10.1159/000028490

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Psychology Press. https://doi.org/10.4324/9781410611147

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44,* 314–324. https://doi.org/10.3758/s13428-011-0168-7

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94,* 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology, 54*(8), 1472–1491. https://doi.org/10.1037/dev0000542

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33–B42. https://doi.org/10.1016/S0010-0277(02)00157-9

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously. *Cognition, 7*(4), 323–331. https://doi.org/10.1016/0010-0277(79)90020-9

Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *The Journal of the Acoustical Society of America, 115*(4), 1777–1790. https://doi.org/10.1121/1.1651192

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America, 130*(1), 461–472. https://doi.org/10.1121/1.3593366

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*(2), 191–215. https://doi.org/10.1006/jpho.2001.0134

R Development Core Team. (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language, 110,* Article 104038. https://doi.org/10.1016/j.jml.2019.104038

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2019). *afex: Analysis of Factorial Experiments* (R Package Version 0.23-0) [Computer software]. https://CRAN.R-project.org/package=afex

Snow, C. E., & Hoefnagel-Höhle, M. (1977). Age differences in the pronunciation of foreign sounds. *Language and Speech, 20*(4), 357–365. https://doi.org/10.1177/002383097702000407

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development, 49*(4), 1114–1128. https://doi.org/10.2307/1128751

Sodoro, J., Allinder, R. M., & Rankin-Erickson, J. L. (2002). Assessment of phonological awareness: Review of methods and tools. *Educational Psychology Review, 14*(3), 223–260. https://doi.org/10.1023/A:1016050412323

Venables, W. N., & Ripley, B. D. (2002). Time series analysis. In *Modern applied statistics with S. Statistics and computing* (pp. 387–418). Springer. https://doi.org/10.1007/978-0-387-21706-2_14

Wagner, R., Torgesen, J., Rashotte, C., & Pearson, N. A. (1999). *Comprehensive Test of Phonological Processing–Second Edition (CTOPP-2)*. Pro-Ed.

Wang, Y., & Kuhl, P. K. (2003). *Evaluating the "Critical Period" Hypothesis: Perceptual learning of Mandarin tones in American adults and American children at 6, 10 and 14 years of age* [Paper presentation]. 15th International Congress of Phonetic Sciences, Barcelona, Spain.

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology, 66,* 173–196. https://doi.org/10.1146/annurev-psych-010814-015104

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7*(1), 49–63. https://1doi.org/10.1016/S0163-6383(84)80022-3

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Riverside Publishing.

Zevin, J. D. (2012). A sensitive period for shibboleths: The long tail and changing goals of speech perception over the course of development. *Developmental Psychobiology, 54*(6), 632–642. https://doi.org/10.1002/dev.20611