frontiers in Communication

ORIGINAL RESEARCH
published: 19 September 2018
doi: 10.3389/fcomm.2018.00035

Contextual Influences on Phonetic Categorization in School-Aged Children

Jean A. Campbell[1], *Heather L. McSherry*[1] *and Rachel M. Theodore*[1,2]*

[1] Department of Speech, Language, and Hearing Sciences, University of Connecticut, Storrs, CT, United States,
[2] Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut, Storrs, CT, United States

Perceptual stability in adult listeners is supported by the ability to process acoustic-phonetic variation categorically and dynamically adjust category boundaries given systematic contextual influences. The current study examined the developmental trajectory of such flexibility. Adults and school-aged children (5–10 years of age) made voicing identification decisions to voice-onset-time (VOT) continua that differed in speaking rate and place of articulation. The results showed that both populations were sensitive to contextual influences; the voicing boundary was located at a longer VOT for the slow compared to the fast speaking rate continuum and for the velar compared to the labial continuum, and the magnitude of the displacement was slighter greater for the adults compared to the children. Moreover, the two populations differed in terms of the absolute location of the voicing boundaries and the categorization slopes, with slopes becoming more categorical as age increased. These results demonstrate that sensitivity to contextual influences on speech perception emerges early in development, but mature perceptual tuning requires extended experience.

Keywords: speech perception, phonetic variability, context effects, language development, phonetic adaptation

OPEN ACCESS

Edited by:
Judit Gervain,
Centre National de la Recherche Scientifique (CNRS), France

Reviewed by:
Hans Rutger Bosker,
Max-Planck-Institut für Psycholinguistik, Netherlands
Sandrien Van Ommen,
UMR8242 Laboratoire Psychologie de la Perception (LPP), France

***Correspondence:**
Rachel M. Theodore
rachel.theodore@uconn.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *03 April 2018*
Accepted: *24 July 2018*
Published: *19 September 2018*

Citation:
Campbell JA, McSherry HL and Theodore RM (2018) Contextual Influences on Phonetic Categorization in School-Aged Children. Front. Commun. 3:35. doi: 10.3389/fcomm.2018.00035

INTRODUCTION

A central goal of research within the domain of speech perception is to describe the mechanisms supporting listeners' ability to reliably perceive speech segments given a lack of invariance between the acoustic signal and any given consonant or vowel. For example, consider the acoustic information used to specify just one class of sounds, stop consonants. The acoustic information produced for a given stop consonant varies widely due to a host of factors including speaking rate (Miller et al., 1986; Nagao and de Jong, 2007), following phonetic context (Delattre et al., 1955), place of articulation (Volaitis and Miller, 1992), and even who in particular is producing the stop consonant (Theodore et al., 2009). In order to achieve reliable perception of stop consonants (and other speech sounds), perceptual mechanisms must support the successful mapping of many acoustically distinct elements to a single phonetic category. For the literature reviewed below, the language of the stimuli and participants is English unless otherwise specified.

Results of some of the earliest investigations in the field suggest that listeners achieve perceptual constancy by ignoring fine phonetic variation in the signal to some degree, instead processing variability in terms of broad categories (Liberman et al., 1961). One acoustic-phonetic property that has been widely examined in this respect is voice-onset-time (VOT). VOT reflects the time between the release of a stop consonant and the onset of voicing for a following vowel (Lisker and Abramson, 1964).

VOT is an important characteristic of these sounds because it is one of the cues that marks the voicing contrast in word-initial stop consonants; voiced stops have relatively shorter VOTs than voiceless stops (Lisker and Abramson, 1964; examined across 11 languages, including English). Though VOT is not the only cue to the voicing contrast (e.g., Van Alphen and Smits, 2004; examined in Dutch), evidence for categorical perception of VOT comes from studies that present a range of VOTs to listeners who are asked to identify each VOT as a voiced or voiceless stop consonant (e.g., /g/ or /k/). Results from these studies show that listeners' responses are not linearly related to VOT. Rather, listeners' responses are categorical in that a range of VOTs are identified as voiced, a different range of VOTs are identified as voiceless, and there is an abrupt discontinuity between the two ranges (Liberman et al., 1967). Thus, listeners appear to have a perceptual boundary at some particular VOT that marks the voicing distinction for stop consonants.

Converging evidence for these categorical effects comes from discrimination paradigms. In these tasks, listeners are presented with pairs of stimuli across a VOT continuum, for example, and are asked to identify whether the members of the pair are the same or if they are different. Results from these types of discrimination tasks show heightened discrimination for pairs that span the voicing boundary (as measured in identification tasks), with less sensitive discrimination for pairs that fall within either side of the boundary, even though the acoustic difference between the pairs is identical (Liberman et al., 1957). Such categorical perception of acoustic-phonetic variation has been shown for a host of phonetic properties that cue many different phonemic contrasts (Stevens et al., 1969; Miller and Liberman, 1979; MacKain et al., 1981; Volaitis and Miller, 1992), though the degree to which phonetic variability is processed categorically varies among specific acoustic-phonetic properties (e.g., Fry et al., 1962) and the specific test task (e.g., Gerrits and Schouten, 2004; examined in Dutch). Moreover, evidence of categorical perception in speech identification and discrimination tasks does not reflect a complete insensitivity to within-category variation, as other findings from both behavioral and fMRI neuroimaging paradigms have shown that speech sound categories have a graded internal structure, with some members considered more prototypical than others (e.g., Miller, 1994; Myers, 2007; Miller et al., 2011, in French; Theodore et al., 2015; Myers and Theodore, 2017).

Though categorical perception has been considered one hallmark of efficient speech perception, many findings have demonstrated that category boundaries are not absolute with respect to the particular value of any acoustic-phonetic dimension. Rather, phonetic categorization is highly sensitive to systematic acoustic-phonetic variation; precise boundaries between phonetic categories shift as a function of context. For example, research in speech production has shown that VOTs are influenced by speaking rate, with VOTs systematically increasing as rate slows (e.g., Miller et al., 1986; Kessinger and Blumstein, 1998; Nagao and de Jong, 2007). Perception studies have shown that listeners adjust the boundary between voiced and voiceless stops such that the voicing boundary for a slow speaking rate continuum is located at a longer VOT than the voicing boundary for a fast speaking rate continuum (Summerfield, 1981; Green and Miller, 1985; Miller and Volaitis, 1989). Place of articulation also influences the location of the stop voicing category boundary. In speech production, VOTs systematically increase as place of articulation moves from anterior to posterior in the vocal tract (e.g., Lisker and Abramson, 1964; Cho and Ladefoged, 1999; Theodore et al., 2009). As reviewed in Cho and Ladefoged (1999), who examined speakers of 18 languages including English, this difference in production may reflect general laws of aerodynamics or motor timing of the articulators. This contextual influence is observed for perception of the voicing contrast, with the VOT boundary for a velar continuum located at longer values compared to the VOT boundary for a labial continuum (Liberman et al., 1967; Volaitis and Miller, 1992). It is important to note that while both speaking rate and place of articulation influence the location of the VOT voicing boundary, they reflect distinct contextual influences. The influence of speaking rate reflects perceptual normalization that occurs in real-time and may be the consequence of general auditory processing mechanisms (e.g., Dent et al., 1997), whereas the influence of place of articulation may be the consequence of category formation and thus may reflect distributional learning mechanisms (e.g., Pierrehumbert, 2003).

Thus, categorical processing of a given acoustic-phonetic dimension is just one component of healthy perceptual processing. In order to achieve robust perception of speech segments, listeners must also be able to dynamically adjust the category boundary to accommodate systematic acoustic-phonetic variation. Explicating the developmental trajectory of sensitivity to contextual influences in speech perception is important to form a complete model of speech perception and because an understanding of the typical trajectory can be used to identify children at risk for atypical speech and language development. Studies examining the development of categorical perception for a single context have revealed that categorical perception emerges very early in development, with infants as young as one month of age showing discontinuous sensitivity to VOT that approximates adults' category boundaries (Eimas et al., 1971). Wolf (1973) examined perception of the stop voicing contrast in 5- and 7-year-old children for VOT continua specifying a /ba/-/pa/ contrast and, with different children of the same age, a /da/-/ta/ contrast. She found no difference in the location of the VOT voicing boundary as measured using an identification task between the two groups of children, nor between the children and a group of adults, though a discrimination task did show evidence of less categorical processing in the children compared to the adults. Other studies, however, have revealed differences in the perception of VOT in children compared to adults with respect to the VOT voicing boundary. For example, Zlatin and Koenigsknecht (1975) compared perception of the stop voicing contrast among 2-year-olds, 6-year-olds, and adults. Participants completed a two-alternative forced-choice (2AFC) phonetic decision task for VOT variants of labial (*bees-peas*, *bear-pear*), alveolar (*dime-time*), and velar (*goat-coat*) continua. The VOT voicing boundary was calculated as the crossover point along the VOT continuum corresponding to 50% responses for each voicing category.

Performance among the three age groups was examined for each continuum, with the results showing that the 2-year-olds showed longer VOT boundaries for the velar continuum compared to the 6-year-olds and adults, with all three groups showing equivalent crossover boundaries for the other continua.

Though limited effects of age were observed for the absolute boundary, additional influences of age were evident when the boundary region was examined (Zlatin and Koenigsknecht, 1975). The boundary region was quantified as the range of VOTs spanning identification as voiced 75% of the time and voiceless 75% of the time; in other words, the boundary region reflected the range of VOTs that straddled the crossover boundary for which identification was not consistent. For this measure, the 2-year-olds showed a wider boundary region compared to the adults (for all four continua) and the 6-year-olds (for two of the continua); the 6-year-olds showed a wider boundary region compared to the adults for the velar continuum. These results are consistent with Burnham and colleagues (Burnham et al., 1991), who examined perception of a labial VOT continuum in infants, two-year-olds, six-year-olds, and adults. They found that although infants showed categorical perception of the continuum, performance became systematically more categorical as age increased (Burnham et al., 1991). Indeed, it has been suggested that boundary precision, defined as the degree to which response patterns to variation in the input are categorical, reflects general cognitive maturation (Hoonhorst et al., 2011). Hoonhorst et al. (2011) examined perceptual organization in children (6–8 years of age) and adults for perception of speech, color, and facial expression variability. Participants made identification (and discrimination) decisions for a voicing contrast (specified by VOT), a color contrast (hues ranging from yellow to green), and a facial expression contrast (a visual face continuum that morphed from happy to sad affect). Their results revealed no difference in the absolute location of the category boundary among the age groups, but did reveal differences in the identification slope. Specifically, the 6-year-olds showed less categorical slopes compared to the other age groups (which were equivalent) for the speech task, and the adults showed more categorical slopes compared to all child age groups for the facial expression task. Flege and Eefting (1986) compared perception of the stop voicing contrast between children and adults for a VOT continuum specifying /da/ - /ta/. The children in their study had a median age of 9 years; half of the children and adults were native speakers of English, and the other half were native speakers of Spanish with limited English experience. In contrast to the studies reported above, they found no evidence that the slope of the voicing identification function differed between the children and adults, but—for both language groups—the children showed voicing boundaries located at shorter VOTs compared to the adults. A follow-up study tested three additional groups of monolingual English children with mean ages of 11, 13, and 17 years, respectively. The results showed that all three age groups had voicing boundaries located at shorter VOTs compared to those of the adults, and the VOT voicing boundary of the 17-year-olds was statistically longer than that of the 9-year-olds.

As a whole, the reviewed literature examining the perception of VOT in children compared to adults does not converge with respect to the absolute location of the VOT voicing boundary; some studies find children to have boundaries at shorter VOTs (e.g., Flege and Eefting, 1986), some report children to have boundaries at longer VOTs (e.g., Zlatin and Koenigsknecht, 1975), and other studies find no difference in the precise location of the VOT voicing boundary between children and adults (e.g., Wolf, 1973; Hoonhorst et al., 2011). The lack of convergence across the literature may reflect any one of the methodological differences, including aspects of the stimuli or the specific task. Of relevance to the current work, however, is the general convergence in past work showing that that VOT variation is processed categorically (and not linearly) in both populations, and that multiple studies have found evidence that perception of VOT is less categorical in children compared to adults (e.g., Zlatin and Koenigsknecht, 1975; Burnham et al., 1991; Hoonhorst et al., 2011). These findings suggest that while the broad strokes of perceptual organization for accommodating contextual influences are in place early in development, experience plays a role in fine-tuning the mapping between the acoustic signal and speech segments for many years. This view is consistent with statistical learning accounts of speech sound categories which posit that stable speech sound categories are refined over an extended period of development in which the listener tracks distributional cues in the environment in order to learn multi-dimensional categories (e.g., McMurray et al., 2009; Galle and McMurray, 2014). Indeed, protracted periods of development have been identified in school-aged children for many speech sounds (e.g., Nittrouer and Studdert-Kennedy, 1987; Burnham et al., 1991; Nittrouer, 1992, 2002). With respect to stop consonant perception, there have been limited investigations that simultaneously examine how contextual influences, such as speaking rate and place of articulation, interact with perceptual tuning across the developmental trajectory.

There is some evidence indicating that rate-dependent processing of phonetic variation is present in young infants. Miller and Eimas (1983) used a discrimination paradigm to examine 3- and 4-month-old infants' perception of the /b/-/w/ contrast. The manner contrast for /b/-/w/ is cued in part by duration of the F1 transition; short transitions cue perception of /b/ whereas relatively longer transitions cue perception of /w/, and transition duration for both segments increases as speaking rate slows (Gay, 1978). Building on previous findings showing that adults require a longer transition duration to perceive /w/ in slow compared to fast speaking rates (Miller and Liberman, 1979), Miller and Eimas examined infants' perception of two pairs of CV stimuli from a /ba/-/wa/ continuum that differed in duration of the F1 transition. One pair consisted of tokens with transition durations of 16 and 40 ms, and the other pair consisted of tokens with transition durations of 40 and 64 ms. The duration pairs were presented in syllables that were either 80 ms in duration (i.e., fast speaking rate) or 296 ms in duration (i.e., slow speaking rate). For the fast speaking rate, infants showed discrimination of the 16/40 ms pair but not the 40/64 ms pair, with the opposite pattern of discrimination observed for the slow speaking rate. In other words, infants showed evidence of rate-dependent processing such that the /ba/-/wa/

boundary was shifted toward longer transition durations for the slow compared to the fast speaking rate. That infants showed rate-dependent processing of temporal phonetic properties of speech is perhaps not surprising given evidence indicating that perceptual accommodation for speaking rate may reflect general auditory processing mechanisms including neural entrainment (Bosker, 2017; examined in Dutch), general cue-integration principles (Toscano and McMurray, 2012), and durational contrast effects (Wade and Holt, 2005); indeed, rate-dependent processing is not unique to the human species (Dent et al., 1997; Welch et al., 2009). The general auditory-based accounts of speaking rate accommodation would predict that linguistic experience is not necessary in order to modify the VOT voicing boundary as a function of speaking rate; that is, this class of theories does not predict a need for perceptual learning of the relationship between speaking rate and VOT variation. This theory thus predicts that the magnitude of the speaking rate effect in children and adults would be equivalent. Though the Miller and Eimas (1983) results are consistent with this prediction, it was not directly tested to the degree that the coarse resolution of the stimulus pairs presented for discrimination to the infants did not promote direct comparison of the precise F1 transition boundary between infants and adults. In contrast to mechanisms that support rate-dependent processing, far fewer studies have been conducted to examine the developmental trajectory of the contextual influence of place of articulation on VOT perception. To our knowledge, Zlatin and Koenigsknecht (1975) provided the only case where this effect could have been examined due to the within-subjects manipulation of place of articulation for the children and adults tested in their study; however, the effect of place of articulation (i.e., the difference in VOT crossover for the labial and velar continua) was not directly compared between the child and adult participants.

Here we report the results of a behavioral experiment that examined the contextual influences of speaking rate and place of articulation on phonetic categorization for adults and children ages 5–10 years. This age range was motivated by the range examined in previous work (e.g., Nittrouer, 2002; Hoonhorst et al., 2011) and also because adopting a wide age range allows examination as to whether performance becomes closer to the adult norm as age increases. That is, should population-level differences be observed between children and adults, age of the children can be analyzed continuously. All participants were presented with VOT continua spanning values that cued voiced and voiceless stops. Speaking rate (fast vs. slow) and place of articulation (labial vs. velar) were manipulated across the continua. Based on previous work (Volaitis and Miller, 1992), we predicted that the adults would show sensitivity to both contextual influences, with the boundary shifting toward longer VOTs for the slow compared to the fast speaking rate, and for the velar compared to the labial continuum. If the contextual influences of speaking rate and place of articulation on VOT perception become adult-like in infancy, as predicted by general auditory-based accounts of context effects, then we predicted no difference between the adults and children with respect to the influences of speaking rate and place of articulation on phonetic categorization. However, if perceptual sensitivity to systematic acoustic-phonetic variation continues to be fine-tuned throughout development, as predicted by statistical learning accounts of speech sound development, then the children examined here will not show adult-like sensitivity in terms of the slope of the identification functions and the magnitude of the context effect. A third possibility is that a difference between children and adults may be observed for place of articulation but not speaking rate, reflecting an increased role for experience-driven learning in the former compared to the latter.

MATERIALS AND METHODS

Participants

Twenty-four adults (ages 19–24 years) and twenty-five children (ages 5–10 years; 13 males, 12 females) were recruited for participation from the University of Connecticut community. All listeners were monolingual speakers of American English with no history of speech, language, or hearing disorders according to self-report or parent-report, save for one adult who had a history of /r/ articulation disorder that was successfully remediated in childhood. All listeners passed a pure-tone hearing screen bilaterally at 20 dB for 500, 1,000, 2,000, and 4,000 Hz on the day of testing. All children participated in a standardized assessment battery in order to assess speech abilities (*Goldman-Fristoe Test of Articulation–Second Edition*; GFTA-2*;* Goldman, 2000), language abilities (*Clinical Evaluation of Language Fundamental—4 Screening Test*; CELF-4 Screening Test; Semel et al., 2004), and nonverbal intelligence (*Raven's Coloured Progressive Matrices* (Raven's; Raven et al., 1998). All children scored within one standard deviation of the mean standard score on the GFTA-2 and above criterion on the CELF-4 Screening Test, indicating within normal limits for speech and language abilities. Moreover, all children scored within normal limits on the Raven's test, indicating typical nonverbal intelligence. Adult listeners were paid for their participation in the experiment; children received trinkets (e.g., stickers, rubber duckies, pompom critters) for their participation and families were provided with monetary compensation to offset transportation costs. This protocol was approved by the Institutional Review Board of the University of Connecticut. All adult participants gave written informed consent. All child participants gave written informed assent; parents of child participants gave written informed consent for their child's participation.

The experiment was stopped for three children who through explicit statement or experimenter judgment were no longer interested in the activity. As a consequence, one child was excluded from both the speaking rate and place of articulation analyses due to insufficient data for either analysis, one child was excluded from the speaking rate analysis due to a lack of data for the fast velar continuum, and one child was excluded from the place of articulation analysis due to a lack of data for the slow labial continuum. Of the 22 children who completed the experiment in full, one was excluded from both analyses due to a failure to respond to over 50% of the trials for the slow velar continuum (which is used in both the speaking rate and place of articulation analyses). Thus, each analysis reported below

includes data from 24 adults and 22 children. For the 22 children included in the speaking rate analysis, there were three 5-year-olds, seven 6-year-olds, four 7-year-olds, two 8-year-olds, four 9-year-olds, and two 10-year-olds (mean age in months = 91, range = 60–128 months). For the 22 children included in the place of articulation analysis, there were four 5-year-olds, seven 6-year-olds, four 7-year-olds, two 8-year-olds, four 9-year-olds, and one 10-year-old (mean age in months = 88, range = 60–128 months).

Stimuli

The stimulus set consisted of three VOT continua: *goal-coal* at a fast speaking rate, *goal-coal* at a slow speaking rate, and *bowl-pole* at a slow speaking rate. The stimuli can be retrieved at http://osf.io/4tq82. Stimuli were created following the methods outlined in Theodore and Miller (2010), to which the reader is referred for comprehensive details on stimulus construction. In brief, the continua were based on natural productions of the voiced-initial endpoints, *goal* and *bowl,* from a female monolingual speaker of American English. The selected *goal* and *bowl* tokens were equated for duration (478 ms for *goal*, 479 ms for *bowl*) by removing energy from the word offset, a cosine ramp was applied to the final 30 ms of each token in order to simulate the naturally-occurring decrease in amplitude at word-offset, and the two tokens were equated for amplitude. A synthesized version of the selected *goal* and *bowl* tokens was created using LPC-based speech synthesis software (Analysis Synthesis Laboratory, Kay PENTAX). These tokens served as the voiced-initial endpoint of the slow velar and slow labial continua, respectively. To create successive steps on each continuum, parameters of the LPC analysis were modified on a frame-by-frame basis (each frame corresponds to one vocal fold cycle) to replace the periodic source with a noise source and to scale peak amplitude. This procedure yielded, for each continuum, a series of tokens that incrementally increased in VOT in approximately 4 ms steps while preserving word duration and filter characteristics of the original token. VOTs of the velar continuum ranged from 15 to 80 ms and VOTs of the labial continuum ranged from 10 to 75 ms; this offset is in accord with how these values pattern in natural speech (e.g., Theodore et al., 2009). To create the velar continuum for the fast speaking rate, we systematically modified the velar continuum that was used for the slow speaking rate. Specifically, 160 ms of vocal energy was deleted from the steady-state portions of the vowel and final liquid at identical points for each token on the continuum. This procedure yielded a *goal-coal* continuum that had shorter token durations (318 ms), and thus a perceptually faster speaking rate, but was in all other respects identical to the slow velar continuum, most critically with respect to word-initial VOT. **Figure 1** shows spectrograms for step 1 (the voiced endpoint), step 8 (the midpoint), and step 15 (the voiceless endpoint) for each of the three continua; the spectrograms were generated using the Praat software (Boersma, 2001).

Procedure

Each listener was tested individually in a sound-attenuated booth. Auditory stimuli were presented at a comfortable listening level that was consistent across participants. Instructions were presented on the computer monitor as well as delivered verbally by the experimenter. All listeners participated in three blocks of trials, one for each test continuum. Block order was counter-balanced across listeners. Each block consisted of six randomizations of the particular continuum. On each trial, listeners were asked to indicate whether the word began with the specific voiced or voiceless stop consonant by pressing an appropriately labeled button on the response box. The labels on the response box were pictures of the target words (i.e., *bowl*, *pole*, *goal*, *coal*). Stimuli were presented over headphones (Sony MDR-V6) for the adults and from a speaker (KRK RP5G2) placed approximately 24 inches in front of the children; amplitude was set at a comfortable listening level that was held constant across participants. The children were tested using a speaker for stimulus presentation because we had difficulty achieving headphone compliance in pilot testing with this population. Accordingly, we chose to present stimuli to the children via a speaker so that the mode of stimulus presentation held constant within each population. We acknowledge that this potentially creates a methodological confound between the two populations, a point that is considered further in the discussion.

Trials were presented with an ISI of 4,000 ms timed from the listener's response; the next trial advanced if the listener failed to respond within 5,000 ms of the onset of each trial. Stimulus presentation and recording of button presses was controlled via the SuperLab software (version 4.5) running on a Mac (OS X). Prior to beginning the experiment, the researcher checked to ensure that the participant understood each target response (*bowl*, *pole*, *goal*, *coal*) and its associated picture on the button box. This process consisted of pointing to each picture, labeling it, providing a short description of the word, and then asking the participant to point to the pictures as the examiner named them. All children demonstrated that they could match each picture to the appropriate target. The entire procedure lasted approximately 30–45 min.

RESULTS

Speaking Rate

The raw data and analysis script for all analyses reported in this manuscript can be retrieved at http://osf.io/4tq82. The top panel of **Figure 2** shows mean proportion voiceless responses as a function of VOT for each speaking rate (fast vs. slow) and each population (child vs. adults); error bars indicate standard error of the mean. To derive these values, mean proportion voiceless responses was first calculated for each participant separately for each continuum and each VOT. Visual inspection of **Figure 2** shows four primary patterns: (1) both populations show sensitivity to the rate manipulation, in that the identification function for the slow speaking rate is displaced to longer VOTs compared to the identification function for the fast speaking rate, (2) for both speaking rates, the adults show voicing boundaries located at shorter VOTs compared to the children, (3) the identification functions for the adults appear more categorical (i.e., show a steeper slope) than those of the children, and (4) the magnitude of the rate affect (i.e., the degree to which the fast and

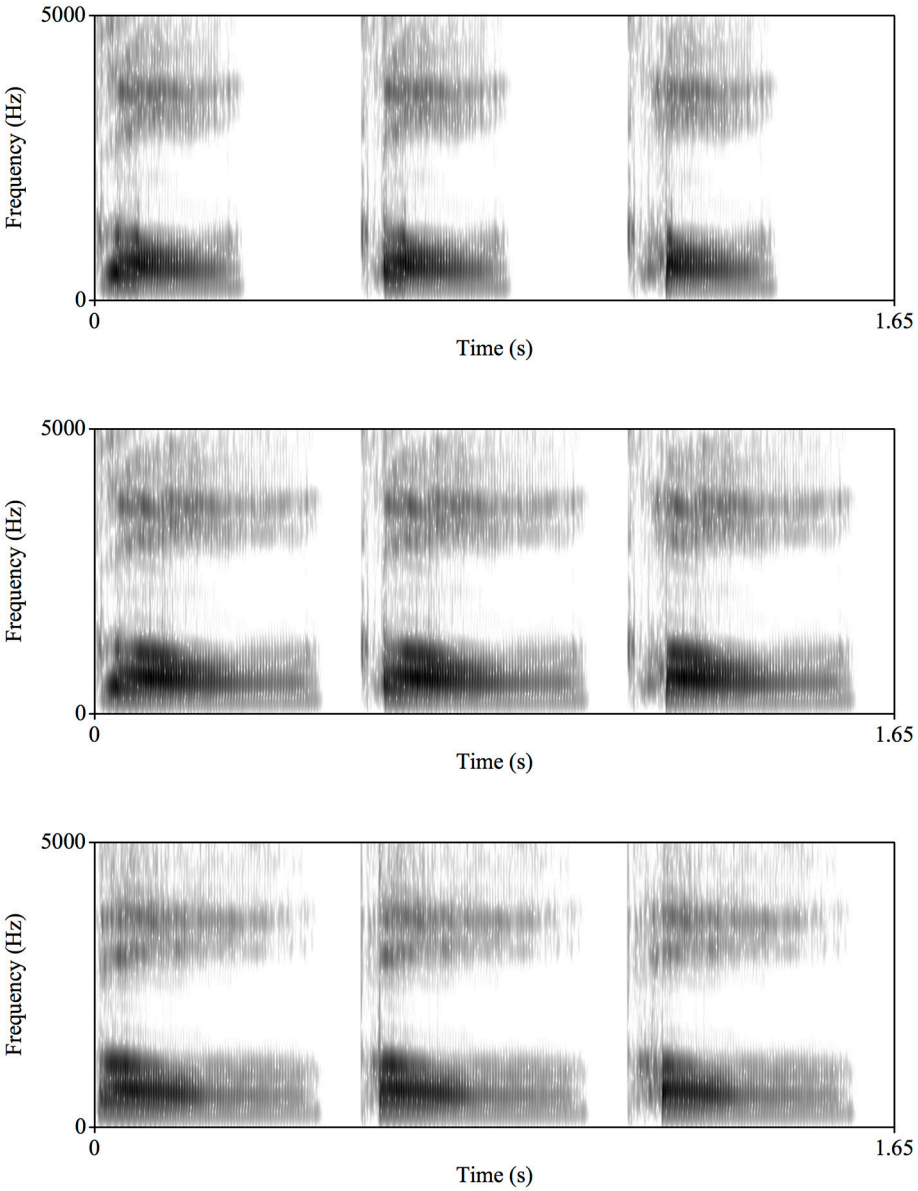

FIGURE 1 | Spectrograms for the voiced endpoint (step 1), midpoint (step 8), and voiceless endpoint (step 15) for the fast velar continuum **(Top)**, slow velar continuum **(Middle)**, and slow labial continuum **(Bottom)**. The stimuli can be retrieved from the link provided in the Data Availability Statement.

slow identification functions are displaced) appears comparable between the two populations.

To analyze the statistical nature of these patterns, trial-level responses (0 = voiced, 1 = voiceless) were fit to a generalized linear mixed-effects model (GLMM) with a logit-link function using the lme4 package (Bates et al., 2015) in R (http://www.r-project.org). The fixed effects were VOT, speaking rate, population, their interactions, and trial number. VOT and trial number were scaled and centered around the mean; speaking rate (fast = −0.5, slow = 0.5) and population (children = −0.5, adults = 0.5) were contrast-coded. The random effects structure included random intercepts by subject, and random slopes by subject for VOT, speaking rate, their interaction, and trial number. The model is visualized in the bottom panel of **Figure 2** in terms of the fixed effects of VOT, speaking rate, and population; this visualization was created using the jtools package in R (Long, 2018). The left side of the panel shows the model faceted by speaking rate; the right side of the panel shows the same model faceted by population. The shaded region around each function shows the 95% confidence interval; the wider shaded region for the children compared to the adults reflects greater variability in the former population.

The results of the model showed a significant main effect of VOT ($b = 4.793$, $SE = 0.290$, $z = 16.515$, $p < 0.001$),

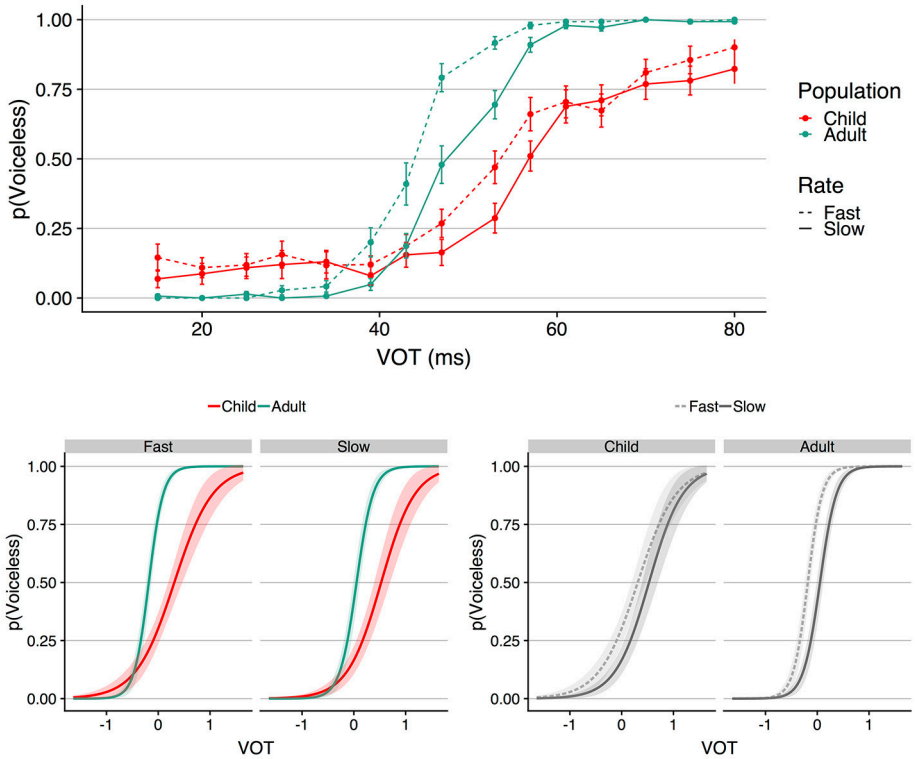

FIGURE 2 | The **(Top)** shows mean proportion voiceless responses as a function of VOT (ms) for each population at each speaking rate. Error bars indicate standard error of the mean. The **(Bottom)** shows the results of the GLMM; VOT (in ms) was entered into the model scaled and centered around the mean, which is shown as 0 on the abscissa for the plots on the bottom panel. At left, speaking rate is faceted to facilitate visualization of the interaction between VOT and population. At right, population is faceted to facilitate visualization of the speaking rate effect in each population. The shaded regions around the functions in the bottom panel indicate the 95% confidence interval region.

indicating that voiceless responses increased as did VOT. The main effect of speaking rate was also significant ($b = -1.207$, $SE = 0.161$, $z = -7.476$, $p < 0.001$), indicating that there were more voiceless responses for the fast compared to the slow speaking rate, indicative of a longer VOT boundary for the slow compared to the fast speaking rate. The model also showed a main effect of population ($b = 1.715$, $SE = 0.207$, $z = 8.295$, $p < 0.001$), with children showing more voiceless responses compared to the adults, in line with a voicing boundary located at longer a longer VOT. The main effect of trial number was also significant ($b = 0.315$, $SE = 0.060$, $z = 5.282$, $p < 0.001$), indicating more voiceless responses as trial number increased.

The model revealed three interactions. First, there was a significant interaction between VOT and population ($b = 3.830$, $SE = 0.552$, $z = 6.941$, $p < 0.001$), with the direction of the interaction indicating steeper identification functions in the adults compared to the children. Second, there was a significant interaction between population and speaking rate ($b = -0.854$, $SE = 0.323$, $z = -2.645$, $p = 0.008$). There was no interaction between VOT and speaking rate ($b = -0.262$, $SE = 0.312$, $z = -0.828$, $p = 0.408$), but a reliable three-way interaction between VOT, speaking rate, and population was observed ($b = -1.265$, $SE = 0.569$, $z = -2.225$, $p = 0.026$).

To explicate the nature of the interaction between VOT, population, and speaking rate, a set of models were constructed to examine the effect of VOT and speaking rate within each population and the effect of VOT and population at each speaking rate. The fixed and random effects structure of these models was identical to the primary model described above, save for the removal of the fixed effect of interest (i.e., the fixed effect of population was removed for the models examining the effect of speaking rate within each population, and the fixed effect of speaking rate was removed for the models examining the effect of population within each speaking rate). These models showed a significant effect of speaking rate for both children ($b = -0.815$, $SE = 0.200$, $z = -4.079$, $p < 0.001$) and adults ($b = -1.738$, $SE = 0.275$, $z = -6.330$, $p < 0.001$), and a significant effect of population for both the fast speaking rate ($b = 1.924$, $SE = 0.225$, $z = 8.552$, $p < 0.001$) and the slow speaking rate ($b = 1.337$, $SE = 0.319$, $z = 4.189$, $p < 0.001$). Moreover, these models showed that the interaction between VOT and population was significant for both the fast ($b = 4.668$, $SE = 0.671$, $z = 6.961$, $p < 0.001$) and slow ($b = 3.092$, $SE = 0.642$, $z = 4.817$, $p < 0.001$) speaking rates. Thus, the interaction between VOT, population, and speaking rate observed in the main model does not reflect the absence of a rate effect in either population, the absence of a population effect at either speaking rate, nor the absence of an

interaction between VOT and population at either speaking rate. Instead, the interaction between speaking rate and population appears to reflect a difference in the magnitude of the speaking rate effect between the two populations. For the adults, mean proportion voiceless responses for the fast and slow speaking rates were 0.56 and 0.49, respectively, resulting in a speaking rate effect of 0.07 as measured in terms of the difference in proportion of voiceless responses. For the children, the speaking rate effect was 0.05, reflecting the difference between mean proportion voiceless responses in the fast and slow speaking rates (0.42 and 0.37, respectively).

Place of Articulation

The top panel of **Figure 3** shows mean proportion voiceless responses as a function of VOT for each place of articulation (labial vs. velar) and each population (child vs. adults); error bars indicate standard error of the mean. To derive these values, mean proportion voiceless responses was first calculated for each participant separately for each continuum and each VOT. Visual inspection of **Figure 3** suggests that both populations were sensitive to the place of articulation manipulation; for both children and adults, the identification function for the velar continuum is displaced toward longer VOTs compared to the labial continuum. As was observed for the speaking rate continua, the children show shallower identification functions compared to adults and voicing boundaries located at longer VOTs.

Trial-level responses (0 = voiced, 1 = voiceless) were fit to a GLMM following the procedures outlined for the speaking rate analysis. The fixed effects were VOT, place of articulation, population, their interactions, and trial number. VOT and trial number were scaled and centered around the mean; place of articulation (labial = −0.5, velar = 0.5) and population (children = −0.5, adults = 0.5) were contrast-coded. The random effects structure included random intercepts by subject, and random slopes by subject for VOT, place of articulation, their interaction, and trial number. The model is visualized in the bottom panel of **Figure 3** in terms of the fixed effects of VOT, place of articulation, and population; the left side of the panel shows the model faceted by place of articulation; the right side of the panel shows the same model faceted by population.

The results of the model showed a significant main effect of VOT ($b = 5.421$, $SE = 0.303$, $z = 17.908$, $p < 0.001$), place of articulation ($b = -6.410$, $SE = 0.451$, $z = -14.218$, $p < 0.001$), and population ($b = 3.001$, $SE = 0.399$, $z = 7.530$, $p < 0.001$).

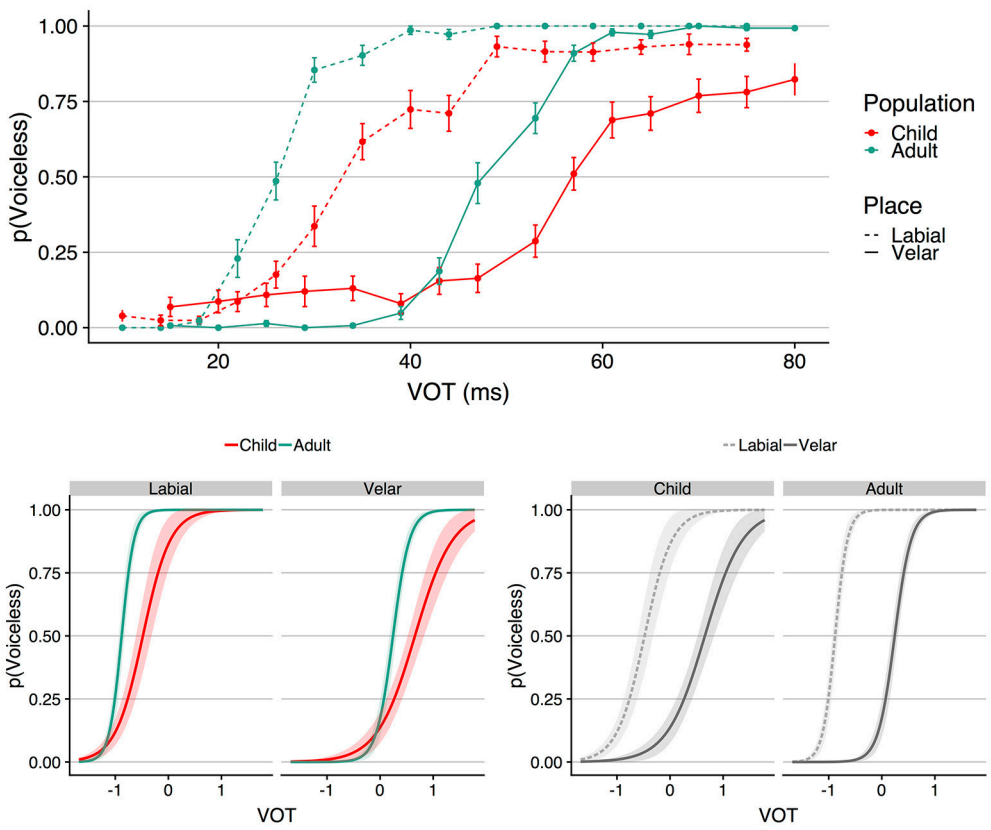

FIGURE 3 | The **(Top)** shows mean proportion voiceless responses as a function of VOT (ms) for each population at each place of articulation. Error bars indicate standard error of the mean. The **(Bottom)** shows the results of the GLMM; VOT (in ms) was entered into the model scaled and centered around the mean, which is shown as 0 on the abscissa for the plots on the bottom panel. At left, place of articulation is faceted to facilitate visualization of the interaction between VOT and population. At right, population is faceted to facilitate visualization of the place of articulation effect in each population. The shaded regions around the functions in the bottom panel indicate the 95% confidence interval region.

These results indicate that voiceless responses increased as VOT increased, the boundary for the velar continuum was located at longer VOTs compared to the labial continuum, and that children showed longer VOT voicing boundaries compared to the adults. The three-way interaction between VOT, population, and place of articulation was not reliable ($b = -1.142$, $SE = 0.725$, $z = -1.576$, $p = 0.115$), but the model did show that all three two-way interactions were significant. First, there was a significant interaction between VOT and place ($b = -1.616$, $SE = 0.410$, $z = -3.937$, $p < 0.001$), with the direction of the interaction indicating that the slope of the labial function was steeper compared to the slope of the velar function. Second, there was a significant interaction between VOT and population ($b = 4.183$, $SE = 0.573$, $z = 7.301$, $p < 0.001$), indicating that the children showed shallower identification curves compared to the adults. This interaction can be robustly observed in the bottom left panel of **Figure 3**.

Third, there was a significant interaction between population and place of articulation ($b = -5.441$, $SE = 0.861$, $z = -6.317$, $p < 0.001$). To explicate the nature of this interaction, a set of models were constructed to examine the effect of place of articulation within each population and the effect of population at place of articulation. The fixed and random effects structure of these models was identical to the primary model except for the removal of the fixed effect of interest in the follow-up model (i.e., the fixed effect of population was removed for the models examining the effect of place within each population, and the fixed effect of place was removed for the models examining the effect of population within place of articulation). These models showed a significant effect of place of articulation for both children ($b = -3.671$, $SE = 0.522$, $z = -7.031$, $p < 0.001$) and adults ($b = -9.556$, $SE = 0.884$, $z = -10.812$, $p < 0.001$), and a significant effect of population for both the labial place of articulation ($b = 5.246$, $SE = 0.788$, $z = 6.660$, $p < 0.001$) and the velar place of articulation ($b = 1.170$, $SE = 0.325$, $z = 3.597$, $p < 0.001$). Thus, the interaction between population and place of articulation observed in the main model does not reflect the absence of a place effect in either population, or the absence of a population effect at either place of articulation. As was observed for speaking rate, the interaction between place of articulation and population appears to reflect a difference in the magnitude of the context effect between the two populations. For the adults, the magnitude of the place effect was a difference in proportion voiceless responses equal to 0.21 (0.70 voiceless responses for labial continuum, 0.49 voiceless responses for velar continuum); this difference was equal to 0.18 in the children (0.55 voiceless responses for labial continuum, 0.37 voiceless responses for velar continuum).

Continuous Effect of Age on Voicing Identification

A final set of analyses was performed in order to further explicate the main effects of population reported above. For both the speaking rate and place of articulation analyses, a significant interaction between VOT and population was observed, indicating steeper identification slopes for the adults compared to the children, which is consistent with the finding of Burnham et al. (1991). To examine this issue further, we performed additional GLMMs on the child data in order to determine whether the slope of the identification function varied as a function of child's age. That is, the primary GLMMs reported above revealed a population-level difference; here we examined whether a more fine-grained relationship among children is observed when age of the child is considered continuously. Two generalized linear mixed-effects models were performed using the glmer() function of the lme4 package (Bates et al., 2015) with the binomial response family, one for the speaking rate data and one for the place of articulation data. Each model was fit to trial-level binomial responses (0 = voiced, 1 = voiceless) for the phonetic categorization task. The fixed effects were VOT, condition, age (in months), their interactions, and trial number. The model included by-subject random slopes and intercepts for VOT, condition, their interaction, and trial number. VOT, trial number, and age were scaled and centered around the mean; contrast coding was used for condition (for the speaking rate model: fast = −0.5, slow = 0.5; for the place of articulation model: labial = −0.5, velar = 0.5).

For the speaking rate model, the results showed a main effect of VOT ($b = 2.792$, $SE = 0.326$, $z = 8.564$, $p < 0.001$), condition ($b = -0.820$, $SE = 0.199$, $z = -4.121$, $p < 0.001$), and trial number ($b = 0.281$, $SE = 0.074$, $z = 3.816$, $p < 0.001$), as was observed in the primary model described above. There was no main effect of age ($b = -0.062$, $SE = 0.127$, $z = -0.485$, $p = 0.628$), nor was there an interaction between condition (i.e., speaking rate) and age ($b = -0.058$, $SE = 0.189$, $z = -0.306$, $p = 0.759$). However, a significant interaction was observed between VOT and age ($b = 0.923$, $SE = 0.264$, $z = 3.491$, $p < 0.001$), with the direction of the interaction indicating that responses became more categorical as age increased. The place of articulation model showed a main effect of VOT ($b = 3.367$, $SE = 0.336$, $z = 10.008$, $p < 0.001$), condition ($b = -3.692$, $SE = 0.478$, $z = -7.715$, $p < 0.001$), and trial number ($b = 0.212$, $SE = 0.068$, $z = 3.137$, $p = 0.002$). There was no main effect of age ($b = 0.275$, $SE = 0.203$, $z = 1.355$, $p = 0.175$), but age did interact with both condition (i.e., place of articulation, $b = -0.963$, $SE = 0.475$, $z = -2.029$, $p = 0.043$) and VOT ($b = 1.010$, $SE = 0.321$, $z = 3.147$, $p = 0.002$). The interactions between age and VOT are visualized in **Figure 4** for the speaking rate model (top panel) and the place of articulation model (bottom panel), which show predicted voiceless responses for the fixed effect of age at the mean age of the children, and at one standard deviation above and below the mean age. The pattern is such that as age increases, the slope of the function relating voiceless responses to VOT becomes steeper, thus moving in line with the steep identification functions that were observed in the adults[1].

[1]Parallel models were constructed representing child's age in years (instead of months), with converging results between the two metrics observed in all cases. In addition, parallel models were constructed for the adult data, which showed no significant interaction between VOT and age for the adult participants. These supplementary models can be viewed in the analysis script provided in the Data Availability Statement.

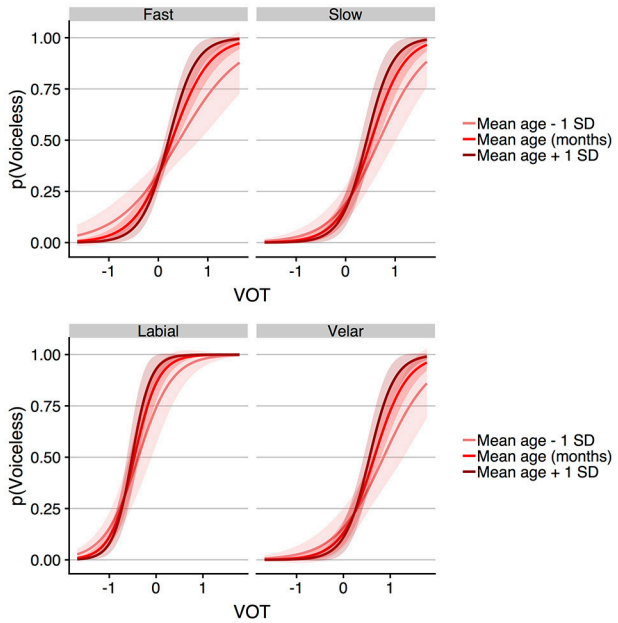

FIGURE 4 | The **(Top)** shows the results of the GLMM predicting voiceless responses in the child participants as a function of VOT, speaking rate, and age (in months). To illustrate the VOT by age interaction, the effect of age is shown at the mean age and at one standard deviation above and below the mean age. The **(Bottom)** shows the parallel GLMM analysis for the place of articulation model. In both panels, the interaction between age and VOT reflects the slope of the predicted identification function becoming steeper as age increases. The shaded regions around the functions in the both panels indicate the 95% confidence interval region. In the GLMMs, VOT (in ms) was entered into the model scaled and centered around the mean, which is shown as 0 on the abscissa in each plot.

DISCUSSION

A comprehensive theoretical account of speech perception must describe the mechanisms that support reliable mapping from the acoustic signal to speech sounds given a lack of invariance between the two. Moreover, such an account must specify how these mechanisms are fine-tuned across the developmental trajectory. Research to date has demonstrated that in the mature system, listeners achieve perceptual constancy, at least in part, by processing acoustic-phonetic variation categorically (Liberman et al., 1961). Moreover, category boundaries are dynamically shifted as a function of context (e.g., Volaitis and Miller, 1992). Infants begin this process very early in development, showing categorical processing of acoustic variation as early as 1 month of age (Eimas et al., 1971). Moreover, categorical processing shows rate-dependency in 3- and 4-month-old infants (Miller and Eimas, 1983). Despite these early indicants of adult-like phonetic categorization, numerous other studies have demonstrated that perceptual tuning persists throughout development (e.g., Burnham et al., 1991; Hoonhorst et al., 2011), and even remains functionally plastic in the mature system (e.g., Norris et al., 2003; Drouin et al., 2016). The current work contributes to a theoretical account of speech perception through examination of contextual influences on speech perception across the developmental trajectory.

Here we examined the effects of speaking rate and place of articulation on phonetic categorization in two populations, adults and school-aged children. For both populations, we found evidence that context influenced phonetic categorization. Specifically, the voicing boundary was located at longer VOTs for the slow compared to fast speaking rate (e.g., Volaitis and Miller, 1992). In addition, the voicing boundary was located at longer VOTs for the velar compared to the labial continuum (e.g., Lisker and Abramson, 1964; Miller et al., 1986; Cho and Ladefoged, 1999). Though the adults and children showed qualitatively similar patterns with respect to how speaking rate and place of articulation influenced phonetic categorization, the children did not show adult-like perceptual organization when performance was quantitatively compared. First, the absolute location of the voicing boundaries was located at longer VOTs in the children compared to the adults. Thus, though both groups processed VOT variation in a categorical manner, the precise perceptual boundary differed between the two groups. Furthermore, when compared to the adults, the children as a group showed less stable categorization for both speaking rate and place of articulation contexts in that phonetic decisions to the VOTs were less categorical. Moreover, the slope of the identification function became more categorical as age of the child increased. This pattern suggests that fine-grained perceptual tuning is still undergoing refinement during the school-age years, consistent with the developmental account of boundary precision put forth by Hoonhorst et al. (2011). Finally, for both speaking rate and place of articulation, the magnitude of the contextual influence was slightly greater in the adults compared to the children. This effect was numerically quite small, however, and may be a consequence of the increased variability and shallower identification slopes of the children compared to the adults. Additional evidence suggesting that the magnitude of the contextual influence adjustment is equivalent between the two populations comes from the follow-up models performed for the children, which showed no reliable interaction between child's age and the speaking rate effect, and only a modest influence of child's age on the place of articulation effect as measured by the beta estimate of the interaction.

In the current study, the voicing boundary was located at a longer VOT for the children compared to the adults in both cases. That is, the children required slightly longer VOTs in order to categorize stimuli as voiceless when compared to the adults. This finding is inconsistent with results from Hoonhorst et al. (2011), who did not observe any difference in the absolute location of the voicing boundary between the child and adult participants. The current results regarding the absolute location of the category boundary also differ from Flege and Eefting (1986), who found that children showed shorter VOT voicing boundaries compared to adults. We consider three possibilities that may explain this difference. First, the resolution of the VOT continuum in Hoonhorst et al. (2011) was 30 ms, but was manipulated in 5 ms increments for the continua used here. The mean difference in category boundary location between the children and adults never exceeded 12 ms; thus, it is possible

that the effect of population on boundary location would not be observed if VOT were manipulated with a coarser resolution. Second, the stimuli in the current study were based on naturally-produced tokens and perceptually sounded like natural speech, whereas previous studies used synthesis procedures that do not sound natural, and there is some evidence voicing identification in children varies as a function of natural vs. artificial speech (Coady et al., 2007). Third, as described in the procedure, stimuli were presented via headphones for the adult listeners but presented via speaker for the children. During pilot testing, we had difficulty establishing headphone compliance for some of the children and, as a consequence, decided to present the stimuli via speaker so that the listening environment would be identical among the children (i.e., we did not want to have some children using headphones while others did not). The testing environment was held constant for both children and adults in all other respects, including being tested in the same sound booth, but it is possible that some difference in the transducers between the headphones and speaker may have contributed to the difference in location of the category boundary.

The most robust difference observed in the current work concerns the slope of the identification functions, which were shallower for children compared to adults, consistent with less categorical processing in the former population. As for the population effect observed for the category boundary, this difference could may also reflect the use of headphones versus speaker between the two populations. However, there was a continuous relationship between the slope of the identification function and age of the children, such that the identification function was tuned to be more categorical as age increased. Because the mode of stimulus presentation was held constant among the children, this relationship among the children cannot be attributed differential modes of stimulus presentation. This result replicates findings from Hoonhorst et al. (2011) and Burnham et al. (1991). As to the mechanisms driving this effect, we consider three possibilities. First, the less stable identification functions may reflect the fact that children do not perform the metalinguistic, 2AFC task used in the current work with the same acuity as adults do. That is, the shallower identification functions may not reflect perceptual differences between the two populations *per se*, but rather may reflect differences in the ability to perform the task. This task is different from the tasks used in the two studies that previously reported no difference in the identification slope between children and adults. In Flege and Eefting (1986), children made decisions by circling "da" or "ta" on an answer sheet, and "the experimenter carefully monitored the subjects to ensure that they gave a response to each stimulus as it was presented." In Wolf (1973), participants indicated their identification decision via a verbal response, having been told to listen to the robot and "when he says /ba/, you say /ba/; when he says /pa/, you say /pa/." In the current study, responses were provided using a button box, with the response options labeled with pictures of target words. After demonstrating to the experimenter that the child could associate the spoken target with the corresponding picture, there was no further monitoring of their responses. Second, the shallower slopes may reflect decreased perceptual acuity in that the children did not recognize the acoustic distinction between continuum steps to the same degree as adults. While this possibility cannot be ruled out in the current behavioral work, it may be an unlikely possibility given that one interpretation of the shallower function is that it reflects heightened sensitivity to within category variation (Galle and McMurray, 2014). A third possibility is that the shallower identification functions for children reflect speech sound categories that are less defined than those in the mature system due to the more limited exposure to the important statistical distributions that allow listeners to learn speech sound categories that are defined by multiple acoustic-phonetic cues, as is the voicing contrast for stop consonants. On this view, the less consistent identification responses may reflect ongoing refinement in children's use of multiple acoustic cues to identify stop voicing (Nittrouer, 1992, 2002; McMurray et al., 2009). The current data set cannot adjudicate between these possibilities, and future examinations that use electrophysiological methods, for example, that do not require explicit responses (e.g., Toscano et al., 2010) will be useful in this respect. Moreover, future research is needed to examine the degree to which the results obtained here hold when the voicing boundary is specified by correlates other than or in addition to VOT (e.g., burst duration and intensity, spectral characteristics of the burst, F1 onset). Studies that manipulate cues other than VOT, or in trading relationships with VOT, will be able to explicate how contextual influences on stop voicing interact with the specific correlates that cue the voicing distinction.

Another important avenue for future research is to consider how children with atypical language development accommodate contextual variation. Impairment in low-level speech perception abilities has been proposed a diagnostic marker and/or locus of impairment in developmental language disorders including specific language impairment (SLI; Joanisse and Seidenberg, 1998, 2003), but the evidence to position perceptual deficits at the locus of SLI is equivocal with some studies showing no differences in categorical perception between typical and clinical populations (e.g., Coady et al., 2005), and others showing striking differences between the two populations (Joanisse et al., 2000). Examining how contextual variation influences phonetic categorization in children with SLI may prove to be an important step in testing hypotheses generated by processing-based theories of this disorder. Future work is aimed at examining this possibility.

In conclusion, healthy perceptual processing results in a perceptual system that is simultaneously stable with respect to phonological knowledge and flexible with respect to mapping acoustic-phonetic variation onto speech sound representations. There is a rich source of evidence indicating that infants begin tuning speech perception to their native language environment very early in development, with fundamental processes for language comprehension such as categorical perception of phonetic variability emerging in the first year of life. The results of the current work demonstrate that these processes continue to be refined even in the school-aged years.

DATA AVAILABILITY STATEMENT

The stimuli and data generated for this study along with the analysis script (in R) can be retrieved at the Open Science Framework: http://osf.io/4tq82.

AUTHOR CONTRIBUTIONS

Portions of this work were completed in conjunction with thesis projects by JC and HM under the direction of RT. JC contributed to stimulus development and experimental programming. JC and HM contributed to data collection. All authors contributed to data analyses and manuscript preparation.

FUNDING

Research was supported by a Faculty Large Grant to RT (College of Liberal Arts and Sciences, University of Connecticut) and NIH NIDCD grant R21 DC016141 to RT. The views expressed here reflect those of the authors and not the NIH or the NIDCD.

REFERENCES

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Boersma, P (2001). Praat, a system for doing phonetics by computer. *Glot Int.* 5, 341–345.Available online at: http://www.fon.hum.uva.nl/praat/manual/FAQ__How_to_cite_Praat.html

Bosker, H. R. (2017). Accounting for rate-dependent category boundary shifts in speech perception. *Attent. Percept. Psychophy.* 79, 333–343. doi: 10.3758/s13414-016-1206-4

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

Cho, T., and Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *J. Phon.* 27, 207–229. doi: 10.1006/jpho.1999.0094

Coady, J. A., Evans, J. L., Mainela-Arnold, E., and Kluender, K. R. (2007). Children with specific language impairments perceive speech most categorically when tokens are natural and meaningful. *J. Speech Lang. Hear. Res.* 50, 41–57. doi: 10.1044/1092-4388(2007/004)

Coady, J. A., Kluendrer, K. R., and Evans, J. L. (2005). Categorical perception of speech by children with specific language impairments. *J. Speech Lang. Hear.Res.* 48, 944–959. doi: 10.1044/1092-4388(2005/065)

Delattre, P. C., Liberman, A. M., and Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *J. Acoust. Soc. Am.* 27, 769–773. doi: 10.1121/1.1908024

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., and Pierce, A. (1997). Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (melopsittacus undulatus). *J. Acoust. Soc. Am.* 102, 1891–1897. doi: 10.1121/1.420111

Drouin, J. R., Theodore, R. M., and Myers, E. B. (2016). Lexically guided perceptual tuning of internal phonetic category structure. *J. Acoust. Soc. Am.* 140, EL307–EL313. doi: 10.1121/1.4964468

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science* 171, 303–306. doi: 10.1126/science.171.3968.303

Flege, J. E., and Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica* 43, 155–171. doi: 10.1159/000261768

Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Lang. Speech* 5, 171–189.

Galle, M. E., and McMurray, B. (2014). The development of voicing categories: A quantitative review of over 40 years of infant speech perception research. *Psychon. Bull. Rev.* 21, 884–906. doi: 10.3758/s13423-013-0569-y

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *J. Acoust. Soc. Am.* 63, 223–230.

Gerrits, E., and Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Percept. Psychophy.* 66, 363–376. doi: 10.3758/BF03194885

Goldman, R. (2000). *Goldman-Fristoe 2: Test of Articulation.* Circle Pines, MN: American Guidance Service.

Green, K. P., and Miller, J. L. (1985). On the role of visual rate information in phonetic perception. *Percept. Psychophy.* 38, 269–279. doi: 10.3758/BF03207154

Hoonhorst, I., Medina, V., Colin, C., Markessis, E., Radeau, M., Deltenre, P., et al. (2011). Categorical perception of voicing, colors and facial expressions: a developmental study. *Speech Commun.* 53, 417–430. doi: 10.1016/j.specom.2010.11.005

Joanisse, M. F., Manis, F. R., Keating, P., and Seidenberg, M. S. (2000). Language deficits in dyslexic children: speech perception, phonology, and morphology. *J. Exp. Child. Psychol.* 77, 30–60. doi: 10.1006/jecp.1999.2553

Joanisse, M. F., and Seidenberg, M. S. (1998). Specific language impairment: a deficit in grammar or processing? *Trends Cogn. Sci.* 2, 240–247.

Joanisse, M. F., and Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: evidence from a connectionist model. *Brain Lang.* 86, 40–56. doi: 10.1016/S0093-934X(02)00533-3

Kessinger, R. H., and Blumstein, S. E. (1998). Effects of speaking rate on voice-onset-time and vowel production: Some implications for perception studies. *J. Phon.* 26, 117–128. doi: 10.1006/jpho.1997.0069

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368. doi: 10.1037/h0044417

Liberman, A. M., Harris, K. S., Kinney, J., and Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and non-speech patterns. *J. Exp. Psychol.* 61, 379–388. doi: 10.1037/h0049038

Lisker, L., and Abramson, A. (1964). A cross-language study of voicing in word-initial stops: acoustical measurements. *Word* 20, 384–422. doi: 10.1080/00437956.1964.11659830

Long, J. A. (2018). *jtools: Analysis and Presentation of Social Scientific Data*. R. Package Version 1. Available Online at: https://cran.r-project.org/package=jtools

MacKain, K. S., Best, C. T., and Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Appl. Psycholinguist.* 2, 369–390. doi: 10.1017/S0142716400009796

McMurray, B., Aslin, R. N., and Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Dev. Sci.* 12, 369–378. doi: 10.1111/j.1467-7687.2009.00822.x

Miller, J. L. (1994). On the internal structure of phonetic categories: a progress report. *Cognition* 50, 271–285.

Miller, J. L., and Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition* 13, 135–165. doi: 10.1016/0010-0277(83)90020-3

Miller, J. L., Green, K. P., and Reeves, A. (1986). Speaking rate and segments: a look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43, 106–115.

Miller, J. L., and Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Percept. Psychophy.* 25, 457–465.

Miller, J. L., Mondini, M., Grosjean, F., and Dommergues, J. Y. (2011). Dialect effects in speech perception: the role of vowel duration in Parisian French and Swiss French. *Lang. Speech* 54, 467–485. doi: 10.1177/0023830911404924

Miller, J. L., and Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Percept. Psychophy.* 46, 505–512. doi: 10.3758/BF03208147

Myers, E. B. (2007). Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: an fMRI investigation. *Neuropsychologia* 45, 1463–1473. doi: 10.1016/j.neuropsychologia.2006.11.005

Myers, E. B., and Theodore, R. M. (2017). Voice-sensitive brain networks encode talker-specific phonetic detail. *Brain Lang.* 165, 33–44. doi: 10.1016/j.bandl.2016.11.001

Nagao, K., and de Jong, K. (2007). Perceptual rate normalization in naturally produced rate-varied speech. *J. Acoust. Soc. Am.* 121, 2882–2898. doi: 10.1121/1.2713680

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *J. Phon.* 20, 351–382.

Nittrouer, S. (2002). Learning to perceive speech: how fricative perception changes, and how it stays the same. *J. Acoust. Soc. Am.* 112, 711–719. doi: 10.1121/1.1496082

Nittrouer, S., and Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *J. Speech Lang. Hear. Res.* 30, 319–329. doi: 10.1044/jshr.3003.319

Norris, D., McQueen, J. M., and Cutler, A. (2003). Perceptual learning in speech. *Cogn. Psychol.* 47, 204–238. doi: 10.1016/S0010-0285(03)00006-9

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Lang. Speech* 46, 115–154. doi: 10.1177/00238309030460020501

Raven, J., Raven, J. C., and Court, J. H. (1998). *Coloured Progressive Matrices*. Bloomington, MN: Pearson.

Semel, E., Wiig, E. H., and Secord, W. A. (2004). *Clinical Evaluation of Language Fundamentals, Screening Test, Fourth Edition*. San Antonio, TX: Pearson.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, and Öhman, S. E. G. (1969). Cross language study of vowel perception. *Lang. Speech.* 12, 1–23. doi: 10.1177/002383096901200101

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *J. Exp. Psychol.* 7, 1074–1095. doi: 10.1037/0096-1523.7.5.1074

Theodore, R. M., and Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *J. Acoust. Soc. Am.* 128, 2090–2099. doi: 10.1121/1.3467771

Theodore, R. M., Miller, J. L., and DeSteno, D. (2009). Individual talker differences in voice-onset-time: contextual influences. *J. Acoust. Soc. Am.* 125, 3974–3982. doi: 10.1121/1.3106131

Theodore, R. M., Myers, E. B., and Lomibao, J. A. (2015). Talker-specific influences on phonetic category structure. *J. Acoust. Soc. Am.* 138, 1068–1078. doi: 10.1121/1.4927489

Toscano, J. C., and McMurray, B. (2012). Cue-integration and context effects in speech: evidence against speaking-rate normalization. *Atten. Percep. Psychophys.* 74, 1284–1301. doi: 10.3758/s13414-012-0306-z

Toscano, J. C., McMurray, B., Dennhardt, J., and Luck, S. J. (2010). Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychol. Sci.* 21, 1532–1540. doi: 10.1177/0956797610384142

Van Alphen, P. M., and Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *J. Phon.* 32, 455–491. doi: 10.1016/j.wocn.2004.05.001

Volaitis, L. E., and Miller, J. L. (1992). Phonetic prototypes: influence of place of articulation and speaking rate on the internal structure of voicing categories. *J. Acoust. Soc. Am.* 92, 723–735. doi: 10.1121/1.403997

Wade, T., and Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Percept. Psychophys.* 67, 939–950. doi: 10.3758/BF03193621

Welch, T. E., Sawusch, J. R., and Dent, M. L. (2009). Effects of syllable-final segment duration on the identification of synthetic speech continua by birds and humans. *J. Acoust. Soc. Am.* 126, 2779–2787. doi: 10.1121/1.3212923

Wolf, C. G. (1973). The perception of stop consonants by children. *J. Exp. Child Psychol.* 16, 318–331.

Zlatin, M. A., and Koenigsknecht, R. A. (1975). Development of the voicing contrast: perception of stop consonants. *J. Speech Lang. Hear. Res.* 18, 541–553.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.