

**--This paper has been accepted for publication in *Language and Speech*, please do not distribute without permission--**

**Comparing phonetic convergence in children and adults**

Melissa Paquette-Smith<sup>1\*</sup>, Jessamyn Schertz<sup>2</sup> & Elizabeth K. Johnson<sup>3</sup>

<sup>1</sup>Department of Psychology, University of California, Los Angeles

<sup>2</sup>Department of Language Studies, University of Toronto Mississauga

<sup>3</sup>Department of Psychology, University of Toronto Mississauga

**Author Note**

We would like to thank Chen Peng, Keren Smith, Sarah Rosen, Caterina Bordignon, Carthy Ngo and Lisa Hotson, as well as the other members of the Child Language and Speech Studies Lab for their support. This work was supported by grants from the Social Sciences and Humanities Research Council, the Natural Sciences and Engineering Research Council, and the Canada Research Chairs program.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

Correspondence should be addressed to Melissa Paquette-Smith, Department of Psychology, University of California, Los Angeles, 1258 Franz Hall, Box 951563, Los Angeles, CA, 90095-1563, USA. E-mail: [paquettesmith@psych.ucla.edu](mailto:paquettesmith@psych.ucla.edu)

### Abstract

Observations by sociolinguists suggest that when children relocate to a new community, they rapidly learn to imitate their peers, adopting the new local accent faster and more effectively than adults. However, few well-controlled laboratory experiments have been conducted comparing speech or accent imitation across ages. Here, we investigated Canadian English-speaking children and adults' imitation of three model speakers: a Canadian English talker, an Australian English talker, and a non-native Mandarin English talker who learned English later in life. The speech of all three talkers was manipulated to have elongated voice onset time (VOT) on word initial stop consonants. The dependent measure was how much participants would lengthen their VOTs after exposure to one of the talkers in two paradigms: delayed shadowing (Experiment 1) and immediate shadowing (Experiment 2). We predicted that overall children would show more imitation than adults, particularly when imitating the Canadian English talker, given previous work on children's social preferences. Although we did not observe age differences in either study, when shadowing was immediate, we found that imitation was influenced by the accent of the speaker, but not in the manner we predicted: both age groups imitated the Mandarin-accented model more strongly than the Canadian model. When shadowing was delayed, we observed no evidence of imitation. We discuss our findings in light of other recent work, and conclude that the development of speech imitation is an area ripe for further investigation.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

*Keywords:* accent/dialect acquisition, speech imitation, speech production, developmental psycholinguistics

### **Comparing phonetic convergence in children and adults**

Adults readily shift the way they speak when encountering a new interlocutor, or moving to a new community (Evans & Iverson, 2007; Munro et al., 1999; Pardo et al., 2012; Sancier & Fowler, 1997). For example, college roommates naturally converge towards each other over the course of a semester (Pardo et al., 2012), and adults who relocate to a new accent community tend to shift their productions towards the locally dominant accent (Evans & Iverson, 2007; Munro et al., 1999). Similar shifting can be seen in the short term, over the course of a single conversation. For example, if talkers are asked to repeat after another talker, they tend to adopt the speaking style of the talker they are shadowing (Babel, 2012; Babel et al., 2014; Shockley et al., 2004). This imitation of another speaker in lab-based tasks – often referred to as phonetic accommodation, or convergence – has been well-studied in adults (see Pardo, 2013 for a review). One of the most interesting characteristics of these studies is that a wide variety of factors, such as frequency of the shadowed words or the social characteristics of the speaker can modulate how readily or how strongly adults converge on the speaking style of others (Babel et al., 2014; Goldinger, 1998; Pardo et al., 2017).

Despite the large and ever-growing body of literature on lab-based phonetic convergence in adults, to date only one study has examined this type of phonetic convergence in children. In this study, American-English speaking preschoolers and third graders appeared to show even stronger convergence towards an American-English speaking model than adults (Nielsen, 2014).

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

However, it is not known whether children's imitation is modulated by the same factors as adults' and whether children will imitate model speakers from different accent communities as readily as they imitated models from the same accent community. In the current study, we further explore children's phonetic convergence towards model talkers in the lab. Our goal is twofold: 1) to revisit the claim that children show greater imitation than adults, and 2) to explore factors that may modulate phonetic convergence by adults versus children. By examining these questions in a controlled laboratory task, we hope to better understand language imitation and phonetic convergence by children and adults in 'the wild'.

The fact that we have very little well-controlled laboratory data on phonetic convergence in young children is surprising given the important role that imitation is thought to play in children's L1 and L2 acquisition. In infancy, children have been shown to imitate the speech of their caregivers (Kuhl & Meltzoff, 1996; see however McRoberts & Best, 1997; Papoušek & Papoušek, 1989 for evidence that caregivers can also converge upon the productions of their infants). Although infants' speech is shaped by the productions of their caregivers initially, by their second birthday there is evidence that infants become highly attuned to the pronunciations that are dominant in their community (Floccia et al., 2012; van der Feest & Johnson, 2016). Once they reach preschool age, children will readily imitate the speaking style of their peers and community members. Indeed, toddlers and young children seem far more adept at adapting to new accents than adults (Chambers, 1992; Smith et al., 2007; Tagliamonte & Molfenter, 2007). Young children's remarkable ability to rapidly adapt their speaking style has been linked to the literature on critical periods for language development (Johnson & Newport, 1989), where after adolescence the ability to acquire a native-like accent in a new language rapidly declines (Piske et al., 2001).



## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

Although sociolinguistic studies suggest that there are age-related changes in imitation, it is difficult to ascertain why these differences occur without systematically studying imitation by children and adults under controlled conditions. Laboratory studies investigating the degree to which adults will converge towards different model talkers more broadly, have played an important role in our understanding of the factors that modulate phonetic convergence. These studies have shown us that information such as the type of imitation task, the social desirability of the model, and the frequency of the words that are imitated all influence the degree to which adults will imitate model speakers (Babel, 2010, 2012; Babel et al., 2014; Goldinger, 1998; Goldinger & Azuma, 2004; Natale, 1975; Pardo, 2013). In sociolinguistic studies of real world accommodation, there is ample evidence that social factors, in particular, have a substantial impact on which speakers will converge over time (Berthele, 2002; Giles et al., 1973; Pardo et al., 2012). For instance, 9-year-olds who were more socially integrated into their elementary school classroom showed greater accommodation towards their peers than children who were more socially isolated (Berthele, 2002). In this view, a speaker's production choices are like small 'acts of identity', signaling to others who the speaker would like to affiliate with. Perhaps the strongest evidence of socially-mediated convergence in the lab comes from work showing that the degree to which adult New Zealanders imitated the productions of an Australian-accented model was modulated by their implicit social attitudes towards Australians. Adults that viewed Australians more positively showed stronger imitation of the Australian speaker's vowels, whereas participants who viewed Australians more negatively showed weaker imitation (Babel, 2010). Taken together this work supports the notion that imitation (even in lab settings) is modulated by the social characteristics of the speaker (Giles, 1973; Giles et al., 1991; Giles &

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

Ogay, 2007). However, it is not known whether children show similar social selectivity in their imitation of others in short term laboratory tasks.

The only study to assess children's phonetic convergence in the lab, compared children and adult's imitation of a model speaker from the same accent community. Here, American-English speaking preschoolers, third graders and adults were prompted with pictures to name objects whose labels began in voiceless stop consonants (e.g., pear, puppy, peanuts; Nielsen, 2014). Participants' productions were recorded to serve as a baseline to compare with subsequent recordings. Then they were exposed to an American English model speaker producing this same list of words with artificially lengthened Voice Onset Time (VOT)<sup>1</sup> on the initial stop consonants. Immediately after exposure to the list containing words with lengthened VOTs, the participants were once again prompted with pictures to repeat the words. Acoustic measurements showed that both children and adults displayed evidence of phonetic convergence, with their post-exposure productions showing longer average VOTs than their baseline productions. When they compared the degree to which children and adults converged towards the model, they found that the child participants converged more than the adult participants. The author concluded that as phonological development increases the propensity to imitate others decreases.

The finding that children show stronger phonetic convergence than adults in short-term laboratory tests dovetails nicely with studies showing how easily toddlers and young children adapt to new accents and speaking styles in the real world (Chambers, 1992; Smith et al., 2007; Tagliamonte & Molfenter, 2007). However, given that previous work has only tested children's lab-based convergence towards a single model speaker, it is unclear whether children show

---

<sup>1</sup> The delay between the release of the stop consonant and the onset of voicing.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

greater imitation of all model speakers or whether their imitation is limited to speakers from the same accent community. Nonetheless, the existing literature enables us to make some predictions. Studies of children's accent-based social preferences suggest that like adults, 5-year-olds are very aware of accent differences and show strong social preferences for talkers that speak a standard (or local) variant of their native language over talkers who speak with a different regional or non-native accent (see Gluszek & Dovidio, 2010; Kinzler et al., 2007, 2009; Paquette-Smith et al., 2019; Souza et al., 2013). Given children's biases against non-local speakers, it is possible that these biases may influence which talkers children choose to imitate.

In the current study, we further investigate the claim that children imitate more strongly than adults by comparing children's and adults' convergence towards three model talkers: a native English talker from the same accent community, a native English talker from a different accent community and a non-native English talker. If, similar to sociolinguistic studies and data on critical periods for language acquisition, children are better imitators than adults, then children should converge their productions more towards the model speakers than adults. Additionally, if children and adults' imitation is influenced by the social status of the model's accent, then both groups may show greater imitation of the local-accented model, but little or no imitation of the non-local accented models.

### Experiment 1

The findings from Nielsen (2014) suggest that American English-speaking children imitated the productions of an American English model more than adults. In Experiment 1, we examine whether this finding can be applied more broadly to different populations and across accent communities. To do this, we compared children and adult's imitation of three different-

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

accented models: A local Canadian-accented model, a regional Australian-accented model, and a non-native Mandarin-accented model<sup>2</sup>. Using a picture naming task modeled after Nielsen (2014), participants were recorded producing a list of words (containing words beginning with /p/ and /k/), before and after exposure to one of the three model speakers. As in Nielsen (2014), for each initial /p/ the VOT of the models' productions were artificially lengthened to elicit observable imitation. If participants imitate the elongated VOTs of the model talkers, then we should observe an increase in the length of their post-exposure VOTs relative to their baseline VOTs.

Although previous work predicts that children will imitate the elongated VOTs of the Canadian model talkers more than adults, it is not known whether children's imitation is impacted by the social characteristics (accent group) of the model. If children's convergence is modulated by social biases, as we expect it will be, then children may imitate the Canadian-accented in-group members more than the Mandarin or Australian-accented out-group members. Although we predict to see main effects of the accent of the model and the age group of the participant (children vs. adults), it is difficult to ascertain how these variables might interact. It is possible that children might imitate certain speakers more than adults (e.g., the Canadian speaker), but imitate other speakers (e.g., the regional or non-native speakers) to the same degree as adults. It might also be the case that children show greater imitation overall, but the patterns of imitation across speakers are similar to adults.

### Method

---

<sup>2</sup> In this paper, "regional accents" refer to different varieties of a language spoken by native speakers in different geographical regions, whereas "non-native accents" refer to varieties of any language learned as a foreign language.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

**Participants.** Forty-eight Canadian English-speaking children (Age range = 5.04 to 6.98 years, Mean age = 6.03 years; 24 females, 24 males) and sixty Canadian English-speaking adults (Age range = 17 to 39 years; Mean age = 18.42; 53 females, 7 Males) were tested in this experiment. Sample size was selected to mirror the number of participants in Nielsen (2014), who tested between 15-18 participants in each age group. Here we tested between 16-20 participants in each of the three accent groups for both children and adults. In order to participate in this study, children had to speak English at least 90% of the time (as reported by their parent or caregiver). Adult participants had to have learned English in North America before the age of six and report that English was their dominant language (currently spoken at least 80% of the time). Participants reported having no diagnosed speech, hearing or vision issues that could interfere with participation. Five additional children and eight additional adults were excluded from the analysis due to technical issues in the running of the experiment (10; e.g., the software freezing), and a failure to follow instructions (3).

**Design.** The design of the study was modelled after Nielsen (2014). In the baseline and the post-exposure phases, participants were recorded naming a set of 59 images presented to them in a random order. The word list contained words beginning with /p/ (e.g., paint, peppers), words beginning with /k/ (e.g., cake, coat) and filler words which did not begin with a voiceless stop consonant (e.g., lion, star, zebra). In the exposure phase, participants listened to a subset of the /p/-initial words produced by either a Canadian-accented model, an Australian-accented model, or a Mandarin-accented model. Participants were randomly assigned to listen to the model produce the words in List 1 (13 of the words beginning with /p/) or List 2 (the other 13

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

words beginning with /p/; see Appendix A)<sup>3</sup>, the VOTs of all exposure tokens were elongated to elicit imitation. The word list was repeated three times (for a total of 39 exposure tokens).

Finally, in the post-exposure phase participants were recorded re-naming the same set of 59 images from the baseline. This design allows us to examine how strongly imitation generalizes to new words that are similar to those heard in the exposure (i.e. new words beginning with /p/) as well as to a different stop consonant (i.e. new words beginning with the segment /k/).

**Production Task: Wordlist.** The production list for the baseline and post-exposure phases consisted of 59 nouns commonly known by preschool-aged children (e.g., strawberry, keys, pants; see Appendix A for a complete list of stimuli). Twenty-six of the nouns began with /p/ (half of which they were exposed to in the exposure phase and half of which they were not), and 13 began with /k/. The remaining 20 nouns did not begin with a voiceless stop and were used as filler items.

**Production Task: Visual Stimuli.** The visual stimuli consisted of images of the 59 nouns used in the study (26 /p/ words, 13 /k/ words and 20 fillers). The images were placed on a white background and were presented to participants on a 26-inch computer monitor. All participants labeled the nouns from images, no text version of the words was provided.

**Exposure Task: Auditory Stimuli.** The auditory stimuli for the exposure phase consisted of the 26 words beginning with /p/ from the production list, recorded by three female “model talkers” varying in accent. Two of the talkers learned English as their first language but differed in dialect: one spoke Canadian English (born, raised, and living in southern Ontario) and the other spoke Australian English (born and raised in Australia and living in southern Ontario).

---

<sup>3</sup> Although there was random assignment to List, the Canadian and Australian conditions were run before the Mandarin Conditions.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

The third talker was born and raised in China and learned Mandarin as a first language. She moved to Canada as an adult and was living in southern Ontario at the time of the study. The Australian- and Mandarin-accented talkers both had clearly detectable accents.

The auditory stimuli were recorded in a double-walled sound-attenuated IAC booth using a Zoom H4N Pro audio recorder and a high-quality external Sennheiser MKE600 shotgun microphone. Each word was recorded two to three times and the clearest, most natural sounding token from each speaker was selected as the token for manipulation. With all else being equal, the token whose duration was most similar to the average duration by all three speakers was chosen, in order to minimize the extent of manipulation that would be done.

The auditory stimuli were then manipulated to have extended VOT (see Table 1 for the mean VOT duration before and after the word duration manipulation). For comparability across the different model accents, we started with baseline tokens that had been equalized for speech rate (see below for details), then manipulated VOT to be identical across all three talkers. This ensured that any differences in participant's imitation of the three models was not due to the fact that a particular speaker had longer VOTs or slower productions. In order to do this, the duration of each word was first equalized to match the mean word duration (for that word) of the three talkers. Then, the mean VOT was calculated for that word, and 50ms was added to this mean to reach the target VOT value. VOT and word durations were word-specific (e.g. the overall VOT and duration could differ between 'paint' and 'peacock'), but the VOT for each word (e.g. 'paint') was identical across the three talkers. The mean VOT values after equalizing word duration were 84ms ( $SD = 17$ ) for the Canadian, 61ms ( $SD = 23$ ) for the Australian, and 67ms ( $SD = 23$ ) for the Mandarin-accented talker. After the VOT manipulation, the mean VOT (for all model talkers) was 120ms.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

*Table 1.* Mean VOT duration for each talker's stimuli set before and after the equating word duration.

Talker	VOT duration before manipulation		VOT duration after equating word duration		VOT duration after VOT manipulation	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Canadian-accented	75	15	84	17	120	15
Australian-accented	62	24	61	23	120	15
Mandarin-accented	64	16	67	20	120	15

Similar to previous work, VOTs were extended by copy-pasting the medial (stable) segments of the aspiration using Praat (following Nielsen, 2014; Shockley et al., 2004). To prevent introducing any audible clicks or pops, all splicing was done at zero crossings. In about a third of cases, extending the medial portion of the aspiration sounded somewhat artificial. Following Nielsen (2014), for these tokens, we used the aspiration from another token of the same word (produced by the same speaker) to extend the VOT.

The six versions of the experiment (3 model talkers X 2 lists of exposure stimuli) were created using Articulate Storyline, which presented the stimuli from each of the three phases (baseline, exposure and post-exposure phase) in a random order. In the exposure phase, the assigned talker's productions of /p/ words from either List 1 or List 2 were presented alongside the images of those items.<sup>4</sup> Exactly 500ms after each image appeared, participants heard the token of the word followed by 1 second of silence. The word list was repeated three times to create the exposure phase.

**Procedure.** For the duration of the procedure, participants sat beside an experimenter in an IAC sound-attenuated booth (see Figure 1). Each participant was positioned approximately 15

<sup>4</sup> Note: 6/48 participants heard a different version of two exposure tokens. The two exposure tokens they heard still had the extended VOT but were not the same tokens that other children in that condition listened to. Data from these 6 children did not look different from the other children in the same condition.



## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

inches away from a Sennheiser shotgun microphone, located just below the computer monitor used to display the images. The microphone recorded the participants' productions at a 48,000Hz sampling rate into a Zoom H4N Pro audio recorder. In the baseline phase, participants were instructed to say the word for each of the pictures on the screen in isolation (i.e., without the addition of any other words like 'a' or 'the'). After the participant produced each word, the experimenter clicked to display the next image. If participants had difficulty labeling the image, the experimenter gave them a hint, but did not produce the words for them. If they still did not know the word after the hint, then the experimenter moved onto the next image. In the exposure phase, participants were told to sit quietly and listen while someone else said some of the words. The exposure stimuli were presented over high-quality Alesis M1Active 520USB speakers. Finally, in the post-exposure phase, participants were asked to complete the baseline task a second time. The baseline and post-exposure task took about 7-10 minutes; the exposure phase took about 1.5 minutes.

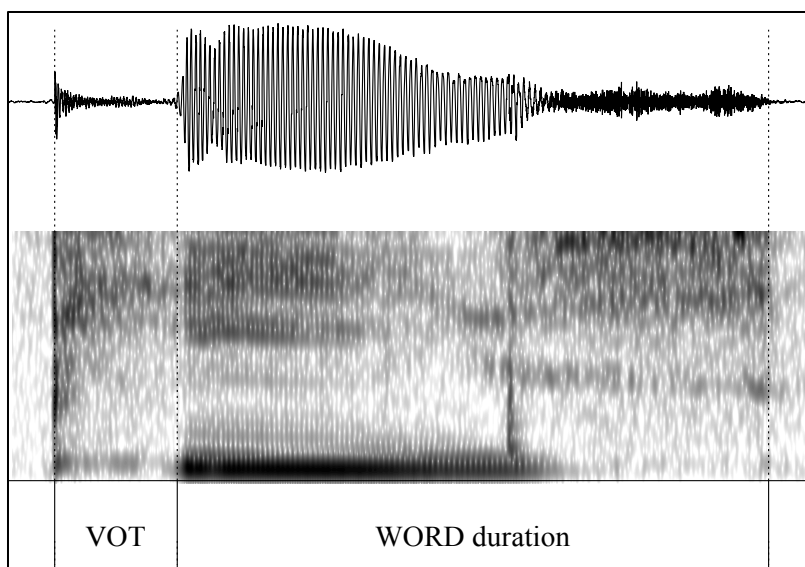


**Figure 1.** Child participating in Experiment 1. The experiment consisted of three phases. In the Baseline phase, the participant was presented with images and asked to produce the corresponding word. In the Exposure Phase, participants listened to one of three model talkers produce a subset of the words beginning with /p/ and in the Post-Exposure phase the child produced the baseline word list a second time.

**Data Analysis.** The VOT of the participants' /p/ and /k/ productions were measured using AutoVOT, an automated VOT measurement software (Keshet et al., 2014; Sonderegger & Keshet, 2012). Prior to analyzing this dataset, the AutoVOT algorithm was trained using 311 hand measured VOT tokens (produced by four adult participants from the current study). All the measurements made using AutoVOT were manually checked and corrected when necessary.

A total of 8,151 tokens were included in this analysis. 3,492 of the tokens were produced by child participants ( $N = 48$ ) and 4,659 tokens were produced by the adult participants ( $N = 60$ ). An additional 252 child tokens were excluded due to the child producing an incorrect label (or no label) for the item ( $n = 226$ ; e.g., producing 'hangaroo' instead of 'kangaroo' or 'nuts' instead of 'peanuts'), the presence of background noise ( $n = 9$ ; e.g., the shuffling of feet, the item being spoken over by the experimenter), or a production that contained a voiceless vowel following the VOT (which made the boundary between the aspiration and the vowel difficult to measure;  $n = 17$ ). An additional 21 tokens were excluded from the adult dataset due to: the participant providing an incorrect label ( $n = 9$ ), the experimenter accidentally producing the target word ( $n = 1$ ), background noise ( $n = 9$ ), or the production contained a voiceless vowel following the VOT ( $n = 2$ ).

The word duration (excluding word initial VOT) was measured from the end of the /p/ or /k/ phoneme to the end of the word. For words that ended in a stop consonant (e.g., cake, pocket) the end of the word was always marked prior to the final stop closure, excluding the closure and release (this was done for consistency, as stop consonants are not always released). In cases where a participant did not produce a word with the plural inflection (e.g., 'pepper' instead of 'peppers'), the final 's' was removed from both the Baseline and Post-exposure token.



**Figure 2.** Measurement of VOT and Word Duration from the word “peas”.

**Statistical Analysis.** A linear mixed-effects model was implemented in R 4.0.0 (R Core Team, 2020) using the lme4 package (Bates et al., 2018)<sup>5</sup>. The dependent variable (or outcome of interest), VOT duration, was modeled as a function of the following fixed factors (reference levels in italics): Age Group (*Child*, *Adult*), Phase (*Baseline*, *Post-exposure*), the Model’s Accent (*Canadian-accented*, *Australian-accented* or *Mandarin-accented*), Word Duration, and Segment (*/p/*, */k/*). The three-way interaction between Age, Phase, and Accent and the two-way interaction between Segment and Phase were also included as they were relevant to our theoretical questions (see explanation below). All categorical factors were simple-coded, and the continuous factor of Word Duration was centered and scaled to z-scores. Along with these fixed factors, the model included random by-participant intercepts and by-participant slopes for Phase, and random by-item (word) intercepts and by-item slopes for Phase. For the fixed effects, we report *b*, standard error, and *t*-values. *P*-values were calculated using Satterthwaite approximations to degrees of

<sup>5</sup> Model structure: `lmer(VOT ~ Phase * Accent * Age + Phase * Segment + Word Duration + (Phase | Participant) + (Phase | Item))`

freedom and implemented using the *lmerTest* package in R (Kuznetsova et al., 2015), and we used an alpha-level of 0.05 as the criterion for significance. Follow-up tests for significant interactions were conducted using the *phia* package in R, using holm-adjusted *p*-values (De Rosario-Martinez, 2015).

If participants imitate the extended VOT of the model talker, we would expect the post-exposure VOTs to be longer than the Baseline (i.e., a main effect of Phase). Our primary question of interest is whether imitation differs by accent (i.e. an interaction of Phase \* Accent), by age (i.e. an interaction of Phase \* Age), and most importantly, whether any accent-based differences differ by age (i.e. the three-way Phase \* Accent \* Age interaction). We also expect an overall main effect of Word Duration (with words with longer durations having longer VOTs), and Segment (with /k/ having longer VOT than /p/), as these are well-attested phonetic effects (e.g. Nielsen, 2011). The Phase \* Segment interaction was included to test whether imitation differs for segments that were not included in the exposure phase (/k/), compared to those that were included (/p/).

### Results and Discussion

Participants' average VOTs were computed for each condition (see Figure 3), and results of the statistical model described above are shown in Table 1. Because all variables were scaled and centered, the beta-estimate for each fixed factor represents the predicted difference in VOT (in ms) across the two levels of a categorical factor, or for a one-standard-deviation change in the continuous factor of word duration, collapsed over all levels of the other factors.

The intercept shows that participants' VOTs were on average 75ms. We observed no main effect of Phase,  $b = 0.19$ ,  $SE = 0.83$ ,  $t(111.09) = 0.23$ ,  $p = .822$ , indicating that as a group

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

there was no change in VOT from the Baseline to the Post-exposure phase. There were, however, significant main effects of Age Group, Word Duration, Segment Type and the Model's Accent. In line with work suggesting that children may produce longer VOTs (Yu et al., 2015), the VOTs produced by the child participants in our study were longer on average than the VOTs produced by the adult participants,  $b = -12.66$ ,  $SE = 2.63$ ,  $t(103.11) = -4.82$ ,  $p < .001$ . Additionally, we replicated the well-attested phonetic effects of word duration and segment on VOT duration. As expected, we found that word duration predicted VOT duration,  $b = 4.78$ ,  $SE = 0.49$ ,  $t(5023.04) = 9.74$ ,  $p < .001$ , and words beginning with /k/ had longer VOTs than words beginning with /p/,  $b = 21.81$ ,  $SE = 3.14$ ,  $t(36.95) = 6.94$ ,  $p < .001$ . Although we had no reason to predict a main effect of Accent, we found that the VOT duration of the participants in the Australian condition tended to be shorter on average than participants assigned to the Canadian condition,  $b = -7.00$ ,  $SE = 3.20$ ,  $t(101.40) = -2.18$ ,  $p = .031$ . This indicates that there may have been systematic differences in the VOTs of children assigned to the Australian condition, however, this does not indicate that there was imitation of the Australian speaker.

*Table 1.* Table of the fixed effects from the Experiment 1 LME model predicting VOT duration. Reference levels are in italics.

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
<b>Intercept</b>	74.75	2.02	36.94	< 0.001
<b>Phase (Post-exposure vs. <i>baseline</i>)</b>	0.19	0.83	0.23	0.822
<b>Model's Accent</b>				
Australian (vs. <i>Canadian</i> )	-7.00	3.20	-2.18	0.031
Mandarin (vs. <i>Canadian</i> )	-2.82	3.20	-0.88	0.380
<b>Age Group (Adult vs. <i>Child</i>)</b>	-12.66	2.63	-4.82	< 0.001
<b>Segment Type (/k/ vs. <i>/p/</i>)</b>	21.81	3.14	6.94	< 0.001
<b>Word Duration</b>	4.78	0.49	9.74	< 0.001
<b>Phase x Model's Accent</b>				
Phase x Australian	0.88	1.97	0.45	0.655
Phase x Mandarin	1.34	1.98	0.68	0.499
<b>Phase x Age Group</b>	-0.19	1.61	-0.12	0.907

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

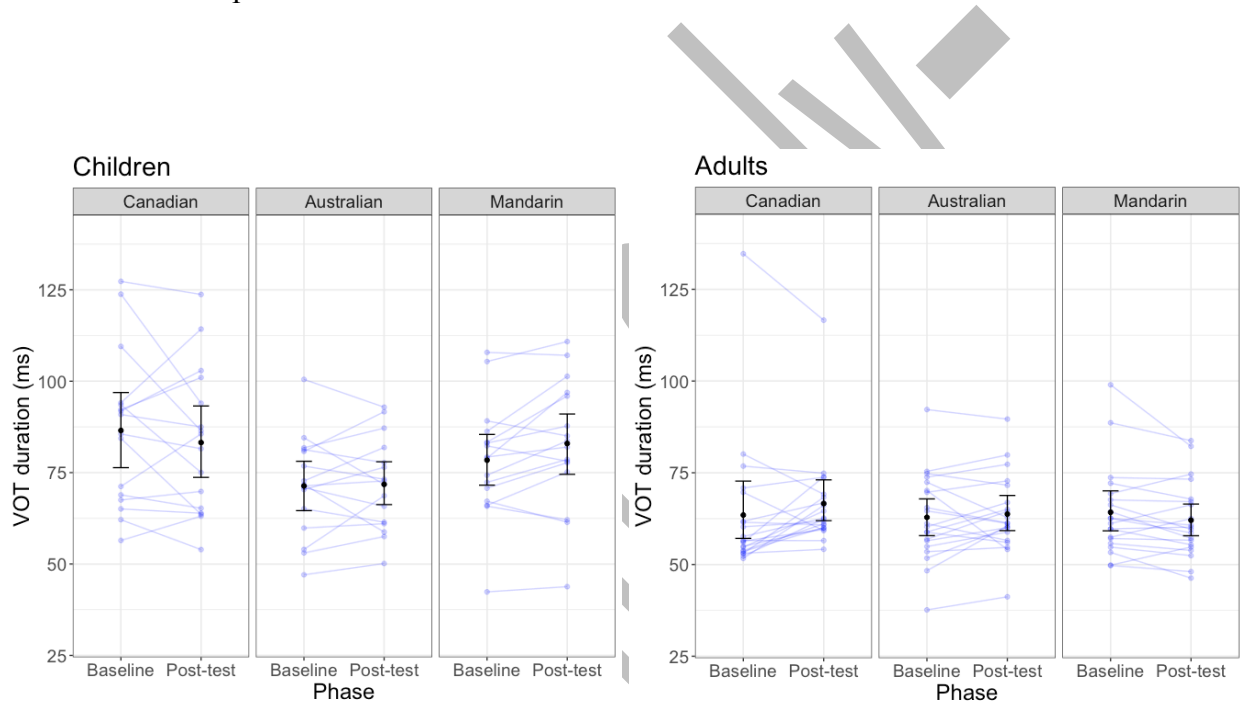
<b>Phase x Segment Type</b>	-0.03	1.04	-0.03	0.979
<b>Model's Accent x Age Group</b>				
Australian x Age	11.65	6.41	1.82	0.072
Mandarin x Age	2.80	6.41	0.44	0.663
<b>Phase x Model's Accent x Age Group</b>				
Phase x Australian x Age	-5.64	3.94	-1.43	0.156
Phase x Mandarin x Age	-12.79	3.96	-3.23	0.002

In order to assess whether Age Group, Segment Type or the Model's Accent influenced the amount of imitation observed we examined the interactions between those variables and Phase (Baseline vs Post-exposure phase). In contrast to previous work (Nielsen, 2014), we found no imitation in either age group (as is evidenced by the non-significant effect of Phase). We also do not find that children imitated the elongated VOTs of the model talkers more than adults,  $b = -0.19$ ,  $SE = 1.61$ ,  $t(101.61) = -0.12$ ,  $p = .907$ . There was no difference in the imitation of /p/ segments compared to /k/ segments in our task,  $b = 0.03$ ,  $SE = 1.04$ ,  $t(1829.95) = -0.03$ ,  $p = .979$ . Finally, the interaction between Phase and Model's Accent was not significant for the Australian versus Canadian,  $b = 0.88$ ,  $SE = 1.97$ ,  $t(101.02) = 0.45$ ,  $p = 0.655$ , or the Mandarin versus Canadian conditions,  $b = 1.34$ ,  $SE = 1.98$ ,  $t(102.11) = 0.68$ ,  $p = .499$ . Thus, as a group, participants did not seem to imitate the elongated VOTs of the Canadian model talkers more than the VOTs of the unfamiliar Mandarin or Australian-accented talkers.

Although there was no main effect of Phase, there was a significant three-way interaction between Phase, Age, and the Mandarin vs. Canadian accent condition. This could be an indication that children and adults showed different accent-based imitation effects, which may have obscured the main effect of Phase. In order to test whether imitation occurred for each model accent, we performed follow-up tests (with holm corrected  $p$ -values) to test the effect of Phase for each Accent/Age group separately. None of these were significant: (Adults: Canadian

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

$\chi^2=1.70, p = .625$ ; Australian  $\chi^2=0.07, p = 1.000$ , Mandarin  $\chi^2=2.01, p = .625$ ; Children: Canadian  $\chi^2=2.83, p = .462$ ; Australian  $\chi^2=0.01, p = 1.000$ , Mandarin  $\chi^2=3.96, p = .279$ ). The significant 3-way interaction in the model above is likely driven by the fact that children showed stronger (although not significant) imitation of the Mandarin-accented talker compared to the Canadian-accented talker, whereas adults showed stronger imitation of the Canadian compared to the Mandarin speaker.



*Figure 3.* Children and adult's mean VOT before and after exposure to the Canadian, Australian and Mandarin-accented models in Experiment 1. Error bars represent the 95% CIs of by-participant means.

Using corrected simple effects tests, we do not see evidence of imitation (i.e., a significant change in VOT length from Baseline to the Post-exposure phase) for any of the model talkers. This is surprising given that the procedure and methodology used in our experiment were similar to Nielsen (2014). Based on this study, we expected to see evidence of imitation in the local accent (i.e., Canadian) condition, as well as greater imitation by children than by adults.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

However, there may be some key differences in how participants responded to our task, which may explain why our task did not elicit imitation. First, the average VOT of our child participants' baseline productions (in the Canadian condition) were substantially longer ( $M$  for /p/ = 80 ms,  $M$  for /k/ = 99) than the Baseline VOTs reported in Nielsen (2014;  $M$  = 60ms for /p/,  $M$  = 73ms for /k/). In fact, the mean baseline VOT in our child participants was even longer than the mean Post-exposure VOT reported in Nielsen's child participants ( $M$  = for /p/ = 71 and  $M$  = for /k/ = 84ms). This could be an indication that the children in our sample hyperarticulated the words in the baseline or just generally have longer VOTs (Yu et al., 2015), which may make it more difficult to observe imitation in this sample.

Alternatively, it may be the case that children do not consistently imitate the extended VOTs of model talkers. Similar to previous work (Lowenstein & Nitttrouer, 2008; Yu et al., 2015), our data suggest that there can be substantial variability in preschoolers' production of voiceless stops, which may make VOT imitation effects particularly difficult to replicate. Indeed, only one study (i.e., Nielsen, 2014), with a relatively small sample, has found evidence of VOT imitation in children. Thus, it is possible that, similar to many of the findings in the imitation literature (see Pardo, 2013 for a discussion), this particular effect may be quite small or difficult to replicate, especially in a population that has relatively long, and variable VOT productions (Yu et al., 2015). Finally, given that the extent of convergence can be modulated by characteristics of the model talker (e.g. Babel et al., 2014), it is possible that the contrasting results of our study and Nielsen's could be due to differences in social characteristics and/or perceived attractiveness of the model talker's speech.

Although we did not find evidence of imitation, we did observe a trend towards children showing stronger (although not significant) imitation of the Mandarin-accented talker compared



## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

to the Canadian-accented talker, whereas adults showed the opposite pattern. This raises the possibility that there could be age-related changes in the imitation of accented speakers or perhaps this finding simply reflects differences in children's and adult's ability to perceive the Mandarin and Australian accents as different from their Canadian dialect. The latter possibility may be more likely given work demonstrating that young children are not as apt as adults in perceiving and categorizing accents (Dossey et al., 2020; Jones et al., 2017). In order to ensure that participants "noticed" the accent of the models in the single-word stimuli, all participants completed a follow-up task, as part of their visit, in which they listened to a subset of the word list produced by the three model talkers<sup>6</sup>. On each trial, participants heard a pair of talkers, and were asked to select "who talks most like you? Like they grew up here?." Despite only hearing isolated words, participants were highly skilled in identifying that the Canadian talker sounded more like them. As a group they correctly selected the Canadian talker over the Australian (90.43% of the time) and the Canadian over the Mandarin (89.36% of the time). There was no difference in performance between the Mandarin and Australian trials,  $b = -0.16$ ,  $SE = 0.57$ ,  $z = -0.28$ ,  $p = .778$ <sup>7</sup>. However, similar to other accent identification tasks (Jones et al., 2017) adult participants were better able to identify the Canadian speakers compared to child participants,  $b = 3.00$ ,  $SE = 0.89$ ,  $z = 3.35$ ,  $p < .001$ . That being said, child participants were quite capable of selecting the Canadian speakers, selecting them 76.32% of the time when paired with the Australian speakers and 78.95% when paired with the Mandarin speakers. Even though both

---

<sup>6</sup> N = 94; 10 children and 4 adults did not complete this task as part of their visit and are not included in this data set.

<sup>7</sup> Here we report the result of a logistic mixed-effects model that included: Age (*Child*, *Adult*), TrialType (*Canadian vs Australian*; *Canadian vs Mandarin*) as fixed effects and with a random intercept per subject.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

children and adults noticed the accents of the speakers, if there really are age-related changes in the imitation of accented speakers, those interactions might have been more pronounced had the talkers produced more than isolated words.

Although only one study has found evidence of VOT imitation in children, there are quite a few studies that have demonstrated VOT imitation in adults (Nielsen, 2011; Shockley et al., 2004). It is surprising that imitation was not observed in any of the adult conditions. In our sample, the baseline VOTs were comparable to the adult Baseline VOTs reported in Nielsen (2014), which makes it less likely that adults started with such long VOTs that they were unable to extend their VOTs further in the Post-exposure phase. Other possible explanations for the lack of adult imitation relate to the design of the task. This task was modeled after the only lab-based study to show imitation in children (Nielsen, 2014). However, the task differs in two key ways from most adult imitation tasks. First, since the study was designed to be completed by both children and adults, almost all the words in the study were high frequency words (e.g., car, coat). Typically low frequency words tend to be imitated more than high frequency words (Goldinger, 1998). Much of the previous work with adults uses words that are lower in frequency than the words selected for this task. Second, unlike much of the adult imitation literature, here participants did not directly shadow (or repeat after) the model (Babel, 2010; Babel et al., 2014; Shockley et al., 2004; Walker & Campbell-Kibler, 2015). It is possible that the delay between when participants heard the token during the exposure phase and when they were asked to produce the same token in the Post-exposure phase, was simply too long to elicit robust imitation. Thus, both the design of the task and the stimuli we used might have decreased the amount of imitation observed in adults.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

In order to examine whether our paradigm is able to elicit VOT imitation, in Experiment 2, we make two major modifications to the task. First, we added a Pre-baseline phase in order to decrease the likelihood that children would hyperarticulate the baseline tokens. Second, we replaced the ‘Exposure Phase’ with a direct ‘Shadowing Phase’ to decrease the delay between the exposure and the speaker’s productions.

### Experiment 2

There can be substantial variation in the types of tasks that have been used to elicit imitation (see Pardo, 2013 for a review). In some studies, there is a delay between when participants listen to the model talker and when they produce the words themselves (i.e., delayed shadowing; Nielsen, 2014); however in most tasks participants repeat directly after the model with little to no delay (immediate shadowing; Babel, 2010; Babel et al., 2014; Shockley et al., 2004; Walker & Campbell-Kibler, 2015). Even a slight 3-4 second delay seems to decrease the amount of imitation observed in adults (Goldinger, 1998). Although only one study has tested children’s imitation of VOT (using a delayed shadowing task), it is possible that given limitations in children’s working memory capacity, they might show more robust evidence of imitation when the shadowing is immediate (see Heath, 2017 for evidence that occupying working memory decreases VOT imitation in adults). In Experiment 2, we modify the paradigm to maximize the likelihood that we will be able to observe imitation.

We do this by making two major changes to the procedure. First, we added a Pre-baseline phase in order to decrease the likelihood that children would hyperarticulate the baseline tokens. Second, we replaced the ‘Exposure Phase’ with a ‘Shadowing Phase’. Instead of simply listening to the productions of the model speaker, participants were instructed to repeat each word immediately after the model’s production. Similar to the predictions for Experiment 1, we

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

predict that children will show greater imitation than adults and that the amount of imitation observed will be impacted by the model talker's accent.

### Method

**Participants.** Forty-eight native Canadian English-speaking children (Age range 5.03 to 6.92 years, Mean age = 5.81 years; 24 females, 24 Males) and forty-eight Native English speaking adults (Age range 18-to 38-years-old, Mean age = 19.67 years; 38 females, 10 Males) were tested in this experiment. As in Experiment 1, in order to participate, children had to speak English at least 90% of the time and adults had to have learned English in Canada before the age of six. Participants did not have any diagnosed speech, hearing or vision issues that could interfere with participation in the study. Four additional children and one additional adult were excluded from this analysis due to technical issues in the running of the experiment (1), failure to complete the experiment (2), or failure to follow the instructions (2).

**Design, Stimuli and Procedure.** The design of Experiment 2 was similar to Experiment 1, except for two important changes. First, we added a Pre-baseline phase. In the Pre-baseline phase, participants were asked to label the entire set of 59 words once (by saying the name of the object out loud), before proceeding to the Baseline phase. This phase was added to decrease children's tendency to hyperarticulate the baseline tokens by familiarizing them with the items.

The second modification to the procedure was to replace the Post-exposure phase with a Shadowing phase. The stimuli used in the Shadowing phase was identical to the Exposure phase used in Experiment 1, but the instructions were different: instead of being asked to "listen" to the model speaker's productions, participants were instructed to directly "repeat" the words after the model. For example, if the model produced the word "paint", the participant would then repeat the word "paint" in the 1.5 s interval between each word.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

**Data Analysis.** The VOT of the Shadowed and Baseline tokens were annotated and measured as in Experiment 1. 2,329 tokens produced by the child participants and 2,464 tokens produced by the adult participants were included in this analysis. An additional 167 tokens from the child set and 32 tokens from the adult set were excluded for the following reasons: the participant labeled the item incorrectly (or failed to produce an audible label;  $n = 58$  child tokens;  $n = 3$  adult tokens), there was an overlap between the participant's productions and the Model's productions (i.e., the participant produced the item at the same time as the model, or the end of the participant's production overlapped with the model's subsequent production,  $n = 81$  child tokens;  $n = 26$  adult tokens), there was background noise ( $n = 17$  child token;  $n = 3$  adult tokens), or the VOT was followed by a voiceless vowel ( $n = 11$  child tokens;  $n = 1$  adult tokens). As in Experiment 1, word duration was measured from the end of the /p/ phoneme to the end of the word. For words that ended in a stop constant (e.g., pirate, pocket) the end of the word was always marked prior to the final stop closure and in cases where a participant did not produce a word with the plural inflection (e.g., 'pepper' instead of 'peppers'), the final 's' was removed from both the Baseline and Shadowed tokens.

**Statistical Analysis.** A linear-mixed effects model was implemented using the lme4 package (Bates et al., 2018) of R 4.0.0 (R Core Team, 2020). Similar to Experiment 1, we modeled the dependent variable (VOT duration) as a function of Age Group (*Child*, *Adult*), Phase (*Baseline*, *Shadowed Production*), the Model's Accent (*Canadian-accented*, *Australian-accented* or *Mandarin-accented*), Word Duration (centered and standardized to z-scores), and the 3-way interaction between Age, Phase and the Model's Accent. Segment type was not included in this model, as all the words in the shadowed phase began with /p/ (i.e., there were no /k/

segments)<sup>8</sup>. All categorical factors were simple-coded, and we used the same structure of random effects from Experiment 1. Follow-up tests for significant interactions were conducted using the *phia* package in R (De Rosario-Martinez, 2015).

## Results and Discussion

The results of the LME Model (see Table 2) indicated that there was a significant main effect of Phase,  $b = 12.19$ ,  $SE = 1.45$ ,  $t(89.86) = 8.42$ ,  $p < .001$  (see Table 2). Overall, we find evidence that participants systematically increased their VOTs after exposure to the Model's productions. Similar to Experiment 1, we replicate the main effects of Word Duration,  $b = 2.97$ ,  $SE = 0.72$ ,  $t(746.34) = 4.13$ ,  $p < .001$  and Age,  $b = -11.43$ ,  $SE = 3.05$ ,  $t(91.80) = -3.75$ ,  $p < .001$ , indicating that longer words had longer VOTs and that children produced longer VOTs than adults. Although children produced longer VOTs overall, we do not see any evidence that children imitated the elongated VOTs of the model talkers more strongly than adults, given the lack of a significant Age by Phase interaction,  $b = -1.24$ ,  $SE = 2.74$ ,  $t(89.30) = -0.45$ ,  $p = .651$ .

In order to assess whether the amount of imitation varied depending on the accent of the model we examined the interactions between the Phase (Baseline vs Shadowed) and the Model's Accent. Although the VOTs in the Australian condition were shorter on average, there was no interaction between Phase and the Australian accent, meaning that, participants imitated the elongated VOTs of both the Canadian and the Australian-accented models equally,  $b = 1.76$ ,  $SE = 3.35$ ,  $t(88.84) = 0.53$ ,  $p = .600$ . There was however, a significant difference in participant's imitation of the Mandarin compared to the Canadian-accented models,  $b = 7.46$ ,  $SE = 3.35$ ,

---

<sup>8</sup> Model structure:  $\text{lmer}(\text{VOT} \sim \text{Phase} * \text{Accent} * \text{Age} + \text{Word Duration} + (\text{Phase} | \text{Participant}) + (\text{Phase} | \text{Item}))$

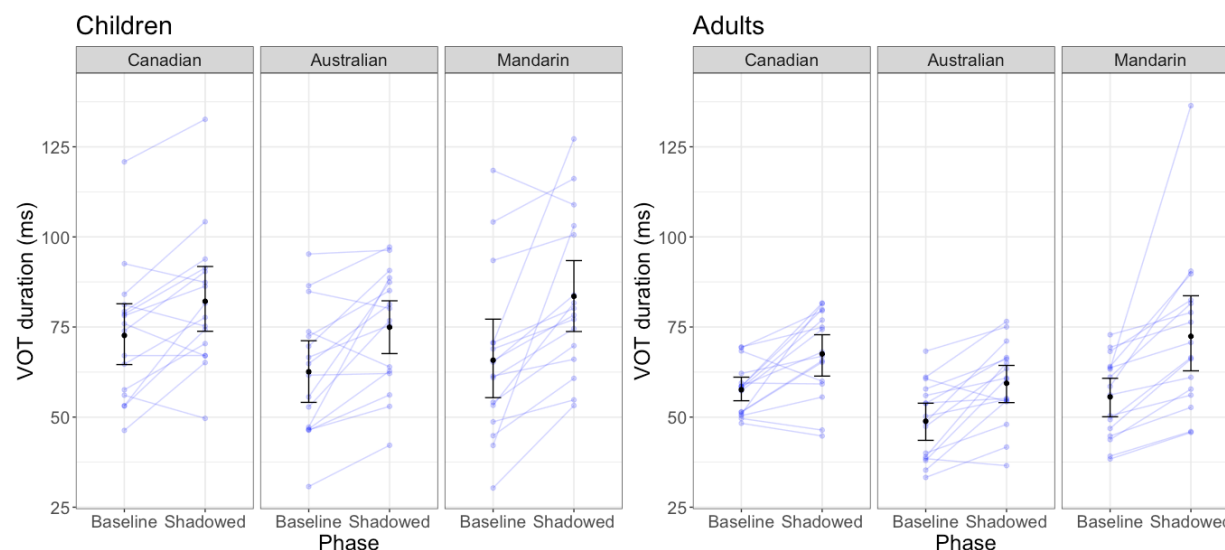
## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

$t(89.26) = 2.22, p = .029$ , with participants showing greater imitation of the non-native Mandarin talker compared to the Canadian talker. Given this significant 2-way interaction, follow-up tests (with holm p-value adjustments) were conducted to examine whether there was a significant increase in VOT from Baseline to Shadowed Phase separately for each Model's Accent. Here we find evidence of imitation of all three model talkers (Canadian  $\chi^2=14.28, p < .001$ ; Australian  $\chi^2=20.26, p < .001$ , Mandarin  $\chi^2=46.78, p < .001$ ), however the significant interaction indicates greater imitation of the Mandarin (vs. the Canadian) talker. There were no 3-way interactions between Age, Phase and Model's Accent, thus there does not appear to be age differences in the degree to which children and adults imitated different accented talkers.

**Table 2.** Table of the fixed effects from the Experiment 2 LME model predicting VOT duration. Reference levels are in italics.

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
<b>Intercept</b>	66.98	2.23	30.05	< 0.001
<b>Phase (Shadowed vs. <i>Baseline</i>)</b>	12.19	1.45	8.42	< 0.001
<b>Model's Accent</b>				
Australian (vs. <i>Canadian</i> )	-8.51	3.69	-2.31	0.024
Mandarin (vs. <i>Canadian</i> )	-0.49	3.69	-0.13	0.895
<b>Age Group (<i>Adult</i> vs. <i>Child</i>)</b>	-11.43	3.05	-3.75	< 0.001
<b>Word Duration</b>	2.97	0.72	4.13	< 0.001
<b>Phase x Model's Accent</b>				
Phase x Australian	1.76	3.35	0.53	0.600
Phase x Mandarin	7.46	3.35	2.22	0.029
<b>Phase x Age Group</b>	-1.24	2.74	-0.45	0.651
<b>Model's Accent x Age Group</b>				
Australian x Age	0.48	7.38	0.07	0.948
Mandarin x Age	3.81	7.38	0.52	0.607
<b>Phase x Model's Accent x Age Group</b>				
Phase x Australian x Age	-1.89	6.70	-0.28	0.778
Phase x Mandarin x Age	-0.73	6.71	-0.11	0.913

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS



*Figure 4.* Children and adult’s mean VOT before and after exposure to the Canadian, Australian and Mandarin-accented models in Experiment 2. Error bars represent the 95% CIs of by-participant means.

Taken together the results of Experiment 2 suggest that participants will imitate the elongated VOTs of regional and non-native accented model talkers (see Figure 4). Similar to previous work, we see stronger evidence of imitation when we used a direct shadowing task, compared to when we used a task with a delay (as in Experiment 1). Surprisingly, in contrast to Nielsen (2014), we found similar levels of imitation in both children and adults.

Consistent with work showing that adults will accommodate the productions of other regional dialects (Babel, 2010; Delvaux & Soquet, 2007; Walker & Campbell-Kibler, 2015), here we find evidence that speakers will imitate the productions of model talkers who speak with a different regional or non-native accent in a shadowing task. Contrary to Kim et al., (2011), which showed less convergence between speakers with different L1s, our results indicated that native-English speaking participants imitated the elongated VOTs of the Mandarin-accented model more strongly than the Canadian-accented model. Similar to Experiment 1, we confirmed (using a forced-choice selection task) that both children and adults were able to identify that the Canadian speaker “talked more like them” than the Australian speaker (89.47% of the time) and



## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

Mandarin speaker (88.42% of the time)<sup>9</sup>. Thus it is unlikely that the observed differences in imitation across accents is driven by an inability to detect the speaker's accent. It is possible, however, that differences in the tasks used (conversational vs. repeating words in isolation) and the method of assessing imitation (perceptual simulating ratings vs. specific phonetic measures) might lead to differences in reported imitation. The imitation observed in conversational tasks (as in Kim, et al., 2011) may be more impacted by participant's social evaluation of the speaker, whereas the imitation that occurs while shadowing may be more automatic.

Although we assume that the differences in imitation we observed across accents are driven by the accents themselves, given the design of the study (in which we have only one model talker per accent), it is possible that these effects may be driven by differences in the characteristics of the speakers we selected (see Babel et al., 2014 for a discussion) or in the properties of the specific model talker's tokens that were used. For example, the model talkers had systematically different baseline VOTs, which resulted in systematically different extents of manipulation. While we have no reason to believe that this should affect the amount of convergence, it remains a possibility that the elongated VOTs may have been more noticeable in some talkers than in others.

### General Discussion

Sociolinguists have observed that when children relocate to a new community, they imitate and eventually adopt the new local speaking style faster and more effectively than adults (Chambers, 1992; Smith et al., 2007; Tagliamonte & Molfenter, 2007). One could speculate that this is because children who have not reached adolescence are still within a critical period of

---

<sup>9</sup> N = 95; 1 adult participant was missing data for this task.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

development when they can very efficiently imitate and eventually acquire new speaking styles (Piske et al., 2001). Additionally, it is possible that motivation and/or social factors may also contribute to children's ability to adapt to new language communities. However, only one published study has examined this imitation process in the lab from a developmental perspective (Nielsen, 2014). Thus, many questions remain regarding how and why children appear to adopt new language varieties more successfully than adults in the real world.

In the current study, we address these questions using a well-controlled laboratory task. Crucially, and contrary to our predictions based on a previous report in the literature, we saw no differences in speech imitation between adults and children in either Experiment 1 or Experiment 2. Although we saw no effect of participant age on imitation, the accent of the model and the type of task (delayed shadowing versus immediate shadowing) impacted the degree to which participants imitated. Participants showed greater imitation of the Mandarin compared to the Canadian-accented models in the immediate shadowing task (Experiment 2) and no imitation of any speakers in the delayed shadowing task (Experiment 1). Taken together this study highlights the importance of considering how variation in the model talkers and variation in the task might influence the degree of imitation observed in the lab. Given the sensitivity of these effects to methodological factors, greater consideration should be given to designing experimental tasks that tap into the type of speech imitation and accent adaptation seen in the real world.

The primary goal of this study was to investigate whether there are age-related differences in speech imitation. To do this, in Experiment 1, we employed the same paradigm used in the only other study that has examined children's imitation of model speakers in the lab. Contrary to Nielsen (2014), neither children nor adults showed lengthened VOTs in a word-naming task after exposure to a model speaker with elongated VOTs. There are a few potential

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

explanations for this. It may be that children's hyperarticulation of the baseline productions prevented them from extending their VOTs further after exposure. It could also be the case that the words we selected (e.g., paint, peppers) were higher in frequency than the words typically selected in tasks with adults (e.g., parcel, forage). However, since Nielsen's (2014) study was susceptible to these same issues, the most likely explanation for our findings may simply be that delayed VOT imitation tasks do not consistently elicit imitation. Indeed, much of the previous work in adults has been conducted using immediate shadowing tasks, which are thought to elicit stronger convergence (Babel, 2010; Babel et al., 2014; Shockley et al., 2004; Walker & Campbell-Kibler, 2015). In general, studies have shown less robust VOT imitation effects in situations that are more taxing on working memory capacity (Heath, 2017), a factor which may have inhibited imitation in a task where the post-exposure productions may occur a few minutes after they heard the speaker's productions.

Although we did not see imitation in the delayed task (Experiment 1), children and adults showed robust imitation of all three speakers in the immediate shadowing task (Experiment 2). Here, when participants were presented with the same stimuli in a controlled environment, there were no differences in the degree to which children and adults imitated. This could be an indication that there are other more complex factors that are driving the differences seen in children and adult's accommodation in the real world. It may not be the case that children automatically imitate more because their phonological representations are less mature.

Given that lab tasks are unnatural, it is important to test how well these tasks scale up to more ecologically valid situations. Indeed, there is some evidence that participants who exhibit strong convergence in shadowed tasks do not necessarily exhibit strong convergence when they are tested using a conversational paradigm (Pardo et al., 2018). And importantly, even among

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

shadowing tasks, our study has demonstrated that small differences in the methodology can elicit differences in imitation. Here, similar to work showing that imitation is greater when there is shorter delay before shadowing (Goldinger, 1998), we find evidence of imitation in the immediate shadowing task (Experiment 2) but not in the delayed shadowing task (Experiment 1). Similarly, studies looking at imitation of specific features (e.g., imitation of increased nasalization) have shown that the ‘way’ in which the feature was imitated differed depending on the task (Zellou et al., 2016). As has been pointed out in the literature (see Pardo et al., 2017), there may also be differences depending on how convergence is measured. Some studies have assessed convergence more globally (using AXB perceptual similarity tasks) and others have measured convergence of specific phonetic features (e.g., VOT, vowels). Although we did not observe age differences in VOT imitation in this study, children may indeed show greater imitation of other features (like intonation, F1, vowel duration) compared to adults. Testing these possibilities is beyond the scope of the current study, but is an important direction for future research.

The secondary goal of this study was to determine whether children’s imitation is influenced by social factors such as the accent of the speaker. Although no laboratory studies have examined the imitation of accented speakers in children, a few studies have examined the imitation of accented speakers in adults. Some of these studies have suggested that participants will imitate model talkers that speak with different regional accents (Babel, 2010; Walker & Campbell-Kibler, 2015), whereas other studies have found no evidence that participants will imitate talkers that speak with different regional or non-native accents (Kim et al., 2011). In our study, we found that children and adults can imitate unfamiliar accented models, but their imitation varied depending on the task. When we used an immediate shadowing task (similar to

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

the task used in Babel, 2010), both children and adults showed greater imitation of the Mandarin model compared to the Canadian model. This accent-based difference was not found in the delayed shadowing task (Experiment 1). It may be the case that immediate shadowing tasks are less likely to elicit socially-based differences in imitation. Babel (2010) also found that social liking (measured using an Implicit Association Test) had a greater impact on whether the model's vowel shifts were maintained in the post-exposure phase compared to whether the model's vowels were imitated in the immediate shadowing block. However, since there was no significant convergence found for any of the models in our delayed shadowing task (Experiment 1), it is difficult to conclude anything about the (lack of) accent-based differences.

In Experiment 2, we found that not only are children and adults capable of imitating model talkers with non-native accents, but they actually imitate the non-native talkers more strongly than the talkers from their own community. This could indicate that imitation may (in this task) be driven more by the distance between the model talker and the participants productions than the social desirability of the speaker's accent. Although the VOT durations that participants listened to across the three model talkers were the same, it is possible that the greater phonetic distance in vowels or other features between the Mandarin talker's productions and the participant's productions might have prompted greater convergence. It could also be the idiosyncrasies of someone's accent are more salient in non-native speech. The results of Experiment 2 could suggest that the lack of imitation of accented speakers in the real world is not due to inability to perceive patterns in accented speech, but a lack of desire or motivation to do so. Although we observed imitation of a non-native accented speaker in a task with very minimal social interaction, it is possible that a different pattern may have emerged in a more naturalistic task (similar to the conversational task used in Kim et al., 2011). Tasks in which

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

there is more prolonged exposure to an accented speaker may better reflect the processes that occur in real world accommodation where the exposure to an accented speaker can occur over the course of weeks, months or even years.

Taken together, this is the first study to compare children and adult's imitation of model talkers with different accents. Although we do not see evidence that children imitate more than adults, the degree to which model talkers were imitated seemed to be influenced by the demands of the task and the accent of the speaker, with participants showing greater imitation of the Mandarin compared to the Canadian accented speaker, and imitation found in the immediate shadowing task but not the delayed task. This raises the possibility that imitation may be less socially selective when shadowing is immediate. These findings highlight the importance of considering how factors like the group status of the speaker may be interacting with the demands of the task to influence the imitation we observe in the lab and in the real world.

References

- Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, 39(04), 437–456. <https://doi.org/10.1017/S0047404510000400>
- Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177–189. <https://doi.org/10.1016/j.wocn.2011.09.001>
- Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, 5(1), 123–150. <https://doi.org/10.1515/lp-2014-0006>
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., & Green, P. (2018). *Linear Mixed-Effects Models using “Eigen” and S4* (1.1-17). <http://lme4.r-forge.r-project.org/>
- Berthele, R. (2002). Learning a second dialect: A model of idiolectal dissonance. *Multilingua - Journal of Cross-Cultural and Interlanguage Communication*, 21, 327–344. <https://doi.org/10.1515/mult.2002.014>
- Chambers, J. K. (1992). Dialect Acquisition. *Language*, 68(4), 673–705. <https://doi.org/10.2307/416850>

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

- De Rosario-Martinez, H. (2015). *phia: Post-Hoc Interaction Analysis. R package* (0.2-1).
- Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, 64(2–3), 145–173.  
<https://doi.org/10.1159/000107914>
- Dossey, E., Clopper, C. G., & Wagner, L. (2020). The Development of Sociolinguistic Competence across the Lifespan: Three Domains of Regional Dialect Perception. *Language Learning and Development*, 16(4), 330-250.  
<https://doi.org/10.1080/15475441.2020.1784736>
- Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: a study of accent change in young adults. *The Journal of the Acoustical Society of America*, 121(6), 3814–3826. <https://doi.org/10.1121/1.2722209>
- Floccia, C., Delle Luche, C., Durrant, S., Butler, J., & Goslin, J. (2012). Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition*, 124(1), 95–100. <https://doi.org/10.1016/j.cognition.2012.03.011>
- Giles, H. (1973). Accent Mobility : A Model and some data. *Anthropological Linguistics*, 15(2), 87–105. <http://www.jstor.org/stable/30029508>
- Giles, H., Coupland, N., & Coupland, J. (1991). Accomodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accomodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge University Press. <https://doi.org/10.1017/CBO9780511663673.001>
- Giles, H., & Ogay, T. (2007). Communication Accomodation Theory. In B. B. Whaley & W. Santer (Eds.), *Explaining communication: contemporary theories and exemplars* (p. 467). Lawrence Erlbaum Associates, Inc., 2007. <https://doi.org/10.4324/9781410614308>



## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

- Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society*, 2(2), 177–192. <https://doi.org/10.1017/S0047404500000701>
- Gluszek, A., & Dovidio, J. F. (2010). The way they speak: a social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, 14(2), 214–237. <https://doi.org/10.1177/1088868309359288>
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279. <https://doi.org/10.1037/0033-295X.105.2.251>
- Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review*, 11(4), 716–722. <https://doi.org/10.3758/BF03196625>
- Heath, J. (2017). How automatic is convergence? Evidence from working memory. *Linguistic Society of America Proceedings*, 2(35), 1–10. <https://doi.org/10.3765/plsa.v2i0.4088>
- 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, 60–99.
- Jones, Z., Yan, Q., Wagner, L., & Clopper, C. G. (2017). *The development of dialect classification across the lifespan*. 60, 20–37. <https://doi.org/10.1016/j.wocn.2016.11.001>
- Keshet, J., Sonderegger, M., & Knowles, T. (2014). *AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction* (0.91). <https://github.com/mlml/autovot/>
- Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, 2(4), 125–156. <https://doi.org/10.1515/labphon.2011.004>

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition.

*Proceedings of the National Academy of Sciences*, 104(30), 12577–12580.

<https://doi.org/10.1073/pnas.0705345104>

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*, 27(4), 623–634.

<https://doi.org/10.1521/soco.2009.27.4.623>

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100(401), 2425–2438.

Kuznetsova, A., Brockhoff, B., & Christensen, R. H. B. (2015). *Tests in Linear Mixed Effects Models* (2.0-29). <https://cran.r-project.org/web/packages/lmerTest/index.html>

Lowenstein, J. H., & Nittrouer, S. (2008). Patterns of acquisition of native voice onset time in English-learning children. *The Journal of the Acoustical Society of America*, 124(2), 1180–1191. <https://doi.org/10.1121/1.2945118>

McRoberts, G. W., & Best, C. T. (1997). Accommodation in mean f0 during mother - infant and father - infant vocal interactions: A longitudinal case study. *Journal of Child Language*, 24. <http://www.haskins.yale.edu/Reprints/HL1085.pdf>

Munro, M. J., Derwing, T. M., & Flege, J. E. (1999). Canadians in Alabama: a perceptual study of dialect acquisition in adults. *Journal of Phonetics*, 27(4), 385–403.

<https://doi.org/10.1006/jpho.1999.0101>

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32(5), 790–804.

<https://doi.org/10.1037/0022-3514.32.5.790>

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

- Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132–142. <https://doi.org/10.1016/j.wocn.2010.12.007>
- Nielsen, K. (2014). Phonetic imitation by young children and its developmental changes. *Journal of Speech, Language, and Hearing Research*, 57(6), 2065–2075. [https://doi.org/10.1044/2014\\_JSLHR-S-13-0093](https://doi.org/10.1044/2014_JSLHR-S-13-0093)
- Papoušek, H., & Papoušek, M. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language*, 9, 137–157. <https://doi.org/10.1177/014272378900900603>
- Paquette-Smith, M., Buckler, H., White, K. S., & Johnson, E. K. (2019). The effect of accent exposure on children's sociolinguistic evaluation of peers. *Developmental Psychology*, 55(4), 809–822. <https://doi.org/10.1037/dev0000659>
- Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, 4, 559. <https://doi.org/10.3389/fpsyg.2013.00559>
- Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, 40(1), 190–197. <https://doi.org/10.1016/j.wocn.2011.10.001>
- Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, and Psychophysics*, 79(2), 637–659. <https://doi.org/10.3758/s13414-016-1226-0>
- Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics*, 69(April), 1–11. <https://doi.org/10.1016/j.wocn.2018.04.001>
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

an L2: A review. *Journal of Phonetics*, 29(2), 191–215.

<https://doi.org/10.1006/jpho.2001.0134>

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4), 421–436.

<https://doi.org/10.1006/jpho.1997.0051>

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66, 422–429. <https://doi.org/10.3758/BF03194890>

Smith, J., Durham, M., & Fortune, L. (2007). “Mam, ma troosers is fa’in doon!” Community, caregiver and child in the acquisition of variation in Scottish dialect. *Language Variation and Change*, 19(1), 63–99. <https://doi.org/10.1017/S0954394507070044>

Sonderegger, M., & Keshet, J. (2012). Automatic measurement of voice onset time using discriminative structured prediction. *The Journal of the Acoustical Society of America*, 132(6), 3965. <https://doi.org/10.1121/1.4763995>

Souza, A. L., Byers-Heinlein, K., & Poulin-Dubois, D. (2013). Bilingual and monolingual children prefer native-accented speakers. *Frontiers in Psychology*, 4(953), 1–6.

<https://doi.org/10.3389/fpsyg.2013.00953>

Tagliamonte, S. A., & Molfenter, S. (2007). How’d you get that accent?: Acquiring a second dialect of the same language. *Language in Society*, 36, 649–675.

<https://doi.org/10.1017/S0047404507070911>

van der Feest, S. V. H., & Johnson, E. K. (2016). Input-driven differences in toddlers’ perception of a disappearing phonological contrast. *Language Acquisition*, 23(2), 89–111.

## COMPARING CONVERGENCE IN CHILDREN AND ADULTS

<https://doi.org/10.1080/10489223.2015.1047096>

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology*, 6, 1–18.

<https://doi.org/10.3389/fpsyg.2015.00546>

Yu, V. Y., De Nil, L. F., & Pang, E. W. (2015). Effects of age, sex and syllable structure on voice onset time: Evidence from children's voiceless aspirated stops. *Language and Speech*, 58(Pt 2), 152–167. <https://doi.org/10.1016/j.clinph.2011.06.006.A>

Zellou, G., Scarborough, R., & Nielsen, K. (2016). Phonetic imitation of coarticulatory vowel nasalization. *The Journal of the Acoustical Society of America*, 140(5), 3560–3575.

<https://doi.org/10.1121/1.4966232>

**Appendix A: Stimuli**

List 1 /p/ words (N=13)	List 2 /p/ words (N=13)	/k/ words (N=13)	Fillers (N= 20)
Paint Pancakes Peach Peacock Pear Penguin Pickle Pig Pinecone Pink Pirate Pocket Puzzle	Pants Peanuts Peas Pencil Peppers Piano Pie Pigeon Pillows Pizza Popcorn Pumpkin Purple	Cake Camel Candy Car Carrot Cat Coat Corn Cow Cup Kangaroo Keys Kite	Ant Apple Egg Fish Fork Frog Hat House Lion Milk Monkey Nose Orange Shoe Snowman Sock Star Strawberry Witch Zebra