Title: VOT imitation in teens versus adults

Accepted for publication in *Journal of Speech, Language, and Hearing Research*

Authors:

1. Jessamyn Schertz, jessamyn.schertz@utoronto.ca, Department of Language Studies, University of Toronto Mississauga; Department of Linguistics, University of Toronto.
2. Elizabeth K. Johnson, elizabeth.johnson@utoronto.ca, Department of Psychology, University of Toronto Mississauga.

Corresponding author:

Jessamyn Schertz, Department of Language Studies, University of Toronto Mississauga

3359 Mississauga Road North

Mississauga, Ontario L5L1C6, Canada

jessamyn.schertz@utoronto.ca

Conflict of interest statement: The authors declare that there are no conflicts of interest.

Funding statement: This research was supported by grants from the Social Sciences and Humanities Research Council of Canada (to EKJ) and the Natural Sciences and Engineering Research Council of Canada (to JS and EKJ).

Abstract

Purpose: We compare teens' and adults' imitation of sentences with shortened and lengthened voice onset time (VOT), in order to test whether purported age-based advantages in phonetic acquisition may be due to differences in imitative ability.

Method: 12- to 14-year-olds (n=39) and adults (n=31) completed an explicit imitation and discrimination task on pairs of sentences characterized by canonical and manipulated (shortened or lengthened) VOT. We assessed extent of imitation using two acoustic metrics (ΔVOT and Proximity), accuracy on the discrimination task, and correlations between imitation and perception.

Results: Teens and adults modified VOT when imitating stimuli with both lengthened and shortened VOT. Adults showed significantly more lengthening than teens (i.e., higher ΔVOT), as well as VOT values that were slightly but significantly closer to the target stimulus values (i.e., lower Proximity). Both age groups showed above-chance discrimination accuracy, and a significant relationship between individual perception and production performance was found for lengthened-VOT sentences.

Conclusions: We found no evidence that teens have greater imitative ability than adults; in fact, adults showed significantly *more* imitation based on both acoustic metrics. Both age groups showed robust imitation of VOT manipulations in both directions, in contrast to previous work showing lack of imitation for shortened VOT. Extent of imitation was predicted by individual perceptual performance, but only to a limited degree, underscoring the importance of other factors in explaining individual variation in imitative ability.

1. Introduction

Younger learners tend to outperform their older counterparts in attaining native-like pronunciation in a foreign language in naturalistic settings (Piske et al., 2001). One potential contributing factor could be that younger language users are better at reproducing the fine-grained phonetic differences that characterize different languages, dialects, or accents. However, the developmental trajectory of imitative ability has not been well-documented, making it hard to assess whether this is indeed the case. While there is a large body of work examining phonetic imitation in adults (see Pardo et al., 2017, for a review), and a growing number of studies with children (Nielsen, 2014; Paquette-Smith, 2021; Wynn et al., 2018), we are not aware of any work targeting teenagers, the crucial transition age where children transform into young adults. This group is particularly interesting to study given that teens arguably have not yet developed a fully mature speech processing system, but lack the performance limitations of children. In this study, we examine how young teens compare with adults in their ability to reproduce systematic phonetic variation, via explicit imitation of sentences characterized by artificially lengthened and shortened voice onset time (VOT), and we test discrimination of these same sentences to determine to what extent imitative ability can be ascribed to accurate perception of the features. As an ancillary question, we probe the source of results from previous work showing lack of imitation of shortened VOT, testing whether this asymmetry persists with a paradigm designed to elicit greater imitation.

Our current knowledge of the developmental trajectory of imitation comes from studies examining children's imitation. Children, like adults, align to pronunciations of speech they are exposed to, as assessed both by their modification of specific acoustic dimensions like f0 or VOT and by more holistic perceptual metrics; however, as with adults, the extent and presence of alignment is not consistent across studies (see review and references in St. Pierre et al., 2021). In the one study we are aware of comparing performance of children to adults, the evidence that exists suggests that younger speakers may show more imitation of phonetic detail than adults: Nielsen (2014) found that 5- and 8-year-old children showed more imitation (i.e., produced longer VOTs following exposure to words with extended VOT) than adults. This effect has not been firmly established, and in fact, a replication and follow-up shadowing task by Paquette-Smith et al. (2021) failed to find any difference in magnitude of imitation by children compared to adults. Furthermore, in the suprasegmental domain, Wynn et al. (2018) found that adults, but not children, entrained to the speech rate heard in exposure prior to a picture description task. Nevertheless, based on the age-based difference found in Nielsen (2014), we predict that young teens, like their younger counterparts, may show more VOT imitation than adults.

At the same time, teens may differ from adults in terms of perceptual sensitivity, which is a necessary component of imitation: accurate perception of the phonetic properties of the target is a prerequisite to accurate reproduction. Despite the well-established sensitivity of infants to phonetic detail, when considering later developmental stages, the existing evidence suggests that, if anything, younger listeners may have *less* perceptual sensitivity. Laboratory studies have found that older learners often outperform younger learners in foreign sound discrimination at early stages of exposure, particularly when controlling for amount of input exposure. For example, Kopečková et al., 2019 found that 9- to 11-year-old German children were less accurate than adults in discrimination of newly-learned Polish sibilant contrasts, although production accuracy did not differ across groups (see also Fuhrmeister et al., 2020). Furthermore, McMurray et al. (2018) showed via an eye-tracking study that sensitivity to fine phonetic detail, as quantified by gradient looking times, increases across late childhood and adolescence (ages 7-8, 12-13, and 17-18). In the current work, we therefore test not only imitation, but also discrimination of the target sentences. In addition to addressing whether perceptual sensitivity to the target distinction differs across age groups, this also allows us to directly examine the role of perceptual acuity in predicting variation in imitative ability – a question that is not often tested, despite the importance of the perception-production relationship in models of phonetic imitation (see Pardo et al., 2017, for review).

Along with the fact that there are gaps in the age ranges targeted in previous work, an additional barrier to a clear picture of the developmental trajectory of imitation is the diverse range of methodologies and analyses used to elicit and quantify imitation, making it difficult to generalize across studies. Previous work has shown that methodological choices, including task and the linguistic status of the feature being imitated, can affect the extent of imitation. Furthermore, the specific metric used to quantify imitation can result in different findings. In the following paragraphs, we discuss the rationale behind our choices of task (explicit imitation), phonetic variable (VOT), and analysis (comparison of two different metrics quantifying imitation).

Since we were interested in what people do when asked to reproduce the details of sounds, as they might be asked to do in a foreign language class, we used an explicit imitation paradigm, asking participants to imitate what they heard, with sentences manipulated to vary in specific, controlled ways (i.e. an "artificial accent," as in Spinu et al. 2020). This stands in contrast to the more implicit tasks used in most previous work: some use shadowing tasks, asking participants to repeat stimuli after they hear them, but with no instruction for imitation (e.g. Paquette-Smith et al., 2021), while others compare participants' productions before and after exposure to target stimuli (as in Nielsen 2011, 2014). Dufour and Nguyen (2013) directly compared imitation of vowels in a shadowing vs. explicit imitation task, and found imitation in both, with larger effects in the explicit task (see also Pardo et al., 2010; Sato et al., 2013 for similar findings). Their results suggest that implicit and explicit imitation may share a common automatic component, but that explicit instruction boosts the imitation effect.

We chose VOT as the target phonetic variable for this study because it has been found to elicit robust imitation by both children and adults in previous work. However, this imitation has been found only in one direction: Nielsen (2011) showed that participants produced longer VOTs after exposure to words with artificially-lengthened VOT, but found no such modification after exposure to shortened VOT. Subsequent work has therefore focused on lengthened VOT (Nielsen 2014; Yu et al., 2013; Paquette-Smith et al., 2021), leaving open the question of whether the source of the asymmetry found by Nielsen (2011) lies in perception, articulation, or some other factor. It could be that reduced-VOT stimuli were less perceptually salient than the lengthened versions. Alternatively, it could be that participants were disinclined to produce tokens with shortened VOT, perhaps because doing so would encroach on a separate phonological category. The current work tests imitation of both extended and shortened VOT, using more extreme manipulated values to increase the expected perceptual salience. If the lack of shortened-VOT imitation in Nielsen (2011) was grounded in perception, we might expect to find imitation in our study. On the other hand, if it was based on articulatory inhibition, we would expect to replicate the lack of reduced-VOT imitation found in Nielsen (2011).

A final methodological consideration is how to quantify imitation. When evaluating imitation of lengthened-VOT stimuli, for example, one possibility is "the more the better," with longer VOT values considered greater imitation. However, another possibility is "the closer the better," with optimal imitation being productions that equal the model's specific VOT value. Our materials were designed with VOT manipulated to fairly extreme values, such that participants would be unlikely to attain those values; in these cases, the two metrics will be equivalent – because the more lengthening there is, the more closely the value will approximate the model. However, as shown below, performance was not always as expected. The question of which of these metrics is "better" is more of a theoretical question than an analytical one and depends on whether the target of imitation is an abstract feature like "lengthening," or a specific value like

“150 ms VOT.”[1] We use two separate metrics to quantify imitation (*Change in VOT* representing “the longer the better,” and *Proximity* representing “the closer the better”)[2], to examine how results and interpretation depend on the metric used.

In sum, in order to address whether the propensity for younger learners to outperform older learners in phonetic acquisition might be due to greater imitative ability, we compare imitation of sentences with artificially lengthened and shortened VOT in young teens and adults, using two different metrics to quantify imitation. We tentatively predict that teens may outperform adults in imitation, based on the finding of greater VOT imitation by children than adults in Nielsen (2014), but that they may show less perceptual sensitivity than adults, based on findings of Kopečková et al. (2019) and McMurray (2018). Our paradigm allows us to test how much of the variability in imitative ability is ascribable to accurate perception of the relevant features, as opposed to other sub-components of imitation, including articulatory factors. Finally, we test whether increasing the perceptual salience of the difference, as well as using an explicit imitation paradigm designed to maximize imitation, elicits imitation of shortened VOT, in an effort to determine the source of asymmetrical findings in past work.

2. Method

Participants: 12-to-14-year-old adolescents and their adult caregivers who live in the Toronto area and who learned to speak only English as a first language were invited to participate. Recordings from 39 adolescents (23 self-reported female/girl, 16 male/boy, 12-14 years old) and 31 adults (28 female, 3 male, 39-56 years old) are analyzed here; additional participants completed the study but

[1] A specific value could either be a raw or normalized value; see Nielsen & Scarborough (2019), for evidence in favor of normalized values in the domain of vowel nasality.

[2] We avoid the most commonly-used metric to assess imitation, the “difference-in-distance” score, which quantifies how much the acoustic distance between the imitator and the model changes from baseline to post-exposure. As shown by MacLeod (2021), this metric is problematic because the values are heavily influenced by participants’ baseline proximity to the target of imitation.

were excluded due to having learned other languages in the home (n=11), diagnosed speech disorders (n=1), or low-quality recordings (n=3).

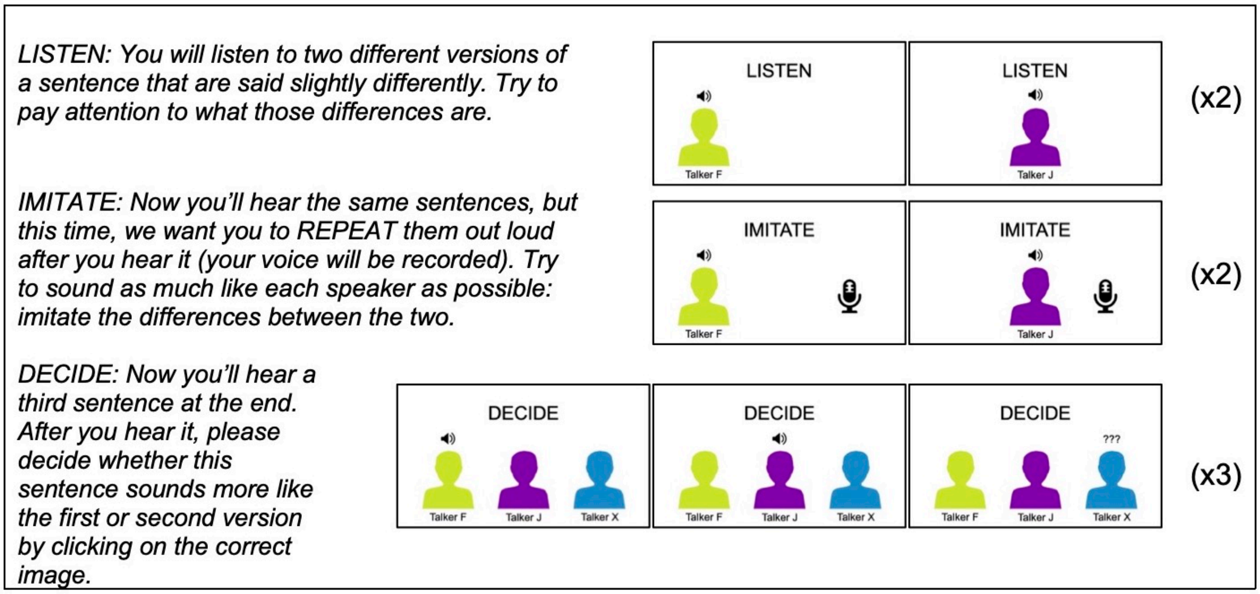
Materials: Stimuli consisted of four English sentences including two target word-initial voiceless stops (e.g. “**C**offee, **t**oast, eggs and cereal are what I ate this morning"), produced by an undergraduate female native English speaker from Ontario, Canada. The VOT of each target stop was manipulated to create three versions differing in VOT duration: shortened, canonical, and lengthened. These values were chosen such that the canonical values roughly matched production values in the natural recordings, with the shortened and lengthened values as extreme as possible without sounding unnatural, and differing equally from the canonical version (e.g. values for shortened/canonical/lengthened /t/ were 15/95/175 ms respectively; see Appendix 1 for complete list of sentences and VOT values).

Design and Procedure: The experiment was completed entirely online, using the platform Gorilla. Participants were asked to wear headphones and use their own recording devices (see Appendix 2 for further details about devices and recordings).

Participants completed 2 blocks, a lengthened-VOT block and a shortened-VOT block, with block order randomized across participants. Each block consisted of four “trial sets,” corresponding to the four sentences described above. The trial set procedure was designed to test both imitation and discrimination of a pair of sentences differing only in VOT, one canonical and one manipulated (either shortened or lengthened, depending on the block). The inclusion of imitation of canonical sentences, not normally used in previous imitation paradigms, provides an additional layer of control in quantifying how precisely the target feature is manipulated. Furthermore, using pairs of sentences allows us to incorporate a discrimination task to directly tests the perceptibility of the target feature.

Each trial set consisted of three phases: exposure, imitation, and discrimination, and proceeded as follows (Figure 1). First, in the *exposure* phase, participants heard two sentences, one after another: the first sentence had canonical VOT, and the second had modified VOT. Each sentence was accompanied by a different silhouette. The exposure phase was repeated twice. Second, the two sentences were presented again, one after another, with a pause after each one and a prompt to *imitate*. The imitation phase was also repeated twice. Finally, three ABX *discrimination* trials were presented, in which participants once again heard the two sentences, followed by a third, and were asked to decide whether the third was more like the first or second. In each of the three discrimination trials, the “X” sentence was identical to either the first or second sentence (i.e. canonical or manipulated, with the choice pseudorandomized). One practice set, with a sentence differing in coda rhoticity (e.g. ‘car’ vs. ‘cah’), preceded the regular trials. All participants successfully discriminated and imitated this difference. The task took about 30 minutes. Each participant imitated 64 stops (2 blocks * 4 trial sets * 2 VOT levels * 2 stops per sentence * 2 repetitions) and completed 24 discrimination trials.

Figure 1: Instructions given to the participants, and visual schematic of the procedure.



Acoustic analysis and measurements: VOT was measured as the duration from the beginning of the stop burst, as visible in the wave form, to the onset of stable upper formants (F2 and above) in the following vowel, as visible in the spectrogram. Tokens were excluded if the target sound and/or entire word was missing due to recording errors or noise (n=74), or if the participant produced a different place or manner of articulation than the target sound (n=42) or omitted the consonant (n=14). 4350 tokens were analyzed.

As discussed in the introduction, we used two different metrics to quantify VOT imitation in order to capture two different conceptualizations of what constitutes "more imitation." *Change in VOT* (ΔVOT) was calculated as the difference between the VOT in a participant's imitation of a manipulated stop, compared to their imitation of the canonical stop, in a given sentence pair (transformed so that positive values indicate the expected direction of imitation). Therefore, the longer a participant's imitation of a lengthened-VOT token, or the shorter the imitation of a shortened-VOT token, the greater their ΔVOT. We also calculated a by-token measure of

Proximity to the model talker's VOT value, which was simply the absolute value of the difference between the participant's VOT and the VOT of the imitated stimulus, with smaller values indicating closer approximation of the model's VOT value. While ΔVOT and Proximity are largely overlapping, they diverge in cases where participants make modifications in the expected direction beyond those present in the stimuli, as will be discussed below.

Statistical analysis: Data were analyzed in R (R Core Team, 2020) with mixed-effects linear (for production) and logistic (for perception) regression models, using lme4 (Bates et al., 2015, with p-values computed using the package *lmerTest*, using an alpha-level of 0.05 for significance; Kuznetsova et al., 2017). In the case of significant interactions, we performed follow-up tests using the *phia* package (De Rosario-Martinez, 2015), with Holm-adjusted p-values, to test whether the effect of interest held at each level of the other factor(s). Full specifications for all models are in the Appendix.

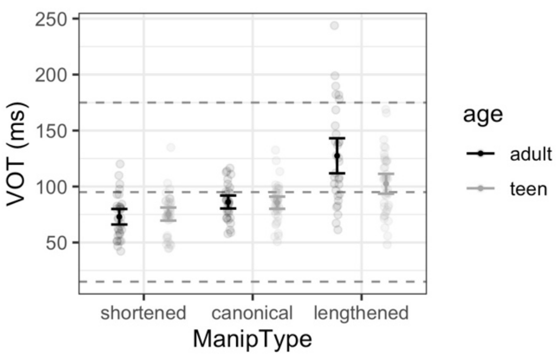
3. Results

3.1. Imitation

The average VOT values in imitation of the three Manipulation Types are shown in Figure 2. Both adults and teens showed imitation: imitations of shortened-VOT stimuli were shorter, and lengthened-VOT longer, than imitations of the canonical-VOT stimuli[3]. Below, we analyze the extent of imitation using the two metrics described above.

[3] In order to determine whether VOT differences could be due to differences in speech rate, we tested whether the duration of the following vowel differed across conditions. Mean vowel durations were 172, 177, and 193 ms for the shortened, canonical, and lengthened conditions respectively, and these differences were significant. Consistent with previous work (Nielsen 2011, Paquette-Smith 2021), these differences were of smaller magnitude than the VOT modifications (shortened: 3% decrease for vowel vs. 13% for VOT; lengthened: 9% increase for vowel vs. 33% for VOT). Following the same logic used in these previous studies, this indicates that the imitation effect cannot be solely attributable to differences in global speech rate. Full statistical details are provided in Appendix 3 (Model 5).

Figure 2: Average VOT values by age and VOT manipulation type. Error bars show 95% confidence intervals of by-participant means; individual dots represent individual participant means. The average stimulus values for shortened, canonical, and lengthened are shown with dotted horizontal lines.



First, we analyzed ΔVOT (difference between imitation of manipulated vs. canonical VOT on a given trial, transformed so a positive value represents the expected direction of change) to test how participants' modification of VOT varies based on age and manipulation type. We used a linear mixed-effects model to test the influence of the predictor variables Age (adult vs. teen) and ManipType (shortened vs. lengthened) and their interaction, on ΔVOT. Both predictor variables were simple-coded (-0.5, 0.5) such that the estimate of the intercept represents the mean difference in VOT (in the expected direction) across all conditions, and the estimate corresponding to each fixed factor represents the difference in VOT modification between the two levels of the comparison. Random intercepts and slopes for ManipType for both Participant and Word were included.

Results for ΔVOT are shown in Figure 3, and statistical results are shown in Table 1. Participants modified VOT in the expected direction (by 18 ms on average), as shown by positive values in all conditions, and by the significant intercept in the statistical model. However, the magnitude of the effect differs across conditions. Based on the significant two-way interaction, we performed follow-up tests for each of the four Age * ManipType groups. Adults showed higher ΔVOT than teens in the lengthened VOT condition (on average, 38 ms ΔVOT for adults vs. 15 ms for teens), but all four conditions were significantly different than zero (see Appendix 3 for statistical results of follow-up tests). Overall, results for ΔVOT indicate that both teens and adults modified their VOT in the expected direction when asked to imitate stimuli containing VOT variation, and that adult participants showed increased modification relative to teens in the lengthened VOT condition[4].

Figure 3: Values for (a) ΔVOT and (b) Proximity by age and VOT manipulation type. Error bars show 95% confidence intervals of by-participant means; individual dots represent individual participant means.

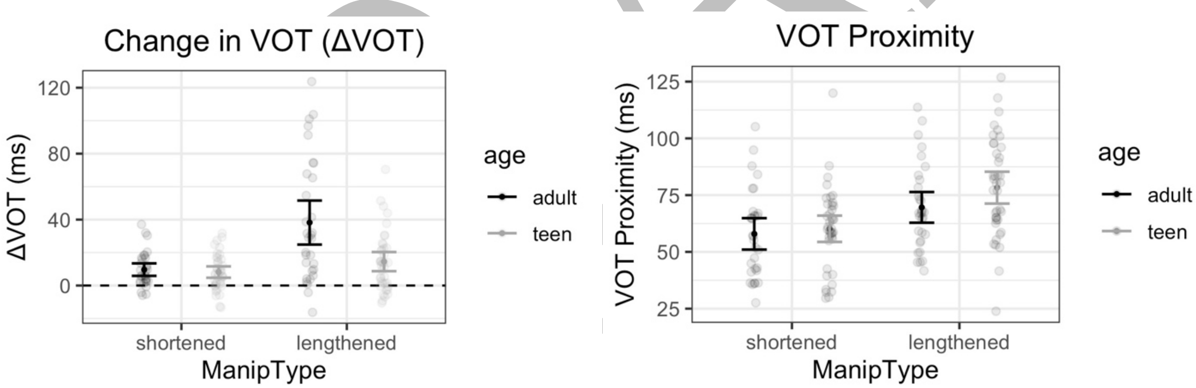


Table 1: Statistical results from all statistical models predicting effects of Age and ManipType on a) Change in VOT (in ms), b) VOT Proximity, and c) Discrimination accuracy. Reference levels are in italics, and significant effects are in bold.

[4] Previous work has reported mixed results about potential effects of gender on imitation (e.g. Pardo et al., 2017). We did not have a priori expectations about this, and our participant sample was not balanced for gender. We looked for gender-based patterns during initial exploratory analyses; failing to find any, we have not included gender as a factor in our analysis.

	Production: ΔVOT			Production: VOT Proximity			Perception: Accuracy		
Factor	**β**	**SE**	**p**	**β**	**SE**	**p**	**β**	**SE**	**p**
Intercept	**17.76**	**2.18**	**< .001**	**66.38**	**2.34**	**< .001**	**1.62**	**0.22**	**< .001**
Age (*adult* vs. teen)	**-12.78**	**3.83**	**0.001**	**5.48**	**2.63**	**0.041**	-0.41	0.28	0.140
ManipType (*short* vs. long)	**17.57**	**3.42**	**< 0.001**	**14.99**	**3.92**	**<.001**	-0.64	0.39	0.097
Age * ManipType	**-22.24**	**6.84**	**0.002**	6.45	7.83	0.414	0.28	0.39	0.474

We then turned to our second metric for quantifying convergence, using a linear mixed-effects model to test the influence of the predictor variables Age (*adult* vs. teen), ManipType (*shortened* vs. lengthened),[5] and their interaction on Proximity. The model structure was the same as above; the estimate of the intercept represents the mean difference in VOT (in the expected direction) across all conditions, and the estimate corresponding to each fixed factor represents the difference in VOT modification between the two levels of the comparison.

Results for Proximity are shown in Figure 3, and statistical results are shown in Table 1. There was a significant effect of age, with adults showing closer Proximity to the stimuli than teens, although this difference was very small (mean 64 ms for adults vs. 69 ms for teens). There was also a significant effect of Manipulation Type, where imitations of shortened-VOT stimuli were closer to the target than imitations of lengthened-VOT stimuli (59 vs. 74 ms), a result that is likely attributable to the fact that participants' canonical productions – which we can assume may reflect natural production values – were closer to the shortened than the lengthened model targets. There was no significant interaction between Age and ManipType.

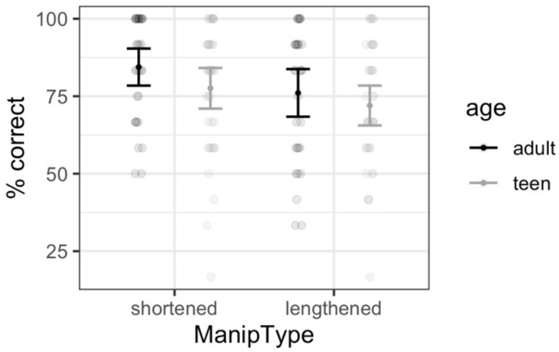
[5] For consistency with the analysis of ΔVOT, we only analyze Proximity for manipulated tokens. Proximity for the canonical tokens, not reported above, did not differ significantly between adults (mean 24 ms) and teens (mean 22 ms).

In sum, teens and adults imitated both shortened and lengthened VOT. Furthermore, by both quantification metrics, adults showed more imitation than teens, producing longer VOT values when presented with lengthened-VOT stimuli, and closer approximation to the model's values than teens. The age-based difference is greater for ΔVOT than for Proximity; this can be attributed to the fact that some adults lengthened VOT beyond the manipulated value present in the model stimuli, as indicated by the very high average values shown by some adult participants in Figure 2.

3.2. Perception

To test perception, each trial set included three AXB discrimination questions: participants heard the two exposure sentences, followed by a new sentence, and decided whether the new sentence matched the canonical or manipulated (shortened/lengthened) sentence. We then used a logistic mixed-effects model to test the influence of Age and Manipulation Type on Accuracy, using the same model structure as for the production models (with the item-based random intercept of SentencePair instead of Word). Results are shown in Figure 4, and statistical results are shown in Table 1. Participants were well above chance in discrimination accuracy (mean 77% correct), and this did not differ significantly by Age or by ManipType.

Figure 4: Perception accuracy by age and VOT manipulation type. Error bars show 95% confidence intervals of by-participant means; individual dots represent individual participant means.



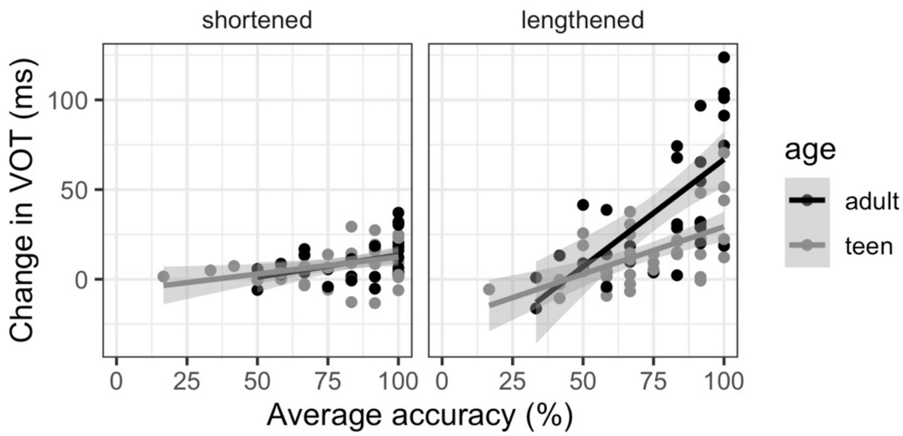
Finally, we tested the relationship between individual-level perception and production: whether listeners who were in general better at noticing the difference between the sentences would also show more imitation, and whether this differed based on Age and ManipType. We calculated individual indices for perception (individuals' mean discrimination accuracy) and for both production metrics (individuals' mean ΔVOT and mean Proximity) for each ManipType separately. We then used linear regression models to test the effect of Individual Perception (along with Age, ManipType, and interactions between all factors) on Individual ΔVOT (Model 4a) and Individual Proximity (Model 4b). Full statistical models and results are provided in Appendix 3.

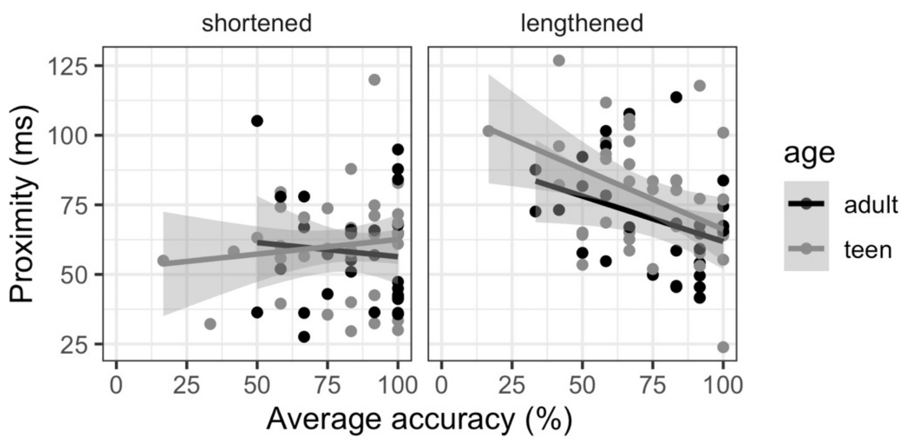
Scatterplots showing the correspondence between perception and both production metrics, broken down by Age and ManipType, are shown in Figure 5. For each production metric, our primary question was whether there was a significant effect of Perception, and whether there was a significant interaction of Perception and Age and/or ManipType, which would indicate differences in the strength of the relationship based on these factors.

The ΔVOT model showed a significant three-way interaction between all three predictors. Follow-ups testing the effect of Perception on ΔVOT for each of the four groups separately indicate that there is a significant effect of Perception on ΔVOT in the lengthened-VOT condition for both age groups, but that this effect was stronger for adults than for teens. The Proximity model showed a significant interaction between Perception and ManipType, with follow-ups indicating a significant effect of Perception on Proximity for the lengthened-VOT condition. These statistical results reflect the patterns seen in the graphs: for both production metrics, there is a clear relationship between perception and production (i.e. a non-zero slope) in the lengthened, but not the shortened, condition, and for ΔVOT, the slope is greater for adults than teens.

These results indicate that individuals who were more accurate in discrimination of sentences with canonical vs. lengthened-VOT stops showed more imitation of the same sentences, by both production metrics. We do not have evidence of such a relationship for sentences with shortened-VOT stops, although this should be interpreted with caution: it is possible that the same relationship exists but is just too small to be detected with our sample size (particularly since the correlation, although not significant, is in the expected direction). The age-based difference in strength of the relationship between perception and ΔVOT can be attributed to the fact that adults lengthened more than teens, as discussed above, and is therefore not necessarily indicative of any difference in the perception-production relationship between the two age groups.

Figure 5: Relationship between by-participant average discrimination accuracy and by-participant imitation, as measured by (a) ΔVOT and (b) Proximity, for shortened and lengthened conditions separately.





4. Discussion

A clear picture of the developmental trajectory of imitative ability is important for testing hypotheses about the source of age-based differences in the acquisition of new languages and dialects, as well as the development of the perception-production link. Furthermore, an understanding of adolescents' imitative ability is of practical importance since the teenage years are often characterized by formal language instruction. This work examined the phonetic imitation of teens, an age group that has not been targeted in previous work, but for whom perception and production abilities may still be in flux. Using a novel paradigm, we compared teens' and adults' explicit imitation and discrimination of sentences characterized by lengthened or shortened VOT.

Both age groups showed imitation in the expected directions for both extended and shortened VOT, but teens did not outperform adults. On the contrary, adults were more imitative by two separate metrics: they produced more extreme VOT modifications than teens (greater ΔVOT) when imitating lengthened VOT, and they produced VOT values that were slightly but significantly closer to the raw acoustic value of the model stimulus (lower Proximity). In perception, we found high-accuracy discrimination in both groups, with no age-based differences. Individuals with more accurate discrimination showed more imitation, although this perception-production relationship was only significant for lengthened, not shortened, VOT.

Based on the relative ease of acquisition of new accents by younger vs. older speakers (Piske et al., 2001), as well as some evidence of children showing more VOT imitation than adults (Nielsen 2014), we had speculated that teens might outperform adults in our imitation task. This was not the case; in fact, adults showed both larger adjustments and more fidelity to the model than teens, although the difference was very small. This could be interpreted as showing less imitative ability by teens, potentially consistent with evidence from previous work of less sensitivity to phonetic detail in L2 sounds (Fuhrmeister 2020), and phonetic gradience in general (McMurray et al. 2018), by younger listeners. However, the differences found in our imitation task cannot be straightforwardly ascribed to differences in perceptual acuity, or willingness/ability to perform experimental tasks, given that there were no age-based differences in our perception task.

As discussed above, quantifying the extent of imitation is not straightforward, in part because doing so first requires defining the target of imitation. If the target of imitation is an abstract property, e.g. “a token that has lengthening”, then more extreme values of that property constitute more imitation, even if they diverge from the raw values of the target stimulus. On the other hand, if the target of imitation is a specific acoustic token itself, then productions that more closely approximate the properties of that specific target constitute more imitation. Future work could examine this question directly, along the lines of work by Nielsen & Scarborough, 2019 and Zellou et al., 2016, who provided experimental evidence that the target of imitation of vowel nasalization is better conceptualized as speaker-normalized, rather than raw, values. However, without further experiments, we cannot know which of these is a more accurate depiction of participants' actual targets, so we considered both as possibilities, using the metrics ΔVOT and Proximity.

The relatively extreme VOT values used in our study meant that for the most part, these two metrics overlap. However, even with this carefully controlled design, slightly different patterns emerged: specifically, we found larger age-based differences when using ΔVOT to quantify imitation. Adults showed more lengthening than teens, driven by the fact that some adults hyper-lengthened VOT above and beyond the manipulated stimulus value. On the other hand, when considering Proximity, a more direct measure of acoustic similarity to the specific target stimulus, performance of adults and teens was more similar, though adults were still slightly more imitative.

There are two possible interpretations for the discrepancy in results across the two metrics: first, if the target of imitation is abstract, then we can conclude that adults imitate more. However, another possibility is that the target of imitation differs between adults and teens, with adults more likely to consider abstract properties of the stimulus as the target of imitation. While the data do not allow us to choose between these possibilities, the differential results highlight the importance – both analytical and theoretical – of the choice of quantification metric. In experimental paradigms with less extreme targets of imitation (including naturalistic speech), the two types of metrics are likely to diverge even further, making it even more important to consider multiple analytical possibilities and their implications.

Despite the subtle differences between the two metrics, they were consistent in showing that contrary to expectations, teens did not outperform adults in imitation. If teens are indeed more successful at novel accent acquisition than older adults in naturalistic settings, as often noted anecdotally, this cannot be straightforwardly attributed to better imitative and/or perceptual acuity than older adults.

An ancillary goal of this study was to probe the source of an oft-cited asymmetry in imitation reported by Nielsen (2011): participants modified VOT values after exposure to lengthened, but not shortened, VOT. We tested whether this lack of shortened-VOT imitation was due to perceptual factors, rather than alternative explanations such as articulatory or cognitive constraints on imitation of shortened VOT, using a paradigm and stimuli designed to increase the perceptual salience of the contrast. In our tasks, participants showed robust imitation of shortened, as well as lengthened, VOT, suggesting that the lack of shortened-VOT imitation in Nielsen (2011) was likely due to a failure to perceive the phonetic contrast in the specific stimuli used in that study. However, we cannot rule out the alternative possibility that Nielsen's (2011) findings were indeed due to production-based constraints which were overruled by the explicit instructions to imitate in our task. In either case, the comparison of these findings underscores the need for more work, using multiple paradigms, since even small differences in procedure, materials, and analysis can lead to different results. Particularly when comparing different age groups, trying to generalize results from a small number of methodologically-diverse studies could lead to a distorted view of the development of imitation.

Faithful imitation of a sound is a complex process that requires both accurate perception and accurate production. To arrive at an accurate and predictive model of speech imitation, it is therefore necessary to know the relative roles of the sub-processes. In this study, we included a direct test of one sub-component, perception of the contrast, and compared individual results on this with individual imitation performance. As expected, we found a positive relationship between the tasks; however, this was only significant for the lengthened-VOT condition. Furthermore, the correlation was fairly weak, and it appeared to be driven by the fact that participants with low discrimination accuracy had consistently low imitation, while those with high discrimination abilities showed a range of imitation. Taken together, this indicates that perception of a difference is necessary, but not sufficient, for accurate imitation, such that not all variability in imitation can be attributed to differences in perception. Another possibility is that the tasks might be limited by participant motivation, with the imitation task requiring more motivation than the discrimination task; if so, age-based differences in motivation could play a role in the differences between the results (see also Wynn et al. 2018). Future work could explore these possibilities by directly testing the correspondence between imitation and other factors (e.g. articulatory precision, personality traits, motivation) potentially conditioning individual differences in imitative performance.

In sum, this study explored the question of whether teens might have greater imitative ability than adults, and our data suggest that this is not the case. Instead, performance by the two groups was largely similar, with adults showing slightly *more* imitation than teens, based on two separate metrics quantifying used to quantify imitation. Furthermore, both age groups showed robust imitation of both shortened and lengthened VOT, consistent with the proposal that the lack of shortened-VOT imitation found in previous was due to perceptual, rather than articulatory, factors. Finally, accurate perception of differences is to some extent predictive of successful imitation, but only weakly so, indicating that there is much to be learned about what drives individual differences in imitation. The paradigm presented here offers a useful tool for future investigations of imitative performance, allowing for systematic comparison of different talkers, phonetic features, and listener populations, which we hope will ultimately lead to a more complete understanding of the various subprocesses underlying phonetic imitation.

5. Acknowledgments

Thanks to Fatima Adil and Lisa Hotson for their help with recruitment and data collection. This research was supported by grants from the Social Sciences and Humanities Research Council of Canada (to EKJ) and the Natural Sciences and Engineering Research Council of Canada (to JS).

6. References

Bates, D., Maechler, M., Bolker, B. M., & Walker, S.C. (2015). Lme4: Linear mixed-effects models using eigen and S4. https://cran.r-project.org/web/packages/lme4/index.html

De Rosario-Martinez, H. R. (2015). "Package 'Phia,'" https://CRAN.R-project.org/package=phia

Dufour, S., & Nguyen, N. (2013). How much imitation is there in a shadowing task? Frontiers in Psychology, 4, 346. https://doi.org/10.3389/fpsyg.2013.00346

Fuhrmeister, P., Schlemmer, B., & Myers, E. B. (2020). Adults Show Initial Advantages Over Children in Learning Difficult Nonnative Speech Sounds. *Journal of Speech, Language, and Hearing Research: JSLHR*, *63*(8), 2667–2679. https://doi.org/10.1044/2020_JSLHR-19-00358.

Kopečková, R., Dimroth, C., & Gut, U. (2019). Children's and adults' initial phonological acquisition of a foreign language. *Journal of Second Language Pronunciation,* 5(3), 374-401. https://doi.org/10.1075/jslp.18033.kop.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

MacLeod, B. (2021). Problems in the Difference-in-Distance measure of phonetic imitation. Journal of Phonetics, 87, 101058. https://doi.org/10.1016/j.wocn.2021.101058

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

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142.

Nielsen, K. (2014). Phonetic imitation by young children and its developmental changes. *JSLHR*, *57*(6), 2065–2075.

Nielsen, K., Scarborough, R. (2019). Perceptual target of phonetic accommodation: a pattern within a speaker's phonetic system or the raw acoustic signal? *ICPhS.*

Paquette-Smith, M., Schertz, J., & Johnson, E. K. (2021). Comparing Phonetic Convergence in Children and Adults. *Language and Speech*. https://doi.org/10.1177/00238309211013864

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception & Psychophysics*, *72*(8), 2254–2264. https://doi.org/10.3758/bf03196699

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, 79(2), 637-659. https://doi.org/10.3758/s13414-016-1226-0

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics,* 29(2), 191-214. https://doi.org/10.1006/jpho.2001.0134.

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.R-project.org

Sato, M., Grabski, K., Garnier, M., Granjon, L., Schwartz, J.-L., & Nguyen, N. (2013). Converging toward a common speech code: imitative and perceptuo-motor recalibration processes in speech production. *Frontiers in Psychology*, *4*, 422. https://doi.org/10.3389/fpsyg.2013.00422

Spinu, L., Wang, J., Pincus, N., and Vasilita, M. (2020). Exploring the use of an artificial accent of English to assess phonetic learning in monolingual and bilingual speakers. *Proceedings of Interspeech 2020.*

St. Pierre, T., Cooper, A., & Johnson, E. K. (2021). Cross-generational Phonetic Alignment between Mothers and Their Children. Language Learning and Development: The Official Journal of the Society for Language Development, 1–22. https://doi.org/10.1080/15475441.2021.1979401

Wynn, C. J., Borrie, S. A., & Sellers, T. P. (2018). Speech Rate Entrainment in Children and Adults With and Without Autism Spectrum Disorder. American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association, 27(3), 965–974. https://doi.org/10.1044/2018_AJSLP-17-0134

Yu, A.C.L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: subjective attitude, personality, and 'autistic' traits. *PLoS ONE*, 8(9), e74746. https://doi.org/10.1371/journal.pone.0074746.

Zellou, G., Scarborough, R., Nielsen, K. (2016) Phonetic imitation of coarticulatory vowel nasalization. JASA 140, 3560-3575.