Phonetic imitation of stops and vowels: Individual stability and perceptual underpinnings

Hanna Zhang[1,2], Jessamyn Schertz[1,2]

[1]University of Toronto, [2]University of Toronto Mississauga
hanna.zhang@mail.utoronto.ca, jessamyn.schertz@utoronto.ca

ABSTRACT

This study examines whether individuals who show greater-than-average imitation of phonetic differences in a given feature also show greater-than-average imitation of other features, as might be expected under the view that general cognitive or social factors play a primary role in predicting individual differences in imitation. An explicit imitation task tested the extent to which English speakers imitated minimal differences in voice onset time (VOT) of voiceless stops and F2 of the vowel /u/, and an ABX discrimination task tested sensitivity to the same differences. Participants imitated differences in both features, and individuals' perceptual acuity for a given feature predicted the extent of imitation of that feature. However, the extent to which an individual imitated one feature did not predict the extent to which they imitated the other, and the same dissociation was found in discrimination. These findings highlight the importance of considering multiple features, and the role of perception, in studies of individual variability in imitation.

Keywords: phonetic imitation, discrimination, individual differences, VOT, F2

1. INTRODUCTION

Phonetic convergence or imitation is the process by which speakers adapt their speech to match more closely with the phonetic properties of the incoming signal. Various social [5, 7], general cognitive [1], and psycholinguistic [12, 15] factors have been shown to influence the degree of convergence. However, there have been few direct tests of whether individual variation in extent of imitation is stable across different features: in other words, do individuals who show greater-than-average imitation of phonetic differences in (for example) consonants also show greater-than-average imitation of vowels?

Previous work has found that extent of convergence in implicit shadowing or conversational tasks is related to several specific personality/cognitive traits, including the Big Five Inventory personality traits of Openness and Neuroticism, Attention-Switching, and rejection sensitivity [27, 13, 2]. An individual's attitude towards the talker has also been shown to play a role, with more positive attitudes eliciting more convergence [4, 27]. Work on explicit imitation, mostly in the domain of second-language sound acquisition, has also found correspondences between non-linguistic traits and phonetic imitation ability, including neurocognitive flexibility [21] and musical talent [8]. This is in line with proposals that there may be innate "talent" for phonetic imitation, with identifiable neurocognitive markers [9, 20].

If non-language-specific characteristics are indeed strong predictors of imitation, they would be expected to be relatively stable and to apply similarly to different phonetic features; in other words, an individual's relative extent of imitation should be consistent over time, across paradigms, and across different features. There is some evidence for this: individual differences in convergence to lengthened VOT during a shadowing task were found to remain relatively stable across time (between two sessions 1-2 weeks apart), and across two different talkers [24]. However, this work examined convergence to a single feature (lengthened VOT) in a single task, leaving open the question of whether this stability would hold across different methodologies or different unrelated features.

Phonetic convergence has been found across a broad range of features, including f0, speaking rate, vowel formants, and VOT (see [17] for a review). Previous work has shown that the extent of convergence or imitation may be influenced by characteristics of the feature itself. Differences in the extent of convergence across two different features have been attributed to differences in phonological relevance [15], perceptual salience [18]. That said, even if there are systematic feature-based differences in extent of convergence, we would still expect that, all else being equal, individuals who show greater-than-average convergence for one feature would also show greater-than-average convergence for other features.

However, there is little evidence that this is the case. Studies exploring the extent to which individuals converge along specific phonetic dimensions have for the most part targeted a single sound or a set of related sound (e.g., VOT of

voiceless stops). Two studies we are aware of have directly examined the stability of individual levels of convergence across prosodic features (e.g. f0 and speech rate) and lexical factors in conversation [22, 25]; neither found evidence that individual levels of convergence were correlated across features. However, given the uncontrolled nature of conversational tasks, and the fact that the amount of possible convergence depends on the distance between the production and the target (such that if the imitator and their interlocuter naturally produce the feature of interest with the same acoustic characteristics, no convergence is possible), it is possible that there is simply too much other variability to reveal individual stability even if it does exist. Therefore, it is possible that more stable patterns would be found in a more controlled task.

1.1. Current Study

In this study, we test whether individual differences in explicit imitation are stable across two distinct phonetic features in English: VOT of voiceless stops and backness of the vowel /u/. Imitation of lengthened VOT has been particularly well-studied and has been shown to be robust [23, 24, 27]. Overall convergence to vowel formant differences has also been found [4, 10], but no studies we are aware of have targeted /u/ backness specifically.

Imitation was tested using an explicit imitation paradigm in which participants are explicitly instructed to imitate minimal differences in the relevant features. Sensitivity to the same differences was tested in an ABX discrimination task. We chose this methodology as it provides a completely controlled paradigm in which it is possible to calculate a measure of imitation that is independent of participants' baseline values, as well as a direct test of perception of the target differences.

Our analysis is structured as follows. First, we test whether there is imitation of, and sensitivity to, differences in each feature: previous work leads us to expect robust imitation of VOT, but we are not aware of studies that directly test sensitivity to, or imitation of, differences in the backness of /u/. We then test whether there is a relationship between individuals' extent of imitation of the two features, as would be expected if the degree to which speakers imitate phonetic differences is a stable property within individuals. Finally, we test the relationship between individual imitation and discrimination in order to determine whether any differences in imitation of the features might be attributable to differences in perceptual salience, since this has been proposed to underlie differences in imitation in previous work [11, 12, 8].

2. METHODS

2.1. Participants

Data from 81 participants was analyzed (mean year-of-birth = 1996 (range = 1991-2004), women: 43, men: 32, non-binary or genderqueer: 6). Participants were recruited via Prolific [19], and all were native speakers of English, born and residing in the US or Canada, and reported learning English and no other languages in the home. An additional 18 participants completed the study but were excluded because they reported early exposure to another language in the home (n=17) or because their recordings were not of sufficient quality for phonetic measurement (n=1).

2.2. Stimuli

Target stimuli consisted of pairs of English words differing minimally in VOT (stop words) or F2 of /u/ (/u/ words). Four stop words (*carrots*, *kale*, *parrots*, *tigers*) and four /u/ words (*choose, food, goose*, *tube*) were produced by a phonetically-trained native speaker of English, and were then manipulated using Praat (PSOLA algorithm for VOT duration and LPC resynthesis for vowel formants) [6] to create two versions of each word. For stop words, Version A had the naturally-produced VOT value (average VOT: 81ms) and Version B had VOT artificially lengthened by 60ms (average VOT: 141ms). For /u/ words, the versions differed in F2: Version A had an F2 corresponding to the naturally produced /u/, which was intentionally produced as far back as possible (1400 Hz), while Version B had a higher F2 (2000 Hz), corresponding to a fronter vowel. Nine additional pairs of stimuli targeting different features were used as fillers.

2.3. Procedure

The experiment was completed fully online on Gorilla [3]. Prior to the main task, participants completed an audio screening task to ensure that they were wearing headphones [26] as well as a baseline word-reading task (not reported here).

The main task consisted of a series of "trial sets," each corresponding to one pair of stimuli (e.g., *tigers* with natural and lengthened VOT). Each trial set was composed of three stages: exposure, imitation, and discrimination. In the exposure stage, participants listened to Versions A and B of the word in sequence. This stage was repeated twice. In the imitation stage, participants listened to the two versions in sequence again, but after each word, they were instructed to repeat the word out loud, imitating what they had heard as closely as possible. This stage was also repeated twice. In the

discrimination stage, participants heard a third word (X) and decided whether it matched the first (A) or second (B). Each trial set contained 4 different ABX trials, with X identical to either the A or B token. In total, each participant produced 32 imitations (8 target words * 2 versions * 2 repetitions) and completed 32 discrimination trials (8 target words * 4 ABX trials).

2.4. Phonetic Analysis (imitation task)

VOT of stop words produced in the imitation task was annotated beginning just before the stop burst and ending at the onset of periodicity in the following vowel. For /u/ words, boundaries were identified using the Montreal Forced Aligner [14] and manually corrected if necessary, using the onset and offset of stable second formant to determine the boundaries. F2 was measured a quarter of the way between the onset of the vowel and the onset of the following consonant (referred to as the 25% point) using Praat [6], and all values were manually checked and corrected if necessary.

3. RESULTS

3.1. Imitation

Imitation was quantified by comparing participants' VOT (for stop words) and F2 (for /u/ words) across the two versions of each word. We used two linear mixed-effects regression models, one for each feature, to test for imitation. The predictor variable for both models was *version* (A vs. B, simple-coded: (-0.5, 0.5)). The response variable for the stop model was VOT and for the /u/ model it was F2 frequency measured at the 25% point, with a by-participant random intercept and slope for feature. For stops, there was a significant difference between conditions (Figure 1): participants produced longer VOTs for lengthened-VOT stimuli (mean 104ms) than natural-VOT stimuli (76ms) ($\beta = 26.06$, $t = 7.86$, $p < 0.001$). For /u/, there was also a significant difference between conditions (Figure 2): participants produced greater F2s for fronted stimuli (1890Hz) than natural stimuli (1669Hz) ($\beta = 208.15$, $t = 2.98$, $p = 0.009$). To summarize, imitation was found for both features.

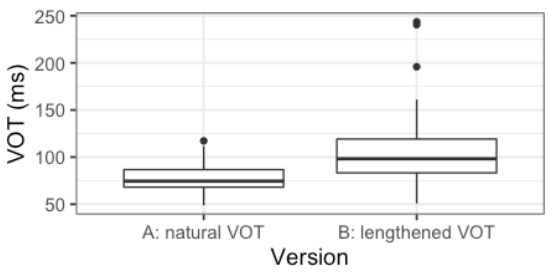

Figure 1: VOT of imitated stop words. Boxplots show distributions of by-participant means.

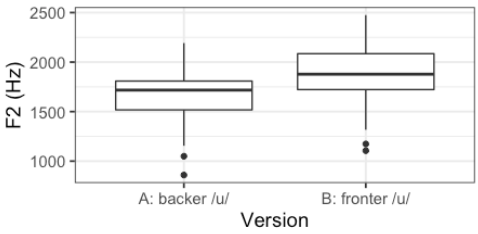

Figure 2: F2 of imitated /u/ words. Boxplots show distributions of by-participant means.

3.2. Discrimination

We used a logistic mixed-effects regression model to test whether participants were able to discriminate differences in both target features and if there were differences in the discriminability of the two features. The predictor variable was *feature* (levels: stop, /u/, simple-coded (-0.5, 0.5)) and the response variable was accuracy on the ABX task, with a by-participant random intercept and slope for feature. Results (Figure 3) showed that overall, discrimination accuracy was significantly above chance (intercept: $\beta = 1.62$, $p < 0.001$) and that there was no significant difference in mean discrimination accuracy between stop words (78%) and /u/ words (81%) ($\beta = 0.07$, $z = 0.126$, $p = 0.900$).

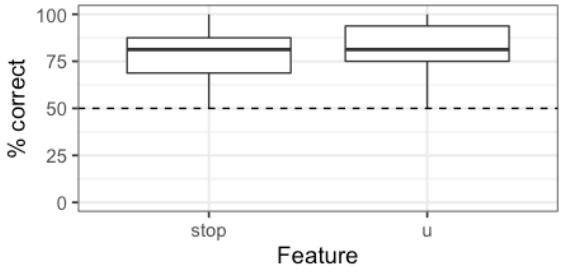

Figure 3: Discrimination accuracy across features. Boxplots show distributions of by-participant means.

3.3. Individual Correlations

To compare individual performance across features and tasks, we calculated by-participant imitation and discrimination scores for each feature. An individual's imitation score was the average difference in VOT (for stop words) or F2 (for vowel words) between imitations of the two versions, with positive values corresponding to differences in the expected direction. An individual's discrimination score for a given feature was the mean accuracy for trials targeting that feature in the discrimination task.

Linear regression models were used to compare correlations between individuals' (a) extent of imitation of stop versus /u/ words, (b) discrimination accuracy of stop versus /u/ words, (c) discrimination

vs. imitation of stop words, and (d) discrimination vs. imitation of /u/ words.

In terms of our primary question about the stability of individual imitation across features, results showed that the extent to which a participant imitates VOT differences does not predict the extent to which they imitate /u/ F2 differences (Figure 4, left: $\beta = 1.24$, $t = 1.62$, $p = 0.110$). In other words, even though most participants did show imitation in the expected direction, as shown by primarily positive imitation scores, there is no evidence that individuals' extent of imitation is stable across these two features. There was also no significant correlation between discrimination of the two features (Figure 4, right: $\beta = -0.02$, $t = -0.211$, $p = 0.833$).

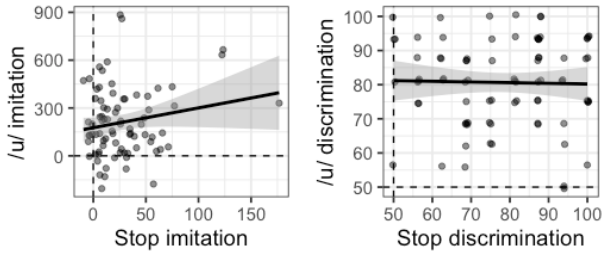

Figure 4: Left: individual imitation of stops vs. /u/; Right: individual discrimination of stops vs. /u/. Dashed lines indicate no imitation or at-chance discrimination.

On the other hand, the extent to which a participant discriminates differences in stop VOTs *does* predict the extent to which they imitate stop VOT differences (Figure 5, left: $\beta = 0.79$, $t = 3.72$, $p < 0.001$), and the same was found for discrimination versus imitation of /u/ fronting (Figure 5, right: $\beta = 6.05$, $t = 3.29$, $p = 0.002$).

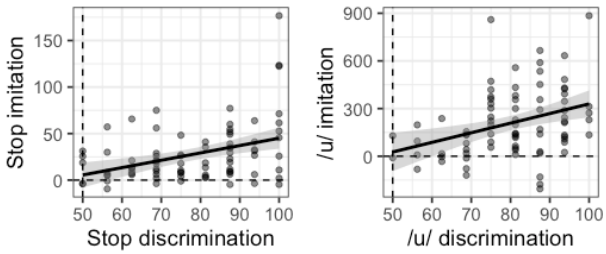

Figure 5: Left: individual discrimination vs. imitation of stops; Right: individual discrimination vs. imitation of /u/.

4. DISCUSSION

Our results suggest that individuals' relative extent of imitation may not be consistent across different features, even in a highly controlled task. English speakers imitated differences in both VOT and voiceless stops and F2 of the vowel /u/ in an explicit imitation task, and considerable individual variability was found for both features. While imitation of within-category VOT differences was expected based on previous work [23, 24, 27], it was less clear as to whether differences in /u/ backness would also be imitated. These results provide novel evidence that speakers are capable of manipulating specific spectral properties of vowels (in this case, F2 of /u/) in an explicit imitation task.

In terms of our primary question of interest, we found no evidence of stability in individual performance across the two features. The extent of an individual's imitation of each feature was, however, related to that individual's perception of the same feature. This extends earlier work examining extent of convergence across features that also found a lack of constancy in convergence during conversational tasks [22, 25]. Furthermore, the correspondence between perception and imitation is consistent with previous proposals for the role of perceptual salience in imitation [11, 12]. While perhaps unsurprising, given that accurate perception is a prerequisite for accurate imitation, this highlights the importance of considering, and ideally, directly testing, the role of perception when trying to account for individual variability in imitation.

Our findings are not consistent with a view that non-linguistic cognitive or social characteristics are the strongest predictors of imitation; if this were the case, we would expect that participants who showed the greatest imitation of stops would also show the greatest imitation of /u/. Our results do not rule out the possibility that these traits play a role in imitation, just that this role may be more minor than that of other factors (including, as shown here, feature-specific individual perceptual acuity). This suggests that it may be problematic to consider results from a task targeting a single feature as an index of an individual's overall imitative ability.

In conclusion, this study found that individual differences in patterns of imitation found for one feature do not necessarily generalize to other features, even in a highly controlled paradigm with substantial individual variability in both perception and production of each feature. Instead, differences in individual perceptual acuity appeared to be the primary driver of the different patterns of imitation in this task. It is therefore imperative to test imitative performance across multiple features, and to control for the role of perception, in work aiming to quantify the role of the many cognitive, social, and linguistic factors underlying phonetic imitation.

5. REFERENCES

[1] Abel, J., & Babel, M. (2017). Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech, 60*(3), 479–502.

[2] Aguilar, L., Downey, G., Krauss, R., Pardo, J., Lane, S., & Bolger, N. (2016). A Dyadic Perspective on Speech Accommodation and Social Connection: Both Partners' Rejection Sensitivity Matters: A Dyadic Perspective on Speech. *Journal of Personality, 84*(2), 165–177.

[3] Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407.

[4] Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*, 437–456.

[5] Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40*(1), 177–189.

[6] Boersma, P. & Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.11, retrieved March 2022 from http://www.praat.org/.

[7] Cohen Priva, U., & Sanker, C. (2020). Natural Leaders: Some Interlocutors Elicit Greater Convergence Across Conversations and Across Characteristics. *Cognitive Science, 44*(10), e12897.

[8] Coumel, M., Christiner, M., & Reiterer, S. M. (2019). Second Language Accent Faking Ability Depends on Musical Abilities, Not on Working Memory. *Frontiers in Psychology, 10*, 257–257.

[9] Dogil, G. & Reiterer, S. (2009). Language Talent and Brain Activity. Berlin, New York: De Gruyter Mouton.

[10] Dufour, S., & Nguyen, N. (2013). How much imitation is there in a shadowing task? *Frontiers in Psychology, 4*, 346–346.

[11] Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience, 34*(6), 769–786.

[12] Large, N. R., Frisch, S., & Pisoni, D. B. (1998). Perception of wordlikeness: Effects of segment probability and length on subjective ratings and processing of non-word sound patterns. *Research on Spoken Language Processing: Progress Report, 22*, 95–125.

[13] Lewandowski, N., Jilka M. (2019). Phonetic Convergence, Language Talent, Personality and Attention. *Frontiers in Communication, 4*.

[14] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., Sonderegger, M. (2017) Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Proceedings Interspeech 2017,* 498-502,

[15] Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from a shadowing task. *Cognition, 109*, 68–173.

[16] Nielsen, K.Y., & Scarborough, R. (2015). Perceptual asymmetry between greater and lesser vowel nasality and VOT. *Proceedings of ICPhS*.

[17] Pardo, J. S., Pellegrino, E., Dellwo, V., & Möbius, B. (2022). Special issue: Vocal accommodation in speech communication. *Journal of Phonetics, 95*, 101196.

[18] Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The Distinctness of Speakers' Productions of Vowel Contrasts Is Related to Their Discrimination of the Contrasts. *The Journal of the Acoustical Society of America, 116*(4), 2338–2344.

[19] Prolific (2014). https://www.prolific.co

[20] Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., Winkler, S., & Ackermann, H. (2011). Individual differences in audio-vocal speech imitation aptitude in late bilinguals: functional neuro-imaging and brain morphology. *Frontiers in Psychology, 2*, 271–271.

[21] Reiterer, S. M., Hu, X., Sumathi, T. A., & Singh, N. C. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Frontiers in Psychology, 4*, 782–782.

[22] Sanker, C. (2015). Comparison of phonetic convergence in multiple measures. *Cornell working papers in phonetics and phonology 2015*, 60–75.

[23] Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*(3), 422–429.

[24] Wade, L., Lai, W., & Tamminga, M. (2020). The Reliability of Individual Differences 456 in VOT Imitation. *Language and Speech*, *64*(3), 576-593.

[25] Weise, A., & Levitan, R. (2018). Looking for structure in lexical and acoustic-prosodic entrainment behaviors. Proceedings of the 2018 *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2*, 297–302.

[26] Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics, 79(*7), 2064–2072.

[27] 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.