PHONETIC IMITATION IS NOT CONDITIONED BY PRESERVATION OF PHONOLOGICAL CONTRAST BUT BY PERCEPTUAL SALIENCE

Václav Jonáš Podlipský and Šárka Šimáčková

Department of English and American Studies, Faculty of Arts, Palacký University Olomouc, Czech Republic
vaclav.j.podlipsky@upol.cz, sarka.simackova@upol.cz

ABSTRACT

This study addresses disagreement between previous studies about the hypothesis that phonetic imitation does not occur if it would threaten a phonological contrast. Using a within-subject pretest-shadowing-posttest design, we tested imitability of reduction and extension of prevoicing and vowel duration in Czech, a vowel-quantity language with prevoiced-vs-unaspirated stop contrasts. Results showed imitation of extended but not of reduced prevoicing. This is compatible with the contrast-preservation hypothesis, but may also be ascribed to the lower perceptual salience of prevoicing reduction than its extension in the presence of other voicing cues. In contrast, reducing duration of a Czech long vowel is salient to native listeners. Indeed, both directions of (natural-sounding) vowel duration manipulation were imitated, even though reduction decreased the distance between phonologically long and short categories. We conclude that contrast preservation does not necessarily preclude imitation and that the likelihood of imitation of a given feature depends on its perceptual salience.

Keywords: Phonetic imitation, contrast preservation, perceptual salience, prevoicing, vowel duration.

1. INTRODUCTION

Despite much attention in recent years, the phenomenon of phonetic imitation (also referred to as, or overlapping with, convergence or alignment) is not fully understood (see [22] and [12] for recent reviews). While some accounts of imitation as an automatic process exist, e.g. [9, 25], it is clear that imitation is not automatic in that it would always occur. It is modulated by social factors [3, 21, 23, 24] and possibly also linguistic closeness between speakers [16] (but cf. [1–3]). In addition, imitation seems to be guided by phonological patterns of the language. This is obvious from findings that an imitation-induced change can generalize from one member of a phonological contrast to another, which was shown for imitation of modified VOT for a place-of-articulation contrast [19, 20] and a voicing contrast [26]. However, there are also discrepancies between different studies with respect to the phonological influence on imitation. At the suprasegmental level, Cole and Shattuck-Hufnagel [8] suggest that speakers imitate the prosody of repeated utterances in its phonological structure more accurately than in its phonetic detail. However, D'Imperio et al. [10] recently documented imitation of intonation both in terms of phonological representations and phonetic detail. At the segmental level, Mitterer and Ernestus [18] did not find imitation of longer prevoicing (i.e. negative VOT) in Dutch and concluded that this was because the exact duration of prevoicing was "phonologically irrelevant" (171, 173). However, most other studies of VOT imitation did show imitation of sub-phonemic detail [11, 19, 20, 27, 28]. These discrepancies have been explained as due to various methodological differences by different authors [1, 10, 13, 15].

Besides whether or not imitation can be sub-phonemic, Mitterer and Ernestus's (M&E) study [18] disagrees with Nielsen's [19] findings in another way. Nielsen found imitation of artificially extended but not of reduced positive VOT in English, and as one possible cause of this asymmetry she suggested another kind of influence of phonology on imitation, namely a selective suppression of imitation if such imitation would result in a reduced distance between phonological categories. M&E found no difference in duration of prevoicing produced by Dutch speakers when shadowing syllables with medium and longer prevoicing, as mentioned above, but did find shorter prevoicing when speakers were shadowing voiced stops without any prevoicing. This is in direct contradiction to Nielsen's hypothesis because a shift away from the boundary between phonological categories was not imitated and a shift towards it was. Similarly, the intonation study by D'Imperio et al. [10] reports imitation-induced peak alignment shifts towards a pitch accent category. The disagreement between M&E's and Nielsen's results may be due to the substantial methodological differences: M&E used a speeded shadowing task with extended-VOT and reduced-VOT trials intermixed, whereas Nielsen's participants, different for VOT extension and reduction, did not shadow words during exposure at all. Another explanation could be that aspiration (Nielsen) and prevoicing

(M&E), even if they are measured along one temporal dimension of VOT, are in fact acoustically different events of inherently different perceptual salience, cf. [30]. Zellou et al. [34] tested the imitability of reduced and increased (allophonic) nasality in pre-nasal English vowels and found that both directions of manipulation resulted in nasality shifts during shadowing, but only reduction persisted into posttest. This asymmetry was ascribed by the authors to a difference in perceptual salience of the two kinds of manipulation. Findings by Babel and colleagues [2, 3] suggest that imitation is facilitated by exposure to novel (and hence more salient) stimuli. Perceptual salience could result in greater attention and, in turn, a higher degree of imitation. Speaking of attention globally, Yu et al. [33] found that individuals more attentive to, and more engaged with, the exposure materials demonstrated a greater degree of imitation.

The present study assesses imitation of two manipulated temporal features in order to test the hypothesis that speakers selectively inhibit imitation of a shift towards a contrasting phonological category. Using a methodology that enables comparisons with both Nielsen [19] and M&E [18], we measure imitation of reduced and extended prevoicing in Czech, a language with prevoiced stops and unaspirated voiceless stops [29]. Further, we examine imitation of reduced and extended vowel duration, the essential cue to vowel quantity in Czech [29]. We expect reduced duration of a long vowel (in stimuli still sounding natural) to be more perceptually salient to Czech listeners than the reduction of prevoicing, given the multiple other cues to voicing [31]. We specifically ask whether contrast preservation constrains even the imitation of a perceptually salient feature.

2. METHOD

This study uses a pretest-shadowing-posttest design with only within-subject comparisons, in light of variability between participants found in previous imitation studies, see [33]. Each participant was exposed to reduction and extension of duration of prevoicing in Czech /d/, and duration of the phonologically long Czech vowel /uː/.

2.1. Materials

The material consisted of 69 existing Czech disyllabic words. (Stress is always initial in Czech.) Twenty words were /d/-initial, with non-high vowels after /d/. There were no other stops in the whole target-word set, except for a few medial voiced and final voiceless ones. Nineteen other words contained /uː/ in the first syllable. Consonants adjacent to /uː/ were obstruents. The remaining words were fillers, 15 with phonologically short vowels in both syllables and 15 with a long non-high vowel in the first syllable. The extent of imitation depends on lexical frequency, e.g. [14, 19]. Low-frequency words were used, as determined from [4]: mean frequency was 1.54 (1.28 SD) and 1.62 (4.76 SD) instances per million positions for the /d/-words and /uː/-words respectively.

An uninformed female Czech speaker recorded all 69 words 3 times following the pretest procedure described below. One token of each word was selected and scaled to equal intensity. Using PSOLA in Praat [5], the duration of each /d/-prevoicing was extended and reduced by factors of 1.44 and 0.255 respectively. The duration of each /uː/ was extended and reduced by factors of 1.277 and 0.763 respectively. The magnitudes of extension and reduction were chosen so that the resulting stimuli sounded natural (though perhaps carefully pronounced) to 22 uninformed native Czech listeners (none taking part in the imitation experiment itself). Larger manipulations than these were noticeable, except for prevoicing reduction: even complete removal of prevoicing went unnoticed. Means of original and resulting durations of prevoicing and /uː/ are given in Table 1.

Table 1: Mean durations of /d/-prevoicing and /uː/ in exposure stimuli before and after manipulation (in ms). Standard deviations are in parentheses.

	Reduced	Original	Extended
/d/-prevoicing	26 (5.1)	104 (16.0)	149 (23.0)
/uː/	127 (16.1)	166 (20.1)	212 (25.4)

2.2. Procedure

Testing was divided into 2 sessions, separated by at least 24 hours. Participants received exposure to both manipulated segments, /d/ and /uː/, in the same session, but exposure to the reduction of a segment and its extension was always in different sessions. The 4 possible orders and combinations of reduction and extension of /d/ and of /uː/ in the 2 sessions (e.g. session 1: extended /d/, reduced /uː/; session 2: reduced /d/, extended /uː/; etc.) were fully counterbalanced across participants.

Session 1 comprised silent familiarization with the target words, the pretest eliciting baseline productions, the shadowing task, and the posttest. The parts were separated by breaks, there were short breaks halfway through each part, and speakers could interrupt and resume at any time. In the pretest and posttest, speakers read words appearing automatically one by one on a computer screen. In shadowing, they repeated each word right after they

heard it. Session 2 was the same in form but there was no new pretest.

To reduce speech rate differences between parts, stimuli, whether visual or auditory, were always presented at 2.7 s intervals. Within as well as between participants, the 69 words were never ordered the same way twice: there were 23 triplets of a random /d/-word (or filler), a random filler, and a random /uː/-word (or filler). A Praat Demo window script was used for presentation of all stimuli.

Testing took place individually in a sound booth. Stimuli were presented via Sennheiser HD 280 pro headphones, which participants never took off (to prevent Lombard-effect differences between parts). A Zoom H4n digital recorder with a 44.1 kHz sampling rate and 24 bit quantization was used.

2.3. Participants

Sixteen female native Czech speakers with normal hearing, aged between 19 and 24, took part in the study. They could speak some foreign languages but had never been abroad for a period longer than 3 weeks. They were (or had been) university students (not majoring in linguistics) and were paid for participating.

2.4. Measurements

A phonetically trained research assistant manually labelled boundaries of prevoicing, and of the /d/-word, in Praat with constant criteria viewing waveforms and spectrograms (with a 50-dB dynamic range). Boundaries of /uː/, and of each /uː/-word, were labelled by the first author using constant criteria based on [17].

3. RESULTS

First, mean word durations were submitted to a repeated-measure analysis of variance (RM ANOVA) with Manipulated segment (/d/, /uː/) and Part (Pretest, Shadowing of extended, Posttest after extended, Shadowing of reduced, Posttest after reduced) as within-subject factors. A significant main effect of Part was found (F[4, 60] = 10.55, p < .001). There was no interaction between Manipulated segment and Part (p > .4). A post-hoc Tukey HSD test showed that both /d/- and /uː/-words were longer in Shadowing of extended and Posttest after extended than in Pretest.

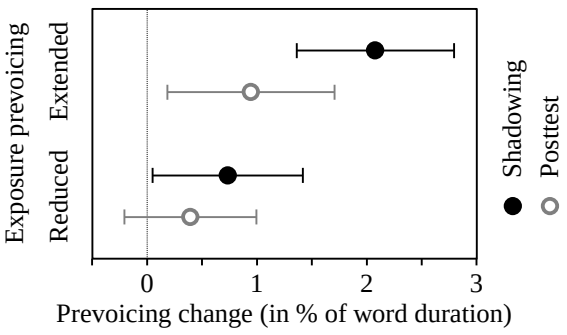
Because of the differences in word duration between parts, subsequent analyses never used absolute duration of the target segment (/d/-prevoicing or /uː/) but normalized duration, defined as the ratio of the duration of the segment to the duration of the word it occurred in.

3.1. Imitation of manipulated /d/-prevoicing

Overall, in about 3% of the recorded /d/-words prevoicing was absent (which constituted missing data). The numbers of cases of absent prevoicing for each participant were submitted to a RM ANOVA with Part as the 5-level within-subject factor. No significant effect of Part was found (p > .2).

For each /d/-word of each participant and each treatment condition (i.e. Shadowing of extended, Posttest after extended, Shadowing of reduced, Posttest after reduced), we computed the differences of normalized prevoicing duration from pretest baseline production (henceforth "prevoicing change"). (Missing data were replaced by means across words for the given condition.) Each participant's prevoicing changes (pooled across words) were submitted to a RM ANOVA with Exposure prevoicing (Extended vs Reduced) and Task (Shadowing vs Posttest) as within-subject factors. Both Exposure prevoicing (F[1, 15] = 17.88, p < .001) and Task (F[1, 15] = 7.60, p <.05) had a significant effect, with a greater positive prevoicing change (i.e. lengthening) for Extended than Reduced and for Shadowing than Posttest. The interaction between the two factors approached significance (F[1, 15] = 3.99, p = .064). Inspection of confidence intervals, as seen in Fig. 1, indicated that prevoicing change in all conditions was significantly higher than 0 (i.e. pretest), except for Posttest after reduced which did not differ from 0 significantly. A post-hoc Tukey HSD test showed that prevoicing change during Shadowing of extended was significantly higher than changes during the three remaining conditions (p < .01) which were comparable.

Figure 1: Mean prevoicing change from pretest plotted by Exposure prevoicing and Task. Error bars indicate .95 confidence intervals.

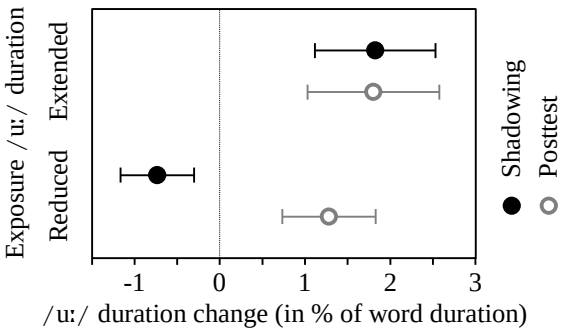


3.2. Imitation of manipulated /uː/ duration

Analogically to prevoicing changes, /uː/ duration changes in each condition were computed. Means are plotted in Fig. 2. A RM ANOVA on each participant's /uː/ duration changes, pooled across

words, revealed a significant effect of both Exposure /uː/ duration (F[1, 15] = 25.19, p < .001) and Task (F[1, 15] = 20.39, p <.001), with higher values of /uː/ duration change for Extended than Reduced and for Posttest than Shadowing. The factors also interacted significantly (F[1, 15] = 74.11, p <.001). A post-hoc Tukey HSD test found that /uː/ duration changes occurring in the two Extended conditions were comparable and were significantly higher than those in the Reduced conditions (p < .05), and also that /uː/ duration change in Shadowing of reduced was lower than changes in the three remaining conditions (p < .001), see Fig. 2. Examination of confidence intervals showed that /uː/ duration change differed significantly from 0 (i.e. pretest) for all conditions, being negative for Shadowing of reduced and positive otherwise.

Figure 2: Mean /uː/ duration change from pretest plotted by Exposure /uː/ duration and Task. Error bars indicate .95 confidence intervals.



3.3. Correlations between /d/ and /uː/ imitation

The design of our study allowed us to make comparisons between imitation of prevoicing and imitation of /uː/ duration within participants. We calculated correlations between prevoicing change in each condition and /uː/ duration change in each condition. The only significant correlation found was between prevoicing change in Shadowing of reduced and /uː/ duration change in Shadowing of reduced, and it was negative (r = -.56, p < .05).

4. DISCUSSION

To summarize the results, we found that participants imitated the lengthening of /d/-prevoicing during shadowing as well as in the subsequent posttest (though less so), whereas they did not shorten prevoicing during or after exposure to words with reduced prevoicing. On the other hand, speakers' /uː/'s were longer after exposure to extended /uː/ duration (both in shadowing and posttest) *as well as* shorter during shadowing of words with reduced /uː/, though in fact longer in the subsequent posttest. (For the last-mentioned finding we can offer no explanation, apart from noting that lack of transfer of imitation from shadowing to posttest has been attested, e.g. [2, 34].) We found no positive correlation between the degrees of /d/ imitation and /uː/ imitation across participants. This suggests that in repetition tasks such as ours, which generally facilitate imitation, differences between individuals in imitative behavior [33] may become obscured, just like socially based effects (as noted e.g. in [32]). Our ancillary finding is that word lengthening, both of the /d/- and the /uː/-words, was imitated which corroborates previous research on speaking rate imitation, e.g. [6, 7].

Turning now to the aims of this study, we investigated the imitability of shifts of Czech /d/ and /uː/ towards and away from their phonological counterparts, /t/ and /u/, in order to test the hypothesis that imitation is constrained by preservation of contrast. Our results for /d/-prevoicing (negative VOT) are comparable to Nielsen's [19] for positive VOT imitation in English, and disagree with Mitterer and Ernestus's [18] findings about prevoicing imitation by Dutch speakers. Thus, our results for /d/-prevoicing are interpretable in terms of the contrast preservation hypothesis.

However, the absence of imitation of a phonetic property could also be attributed to its low perceptual salience, as noted e.g. in [1]. An independent test confirmed our expectation that the reduction of /d/-prevoicing is less salient to Czech listeners than its extension, probably owing to other available voicing cues. By contrast, there is no apparent asymmetry in the salience of reduction and extension of /uː/ duration, and manipulation of /uː/ duration thus represents a better test case for the contrast preservation hypothesis. Recall that imitation of both directions of /uː/ duration manipulation was indeed found (albeit only during shadowing). Therefore, we conclude that contrast preservation does not necessarily preclude imitation. This is in line with D'Imperio et al.'s [10] findings about imitation of intonation. Our second conclusion, also in line with previous literature (e.g. [2, 3, 34]), is that the likelihood of imitation of a given phonetic property depends on its perceptual salience.

5. ACKNOWLEDGEMENTS

The work of the first author was supported by an internal research grant from the Faculty of Arts of Palacký University Olomouc.

6. REFERENCES

[1] Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J. Phonetics* 40, 177–189.

[2] Babel M., McAuliffe, M., Haber, G. 2013. Can mergers-in-progress be unmerged in speech accommodation? *Front. Psychol*. 4, art. no. 653.

[3] Babel, M., McGuire, G., Walters, S., Nicholls, A. 2014. Novelty and social preference in phonetic accommodation. *Lab. Phonol.* 5, 123–150.

[4] Benešová, L., Křen, M., Waclawičová, M. 2013. *ORAL2013: a representative corpus of informal spoken Czech*. Institute of the Czech National Corpus. http://www.korpus.cz/.

[5] Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer [Computer program]. Version 5.4.01, retrieved 9 November 2014, http://www.praat.org/.

[6] Bosshardt, H. G., Sappok, C., Knipschild, M., Hölscher, C. 1997. Spontaneous imitation of fundamental frequency and speech rate by nonstutterers and stutterers. *J. Psycholinguist. Res.* 26, 425–448.

[7] Brouwer, S., Mitterer, H., Huettig, F. 2010. Shadowing reduced speech and alignment. *J. Acoust. Soc. Am*. 128, EL32–EL37.

[8] Cole, J., Shattuck-Hufnagel, S. 2011. The phonology and phonetics of perceived prosody: What do listeners imitate? *Proc. Interspeech 2011* Florence, 969–972.

[9] Delvaux, V., Soquet, A. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64, 145–173.

[10] D'Imperio, M., Cavone, R., Petrone, C. 2014. Phonetic and phonological imitation of intonation in two varieties of Italian. *Front. Psychol.* 5, art. no. 1226.

[11] Fowler, C. A., Brown, J. M., Sabadini, L., Weihing, J. 2003. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *J. Mem. Lang.* 49, 396–413.

[12] Gambi, C., Pickering, M. J. 2013. Prediction and imitation in speech. *Front. Psychol.* 4, art. no. 340.

[13] German, J. S., Carlson, K., Pierrehumbert, J. B. 2013. Reassignment of consonant allophones in rapid dialect acquisition. *J. Phonetics* 41, 228–248.

[14] Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279.

[15] Honorof, D. N., Weihing, J., Fowler, C. A. 2011. Articulatory events are imitated under rapid shadowing. *J. Phonetics* 39, 18–38.

[16] Kim, M., Horton, W. S., Bradlow, A. R. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Lab. Phonol.* 2, 125–156.

[17] Machač, P., Skarnitzl, R. 2009. *Principles of Phonetic Segmentation*. Prague: Epocha.

[18] Mitterer, H., Ernestus, M. 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* 109, 168–173.

[19] Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *J. Phonetics* 39, 132–142.

[20] Nielsen, K. 2014. Phonetic imitation by young children and its developmental changes. *J. Speech Lang. Hear. R.* 57, 2065–2075.

[21] Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am*. 119, 2382–2393.

[22] Pardo, J. S. 2013. Measuring phonetic convergence in speech production. *Front. Psychol.* 4, art. no. 559.

[23] Pardo, J. S., Jay, I. C., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C., Krauss, R. M. 2013. Influence of role-switching on phonetic convergence in conversation. *Discourse Processes* 50, 276–300.

[24] Pardo, J. S., Jay, I. C., Krauss, R. M. 2010. Conversational role influences speech imitation. *Attent. Percept. Psycho.* 72, 2254–2264.

[25] Pickering, M. J., Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–190.

[26] Podlipský, V. J., Šimáčková, Š., Chládková, K. 2013. Imitation interacts with one's second-language phonology but it does not operate cross-linguistically. *Proc. Interspeech 2013* Lyon, 548–552.

[27] Sanchez, K., Miller, R. M., Rosenblum, L. D. 2010. Visual influences on alignment to voice onset time. *J. Speech Lang. Hear. R.* 53, 262–272.

[28] Shockley, K., Sabadini, L., Fowler, C. A. 2004. Imitation in shadowing words. *Percept. Psychophys.* 66, 422–429.

[29] Šimáčková, Š., Podlipský, V. J., Chládková, K. 2012. Czech spoken in Bohemia and Moravia. *J. Int. Phon. Assoc.* 42, 225–232.

[30] Simon, E. (2009). Acquiring a new second language contrast: an analysis of the English laryngeal system of native speakers of Dutch. *Second Lang. Res.* 25, 377–408.

[31] Skarnitzl, R. 2011. Znělostní kontrast nejen v češtině [The voicing contrast (not only) in Czech]. Prague: Epocha.

[32] Wilson, C., Davidson, L., Martin, S. 2014. Effects of acoustic–phonetic detail on cross-language speech production. *J. Mem. Lang.* 77, 1–24.

[33] 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, e74746.

[34] Zellou, G., Scarborough, R., Nielsen, K. 2013. Imitability of contextual vowel nasalization and interactions with lexical neighborhood density. *Proc. of Meetings on Acoustics* 19, art. no. 060083.