Imitation interacts with one's second-language phonology but it does not operate cross-linguistically

Václav Jonáš Podlipský [1], *Šárka Šimáčková* [1], *Kateřina Chládková* [2]

[1] Department of English and American Studies, Palacký University, Olomouc, Czech Republic
[2] Amsterdam Center for Language and Communication, University of Amsterdam, The Netherlands
vaclav.j.podlipsky@upol.cz, sarka.simackova@upol.cz, k.chladkova@uva.nl

Abstract

This study explored effects of simultaneous use of late bilinguals' languages on their second-language (L2) pronunciation. We tested (1) if bilinguals effectively inhibit the first language (L1) when simultaneously processing L1 and L2, (2) if bilinguals, like natives, imitate subphonemic variation, (3) if bilinguals' imitation operates cross-linguistically, and (4) if imitation interacts with phonological structure. Sixteen L1-Czech L2-English speakers heard stimuli with two factors manipulated: language (Czech, English) and Voice Onset Time (VOT) in /p, t, k/ (short, long). They subsequently pronounced English /t/- and /d/-initial words. Speakers' VOTs in the Czech-Short-VOT, Czech-Extended-VOT, and English-Reduced-VOT conditions were comparable, but VOTs were more English-like after exposure to English-Long-VOT, which applied to both /t/ and /d/. The conclusions are as follows. (1) Bilinguals' potentially ineffective L1 inhibition did not affect their L2 production, since exposure to Czech did not lead to VOT reduction. (2) Imitation is not limited to native speech, since bilinguals increased their VOTs following exposure to English-Long-VOT. (3) Imitation did not operate cross-linguistically, since bilinguals' English productions following Czech-Short-VOT and Czech-Extended-VOT did not differ. Finally, (4) imitation does interact with phonology, since exposure to English long-VOT /t/ resulted in a reduction in prevoicing of its voiced counterpart, /d/.

Index Terms: phonetic imitation, L1 inhibition, bilinguals

1. Introduction

Bilinguals often find themselves in situations when both their languages are in use. They can listen to one language being spoken and then speak themselves in the other, they can converse with two people, with each in a different language, and so on. Under such circumstances, especially late bilinguals may let first-language (L1) phonetic properties permeate into their second-language (L2) productions to a greater degree than they do when exclusively communicating in the L2. As a consequence, L2 speech may be more L1-accented when the L1 is also in use than when it is not.

One potential cause of the L1 interference in L2 speech production is that L1 sound categories and processes cannot be effectively suppressed or inhibited during L2 productions if the L1 has just been used. While perception studies of code-switches suggest that the 'precursor' language affects the phonetic perception of the 'guest' word (e.g. [1]), the few available production studies found no such effects (e.g. [2], [3]). However, as Bullock [4] points out, this may well have been due to limitations of the design (typically using interlingual homophones) and due to group averages obscuring individual differences. A recent study [5] found that bilinguals more skilled in inhibiting the language they are not just using showed a lower degree of cross-language interference, specifically, more distinct Voice Onset Time (VOT) values in stop consonants in each language.

At the same time, it is well known that, in single-language contexts, speakers tend to adopt phonetic properties of the speech they hear. Interlocutors' accents converge not only over longer time spans [6] but even within one conversation [7]. It is not yet well understood what factors contribute to this effect and how, and to what extent such convergence is automatic. However, it is clear that articulatory imitation occurs even in non-social settings [8] and in the absence of conscious attention [9]. If imitation is automatic and if the languages of a bilingual are interconnected, then imitation may even occur language-independently. That is, a bilingual listening to language A and subsequently speaking in language B may imitate the just-heard features of A in her own productions of B. Cross-dialectal imitation has been attested [10], but as far as we know, no one has explicitly studied cross-language imitation.

To investigate the possibility of cross-language imitation is only one of the objectives of the present study. In an earlier experiment [11] with 22 L1-Czech L2-English participants we found that when speakers read short answers in English to questions in Czech, the VOT of word-initial /t/ got progressively more Czech-like, i.e. shorter, which did not happen when the questions were also in English. However, the methodology did not allow us to decide whether these findings resulted from cross-language imitation or inefficient L1 inhibition. The present study was designed to pit these two factors against each other.

The method of the present study is partially inspired by that of Nielsen's study [8] with several important differences. Nielsen exposed English listeners to recordings of English words with extra-long VOT in initial /p/ and found that VOT lengthening was imitated in subsequent production and generalized to other words and another phoneme, /k/. In the current study, English word-initial /t/s and /d/s were elicited from L1-Czech L2-English speakers under four conditions: after listening to words starting with /p, t, k/ that were (1) Czech and had the natural, short VOT, that were (2) Czech but had artificially extended VOT, that were (3) English and had the natural, long VOT, and that were (4) English but had artificially reduced VOT. For the natural-VOT conditions alone, both inefficient L1 inhibition and cross-language imitation predict VOT shifts in L2-English production towards the values of the exposure language. However, with the edited-VOT conditions included, the predictions differ: *inefficient L1 inhibition* predicts VOT shifts towards Czech-like values in production following exposure to Czech, even Czech with extended VOT, whereas *imitation* predicts shifts towards exposure VOT values, regardless of the exposure language.

We elicited productions of /d/-initial as well as /t/-initial words to test whether any exposure-induced shifts in VOT values of /t/ would be reflected in the VOT values of the other member of the phonological voicing contrast, i.e. /d/. Note that /d/ was not included in the listening materials.

To summarize, our research questions are: (1) Can bilinguals effectively inhibit the L1 during L2 productions when the L1 has just been processed? Specifically, will speakers have more L1-Czech-like VOTs in the English words they produce after exposure to Czech? (2) Do bilinguals, like native speakers, imitate the phonetic properties of recently heard speech? Specifically, will speakers have longer VOTs after exposure to naturally long English VOTs than after exposure to English reduced VOTs? (3) Do bilinguals imitate cross-linguistically? Specifically, will speakers have shorter VOTs in English after exposure to Czech short VOT than after Czech extended VOT? (4) Does imitation interact with phonological structure? Specifically, will any post-exposure shifts in VOT of /t/ be accompanied by equivalent shifts in VOT of its voiced counterpart, /d/?

2. Method

2.1. Choice of stimuli and target words

Perception stimuli and production targets were English and Czech disyllabic words with initial stress. Czech-English cognates and interlingual homophones were excluded. The perception stimuli began in a prevocalic /p/, /t/, or /k/ and contained no other voiceless stops; voiced stops occurred only medially. In each condition listeners heard 12 /p/-initial words, 12 /k/-initial words, 24 /t/-initial words and 24 fillers, all repeated once (i.e. 2 x 72 words). The production targets were 12 new /t/-initial and 12 /d/-initial words complemented by 6 new fillers. All perception and production fillers contained only sonorants, fricatives and an occasional non-initial voiced stop.

The 72 Czech and the 72 English stimuli were the same lexical items for both Czech conditions and both English conditions respectively, and the 30 targets were the same in all four conditions. This was for two reasons. First, this eliminated lexical-frequency differences between words in different conditions. As predicted by exemplar theories and attested empirically (e.g. [12], [13], [8]) imitation is more likely in low-frequency words, of which speakers store fewer exemplars. This is why we also controlled the mean lexical frequency of the /t/- and /d/-initial targets, which was 11,485 and 11,574 per million respectively (std. dev. 6324 and 6814 respectively), as determined by consulting the Corpus of Contemporary American English [14], (the difference was not significant, $t[22] = 0.033$, $p > .97$). Second, we used the same lexical items as stimuli and targets in each condition to avoid differences in the words' segmental make-up, as VOT of initial stops is affected by articulatory and acoustic properties of following segments (e.g. vowel height [15]). To factor out potential effects of repetition then, the order of conditions was fully counterbalanced across participants and both stimuli and targets were randomized within conditions differently for every participant (preserving proportions in each of 6 blocks, see 2.3 below).

2.2. Stimulus preparation

A female speaker of each language was recorded using a Zoom H4n in a sound booth, reading the words in a random order, each 3 times. The second author selected the best-sounding token while inspecting waveforms and spectrograms, especially to avoid multiple bursts in [k]. The chosen tokens became the stimuli for the English natural Long-VOT condition and the Czech natural Short-VOT condition.

In addition, the Czech speaker (also proficient in English) was trained to pronounce the Czech words with aspiration in the initial stops and one of three renditions of each word was again selected. The Czech Extended-VOT stimuli were created by splicing the burst and aspiration from the extra aspirated tokens with the rest of the natural Czech words. The duration of the extended VOT, l (in ms), was determined as

$$l = 1.35s + 25 \qquad (1)$$

where s is the original VOT, which resulted in a VOT increase of about 30 ms on average (see Table 1). The amount of lengthening was empirically determined by identifying the maximum increase that would still go unnoticed by 5 uninformed native listeners. Further, to minimize the salience of the VOT manipulation, we prepared another set of the Czech stimuli with VOT values equidistant on a logarithmic scale between s and l. This set was used in the 'accommodation' block of the Czech Extended-VOT condition, i.e. the first of 6 exposure blocks in that condition. Productions after the first block were not measured.

The English Reduced-VOT stimuli were edited copies of the original English words. They were created in the following way: in every original English word, the stable portion of the aspiration noise was identified (from and to zero-crossings) and its final part was truncated reducing VOT to a third of its original duration (except for a few [p]-initial tokens where truncating the whole ⅔ would have led to unnatural sounding words).

Table 1 shows mean VOT values for the /p/-, /t/-, and /k/-initial words used as stimuli in each condition, as well as means for all stimuli in a given condition.

Table 1. *Mean VOT values of the initial stops in the perception stimuli used in the four conditions (in ms). Standard deviations are in parentheses.*

Initial stop	Condition			
	English Reduced-VOT	English Long-VOT	Czech Short-VOT	Czech Extended-VOT
/p/	26.3 (5.9)	78.4 (18.2)	7.0 (1.5)	34.4 (1.7)
/t/	30.1 (4.1)	90.7 (12.0)	14.1 (4.1)	43.7 (4.9)
/k/	33.9 (5.4)	101.3 (16.1)	23.1 (4.5)	55.9 (6.1)
All stimuli	30.1 (5.5)	90.3 (16.6)	14.6 (6.8)	44.4 (9.0)

2.3. Procedure

Each participant took part in all 4 sessions corresponding to the 4 conditions. Everyone completed two sessions in one day and the other 2 sessions at least 24 hours later (with the exception of one participant who took them only 9 hours later). Consecutive sessions were separated by at least a 5-

minute break. As mentioned above, the order of conditions was fully counterbalanced between participants, providing that a participant never started with the same language on both days and never had the same language twice in one day. Each day started with a recording of 16 words randomly selected from the filler stimuli. This recording was done for 3 reasons. First, to serve as a warm-up reducing initial nervousness and/or hyper-articulation, second, to distract attention from word initial stops, and third, to introduce and practice the format of the production task: a target word always appeared 3 times at 2 s intervals on the screen of a computer, and participants pronounced it 'as naturally as possible' each time it appeared. The timing was exactly the same in all conditions to equalize speech tempo as much as possible, as differences in tempo could affect VOT. Participants were tested individually in a sound booth, using Sennheiser HD 280 Pro headphones for presentation of stimuli, and the Zoom H4n recorder.

Rather than collecting only post-exposure data, we elicited stop-initial words at 6 points during the exposure within each session with the aim to track potential changes in the VOT values over time. That is, each session comprised 6 equally long blocks, in which an exposure to 24 stimuli was followed by a production of 5 target words. The blocks were constructed (uniquely for each participant) as follows: the 72 stimuli prepared for a particular condition were assigned randomly to blocks 1-3, and then again in a different order to blocks 4-6 so that each block contained 8 /t/-, 4 /p/-, 4 /k/-initial words, and 8 fillers in a random order. The stimuli had equalized intensity and were presented with the inter-stimulus interval of 1.5 s. In the production part of each block, participants pronounced (3 times, exactly like in the warm-up recording) 2 /t/-, 2 /d/-initial words, and 1 filler, randomly selected from the pool of targets and randomized within the block. Moreover, in order to ensure activation of the exposure language, the production of each target word was immediately preceded by silent reading in that language. Short lines of about 8 words were displayed for 3.5 s before each target word appeared. They formed a coherent text (always different in each condition) and participants were motivated to pay attention to the text by the necessity to answer content questions at the end of each session.

2.4. Participants

Sixteen female university students, aged 19-25, participated in the study. They began learning English at 8 years of age on average (std. dev. 2.8), the mean time spent in English-speaking countries was 7.2 months, (std. dev. 7.7).

As mentioned above, all took part in all four sessions to allow us to make within-speaker comparisons. This was important because of possible differences in task strategies between the L2 speakers (cf. [4]).

Speakers of the same sex in live interaction have been found to show a greater degree of phonetic convergence than speakers of opposite sex (e.g [7]). Although our task did not involve live interaction, we eliminated sex as a factor eliciting stimuli from two female speakers and recruiting only female participants.

2.5. VOT measurement

All measurements were performed in Praat [16]. As mentioned above, the first production block was never measured. The second of the 3 productions of a given word was measured, with the exception of a small number of cases when the first production was measured instead (mostly because of non-modal phonation). VOT was measured from the moment of burst (in the infrequent case of multiple bursts, the first one was counted) to a zero-crossing nearest to the onset of periodicity, which was determined from the waveform (voice pulses) and spectrogram (voicebar). One person (the first author) measured all tokens adhering to these criteria.

3. Results

Every participant's VOT values were averaged for /t/ and for /d/-initial words separately. For the purposes of statistical analysis, data from blocks 2 and 3, as well as data from blocks 4–6 were pooled, and are referred to below as part 1 and part 2 respectively. Block 1 was not included in part 1, because in the Czech Extended-VOT condition it served as the accommodation block (see 2.2).

We ran a repeated-measures MANOVA on the VOT values from part 1 and part 2, with phoneme (/t/, /d/), exposure language (Czech, English), and exposure VOT-length (short, long) as the within-subject factors. The level 'short' of the factor VOT-length represented the Czech Short-VOT and the English Reduced-VOT conditions, the level 'long' represented the English Long-VOT and the Czech Extended-VOT conditions. The analysis revealed a main effect of phoneme on VOT in both parts of the recording (part 1: $F[1,15] = 79.920$, $p < .001$; part 2: $F[1,15] = 76.293$, $p < .001$). Unsurprisingly, /d/ had a shorter VOT than /t/ in both parts. More importantly, the analysis also revealed a marginally-significant interaction effect of exposure language and exposure VOT-length on VOT in part 1 ($F[1,15] = 4.428$, $p = .053$), and a nearly-significant main effect of exposure language on VOT in part 1 ($F[1,15] = 3.796$, $p = .070$). Table 2 shows the mean VOT values measured in part 1 and part 2 averaged across /t/ and /d/.

Table 2. *VOT values averaged across /t/ and /d/ split by condition and part (in ms). Standard errors are in parentheses.*

	Part 1	Part 2
English Reduced-VOT	20.7 (5.5)	23.9 (5.3)
English Long-VOT	28.0 (6.5)	21.2 (7.2)
Czech Short-VOT	22.6 (6.5)	22.3 (7.0)
Czech Extended-VOT	19.8 (5.5)	22.2 (6.4)

To further inspect the two-way interaction of exposure language and exposure VOT-length in part 1, we carried out four paired-samples *t*-tests, comparing VOT productions after exposures to short-Czech versus long-Czech, short-English versus long-English, short-Czech versus short-English, and long-Czech versus long-English. The *t*-tests revealed a significant difference between short-English and long-English exposure conditions ($t[15] = 2.420$, p [two-tailed] = .029), and between long-Czech and long-English conditions ($t[15] = 2.923$, p [two-tailed] = .010). Figure 1 shows the VOT values elicited after each exposure language and exposure VOT-length in part 1. As can be seen, speakers' VOTs were longer after exposure to naturally long English VOTs than after exposure to English reduced VOTs, and than after exposure to Czech extended VOT.

To determine whether the longer VOT produced in the English Long-VOT condition in part 1 was a lengthening or if the shorter VOTs in the other conditions in part 1 were the result of shortening, we compared part 1 with part 2 within each condition using a paired-samples *t*-test. Only the difference between the two parts of the English Long-VOT condition approached significance (*t*[15] = 1.943, *p* [two-tailed] = .071; for all other conditions *p* > .1). In the English Long-VOT condition speakers tended to have longer VOTs in part 1 than in part 2. This means that the long VOT in the first part of this condition is likely to have been a lengthening of the speakers' interlanguage VOT.

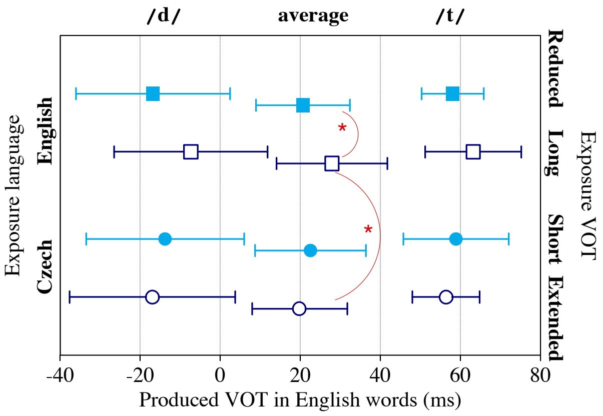

Figure 1: *VOT produced in /t/- and /d/-initial English words in the 1st part of each of the 4 sessions. Mean VOT values for /t/ and /d/ are shown, as well as the average VOT over the two phonemes. Circles represent Czech and squares English exposure. Filled shapes represent short/reduced VOT, while empty shapes represent long/extended exposure VOT. Whiskers show 95% confidence intervals. Asterisks mark significant differences (p < .05) between means connected by arcs.*

4. Discussion

In this study we examined the influence of simultaneous use of the two languages of late-bilingual speakers on their L2 productions. We assessed the effects of two factors: the potentially impoverished inhibition of L1 due to its recent activation, and the potentially cross-linguistic imitation of recently heard speech.

The first research question was whether speakers would produce more L1-like VOTs after exposure to the L1, irrespective of the actual VOTs in the L1 stimuli, which would be indicative of inefficient L1 inhibition. As mentioned above, in our earlier study [11] (different) L1-Czech L2-English speakers did reduce mean VOT of English /t/ when responding to Czech questions. In that study, however, we did not control for lexical frequency of the targets. The present study found no such reduction and, using a different design excluding interlingual homophones, it thus supports the conclusions of production studies of code-switching (e.g. [2], [3]) suggesting that in production, unlike in perception, bilinguals can switch completely between the phonetics of their two languages.

Next, we asked if the non-native speakers of English would imitate VOT shifts in the exposure stimuli (and, if so, in which direction). We found that exposure to English long VOTs lead to lengthened VOTs in new English lexical items produced by the L2 speakers, albeit only in the first part of the session. This result suggests that imitation is not limited to repeated words (cf. [8]), and more importantly, that it is not a process specific to native speakers: also speakers of a second language imitate the phonetic characteristics of recently heard L2 speech. Evidently, the L2 categories, i.e. categories that in the view of exemplar theory [12] are based on fewer exemplars, are not so firmly established and are likely to shift towards exposure values. Moreover, our L2 speakers were like Nielsen's [8] native speakers in that VOT reduction was not imitated.

While Nielsen [8] showed that VOT imitation generalized from /p/ to /k/, we found that the increase of VOT in /t/-initial words in the English Long-VOT condition was accompanied by a decrease of prevoicing in /d/s. This indicates that imitation has an impact on phonological categories.

At the same time, it seems that the phonetic imitation was constrained by the bilinguals' interlanguage phonology. This is because, as just mentioned, our L2 speakers imitated VOT lengthening, shifting /t/ and /d/ towards the target (i.e. English) phonology, but did not imitate VOT shortening which would have shifted /t/ and /d/ away from the acquisition target and towards the L1.

Our experimental design also allowed us to examine the possibility that bilingual speakers imitate even across their two languages. We did not find any evidence of such cross-language imitation, since speakers' English VOTs after exposure to Czech short VOTs and to Czech extended VOTs did not differ significantly. Imitation thus seems to be constrained by language-specific linguistic structure.

5. Conclusions

First, as in previous studies, we did not find any influence of recent L1 activation on L2 production. Second, we found that phonetic imitation is not exclusive to monolinguals but it occurs in L2 speakers as well. Next, from our finding that post-exposure shifts in /t/ were paralleled by shifts in /d/, we concluded that imitation interacts with phonological structure. Finally, since imitation did not operate cross-linguistically in bilinguals, we concluded that the interaction between imitation and phonology is phonology-specific, in the sense that exposure modulates subsequent production only within the same language.

6. Acknowledgements

The work of the first and the second author was supported by the ESF project CZ.1.07/2.3.00/20.0061, the work of the third author by the Netherlands Organization for Scientific Research (NWO) grant 277.70.008 awarded to Paul Boersma (University of Amsterdam). We are grateful to Kateřina Braunová for help with the data collection.

7. References

[1] Elman, J. L., Diehl, R. L. and Buchwald, S. E., "Perceptual switching in bilinguals", J. Acoust. Soc. Am., 62:971-974, 1977.

[2] Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B. and Carbone, E., "The acquisition of a new phonological contrast: The case of

stop consonants in French-English bilinguals", J. Acoust. Soc. Am., 54:421-428, 1973.

[3] Grosjean, F. and Miller, J. L., "Going in and out of languages: An example of bilingual flexibility", Psychol. Sci., 5(4):201-206, 1994.

[4] Bullock, B. E., "Phonetic reflexes of code-switching", in B. E. Bullock and A. J. Toribio [Eds], The Cambridge Handbook of Linguistic Code-switching, 163-181, Cambridge University Press, 2009.

[5] Lev-Ari, S. and Peperkamp, S., "Low inhibitory skill leads to non-native perception and production in bilinguals' native language", submitted.

[6] Pardo, J., Gibbons, R., Suppes, A., and Krauss, J., "Phonetic convergence in college roommates", J Phonetics, 40:190-197, 2012.

[7] Pardo, J., "On phonetic convergence during conversational interaction", J. Acoust. Soc. Am., 119:2382-2393, 2006.

[8] Nielsen, K., "Specificity and abstractness of VOT imitation", J Phonetics, 39:132-142, 2011.

[9] Shockley, K., Richardson, D., and Dale, R., "Conversation and coordinative structures", Topics in Cognitive Science, 1:305-319, 2009.

[10] Delvaux, V., and Soquet, A., "The influence of ambient speech on adult speech productions through unintentional imitation", Phonetica, 64:145-173, 2007.

[11] Podlipský, V. J. and Šimáčková, Š., "Automatic imitation or poor first-language inhibition? Why foreign-accentedness increases during interpreting", in J. Zehnalová, O. Molnár and M. Kubánek [Eds], Trends and Tradition in Trans-Language Communication, Palacký University, in press.

[12] Goldinger, S. D., "Echoes of echoes? An episodic theory of lexical access", Psychol Rev, 105:251-279, 1998.

[13] Goldinger, S. D., and Azuma, T., "Episodic memory reflected in printed word naming", Psychon. B. Rev., 11:716-722, 2004.

[14] Davies, M., "The Corpus of Contemporary American English: 450 million words, 1990-present", 2008-2013. Online: http://corpus.byu.edu/coca/, accessed in Jan 2013.

[15] Klatt, D. H., "Voice onset time, frication, and aspiration in word-initial consonant clusters", J. Speech Lang. Hear. R., 18(4): 686-706, 1975.

[16] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer [Computer program]", Version 5.3.34, 2012. Online: http://www.praat.org/, retrieved on 21 Nov 2012.