

Phonetic drift reveals interconnected phonological representations in simultaneous bilinguals: A case study of English and Czech stop consonants

International Journal of Bilingualism

1–11

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1367006920932211

journals.sagepub.com/home/ijb**Václav Jonáš Podlipský** 

Palacký University Olomouc, Czech Republic

Šárka Šimáčková

Palacký University Olomouc, Czech Republic

Kateřina Chládková

Czech Academy of Sciences, Czech Republic Charles University, Czech Republic

Abstract

Purpose: The interconnectedness of phonological categories between the two languages of early bilinguals has previously been explored using single-probe speech production and perception data. Our goal was to tap into bilingual phonological representations in another way, namely via monitoring instances of phonetic drift due to changes in language exposure.

Design: We report a case study of two teenage English–Czech simultaneous bilinguals who live in Canada and spend summers in Czechia (Czech Republic). Voice onset time (VOT) of word-initial voiced and voiceless stops was measured upon the bilinguals' arrival to and before their departure from a two-month stay in Czechia.

Data and Analysis: Each bilingual read the same set of 71 Czech and 58 English stop-initial target words (and additional fillers) at each time of measurement. The measured VOT values were submitted to linear mixed effects models, assessing the effects of target language, measurement time, and underlying voicing.

Findings/Conclusions: After the immersion in a Czech-speaking environment, for both speakers the count of voiced stops realized as prevoiced (i.e., having negative VOT) increased and the measured VOT of voiced stops (appearing different for English and Czech initially) drifted towards more negative (more Czech-like) values in both languages, while no change was detected for the voiceless stops of either English (aspirated) or Czech (unaspirated). The results suggest that the bilinguals maintain three-way VOT distinctions, differentiating voiceless aspirated (English), voiceless unaspirated (Czech), and voiced (English–Czech) stops, with connected bilingual representations of the voiced categories.

Corresponding author:

Václav Jonáš Podlipský, Department of English and American Studies, Palacký University Olomouc, Křížkovského 10, Olomouc, 77180, Czech Republic.

Email: vaclav.j.podlipsky@upol.cz

Originality: Data on phonetic drift in simultaneous bilinguals proficient in their two languages have not previously been published.

Significance/Implications: We show that observing phonetic shifts due to changes in the ambient linguistic environment can be revealing about the organization of phonological space in simultaneous bilinguals.

Keywords

Speech production, phonological representation, phonetic drift, English–Czech simultaneous bilinguals, voice onset time

Introduction

The linguistic competence of a bilingual speaker is not the sum of two separate monolingual competences (Grosjean, 1989). Instead, it forms a unified system (Cook, 1992). This is seen in cross-language interactions observed both on the level of linguistic representations (e.g. Barlow et al., 2013; Flege & Eefting, 1987; Sebastián-Gallés et al., 2005), and on the level of online speech processing (e.g., Goldrick et al., 2014; Jacobs et al., 2016; Olson, 2013). Research has documented cross-language influences indicating that a bilingual's two languages are not isolated from each other even for early sequential and/or simultaneous bilinguals. Even though such bilinguals often display a differentiation between speech sounds of one language and similar sounds of their other language (Barlow et al., 2013; Simonet, 2010; Yusa et al., 2010), they still differ in phonetic realization from monolingual speakers of each language (e.g., Flege, 1987; Fowler et al., 2008; Hazan & Boulakia, 1993).

Reports of a null difference between bilingual and monolingual performance are less common. For example, Antoniou et al. (2010) found that early Greek–Australian-English bilinguals could produce voice onset times (VOTs) of stop consonants, an acoustic correlate of voicing, indistinguishable from the Greek monolinguals or from the Australian-English monolinguals if the language context during data elicitation was kept strictly monolingual. However, even when a study finds no difference between bilingual and monolingual performance, that does not guarantee that these bilinguals have independent underlying phonological representations for each language. Sundara and Polka (2008) argue that evidence of a bilingual differentiating similar sounds in production must be matched by evidence from perception. They compared perceptual discrimination of the (French) dental and (English) alveolar voiced stops by Canadian French–English simultaneous bilinguals and early sequential bilinguals, monolingual speakers of each language, and native speakers of Hindi (where dental and retroflex stops contrast). They found the early sequential bilinguals' performance to be based on a merged category. The simultaneous bilinguals appeared to have a separate stop category for each language, even in perception. Interestingly, their performance was similar to that of the Hindi speakers but differed from that of the English monolinguals.

Another way of tapping into phonological representations is via monitoring phonetic drift, that is, shifts that occur in response to a change of the ambient linguistic environment (see Chang (2019), for a review of the phenomenon). Phonetic adjustments induced by recent input have been documented for monolinguals (as shown by studies of perceptual adaptation (Cutler, 2012) and phonetic convergence (Pardo, 2013), for late non-balanced bilinguals in their second language (L2) (Tobin et al., 2017) or both in L2 and their first language (L1) (Sancier & Fowler, 1997), and in L1 even for inexperienced L2 learners exposed to the L2 for a relatively short time (Chang, 2012, 2013). Data revealing phonetic drift in simultaneous or early sequential bilinguals proficient in

their two languages are not available to date. The few longitudinal case studies monitoring L2 and L1 development in early L2 learners which attest phonetic shifts due to exposure are studies only of the early stages of L2 acquisition (the first seven months for Simon's (2010) learner and the first 20 months for the learner in Yang et al. (2015)). The current study measures phonetic drift in teenage simultaneous bilinguals to investigate their representations of similar sounds. We examine the VOT of stop consonants in English and Czech, languages with different phonetic implementation of the voiced and voiceless phonemes (short vs long positive VOT in English, and negative vs short positive VOT in Czech¹). Although studies with children simultaneously acquiring two languages with a difference in the phonetic implementation of stop voicing similar to that between English and Czech report separate phonetic categories for phonologically voiceless but not for voiced stops (e.g., Johnson & Wilson, 2002), by adulthood, bilinguals seem to produce language-specific monolingual-like voiced stops in their two languages (Sundara et al., 2006). However, as suggested above, having non-identical realizations for equivalent L1 and L2 sounds does not mean that these sounds are equally well separated at a higher level of representation. An L1–L2 link at a phonemic level of representation of /b, d, g/ (and possibly /p, t, k/) may be revealed by an L1–L2 concurrent phonetic drifting in response to a recent change in the input, such as traveling to another country for summer holidays. A change in the linguistic environment inducing a shift in the phonetic implementation of stop voicing in both languages (in the direction of the current input language) can be interpreted as evidence for a connected phonological representation of such sounds.

We measured the VOT of word-initial stops of two teenage Canadian English–Czech simultaneous bilinguals (sisters) upon their arrival to and then before their departure from a two-month stay in Czechia (Czech Republic). Our research questions were: (a) What VOTs do our simultaneous (non-balanced) bilinguals produce for phonologically voiced and voiceless stops in each language? (b) Do any of their VOT values change in response to the change of the ambient linguistic environment? and (c) If so, does the change occur only in the language of the current environment or in both languages, reflecting a connection of phonological representations between the bilinguals' languages?

Method

Participants

The speakers volunteered to participate. They were two female simultaneous English–Czech bilinguals dominant in English, two sisters labeled herewith as A and B—A aged 13 and B aged 16 at the time of the recording. They were born and live in Toronto, Canada, with their Canadian father and their Czech mother. Before entering elementary education at the age of 6, each sister had had extensive exposure to Czech from their mother at home, their mother's parents during regular 3- to 4-month visits to Czechia every summer, as well as during the grand-parents' three-week visits to Canada every winter. After the age of 6, they continued speaking Czech primarily with their mother, and also their grand-parents whom they saw in Canada every winter (about three weeks) and in Czechia every summer (two months).

Stimuli

The material consisted of a list of 71 monosyllabic or disyllabic (stress-initial) Czech and 58 monosyllabic English stop-initial words, and 46 Czech and 15 English filler words. The initial stops had different places of articulation (see Table 1 for the exact counts) and were followed by vowels of differing heights (see Table 2).

Table 1. Numbers of stop-initial stimulus words split by places of articulation.

	Czech	English
Labial	37	28
Coronal	24	22
Dorsal	10	8
Total	71	58

Table 2. Numbers of stop-initial stimulus words split by the height of the first vowel.

	Czech	English
High	25	24
Mid	23	10
Low	23	24
Total	71	58

Procedure

Upon arrival to Czechia, the speakers were told that the purpose of the recordings was to monitor changes of their pronunciation over the course of their stay in the Czech-speaking environment. No other details were provided. First, the participants were implicitly familiarized with the stimulus words, hearing them in conversation with the data collector (their Czech cousin) one day prior to the first recording session, which took place two days after their arrival to Czechia.

In the first session itself, the speakers were presented with each of the stimulus words once, on the screen of a computer, one by one in random order but in two blocks split by language, at a fixed interval of 2 seconds for the English and 5 seconds for the Czech words, and they read them out loud, in isolation without any carrier phrase. An immediate repetition was elicited by the data collector in case of unclear pronunciations or mistakes. A Zoom H4n portable recorder was used for recording the speech at 16-bit and 44.1 kHz without compression.

The different fixed trial duration for English versus Czech was motivated by allowing greater time for lexical access in the speakers’ non-dominant language, Czech. Because of this difference, comparisons of measured VOT for Czech versus English have to be made with caution, as the longer trial duration could potentially encourage slower speech tempo and thus lengthen VOT in Czech relative to English. However, the main comparison is between recording sessions within each language and importantly, the same procedure and stimulus words were used again in the second recording session at the end of the speakers’ two-month stay in Czechia (59 days after session 1). The comparability of the two recording sessions within each language allowed for a reliable assessment of the effect of the immersion stay on VOT production in each language.

Measurements and analysis

After excluding noisy or erroneous recordings (eight in total), but including occasional repeated productions of a word, 267 and 268 initial stops could be analyzed for speakers A and B, respectively. Praat (Boersma & Weenink, 2019) was used to manually label their VOT, that is, time from the stop release to the first zero crossing of periodicity detectable in the waveform and broadband spectrogram that either followed (positive VOT) or preceded (negative VOT) the release. The labeling was performed by the data collector and corrected by the first author manually.

Per speaker, the measured VOT values were submitted to a linear mixed effects model, *lme4* and *lmerTest* packages (Bates et al., 2015; Kuznetsova et al., 2017), in R (R Core Team, 2019) with Language, Underlying voicing, and Session as fixed factors with orthogonal sum-to-zero contrasts (Czech -0.5 vs English +0.5, voiced -0.5 vs voiceless +0.5; 1st session -0.5 vs 2nd session +0.5), and Word as a random factor. Comparison of estimated means across factor levels was carried out with the *emmeans* package (Lenth et al., 2018).

Results

Analyzing first the number of occurrences of phonologically voiced stops realized as prevoiced (i.e., having negative VOT), we found that for both speakers and both languages the incidence of prevoiced stops increased between sessions: for speaker B for Czech, the number of prevoicings increased from nine out of 28 stops measured in session 1 to 15 out of 28 stops in session 2 ($\chi^2(1) = 2.63, p = 0.1052$), and for English from four out of 32 to 13 out of 33 ($\chi^2(1) = 6.08, p = 0.0136$). Similarly for speaker A for Czech, the number of prevoicings rose from 12 out of 29 stops in session 1 to 18 out of 28 ($\chi^2(1) = 3.00, p = 0.0834$), and for English from eight out of 31 to 25 out of 32 ($\chi^2(1) = 17.28, p < 0.0001$).

The measured VOT values for each speaker are shown in Figure 1. For both speakers, the linear mixed effects models detected the following significant effects: Intercept, Language, Underlying voicing, Session, Language * Underlying voicing, Underlying voicing * Session; Table 3 gives their estimated effect sizes, *t*-values and *p*-values. Unsurprisingly, the main effects of Language and Underlying voicing show that stops have longer VOT in English than in Czech, and that voiceless stops have longer VOT than voiced ones. The main effect of Session showed that overall VOT was shorter in session 2 than in session 1.

As for the predictors addressing our research questions, the interaction effects of Language and Underlying voicing and the comparison of estimated means listed in Table 4 show that, for both speakers, voiceless stops were produced with different VOT in Czech versus in English, while for voiced stops any between-language difference was smaller or non-existent. The interaction of Underlying voicing and Session, with the estimated means again in Table 4, demonstrates that VOTs differed between session 1 and session 2 for voiced stops for both speakers but to a smaller extent, or not at all, for voiceless ones.

Although the three-way interaction of Language, Session and Underlying Voicing was not significant for either speaker, comparison of estimated means shows that in session 1, the difference in VOT between Czech and English voiced stops was significant for speaker B, and was of a comparable magnitude for speaker A, though not reaching significance (mean Czech minus English difference for speaker B = -18 milliseconds, $t = -2.210, p = 0.028$; for speaker A = -16 milliseconds, $t = -1.482, p = 0.140$), while in session 2 the effect was numerically smaller and no longer significant for speaker B and was even in the opposite direction for speaker A (mean Czech minus English difference for speaker B = -13 milliseconds, $t = -1.613, p = 0.108$; speaker A = +7 milliseconds, $t = 0.664, p = 0.507$). Besides suggesting a somewhat clearer differentiation of the Czech and English voiced stops for speaker B than for speaker A, at least at session 1, these comparisons, as well as visual inspection of the density plots in Figure 1, also show that while it is possible that before the immersion period (session 1) the Czech versus English voiced stops had different VOTs, after the immersion (session 2) a difference between Czech and English voiced stops in VOT was less likely to exist.

Further, a comparison of means shows that for the voiced stops the difference between sessions 1 and 2 was larger for English than for Czech: for speaker A, 63 milliseconds for English ($t = 6.773, p < 0.0001$) vs 40 milliseconds for Czech ($t = 4.138, p < 0.0001$) and for speaker B, 35 milliseconds for English ($t = 4.736, p < 0.0001$) vs 30 milliseconds for Czech ($t = 3.754,$

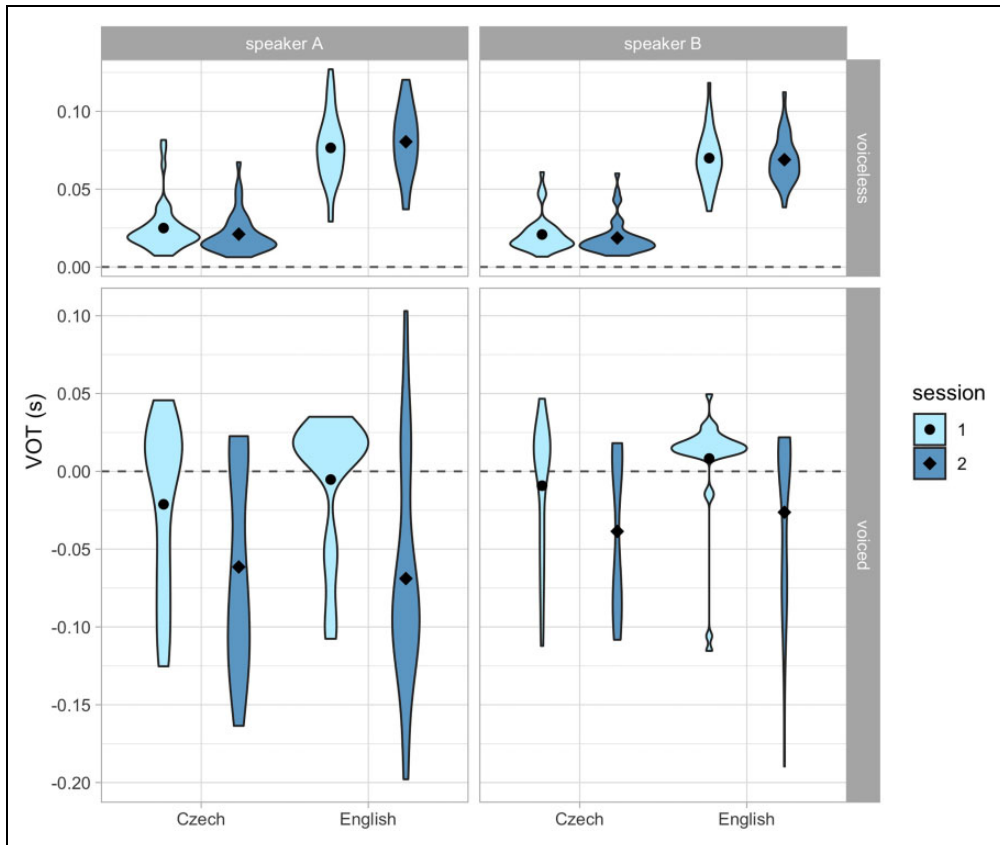


Figure 1. Violin plots of the measured voice onset time (VOT) values (in seconds) of word-initial voiceless and voiced stops per speaker, language, and session. The colored shapes represent density plots and are trimmed at the edges to the range of the data. Filled symbols show means. (Color online.).

$p = 0.0002$), also cf. Figure 1. No such language-specific between-session differences were detected for the underlyingly voiceless plosives in either speaker (all $|t|$ scores < 0.51 , all p values > 0.61).

Finally, Figure 1 suggests somewhat greater cross-linguistic similarity of VOT values for voiced stops for speaker A than for speaker B, especially for session 2. The significant two-way interaction of Underlying voicing and Language suggests that there are language-specific effects on VOT of voiced versus voiceless stops. To inspect those, we carried out pairwise comparisons of the estimated means for each Underlying voicing and Language. The comparisons show that speaker B reliably distinguished the Czech versus English voiced stops (difference = 16 milliseconds, $t = -2.524$, $p = 0.0128$), while speaker A did so to a smaller extent, if at all (difference = 5 milliseconds, $t = -0.526$, $p = 0.600$).

Discussion

Several observations on the results of this study can be made. First, our two speakers produced language-specific VOT values of word-initial stop consonants, as did simultaneous bilinguals in previous studies (Fowler et al., 2008; Simon, 2010; Sundara et al., 2006; Yusa et al., 2010). Both

Table 3. Modeled effects on voice onset time for each speaker. Estimated values are in seconds.

Speaker	Effect	Estimate	t	p
A	Intercept	0.006	2.002	0.048
	Language	0.030	5.180	<0.001
	Underlying voicing (Und. voicing)	0.089	15.616	<0.001
	Session	-0.025	-5.656	<0.001
	Language * Und. voicing	0.050	4.453	<0.001
	Und. voicing * Session	0.051	5.666	<0.001
	Language * Session	-0.007	-0.803	0.423
	Language * Und. voicing * Session	0.031	1.729	0.086
B	Intercept	0.014	6.617	<0.001
	Language	0.033	7.696	<0.001
	Underlying voicing	0.061	14.155	<0.001
	Session	-0.017	-4.612	<0.001
	Language * Und. voicing	0.034	3.995	<0.001
	Und. voicing * Session	0.031	4.193	<0.001
	Language * Session	-0.002	-0.273	0.785
	Language * Und. voicing * Session	0.006	0.413	0.680

Table 4. Estimated means and t-statistics for the two-way interactions.

Speaker	Underlying voicing	Language/Session	Mean	95% confidence interval	t-ratio	p
A	Voiceless	Czech	0.023	0.013 ... 0.033	-7.103	<0.0001
		English	0.077	0.066 ... 0.089		
	Voiced	Czech	-0.041	-0.053 ... -0.029	-0.526	0.6000
		English	-0.036	-0.048 ... -0.025		
B	Voiceless	Czech	0.020	0.012 ... 0.027	-8.596	<0.0001
		English	0.070	0.061 ... 0.078		
	Voiced	Czech	-0.024	-0.033 ... -0.015	-2.524	0.0128
		English	-0.008	-0.017 ... -0.0003		
A	Voiceless	Session 1	0.050	0.041 ... 0.060	-0.031	0.9754
		Session 2	0.050	0.040 ... 0.060		
	Voiced	Session 1	-0.013	-0.024 ... 0.002	7.679	<0.0001
		Session 2	-0.064	-0.075 ... -0.054		
B	Voiceless	Session 1	0.045	0.038 ... 0.053	0.310	0.7567
		Session 2	0.044	0.036 ... 0.051		
	Voiced	Session 1	-0.00006	-0.008 ... 0.008	5.972	<0.0001
		Session 2	-0.032	-0.041 ... -0.024		

speakers pronounced phonologically voiceless stops with longer VOT in English than in Czech at both times of measurement (despite the shorter duration of English than of Czech trials, which may potentially have led to a higher speech tempo, and thus a shortening of VOT, in English overall). The production of word-initial voiced stops was more variable in two senses. First, both speakers realized some voiced stops in both languages with positive and others with negative VOT (as prevoiced). Second, unlike for voiceless stops, the production of voiced stops underwent a change after the immersion period: prevoicing became more frequent and the VOT values became more negative. In other words, phonetic drift was observed for our simultaneous bilinguals, corroborating previous findings of drifting in bilinguals of other types (Sancier & Fowler, 1997; Tobin et al.,

2017) and in second-language learners (Chang, 2012, 2013). Importantly, for both speakers the production of voiced stops changed towards the more Czech-like values not only for Czech but also for English. In fact, in terms of the number of stops realized as prevoiced, and especially for the younger speaker A also in terms of VOT, the drifting had greater magnitude for English than for Czech.

Since the surface phonetic realizations of English versus Czech voiced stops, which initially appeared to be somewhat different (as they did in a previous single-probe case study of an early English–French bilingual child by Mack, 1990), drifted together (and their initial potential difference largely dropped away), we conclude our bilinguals have English–Czech voiced stop categories integrated at a more abstract level of sound representation. Such integration may be caused by a cross-language “equivalence classification” (Flege, 1995) of the corresponding sounds (Chang, 2019, pp. 192–193). Furthermore, the drifting we observed provides insight into underlying representation in another way. Although at the time of the first recording the VOT values of English phonologically voiced stops and of Czech phonologically voiceless (unaspirated) stops had largely overlapping distributions if outliers are excluded (see Figure 1), it is clear that they represent separate categories because drifting affected only the English voiced stops.

While Tobin et al. (2017) found variation in the extent of phonetic drifting for a heterogeneous group of late bilinguals, our two speakers displayed considerable similarity in their productions both before and after the immersion-induced drifting. This is perhaps expectable since they are sisters and thus have received similar input in both languages. One indication of a difference between our speakers was in that the 16-year-old (speaker B) seemed to maintain a somewhat clearer differentiation of the Czech and English voiced stops than her 13-year old sister (in other words that for speaker A the phonetic drift in English voiced stops was somewhat larger in magnitude than for her older sister). This is in line with suggestions in the previous literature (Sundara et al., 2006) that a differentiation of voiced stops in bilinguals whose languages have different VOT settings for stops develops slowly.

The current study was designed so that the main comparison, assessing whether phonetic drift occurred, is between the two recording times within each language. Comparisons between English and Czech VOTs are possible (though not necessary for our main research questions) but only with caution. The reason is twofold. First, as stated above, there was a difference between the two language conditions in trial length, Czech trials lasting longer than English (to provide the speakers with more time to access the lexical items in their non-dominant language). The longer duration of Czech trials could potentially have encouraged slower speech tempo, and thus a general lengthening of the VOT. For voiceless stops that would mean an increased similarity between Czech and English VOT and for voiced stops an increased difference. Even though the differences between English and Czech VOTs are in the expected direction suggesting that no large systematic changes were induced by the differing trial duration, it is still possible that the VOTs measured for English versus Czech would differ somewhat if trial duration had been constant. Second, the counts of the stops with different places of articulation measured (given above in Table 1) were only roughly comparable between the two languages and since VOTs differ universally across places of articulation (Cho & Ladefoged, 1999), this could potentially have introduced some difference between the values we measured for each language. Relatedly, future research, carefully controlling for place of articulation and having a larger data set than the one we could obtain from our speakers, could assess the amount of drift across different places of articulation, as it may not necessarily be the same (as it was not for the novice late second-language learners in Chang, 2012). In addition, future research may determine whether comparable phonetic drift occurs for stops in other positions (word-medial and word-final).

This being a case study, a further limitation is that the number of participants recruited was very small. However, also previous longitudinal studies of early bilinguals have been case studies

(Simon's (2010) and Yang et al.'s (2015) studies both have a single participant). This is because there is a relative scarcity of early bilinguals available for longitudinal investigations. It is possible that the patterns of phonetic drift observed in our two speakers (who are siblings) are typical of simultaneous bilinguals. However, further research with a more representative sample size is necessary to determine with confidence whether this is actually the case or not.

In summary, despite some limitations, our data from two teenage simultaneous English–Czech bilinguals provide evidence of a phonetic drift due to a change in the ambient linguistic environment. We show that phonetic drift can reveal cross-language connections (cf. the English and Czech voiced stops in our study), as well as separations for sounds with similar phonetic realizations in both languages (cf. the English voiced and Czech voiceless stops in session 1 in our study). Therefore, observing phonetic drift can be informative about the organization of phonological space in simultaneous bilinguals.

Acknowledgments

We are grateful to two anonymous reviewers for valuable comments on an earlier version of this paper. We thank David Ryška for technical assistance.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Czech Science Foundation (GACR) (Grant Number: 18-01799 S).

ORCID iD

Václav Jonáš Podlipský  <https://orcid.org/0000-0002-2173-4375>

Note

1. To our knowledge no published data are available for the voice onset time (VOT) of word-initial Czech stops. The existing research by Machač (2006) focused on stops in medial positions where the VOT of voiceless but not of voiced stops can be determined. For the voiceless stops, Machač (2006) provides the following values: 0.017 seconds for /p/; 0.020 seconds for /t/; and 0.032 seconds for /k/. For the voiced stops, our own unpublished measurements of two speakers, pronouncing 53 Czech isolated voiced-stop-initial words each, show that one speaker (male) always realized the voiced stops with prevoicing, with the average VOT of -0.098 seconds, while the other speaker (female) prevoiced the stops 99.6% of the time, with prevoicing lasting -0.076 seconds on average.

References

- Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, 38(4), 640–653.
- Barlow, J. A., Branson, P. E., & Nip, I. S. (2013). Phonetic equivalence in the acquisition of /l/ by Spanish–English bilingual children. *Bilingualism: Language and Cognition*, 16(1), 68–85.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, 67(1), 1–48.

- Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.04. www.praat.org/
- Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40(2), 249–268.
- Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, 41(6), 520–533.
- Chang, C.B. (2019). Phonetic drift. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 191–203). Oxford University Press.
- Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27(2), 207–229.
- Cook, V. J. (1992). Evidence for multicompetence. *Language Learning*, 42(4), 557–591.
- Cutler, A. (2012). Native listening: The flexibility dimension. *Dutch Journal of Applied Linguistics*, 1(2), 169–187.
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15(1), 47–65.
- Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 229–273). York Press.
- Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, 15(1), 67–83.
- Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Hallé, P. (2008). Cross language phonetic influences on the speech of French–English bilinguals. *Journal of Phonetics*, 36(4), 649–663.
- Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychological Science*, 25(4), 1031–1036.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36(1), 3–15.
- Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French–English bilinguals. *Language and Speech*, 36(1), 17–38.
- Jacobs, A., Fricke, M., & Kroll, J. F. (2016). Cross-language activation begins during speech planning and extends into second language speech. *Language Learning*, 66(2), 324–353.
- Johnson, C. E., & Wilson, I. L. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *International Journal of Bilingualism*, 6(3), 271–289.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). R package *emmeans*. Version 1.2.4. Retrieved from <https://cran.r-project.org/web/packages/emmeans/>
- Machač, P. (2006). Temporální a spektrální struktura českých explozív [Temporal and spectral structure of Czech plosives]. Ph.D dissertation, Charles University, Faculty of Arts. <https://dspace.cuni.cz/handle/20.500.11956/7718> [In Czech.]
- Mack, M. (1990). Phonetic transfer in a French–English bilingual child. In P. H. Nelde (Ed.), *Language attitudes and language conflict* (pp. 107–124). Bonn: Dümmler.
- Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, 41(6), 407–420.
- Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, 4, 559.
- R Core Team. (2019). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2016. <http://www.r-project.org/>
- Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4), 421–436.

- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, 52(2), 240–255.
- Simon, E. (2010). Child L2 development: A longitudinal case study on Voice Onset Times in word-initial stops. *Journal of Child Language*, 37(1), 159–173.
- Simonet, M. (2010). Dark and clear laterals in Catalan and Spanish: Interaction of phonetic categories in early bilinguals. *Journal of Phonetics*, 38(4), 663–678.
- Sundara, M., & Polka, L. (2008). Discrimination of coronal stops by bilingual adults: The timing and nature of language interaction. *Cognition*, 106(1), 234–258.
- Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, 9(1), 97–114.
- Tobin, S. J., Nam, H., & Fowler, C. A. (2017). Phonetic drift in Spanish–English bilinguals: Experiment and a self-organizing model. *Journal of Phonetics*, 65, 45–59.
- Yang, J., Fox, R. A., & Jacewicz, E. (2015). Vowel development in an emergent Mandarin–English bilingual child: A longitudinal study. *Journal of Child Language*, 42(5), 1125–1145.
- Yusa, N., Nasukawa, K., Koizumi, M., Kim, J., Kimura, N., & Emura, K. (2010). Unexpected effects of the second language on the first. In K. Dziubalska-Kolaczyk, M. Wrembel, & M. Kul (Eds.), *New sounds 2010: Proceedings of the 6th international symposium on the acquisition of second language speech* (pp. 580–584). Adam Mickiewicz University.

Author biographies

Václav Jonáš Podlipský obtained his PhD at Palacký University Olomouc, where he is currently a lecturer in phonetics and English linguistics. He has published research findings on the phonetics of Czech, on L2 speech perception and production, and on the adaptability of speech-sound categories.

Šárka Šimáčková, MA, University of London; PhD, University of Hawaii, is a lecturer in English phonetics and second language acquisition at Palacký University Olomouc. Her research focuses on L2 speech production and perception and cross-language influences in bilingual speech.

Kateřina Chládková, PhD, University of Amsterdam, is a researcher affiliated to the Czech Academy of Sciences and Charles University in Prague. She is the head of the SPEAKin lab, a research group studying speech learning and adaptation, speech perception and production, and their dynamics across the lifespan.