LANGUAGE PROFICIENCY AND SONORANT DEVOICING IN ENGLISH PLOSIVE-SONORANT CLUSTERS

KLÁRA PŘEČKOVÁ
Department of English Language and ELT Methodology,
Charles University, Prague, Czech Republic
klara.preckova@seznam.cz

PAVEL ŠTURM
Institute of Phonetics, Charles University, Prague, Czech Republic
pavel.sturm@ff.cuni.cz

JAN VOLÍN
Institute of Phonetics, Charles University, Prague, Czech Republic
jan.volin@ff.cuni.cz

Abstract
This article explores the timing of phonetic voicing in plosive-sonorant clusters in English. VOT was measured in two groups of Czech learners with different proficiency levels and a native English control group. The hypothesis was that cross-language differences in the implementation of the voicing contrast would be reflected in lower devoicing by the non-native speakers, modulated by proficiency. 24 participants read a text with plosive-sonorant clusters (such as in *plan* or *troops*). The study found that less proficient speakers exhibited smaller degrees of devoicing compared to more proficient speakers, who however did not differ from the native controls. In line with the absence of devoicing in Czech secondary-school textbooks, the results provide insight into the interplay between language proficiency and pronunciation details in L2 acquisition.

Keywords: voicing, devoicing, aspiration, sonorants, English, Czech, foreign language

1. Introduction

Speech sounds belong to two broad types, voiced and voiceless, depending on whether or not the vocal folds are vibrating during their production. *Phonetic devoicing* occurs when a speech sound that would typically be considered voiced in the language system – such as a /b/, /l/, or a vowel – partially or completely loses its phonetic voicing, and the vocal folds cease to vibrate for some portion of the sound's duration. Many languages exhibit phonetic devoicing in specific positions within a word. In English, extensive research has been conducted on the timing of voicing in obstruents before vowels (e.g., Lisker & Abramson, 1964; Docherty, 1992; Cho & Ladefoged, 1999; Hoole, 1999; Abramson & Whalen, 2017; Chodroff & Baese-Berk, 2019). In stressed syllable onsets, phonologically voiceless plosives are aspirated (*pan* [pʰæn]) while phonologically voiced obstruents are fully or partially devoiced (*ban* [b̥æn]). The devoicing takes place also word-finally (*David* [d̥e͡ɪvɪd̥]) but it should be noted that this phenomenon is different from the final neutralization of voicing in languages such as Czech, Polish or German. Specifically, while *David* is pronounced in Czech as [davɪt], with full phonetic voicing of the initial consonant and no phonetic voicing in the final consonant where the phonological voicing distinction is neutralized entirely, other phonetic correlates of the word-final voicing contrast like vowel duration are preserved in English.

However, what is less commonly noted than devoicing in obstruents is the delayed onset of voicing in a sonorant consonant following a voiceless plosive. In fact, the devoiced [l̥] or [ɹ̥] in words such as *plan* or *troops* is a consequence of aspirating the preceding [pʰ] or [tʰ], respectively, which would in other words (e.g., *pan* or *tool*) be released into a vowel. Empirical data are provided for instance by Tsuchida, Cohn and Kumada (2000), who confirmed that sonorants are strongly devoiced after (aspirated) plosives but not after (unaspirated) fricatives (*plea* vs. *flee*).

Although we often use, due to tradition and also for practical reasons, different terms and transcriptions for the two phenomena, ASPIRATION and SONORANT DEVOICING are phonetically a single phenomenon, namely a delay in voice onset time (VOT). The vocal folds are in both cases abducted and the air flows relatively freely through the glottis. However, due to a tighter constriction, devoiced sonorants may involve additional friction apart from the aspiration noise produced in the vicinity of the glottis. Since vowels are relatively long and sonorants relatively short, only part of the sound is devoiced in vowels (we mark it as aspiration of the plosive), while a large portion, if not the whole sound, is devoiced in sonorants (we conventionally mark it as sonorant devoicing).

This phenomenon is well-documented in academic textbooks or other specialized sources (Docherty, 1992; Hoole, 1999; Volín, 2002; Laver, 2012; Cruttenden, 2014; Ogden, 2017), but is often missing in language teaching, secondary-school textbooks, and L2 English research. For instance, Czech materials provided for primary and secondary school English lessons cover sonorant devoicing inadequately or not at all (Přečková, 2022).

The fact that only some languages exhibit phonetic devoicing (Lisker & Abramson, 1964) is important to consider in the context of cross-linguistic differences in L2 English pronunciation. For instance, Czech, unlike English, retains phonetic voicing in all positions for phonologically voiced obstruents (unless assimilation or word-final neutralization of voicing occurs). Similarly, sonorants in Czech are always voiced, even after voiceless plosives, which is different from English, as mentioned above (thus English *play* [pl̥e͡ɪ] but Czech *plavat* [plavat], 'swim'). These differences in implementation norms suggest that L2 English learners who speak Czech as the mother tongue may produce a lower degree of devoicing compared to native English speakers.

In terms of English plosives in stressed syllable onsets, the main phonetic difference between voiced and voiceless sounds is the variation in VOT (Volín, 2002, p. 71; Laver, 2012, p. 340; Ladefoged & Johnson, 2015, p. 61). The vocal folds are typically not vibrating during the articulatory closure of either type of obstruent; instead, the timing of voice onset in the following vowel or sonorant is what distinguishes /b/ from /p/. Unaspirated voiceless stops [p t k] in Czech or devoiced [b̥ d̥ g̊] in English have a VOT close to zero, indicating that voicing in the following vowel begins shortly after the stop is released. In contrast, aspirated stops [pʰ tʰ kʰ] have a VOT exceeding 30 ms. Once again, when a sonorant consonant and not a vowel follows, this is usually referred to as sonorant devoicing rather than aspiration (compare [pʰæn] and [pl̥æn]). Importantly, failure to produce this feature of voiceless plosives may result in confusion for native English speakers, who may mistake unaspirated /p t k/ for devoiced /b d g/ (Roach, 2009: 28; Cruttenden, 2014: 166). According to Cruttenden (2014, p. 336), sonorant devoicing is equally crucial as initial aspiration for comprehension. Therefore, it is important to pay attention to this pronunciation aspect when learning and using English.

Sonorant devoicing has been observed in several languages. In French or Polish, final clusters typically manifest devoicing, such as in *wiatr* [vjatr̥] 'wind' or *quatre* [katʀ̥] 'four' (Sieczkowska, Möbius & Dogil, 2010), which does not occur in Czech (*Petr* [pɛtr̩]). According to Strycharczuk (2012), the process can also occur in sonorants positioned between a voiceless obstruent and a following phrase boundary, not being restricted to clusters in absolute-final position before a pause. Research on sonorant devoicing in other languages, such as Swedish and Norwegian, is rather scarce, with Hansson (1999) noting the lack of comprehensive studies. Nevertheless, he states that in the two Scandinavian languages devoicing of /r/ is the most common, while devoicing of nasals is the least common (Hansson, 1999, p. 162).

Cho and Ladefoged's (1999) study on VOT variations in obstruents may also shed some light on the universal aspects of the phenomenon. They found that increasing values of VOT were primarily tied to the place of stop articulation moving backwards through the oral cavity, which could have implications for understanding why devoicing is more common in certain sonorant consonants in some languages than in others. Finally, Gonet and Święciński (2012) have shown that stress is another factor to consider, as VOT values tended to be higher in stressed syllables.

Studies on the L2 pronunciation of voiceless plosives have generally found low values of VOT in non-native speakers of English (Mariano, Bassetti, Sokolović-Perović & Cerni, 2018; Chondroff & Berk, 2019; Skarnitzl & Rumlová, 2019; Saud Alharbi, Foltz, Kornder & Mennen, 2022). Generally, we can expect that speakers of non-aspirating languages like Czech need to learn to produce higher VOT values for the aspirated L2 obstruents. Chondroff and Berk (2019) examined VOT variation in a large pool of speakers and found that, while most native speakers of English produced long VOTs, L2 speakers exhibited greater variability in the timing of voicing across speakers than the native group. Moreover, fifteen L2 speakers produced average VOTs below 35 ms, the boundary between aspirated and unaspirated plosives, and most of their native languages are known to have unaspirated stops. These findings support the notion that VOT can be influenced by the phonological characteristics of the speaker's native language.

Furthermore, the ability of a non-native speaker to produce native-like VOT values is influenced by various factors such as motivation, aptitude, and experience among others (Piske, MacKay & Flege, 2001; Hanzawa, 2018; Fahey, 2019). Some non-native speakers may even achieve native-like values, as observed in several speakers in the Chondroff and Berk (2019) study. However, Flege and Hillenbrand (1984) suggest the opposite for L2 learners, as the examined speakers of English did not converge to the VOT values typical for French. Similarly, Fahey (2019) conducted a study on two L2 Korean learners of English (aged 11) and found that even when these speakers were considered native-like in English generally, their measured VOT did not support this claim. The results of these and similar studies suggest that achieving native-like VOT values is a challenging task for L2 speakers, even when they are highly proficient in the language, and highlight the importance of considering individual learner characteristics and language background settings.

The aim of the current study is to investigate the implementation of the voicing contrast in Czech L2 speakers of English, hypothesizing that differences between the two languages will manifest in the amount of devoicing in plosive-sonorant clusters. The study also proposes that language proficiency will have a mediating effect on this relationship. Specifically, it is expected that the low proficiency group, consisting of speakers with a strong Czech accent who do not study English at the university, will exhibit lower VOT values than the high proficiency group, consisting of Czech learners who do study English at a university and who also manifest near-native pronunciation. The highest VOT values are expected from the control group of native English speakers. Overall, this study aims to shed light on the role of language proficiency in the implementation of the voicing contrast in L2 speech.

2. Methodology

2.1. Material and speakers

The material comprises 24 recordings of speakers in their twenties or early thirties, mostly female, and grouped into three categories. One group consisted of eight native English speakers, forming the control group. The remaining participants were Czech learners of English, divided into two groups of eight each. The first group (‘beginner L2’) had low proficiency in English, with no university studies related to the language or any English study programme. The second group (‘proficient L2’) comprised students of English and American studies at Charles University, who had completed a two-semester course in English phonetics in their first year. Their level of English proficiency was C1 or C2, and their pronunciation was rated impressionistically as nearly native by the authors, unlike the beginner group.

The study involved the participants reading a short text based on BBC news reports in the studio of the Institute of Phonetics at Charles University. It is a general-purpose text which is standardly used at the institute for examining English pronunciation. The specific version was selected due to its high incidence of sonorant clusters; it included a total of 35 target clusters, but only 26 were used for further analysis (details can be found in Tab. 1). The excluded clusters consisted of repeated occurrences of certain words in the text (*countries*, *nuclear*, *between*, *president*, *troops*). Additionally, the word *Eritrea* was also excluded as many participants were unfamiliar with it, and the /pj/ cluster in *unscrupulous* was disregarded as most participants did not produce the /j/ sound. One word in the dataset (*expropriation*) involved two clusters, so there were 25 target words for each speaker. In total, the analysis was based on 613 tokens of plosive-sonorant clusters.

Table 1. Tokens of plosive-sonorant clusters used in the analysis (for excluded items see text).

Cluster	Number of tokens per text	Target words
/pl/	3	plunder, powerplant, plan
/pɹ/	7	president, expropriation (2×), approve, price, prices, prudence
/tw/	1	between
/tɹ/	8	country, countries, extreme, troops, traders, try, Patrick, strongly
/kl/	2	nuclear, climbed
/kɹ/	4	secretary, democratic, critics, unscrupulous

2.2. Phonetic segmentation and data processing

The target cluster boundaries were manually determined following Machač and Skarnitzl's (2009) recommendations. The beginning of [p t k] was placed at the onset of closure, including voicing continuation from the vowel, while the final boundary was placed in the middle of the transition area, considering the lower frequency intensity of the sonorant and the onset of the full formant structure of the following vowel. As the boundary between the plosive and approximant was the least confidently placed (as discussed in Přečková, 2022), VOT was used as an alternative measure to reflect the devoicing of the sonorant consonant. The parameter was determined by the onset of the burst in the plosive and the onset of voicing, manifested by the appearance of a low-frequency component in the spectrogram and periodic vibrations in the waveform. Values exceeding the burst duration (typically 20–30 ms) can be regarded as instances of sonorant devoicing. Figure 1 illustrates the segmentation process and VOT measurement.

As VOT also reflects the tempo of the speaker, we normalized this parameter for articulation rate (AR). We selected stretches of speech corresponding to three stress groups in which the cluster occurred and measured the local AR. Native controls had a median AR of 14.3 phones/s, while the two L2 groups had averages of 13.0 and 12.1 phones/s. We normalized the duration data for the analysis with both the local tempo and the speaker's overall median. This step helped eliminate differences that could arise when the speaker pronounced speech segments at different rates of speed. To perform the normalization, we used the following formula: *Normalized VOT = raw VOT * (LAR / SAR) * (SAR / overall median)*, where *LAR* indicates local AR, *SAR* speaker's median, and *overall median* a median of all SARs.

To ensure accurate analysis, it is important to consider stress placement, as stressed sounds tend to have a longer duration and could potentially affect devoicing. Therefore, we relied on the actual realization of lexical stress according to the recording, rather than relying on expectations from canonical pronunciation. We observed that speakers with lower proficiency tended to continuously stress the first syllable, in accordance with the norms of the Czech language. Additionally, we carefully examined any VOT values that deviated markedly from expected values and compared them to the acoustic signal to minimize errors.

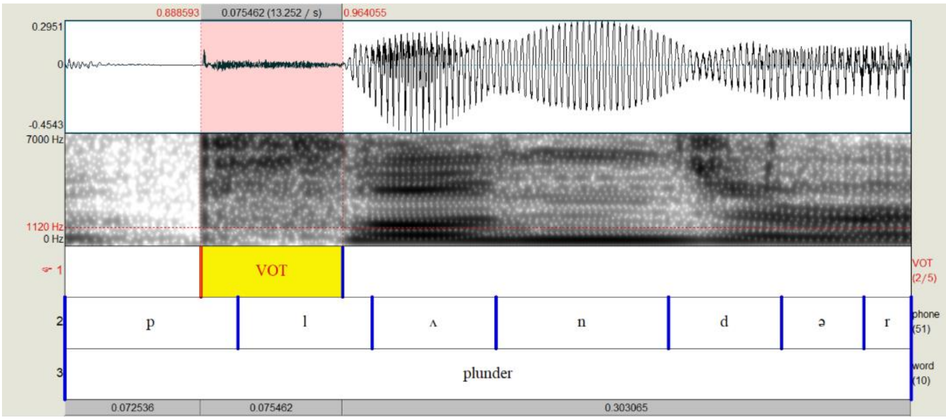

Figure 1. Phonetic segmentation for the target cluster /pl/ and the VOT measurement, illustrated for the word *plunder* (native speaker).

3. Results

3.1. General results

The dataset for this experiment consisted of 613 tokens of plosive-sonorant clusters. Figure 2 presents the results for the three speaker groups. The proficient learners achieved a level of VOT values comparable to that of the native controls, with their VOT values being nearly twice as large as those of the beginner group.

As we already explained, sonorant devoicing is phonetically analogous to aspiration, and /s/ blocks aspiration of plosives before vowels in English (Klatt, 1975; compare *peek* and *speak*). Due to a significant imbalance of this factor in the sample of clusters, we excluded the words that contained clusters preceded by /s/ from further analyses (93 tokens) rather than keeping the factor as an interacting variable. Table 1 compares the two datasets, showing practically no change in the beginner group, but a larger increase in VOT after /s/-cluster exclusion in the proficient L2 group and the native controls (5% and 7%, respectively), which yielded even larger differences between these speakers and the beginner group. As shown in Figure 3, the effect of /s/-clusters was most prominent in the native controls, with their VOT values after /s/ being in the same range as the beginner learners' values for both types of cluster. This indicates that the beginner learners did not produce sonorant devoicing, whereas L2 controls (and proficient learners) did, unless /s/ preceded the target cluster.

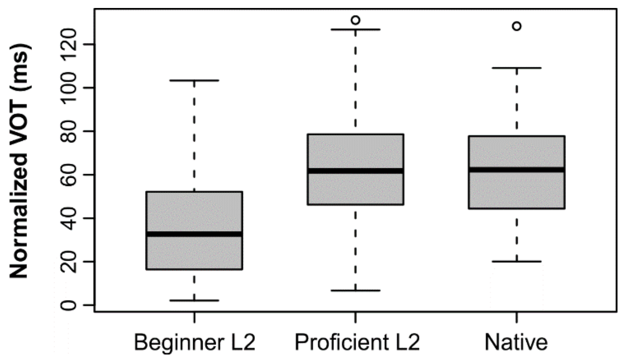

Figure 2. Boxplots of VOT as a function of speaker group.

Table 2. Mean VOT values and standard deviations according to speaker group and inclusion/exclusion of /s/-clusters.

Speaker group	Mean (SD) in ms with /s/-clusters included	Mean (SD) in ms with /s/-clusters excluded
Beginner L2	36 (23)	37 (24)
Proficient L2	60 (24)	63 (24)
Native controls	58 (23)	62 (22)

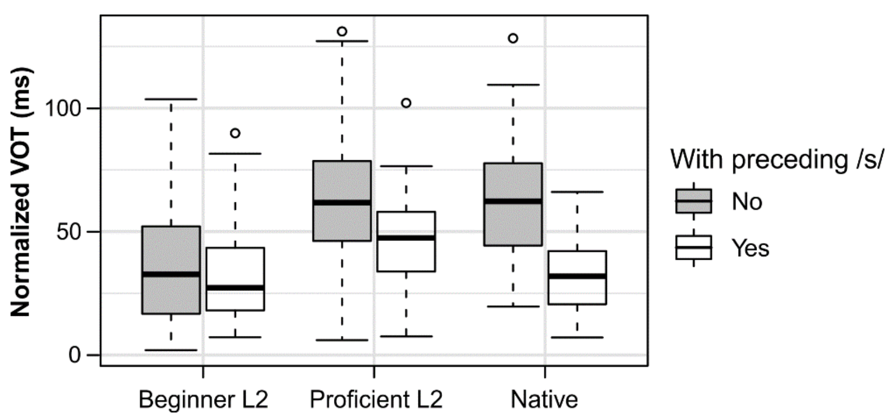

Figure 3. Boxplots of VOT as a function of speaker group and the presence of preceding /s/.

The position of stress in a word is another important factor to take into consideration. Figure 4 illustrates the impact of stress on each of the speaker groups. Interestingly, our findings suggest that the location of the cluster within the stressed or unstressed syllable did not affect the VOT values. This held true for all speaker groups, including the control group.

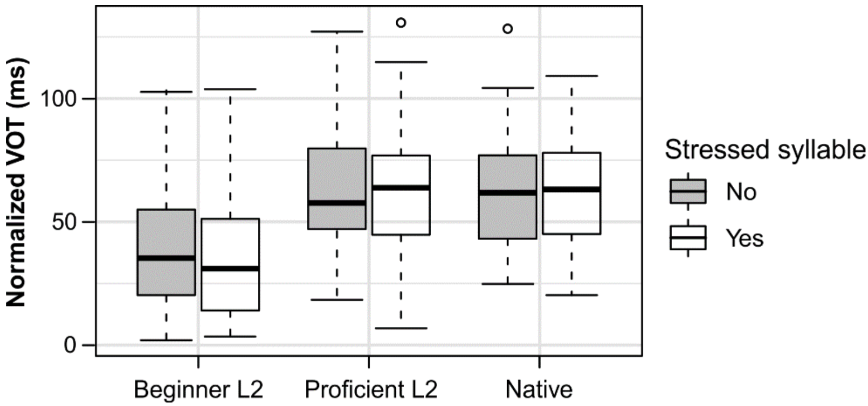

Figure 4. Boxplots of VOT as a function of speaker group and the presence of stress.

The type of cluster was also analyzed, and the results are presented in Figure 5. Overall, the VOT values for the three groups showed similar patterns across all clusters. Furthermore, clusters beginning with /k/ and /p/ exhibited similar behaviour. In contrast, clusters with /t/ were associated with higher VOT values, indicating a greater degree of aspiration.

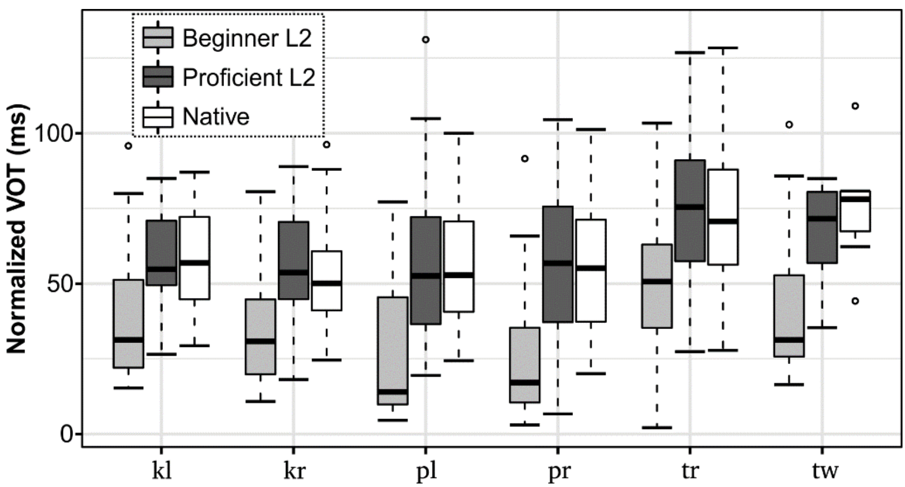

Figure 5. Boxplots of VOT for individual clusters as a function of speaker group.

3.2. Statistical evaluation

The dataset was analyzed using LME models in R (R Core Team, 2022), with the lme4 (Bates, Maechler, Bolker & Walker, 2015) and emmeans (Lenth, 2022) packages. The analysis was conducted on the dataset excluding the /s/-clusters (n=520). The fixed factors included GROUP (beginner L2, proficient L2, native control [= reference level]), STRESS (unstressed [= reference level], stressed) and CLUSTER (/kl/ [= reference level], /kɹ/, /pl/, /pɹ/, /tɹ/, /tw/), and the model also included interactions of GROUP with the other two factors. Random intercepts were included for SPEAKER and WORD but no random slopes could be added due to convergence issues (namely, by-word slope for GROUP and by-speaker slope for CLUSTER). Likelihood ratio tests were used to evaluate the significance of each fixed effect or interaction, comparing the full model to a reduced model without the effect in question.

The statistical model confirmed a lack of significant interactions with GROUP ($\chi^2(2) = 0.1$, $p = 0.947$ and $\chi^2(10) = 10.5$, $p = 0.401$ for STRESS and CLUSTER, respectively). When GROUP was analyzed as an individual factor, it significantly predicted VOT production ($\chi^2(2) = 17.0$, $p < 0.001$). Post-hoc Tukey comparisons showed that the beginner L2 group had significantly smaller VOTs than both the proficient L2 and native control groups (p < 0.001). Compared to the beginner L2 learners, the proficient group produced VOTs larger by 26 ms (SE = 5.9 ms) and the native controls larger by 25 ms (SE = 5.9 ms). These two groups did not differ significantly, confirming that the proficient L2 learners attained native-like values of VOT in plosive-sonorant clusters (cf. Figures 2–5). STRESS was not found to be a significant predictor ($\chi^2(1) = 0.469$, $p = 0.493$), consistent with the findings in Figure 4. On the other hand, CLUSTER was found to have a significant effect on VOT ($\chi^2(5) = 17.2$, $p = 0.004$). Further analysis using post-hoc Tukey comparisons revealed that the effect was mainly driven by the /tɹ/ cluster (see Fig. 5), which was found to be associated with significantly higher VOTs compared to three other clusters (/pɹ/, /pl/, /kɹ/).

3.3. Individual variability

As depicted in Figure 6 in the upper panels, there was noticeable variability in VOT values within each group of speakers. However, the degree of homogeneity varied across the groups, with the beginner learners exhibiting the lowest levels. Two speakers, #4 and #7, produced relatively higher VOT values compared to the rest of the beginner group. In contrast, only speaker #1 in the proficient L2 group displayed non-native-like VOT values, more akin to the majority of the beginner group. Notably, the native English speakers formed the most homogeneous group of speakers, with the exception of speaker #5, who tended to produce slightly lower VOT values.

For instance, stress (unstressed syllable) might lower the VOT values for *secretary*, but not in other clusters, and the opposite level (stressed syllable) can be positioned both higher and lower, apparently at random. Other, unidentified factors – probably contextual or related to frequency – are probably causing the variability.

The lower panels of Figure 6 depict the variability of VOT values for words grouped by clusters. The analysis reveals that there is no identifiable systematic factor that would consistently contribute to increased variability. For instance, the lack of stress on the cluster /kɹ/ in *secretary* may result in lower VOT values, while the effect is less clear-cut in other words (compare /tɹ/ in *Patrick* and *country*). Stress may produce both higher and lower VOT values in a seemingly random manner. It is likely that other factors, possibly related to contextual or frequency predictors, are responsible for the observed variability.

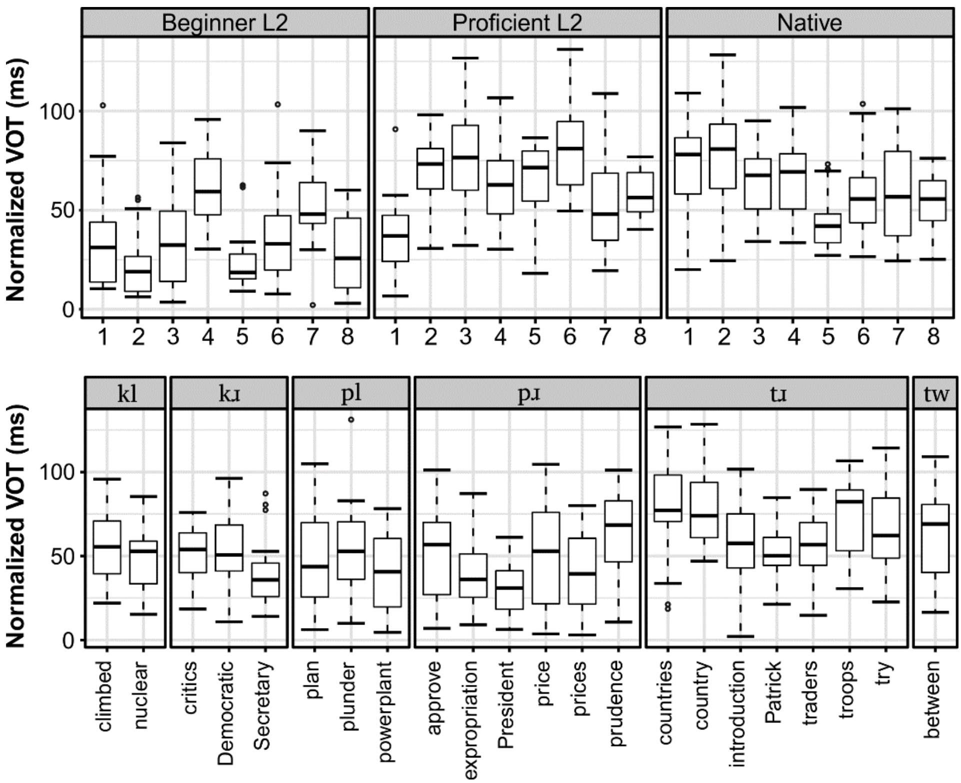

Figure 6. Variability within groups (speaker groups at the top, clusters at the bottom).

4. Discussion

4.1. The effect of proficiency on sonorant devoicing

The present study aimed to investigate the production of VOT in plosive-sonorant clusters by three groups of speakers: beginner L2 learners, proficient L2 learners, and native English speakers. The results revealed significant differences in VOT values between the groups, where beginner learners demonstrated a higher level of heterogeneity and clear deviations from both the proficient L2 learners and native controls. Interestingly, the proficient L2 group displayed no significant differences from the native group regarding VOT. In this respect, our results are inconsistent with previous findings which suggested that greater proficiency and exposure may not lead to more native-like VOTs for most learners. Flege and Hillenbrand (1984, p. 717) stated quite categorically that 'because of interlingual identification, adult learners of a foreign language will never succeed in producing L2 stops with complete accuracy when stops in their native language differ substantially in VOT from those in L2'. Our findings clearly contradict this claim.

When taking into account the methodology of these studies, the discrepancy between our findings and those of Fahey (2019) and Flege and Hillenbrand (1984) may be attenuated. Fahey examined only two subjects, who were 11-year-old children, and the investigated language was Korean, which has a three-way phonemic contrast in VOT. Flege and Hillenbrand used three groups of seven speakers with varying experience with French, which is comparable to our study in terms of the number of participants and language experience. However, their study examined the production of only two French words (/tu/ and /ty/), and the participants were L1 English speakers learning L2 French, which has lower VOT values than English. This situation might not be comparable to speaking in a language with higher VOT values (i.e., L2 English produced by L1 Czech speakers). Moreover, both mentioned studies focused on aspiration before vowels, whereas ours examined sonorant devoicing (aspiration before a sonorant consonant).

Perhaps most importantly, our proficient speakers were strongly motivated and had a rigorous academic background in English, including two semesters of phonetic instruction. This distinguishing factor may differentiate our findings from those of proficient learners who lack formal phonetic training or speakers who acquired the language in a naturalistic environment. Consequently, our study underscores the multifaceted nature of L2 acquisition, where factors such as motivation, instruction, and exposure may all play a significant role.

Our findings also revealed an inclination towards native VOT norms in the treatment of /s/-clusters. In English, the production of /s/-clusters does not involve aspiration after /s/. A comparison of words such as *stop* and *top,* or *school* and *cool* illustrates this phenomenon. Figure 2 in our study demonstrates that both native and proficient L2 speakers are sensitive to the /s/-environment, which is apparent in their production of VOT values. However, this sensitivity was not observed in the beginner group, who appeared to lack the ability to differentiate the presence of /s/ due to their overall lack of aspiration in their speech. Such results further demonstrate that sonorant devoicing is due to aspiration of the preceding plosive and not due to voicing assimilation (which would not be blocked by a preceding /s/).

Finally, the finding that speakers with a strong Czech accent in English lack sonorant devoicing aligns with our analysis of dozens of maturita (‘A-levels’) textbooks, which do not mention the phenomenon at all, and only briefly mention obstruent devoicing in some of them. Given that sonorant devoicing may be crucial for comprehension, as seen in words such as *glue* versus *clue*, where the phonetic difference largely stems from the devoicing of the sonorant consonant, it is crucial to pay more attention to this phenomenon in TEFL. It is important to note, however, that the absence of devoicing in the curriculum does not necessarily indicate a lack of education or learning (students may need to conduct their own research to practice the phenomenon or acquire the pronunciation intuitively or by imitation). Moreover, the dialects of Northern England are known for very low levels of aspiration relative the southern standard, which indicates that intermediate values may be sufficient for comprehension.

4.2. Theoretical implications

In this section, we will briefly examine how our research findings align with the theories of L2 acquisition, particularly in the context of the Speech Learning Model (SLM) developed by Flege and his colleagues (e.g., Flege, 1987). The SLM proposes the concept of *equivalence classification* as a central factor contributing to cross-language phonetic interference. In essence, this model posits that L2 sounds perceived as equivalent to their counterparts in the learner’s L1 will tend to gravitate towards the phonetic values of the native language. Conversely, speech sounds categorized as belonging to a ‘new’ category in the L2 are more likely to be produced in a manner that aligns with the target pronunciation.

Our study revealed interesting findings. It appears that the more proficient group has effectively acquired the pronunciation feature of plosive-sonorant clusters in the L2 ([pl̥]), presumably due to their non-equivalence with the corresponding L1 clusters ([pl]). In contrast, the less proficient speakers seem to treat these clusters as equivalent, as they did not exhibit a marked shift in their pronunciation. However, for a more comprehensive interpretation of these results, additional data is necessary, particularly regarding the context in which plosives precede vowels.

Upon conducting auditory analysis on a randomly selected subset of speakers and items, we observed distinct pronunciation patterns within the more proficient group. For instance, words like *peace*, *report*, *tension* and *country* consistently exhibited aspiration in this group but remained unaspirated in the less proficient group. This implies that higher proficiency in English is associated with the acquisition of the aspiration feature in voiceless (fortis) plosives, resulting in distinct phonetic categories for these plosives in L1 Czech and L2 English (unaspirated vs. aspirated, respectively). Such a pattern was also evident in the pronunciation of both /s/-stop clusters and /s/-stop-sonorant clusters, where the presence of the fricative blocked aspiration and, consequently, sonorant devoicing. In contrast, speakers with lower proficiency appeared to treat all voiceless (fortis) plosives as equivalent.

These findings resonate well with what one might anticipate from Czech learners of English. A parallel case study conducted by Podlipský, Šimáčková and Chládková (2021) on English-Czech bilinguals observed similar cross-language influences on VOT values. Their results indicated a tendency for phonetic drift towards L1 values within the voiced series but not within the voiceless series, implying a shared representation in the former and distinct categories in the latter. A similar asymmetry in cross-language influence also emerged in Polish speakers of English, as observed by Schwartz (2022). Specifically, the interference between L1 and L2 was more prominent in voiced plosives than in voiceless plosives, with aspiration being mastered earlier and to a greater degree than lenis obstruent devoicing. This finding was interpreted as reflecting differences in the encoding of laryngeal contrasts.

In conclusion, our findings align with the predictions of the SLM and are consistent with related research in the field, underscoring the role of equivalence classification and the differential acquisition of phonetic features in L2 acquisition. Moreover, it highlights the importance of aspiration (delay in VOT) in the perception and production of English fortis plosives before vowels and sonorant consonants alike.

4.3. Limitations and future research

While our study represents an improvement over previous research, which often relied on much smaller sample sizes, we acknowledge that our own sample size of 24 participants and 26 words is still relatively small. Nevertheless, we carefully controlled for proficiency levels and used a single text to restrict variation, which allowed for the elicitation of identical words across speakers and a more realistic situation than word lists. Additionally, the text included six out of the ten possible English plosive-sonorant cluster types (Cruttenden, 2014, p. 261), particularly those of high frequency. To further strengthen our findings, future studies could utilize a factorial design that includes a more balanced range of clusters in comparable contexts. This approach would be particularly valuable for examining the effects of stress, which our study did not find to be significant, possibly due to the highly imbalanced distribution of stress in the clusters.

There are several areas for future research that could build on the findings of this study. One possible direction would be to investigate in more detail the role of different contextual and frequency factors that might impact the production of VOT. Another direction could be to explore the effect of different training methods on VOT production in L2 learners, with a focus on how training could potentially enhance the production of native-like VOT values. Alternatively, longitudinal studies could track the development of VOT production over time to determine how changes occur as L2 proficiency increases (see e.g., Hanzawa, 2018). These studies could examine the influence of various factors such as motivation, instruction, and exposure on VOT production. Finally, future research should measure the impact of sonorant devoicing on perceived ‘accentedness’ of the speaker, and also pay attention to the effects on intelligibility and comprehensibility.

5. Conclusion

The current study provides valuable insights into the impact of proficiency level on the production of VOT in English plosive-sonorant clusters. Our findings indicate that beginner learners exhibit distinct VOT values compared to both proficient L2 learners and native English speakers. Importantly, there were no significant differences between the VOT values of proficient L2 learners and native speakers. Our results therefore suggest that L2 learners can achieve native-like VOT values, contradicting the claim made by Flege and Hillenbrand (1984). While the study has some limitations, it contributes to a better understanding of how L2 learners with different proficiency levels produce voiceless stops in English.

Acknowledgements

The research was supported by the European Regional Development Fund-Project ‘Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World’ (No. CZ.02.1.01/0.0/0.0/16_019/0000734). The study was created within the Charles University programme ‘Cooperatio’, scientific field Linguistics.

References

Abramson, Arthur S. and Douglas H. Whalen. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63. 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

Bates, Douglas, Martin Mächler, Ben Bolker and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67 (1). 1–48. https://doi.org/10.18637/jss.v067.i01

Cho, Taehong and Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27 (2). 207–229. https://doi.org/10.1006/jpho.1999.0094

Chodroff, Eleanor Rosalie and Melissa Baese-Berk. 2019. Constraints on variability in the voice onset time of L2 English stop consonants. In Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren (ed.), *Proceedings of the 19th International Congress of Phonetic Sciences*, paper 710. Canberra, Australia: ASSTA.

Cruttenden, Alan. 2014. *Gimson's Pronunciation of English*. 8th edn. London: Routledge. https://doi.org/10.4324/9780203784969

Docherty, Gerry J. 1992. *The Timing of Voicing in British English Obstruents*. Berlin: Foris. https://doi.org/10.1515/9783110872637

Fahey, Danielle. 2019. Voice Onset Times of voiceless stops by bilingual Korean-English children: Phonetic acquisition and sociolinguistic preferences. In Hope Wilson et al. (ed.), *Selected Proceedings of the 2017 Second Language Research Forum,* 79–91. Somerville, MA: Cascadilla Proceedings Project.

Flege, James Emil. 1987. The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15. 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, James Emil and James Hillenbrand. 1984. Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America* 76 (3). 708–721. https://doi.org/10.1121/1.391257

Gonet, Wiktor and Radosław Święciński. 2012. More on the voicing of English obstruents: Voicing retention vs. voicing loss. *Research in Language* 10 (2). 183–199. https://doi.org/10.2478/v10015-011-0035-8

Hansson, Gunnar Ólafur. 2001. Remains of a submerged continent: Preaspiration in the languages of Northwest Europe. In Laurel J. Brinton (ed.), *Historical Linguistics 1999,* 157–173. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.215.12han

Hanzawa, Keiko. 2018. The development of Voice Onset Time (VOT) in a content-based instruction university program by Japanese learners of English: A longitudinal study. *Canadian Modern Language Review* 74. 502–522. https://doi.org/10.3138/cmlr.2018-0196.

Hoole, Philip. 1999. Laryngeal coarticulation. Section A: Coarticulatory investigations of the devoicing gesture. In William J. Hardcastle & Nigel Hewlett (ed.), *Coarticulation: Theory, Data and Techniques*, 105–121. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486395.006

Klatt, Dennis. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* 18. 686–706. https://doi.org/10.1044/jshr.1804.686

Laver, John. 2012. *Principles of Phonetics*. Cambridge: Cambridge University Press.

Ladefoged, Peter and Keith Johnson. 2015. *A Course in Phonetics*. 7th edn. Stamford: Cengage Learning.

Lenth, Russell V. 2022. emmeans: Estimated Marginal Means, aka Least-Squares Means. Available at: https://CRAN.R-project.org/package=emmeans.

Lisker, Leigh and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384–422. https://doi.org/10.1080/00437956.1964.11659830

Machač, Pavel and Radek Skarnitzl. 2009. *Principles of Phonetic Segmentation*. Praha: Nakladatelství Epocha.

Mairano, Paolo, Bene Bassetti, Mirjana Sokolović-Perović and Tania Cerni. 2018. Effects of L1 orthography and L1 phonology on L2 English pronunciation. *Revue Française De Linguistique Appliquée* 23 (1). 45–57. https://doi.org/10.3917/rfla.231.0045

Ogden, Richard. 2017. *An Introduction to English Phonetics*. 2nd edn. Edinburgh: Edinburgh University Press.

Piske, Thorsten, Ian R. A. MacKay and James Emil Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29 (2). 191–215. https://doi.org/10.1006/jpho.2001.0134

Podlipský, Václav Jonáš, Šárka Šimáčková and Kateřina Chládková. 2021. Phonetic drift reveals interconnected phonological representations in simultaneous bilinguals: A case study of English and Czech stop consonants. *International Journal of Bilingualism* 25(3). 789–799. https://doi.org/10.1080/00437956.1964.11659830

Přečková, Klára. 2022. Sonorant Devoicing in Plosive-sonorant Clusters in Czech-accented and Native British English [Unpublished BA thesis]. Faculty of Arts, Charles University.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Roach, Peter. 2009. *English Phonetics and Phonology: A Practical Course*. 4th edn. Cambridge: Cambridge University Press.

Saud Alharbi, Amirah, Anouschka Foltz, Lisa Kornder and Ineke Mennen. 2022. L2 acquisition and L1 attrition of VOTs of voiceless plosives in highly proficient late bilinguals. *Second Language Research* [online first]. https://doi.org/10.1177/02676583221124011

Schwartz, Geoffrey. 2022. Asymmetrical cross-language phonetic interaction: Phonological implications. *Linguistic Approaches to Bilingualism* 12 (2). 103–132. https://doi.org/10.1075/lab.19092.sch

Sieczkowska, Jagoda, Bernd Möbius and Grzegorz Dogil. 2010. Specification in context – Devoicing processes in Polish, French, American English and German sonorants. In Takao Kobayashi, Keikichi Hirose and Satoshi Nakamura (ed.), *Proceedings of Interspeech 2010*, 1549–1552. Makuhari, Japan: ISCA. https://doi.org/10.21437/Interspeech.2010-451

Skarnitzl, Radek and Jana Rumlová. 2019. Phonetic aspects of strongly-accented Czech speakers of English. *AUC Philologica* 2/2019. 109–128. https://doi.org/10.14712/24646830.2019.21

Strycharczuk, Patrycja. 2012. Sonorant transparency and the complexity of voicing in Polish. *Journal of Phonetics* 40 (5). 655–671. https://doi.org/10.1016/j.wocn.2012.05.006

Tsuchida, Ayako, Abigail C. Cohn and Masanobu Kumada. 2000. Sonorant devoicing and the phonetic realization of [spread glottis] in English. *Working Papers of the Cornell Phonetics Laboratory* 13. 167–181.

Volín, Jan. 2002. *IPA-Based Transcription for Czech Students of English*. Praha: Karolinum.