GEOFFREY SCHWARTZ
Adam Mickiewicz University, Poznań
geoff@amu.edu.pl
ORCID: 0000-0002-0728-7820

SPIT OR *SBIT*? ST-TYPE CONSONANT CLUSTERS IN THE SPEECH OF POLISH LEARNERS OF ENGLISH

Abstract

This paper presents two small-scale acoustic phonetic studies investigating the pronunciation of sibilant-stop (ST) consonant clusters in Polish, and in the L2 speech of L1 Polish learners of English. In English, aspiration of fortis stops is not attested in the post-/s/ context. Rather, short-lag voice onset time (VOT) measures are observed in L1 English in post-/s/ stop consonants, a phonetic weakening that renders them phonetically similar in terms of VOT to lenis stops in initial position. In Polish, both voiced and voiceless stops may appear after sibilant fricatives. The acoustic results suggest that (1) L1 Polish does not weaken its stops in ST clusters, and (2) that more L1 Polish speakers exhibit some weakening in their L2 English clusters as a function of proficiency, but do not produce native-like VOTs in ST sequences. Implications of these findings for L2 speech research and the phonological status of ST clusters are discussed.

1. Introduction

The present paper presents acoustic data on the production of stops in word-initial sibilant-stop (ST) consonant clusters in English as a second language (L2) by first language (L1) speakers of Polish. While many studies have been devoted to acquisition of aspirated English fortis stops (for a summary see Zampini 2008), the post-/s/ context is familiar as a position in which aspiration does not occur, and in which the voicing contrast is said to be neutralized. The absence of aspiration in English ST clusters renders post-/s/ stops phonetically similar to /b d g/, with short-lag voice onset time (VOT; Lisker, Abramson 1964) resembling that of lenis stops

(Cho et al. 2014). Thus, English stops in ST clusters may be thought of as weaker than their initial counterparts. Indeed some phonological proposals (see Szigetvári 2020) have claimed that /ptk/ following /s/ should be analyzed as lenis. By contrast, in Polish both voiced and voiceless stops may appear after fricatives, and the voicing of the fricative is dependent on the stop that follows. No phonetic weakening is said to occur.

Sibilant-initial clusters show unusual phonological behaviour. In many languages, including English, they are the only initial consonant sequences to violate the generalization that sonority (e.g. Parker 2002) must increase in syllable onsets (cf. English *twin*, *flash, snack, swing, cry*). Conversely, in Polish, ST is just one of a large number of sonority-violating sequences – the language features many words with sonority decreases (e.g. *lwa* /lva/ 'lion [gen.]', *lnu* 'linen [gen.]'), and sonority plateaus (e.g. *kto*, *gdy* 'who, when'). Clearly, in English ST clusters constitute an unusual structure, while in Polish the sequences are not exceptional.

Taken together, the phonological and phonetic facts suggest that ST clusters in Polish and English may be structurally distinct entities (see 4.1). This paper will present two small-scale acoustic studies of ST clusters produced by Polish speakers designed to investigate this possibility. The first study (Section 2) compares initial singleton voiceless stops with voiceless stops embedded in ST clusters produced by monolingual Polish speakers. This experiment provides a baseline for determining whether the ST context in Polish induces any phonetic changes in the stop. The second study (Section 3) examines the phonetic realization of ST clusters produced by L1 Polish speakers with two different levels of proficiency, in order to establish the degree to which Polish learners of English may acquire the weakening of stops in an ST context. The implications of those findings for L2 speech acquisition more generally will also be discussed (4.2).

2. Experiment 1: VOT in Polish ST clusters

The first study investigates the following research question: are post-/s/ voiceless stops different from initial singleton voiceless stops in L1 Polish? VOT of initial singleton stops in Polish typically falls in range between 20–50 ms (see Schwartz 2020 for a review), with an average of about 30 ms (Wojtkowiak, Schwartz 2019), while L1 English post-/s/ VOT is usually shorter (Cho et al. 2014), with an average of about 14 ms.

2.1. Participants

Data come from recordings of 38 native speakers of Polish, with minimal knowledge of foreign languages. The speakers were all between the ages of 19 and 27. Twenty-six of the speakers were female and twelve were male. None reported any hearing or speech impairments.

2.2. Materials and procedure

Recordings were made in a sound-treated recording booth at a Polish university. The data to be presented are based on productions of two pairs of words, (*Pawła – spawać* 'Paul [acc.]' – 'weld'; *taflę – stawać* 'slab' – 'to stand'), contrasting initial voiceless stops with stops in ST clusters. In each case, the target stop /p/ or /t/ is preceding the first vowel, always /a/, in a two-syllable word. The words themselves were placed in phrase-initial, yet utterance-medial position, preceded and followed by the same number of syllables (8) in the utterance. For this reason, it may be stated that prosodic factors affecting the realization of the stops were carefully controlled for. The two utterances were embedded within a larger set of 48 utterances, presented to participants in pseudo-randomized order as part of an unrelated project. Speakers had a chance to familiarize themselves with the utterances before the recording session. During the session, the utterances were elicited using PowerPoint slides on a monitor housed inside the recording booth. The booth was equipped with a high quality microphone, which was connected to a USB audio interface that allowed for recordings made directly on to a computer. Speakers read two repetitions of each utterance, yielding a total of 304 items (4 words × 38 speakers × 2 repetitions).

2.3. Acoustic and statistical analysis

Acoustic annotation of the recordings was done manually in Praat (Boersma, Weenink 2018). Voice onset time (VOT), defined as the time window from the release of the stop to the onset of periodicity in the waveform, was marked into a Praat text grid, allowing a script to automatically extract the VOT measures. The data were analyzed using the SPSS statistical software (IBM Corporation 2017). A generalized linear mixed model was fitted, with VOT as the dependent variable, an interaction term of Consonant (/p/ or /t/)*Position (initial or post-/s/) as the main predictor of interest, and by-speaker random slopes and intercepts.

2.4. Results

The mean VOT measures were 21.95 ms (SD = 9.6) for singleton stops, and 19.94 ms (SD = 7.6) for post-/s/ stops, a difference of approximately 2 ms. The data sorted for consonant place and position are summarized graphically in the boxplots in Figure 1. The statistical results revealed that VOT of /p/ was approximately 3 ms shorter for post-/s/ stops than singleton stops (contrast estimate = 0.003, SE = 0.001, t = 4.35, p < .001), but the VOT of /t/ did not differ in the two positions (contrast estimate = 0.001, SE = 0.001, t = 1.29, p = .199).

Figure 2 presents VOT measures as a function of individual speaker, grouped for position, with separate panels for the labial and dental stops. In the figure, the blue crosses represent singleton stops, while the red dots represent post-/s/ stops. While the figure shows a significant degree of inter-speaker VOT variability, we can see the effect found in the statistical results by which longer VOTs for singleton stops are found for /p/, but not /t/.
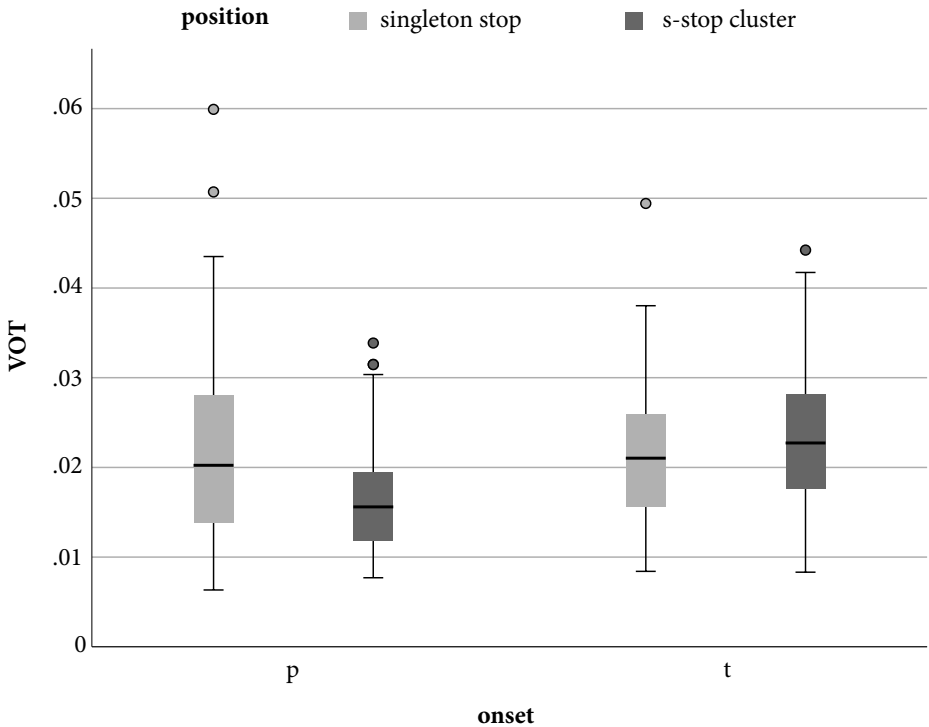
Figure 1. Boxplots of VOT results in L1 Polish as a function of consonant, sorted
         for position

2.5. Discussion

Overall, the results suggest that Polish does not weaken its stops in the post-/s/ posi-
tion. Although /p/ showed slightly shorter VOT (3 ms) in the post-/s/ position, no
difference was observed for /t/. The difference between /p/ and /t/ may be attributed
to the labial place of articulation of the former,[1] since labial consonants are the least
resistant to coarticulation with neighbouring sounds (Pastätter, Pouplier 2017), it is
likely that the /p/ in the ST cluster showed a greater degree of coarticulation with
the preceding /s/ than the /t/ did, inducing the slight VOT shortening. Another
possibility, suggested by the data in Figure 3, is that speakers in fact lengthened
the VOT of singleton bilabial stops. Since bilabials typically have both the lowest
amplitude bursts and the shortest VOTs among labial, coronal, and dorsal places
of articulation (e.g. Stevens 1997), it is possible that speakers produced longer VOT
for singleton /p/ to enhance the perceptibility of the stop, compensating for the
inherent weakness of the labial noise. Regardless of which explanation is correct,
the shorter VOT of /p/ in the clusters relative to singletons is apparently a purely
phonetic effect, rather than any sort of phonological weakening process of the type
that is observed in English ST clusters.

---

[1]   A reviewer notes that /k/ is missing from the data analyzed here. With regard to VOT patterns,
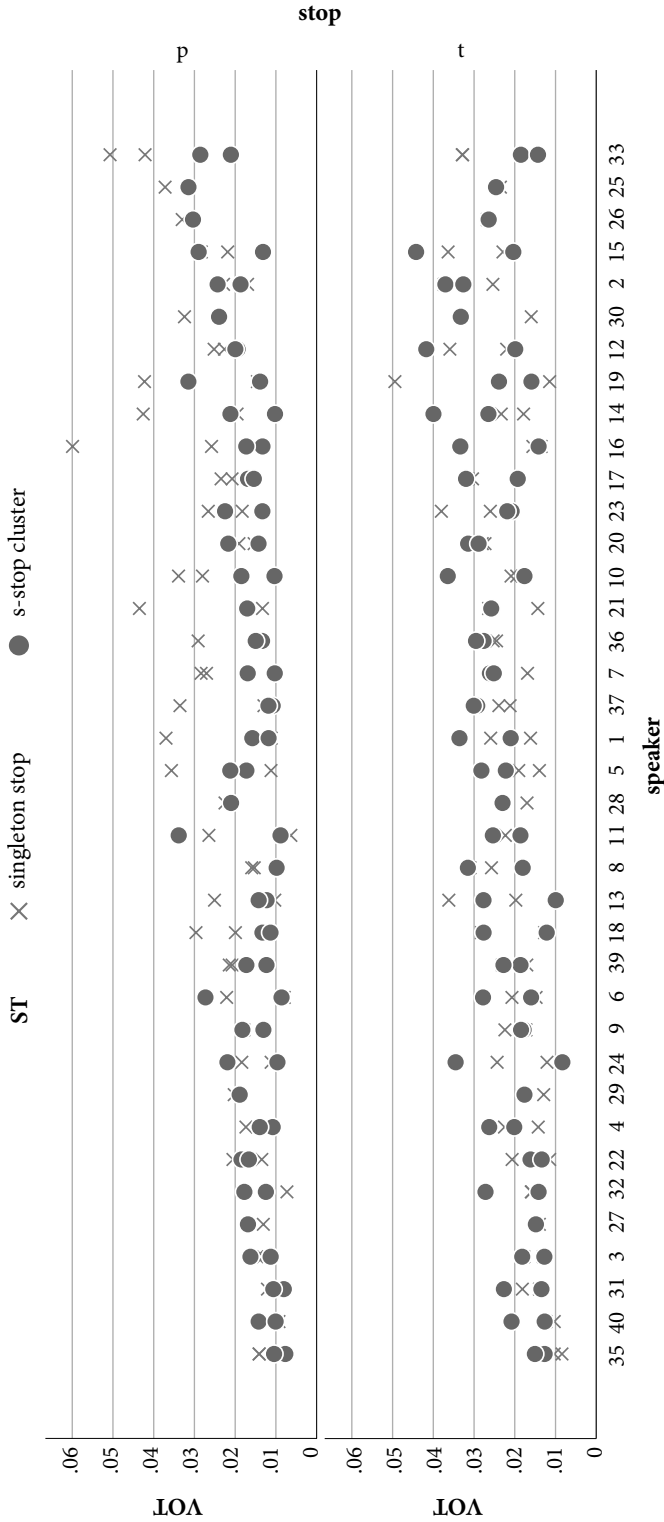   /k/ may be expected to behave in a similar manner to /t/.

Figure 2.   Individual VOT measures of /p/ and /t/ in L1 Polish sorted for position. The top panel represents bilabial /p/. The bottom panel represents dental /t/.

3. Experiment 2: ST clusters in the speech of Polish learners of English

3.1. Participants

The data to be presented here come from recordings of two groups of L1 Polish speakers of English. The first group (Students) was made up of 16 first year university students (12 female, 4 male) between the ages of 18 and 21. The Students were majors in English with a proficiency level of B1 according to the Common European Frame of Reference for languages (CEFR; Council of Europe 2011). The second group (PhD students / Professors) was made up of 16 PhD students and professors (10 female, 6 male) in the same university English department where the students were enrolled. The PhD students / Professors group had C2-level proficiency in English, without easily discernible features of a foreign accent. The students' group had just commenced phonetic training in their English program, while the PhD students / Professors had all completed two years of phonetic training in the program the students were just starting.

3.2. Materials and procedure

The data to be presented come from recordings of four repetitions of two English words, *scat* and *stat*, which were contained in a much larger set of words recorded for unrelated phonetic analyses.[2] The target words were elicited using the Speech Recorder program, which presents slides to experiment participants one at a time in a randomized order, and records the productions directly into WAV files. In this experiment, each slide contained a single word, intended to elicit citation form productions.[3] The recordings were made in the same sound treated booth with the same audio equipment as those in the first study described in this paper. With 32 speakers (16 Students and 16 PhD students / Professors), 2 words, and 4 repetitions, 256 items were collected for analysis.

3.3. Acoustic measures and analysis

VOT was annotated in Praat as in the previous study. The statistical analysis was structured differently, however, since it examined only stops in ST clusters, rather than comparing singleton and cluster-internal stops. A generalized linear mixed model was fit with VOT as the dependent variable, a Group*Consonant interaction term as the main predictor variable of interest, and by-speaker random slopes and intercepts.

---

[2]   The recordings described here were part of an unrelated project carried out before the COVID-19 pandemic, which for the time being has made new recording sessions impossible. For this reason the analysis will be limited to the words *scat* and *stat*. Nevertheless, the number of speakers and repetitions allow for valid statistical comparisons.

[3]   This is in opposition to the first experiment where items were embedded in a carrier utterance. VOT measures of citation forms tend to be longer than in utterances, a difference that is also observed here.

3.4. Results

The mean VOT of voiceless stops in the ST clusters in L2 English was 47 ms (SD = 15) for the Students and 33 ms (SD = 13) for the PhD students / Professors. The data sorted for consonant and Group are summarized graphically in the boxplots in Figure 3. The statistical results revealed that VOT of /k/ was approximately 15 ms shorter for the Teachers (contrast estimate = 0.015, SE = 0.003, $t = 4.39$, $p < .001$), and the VOT of /t/ was approximately 14 ms shorter for them (contrast estimate = 0.014, SE = 0.003, $t = 4.04$, $p < .001$).
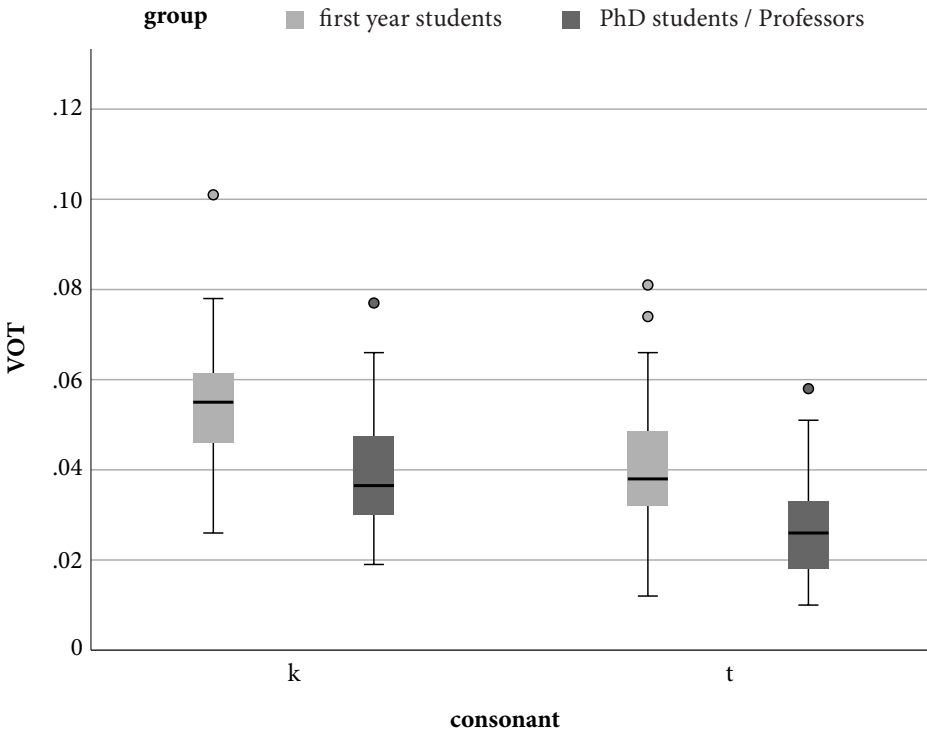

Figure 3. Boxplots of VOT results in L2 English ST clusters as a function of consonant, sorted for speaker group

Figure 4 shows a scatter plot of individual VOT measures for L2 English post-/s/ stops. On the x-axis, speakers are arranged in increasing order according to their mean VOT measures. The light crosses represent the Students' group, and the dark dots represent the PhD students / Professors' group. The individual results clearly show the pattern observed in the group results by which the Students show longer post-/s/ VOTs than the PhD students / Professors. Twelve out of the thirteen highest intra-speaker mean VOT values come from the PhD students /Professors' group, while eight out of the nine lowest intra-speaker mean VOT values from the Students' group.

Figure 4. Individual VOT measures for L2 English post-/s/ voiceless stops as a function of speaker. Speakers on the x-axis are arranged according to intra-speaker mean VOT values (lowest to highest).

3.5. Discussion

The results revealed that the PhD students / Professors' group, with higher proficiency in English, produced shorter VOTs in ST clusters than the Students' group. The most likely interpretation of this finding is that the more proficient L1 Polish speakers have acquired the weakening of post-/s/ fortis stops to a greater degree than the less proficient learners. It should also be noted that the VOT measures for the post-/s/ stops for both speakers groups are longer than those observed in L1 English (e.g. Cho et al. 2014). Thus, neither group produces "native-like" short VOT of unaspirated English stops. This fact is surprising from the point of view of phonological "markedness" by which voiceless stops with short-lag VOT are claimed to be the easiest category to acquire (e.g. Eckman 2008). The failure to produce the "unmarked" short-lag VOT on voiceless stops suggests deeper structural differences between ST clusters in English and Polish. These issues are discussed in the following section.

4. Implications for phonology and L2 speech acquisition

4.1. The phonology of ST clusters revisited

On the basis of both segmental transcription as defined by the International Phonetic Alphabet, there is no reason to expect ST clusters in two different languages to exhibit different structural properties. Yet that is what the data described in this paper indeed suggest. In English, the post-/s/ context yields a weakening of the stop, and apparent neutralization of the fortis-lenis contrast. In Polish, post-fricative stops exhibit normal laryngeal contrast, which causes regressive voicing or devoicing assimilation to the preceding fricative.

In what follows, a brief presentation of a representational proposal encoding this difference, formulated within the Onset Prominence framework (OP; Schwartz 2016, 2018), will be presented, as well as a comparison with more traditional phonological approaches. The key aspect of the OP model is that prosodic structure and segmental representations are built from a single representational building block: a hierarchy of phonetic events associated with a stop-vowel CV sequence. The OP hierarchy is shown in Figure 5. For thorough discussion, see Schwartz (2016).
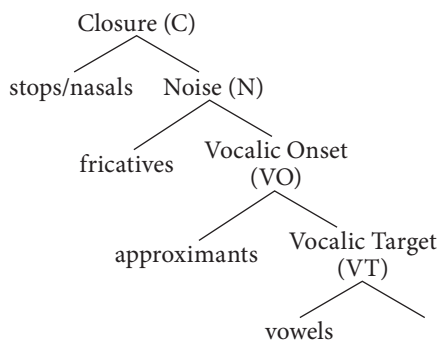
Figure 5. The Onset Prominence representational hierarchy (after Schwartz 2016)

In OP, a CV sequence is not a linear string of C+V. Rather, it is a single unit, from which the individual consonant and vowel are extracted. Additionally, the hierarchy encodes rising sonority in onsets, since sonority increases at lower levels in the hierarchy. This postulate is shown in Figure 4 as the labels of the terminal nodes. Stops and nasals are at the top,[4] then fricatives, then approximants, then vowels. For more complete discussion, see Schwartz (2016).

Since the basic building block in OP is a CV unit, additional mechanisms are needed to describe how consonant clusters may form, as well as the origins of the "codas". While rising sonority clusters can be expressed in a single iteration of the tree structure in Figure 5, the formation of ST clusters in the OP model requires additional phonotactic mechanisms, the results of which are shown in the two trees in Figure 6. One mechanism produces a configuration in which both consonants are contained in a single tree, but the /t/ is *submerged* under the /s/. This is shown on the left. The other mechanism, shown on the right, yields a configuration in which the two consonants are *adjoined* at a higher level of structure; their structures are not combined under the level of Closure that contains a single iteration of the hierarchy. The submerged configuration facilitates weakening of the stop, since the stop is lower in the hierarchy than the fricative. In the adjoined configuration, the fricative and stop are at the same hierarchical level and no weakening is expected. Thus, on the basis of what we have observed in this paper, the submerged configuration is posited for English, and the adjoined structure obtains in Polish. For more details on the submersion and adjunction mechanism, and other arguments in favour of these postulates for English and Polish, see Schwartz (2016).

In traditional phonological models, cross-language differences in ST cluster behaviour are quite problematic to represent.[5] Descriptively, the weakening process in English might be thought as the loss of a laryngeal feature ([fortis] / [spread glottis] / [-voice]) after /s/. Two objections to this approach may be raised. First, since the /s/ in ST clusters may be assumed not to be part of a syllable onset (see Kaye 1992; Goad 2012), the stops would have to be represented as singleton onsets. This means that in English we would have to posit the loss of a contrastive phonological specification in "onset" position, quite an unusual process typologically. The onset position is one where contrasting features are observed most consistently – neutralization is generally associated with "codas". In the OP approach, "onset" and "coda" are epiphenomena that are not formally encoded, so no such controversial claims about onset neutralization are made.

---

[4]   The place of nasals in sonority sequencing is debatable. From the point of view of consonant clusters, nasals often behave similarly to stops. However, nasals are among the most commonly found "syllabic" consonants, so theories equating sonority with syllabic peaks would describe them as quite sonorous. Krämer and Zec (2020) propose that there are two types of nasals – high sonority and low sonority.

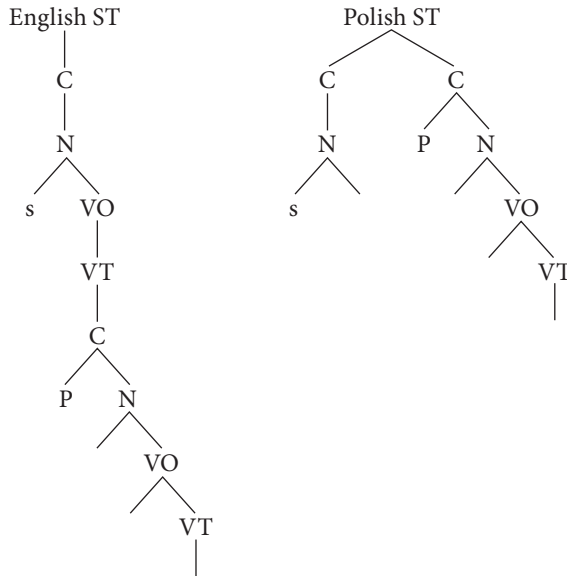[5]   Space restrictions prevent a thorough discussion.

Figure 6. OP configurations for English and Polish ST clusters

4.2. ST and the acquisition of L2 English voice contrasts

Clearly, for the Polish learners in the L2 study described here, the post-/s/ stops are not equivalent to English lenis stops, as evidenced by the relatively long VOTs relative to those of native speakers. The data from Experiment 2 showed cross-proficiency differences in post-/s/ weakening, but also revealed that even very proficient L2 users did not produce native-like VOTs. This result may be thought of in terms of the principle of *equivalence classification* (Flege 1987) between L1 and L2 phonetic categories. L1 Polish learners of English apparently perceptually link unaspirated fortis stops in L2 English with their L1 fortis counterparts, rather than with L2 lenis stops, despite the similar VOT observed in post-/s/ fortis stops and initial /bdg/ in L1 productions. As a result, it appears as if the equivalence classification that hindered acquisition was based on cues *other* than VOT.

An important implication of the present study relates to the question of how L2 English voicing contrasts are acquired by Polish learners. Recent studies (e.g. Schwartz 2020) have found that L1 Polish speakers are generally more successful in their acquisition of aspirated fortis stops than unaspirated but unvoiced lenis stops. In other words, while Polish learners appear to acquire aspiration relatively easily, even much more proficient speakers show continued L1 interference by substituting fully voiced /bdg/ in their L2 speech. At first glance, this may seem like a minute phonetic detail, rather than a serious error. However, problems arise when fully voiced Polish-style /bdg/ appear in obstruent clusters, such as in words like *Facebook* and *update*. In these contexts the fully voiced stops yield regressive voicing and we often observe pronunciations like [fejzbuk] and [abdejt] in Polish-accented speech (Gonet 2001).

This type of pronunciation indeed stands out as a major contributing factor to foreign accentedness ratings by L1 English listeners (Gonet, Pietroń 2004).

By drawing attention to the production of unvoiced, unaspirated stops, the present study may have pedagogical implications for preventing this type of error. In the twenty-first century, with computer-based sound editing tools it is a very simple matter to show the phonetic relationship between unvoiced /b/ in *bill* and unaspirated /p/ in *spill*. Teachers can use the post-/s/ context as a tool for teaching students how to produce unvoiced English lenis stops, a strategy may then be extended to words with clusters. In other words, to produce a word like *baseball* without transferring L1 Polish regressive voicing, it may be pointed out to learners that the word is in fact pronounced *bay-spall* – the /b/ is for all intents and purposes equivalent to a /p/ in the post-/s/ position.

References

Boersma P., Weenink D. 2018. Praat: Doing phonetics by computer (version 6.0.29). [Computer program; http://www.praat.org/; accessed March 2019].

Cho T., Lee Y., Kim S. 2014. Prosodic strengthening on the /s/-stop cluster and phonetic implementation of an allophonic rule in English. – *Journal of Phonetics* 46: 128–146.

Council of Europe. 2011. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Press Syndicate of the University of Cambridge.

Eckman F. 2008. Typological markedness and second language phonology. – Hansen Edwards J.G., Zampini M. (eds.). *Phonology and second language acquisition*. Amsterdam: John Benjamins: 95–116.

Flege J. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. – *Journal of Phonetics* 15: 162–177.

Goad H. 2012. sC clusters are (almost always) coda-initial. – *The Linguistic Review* 29.3: 335–373.

Gonet W. 2001. Obstruent voicing in English and Polish: A pedagogical perspective. – *International Journal of English Studies* 1: 73–92.

Gonet W., Pietroń G. 2004. The Polish tongue in the English ear. – Waniek-Klimczak E., Sobkowiak W. (eds.). *Materiały z konferencji "Dydaktyka fonetyki języka obcego"*. Konin: Wydawnictwo PWSZ w Koninie: 55–65.

IBM Corporation. 2017. *IBM SPSS Statistics for Windows*. [Version 25.0]. Armonk (NY): IBM Corp.

Kaye J. 1992. Do you believe in magic? The story of *s*+C sequences. – *SOAS Working Papers in Linguistics and Phonetics* 2: 293–313.

Krämer M., Zec D. 2020. Nasal consonants, sonority and syllable phonotactics: The dual nasal hypothesis. – *Phonology* 37: 27–63.

Lisker L., Abramson A.S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. – *Word* 20: 384–422.

Pastätter M., Pouplier M. 2017. Articulatory mechanisms underlying onset-vowel organization. – *Journal of Phonetics* 65: 1–14.

Parker S. 2002. *Quantifying the sonority hierarchy.* [PhD dissertation, University of Massachusetts at Amherst].

Schwartz G. 2016. On the evolution of prosodic boundaries – Parameter settings for Polish and English. – *Lingua* 171: 37–73.

Schwartz G. 2018. *Towards a typology of consonant synchronicity.* [Paper presented at the "Representing Phonotactics" session, LabPhon 18, Lisbon, Portugal].

Schwartz G. 2020. Asymmetrical cross-language phonetic interaction – phonological implications. – *Linguistic Approaches to Bilingualism.* [https://doi.org/10.1075/lab.19092.sch].

Stevens K.N. 1997. Articulatory-acoustic-auditory relationships. – Hardcastle W.J., Laver J. (eds.). *The Handbook of Phonetic Sciences.* Oxford: Blackwell: 462–506.

Szigetvári P. 2020. Emancipating lenes – A reanalysis of English obstruent clusters. – *Acta Linguistica Academica* 67.1: 39–52. [https://akjournals.com/view/journals/2062/67/1/article-p39.xml].

Wojtkowiak E., Schwartz G. 2019. Prosody-segment interactions in the acoustics of Polish plosives. – Calhoun S., Escudero P., Tabain M., Warren P. (eds.). *Proceedings of the 19th International Congress of Phonetic Sciences,* Melbourne, Australia 2019. Canberra. [https://assta.org/proceedings/ICPhS2019/papers/ICPhS_3334.pdf].

Zampini M. 2008. L2 speech production research: Findings, issues, and advances. – Hansen Edwards J.G., Zampini M. (eds.). *Phonology and second language acquisition.* Amsterdam: John Benjamins: 219–249.