Asymmetrical cross-language phonetic interaction

Phonological implications

Geoffrey Schwartz
Adam Mickiewicz University in Poznań

Two acoustic studies were carried out with L1 Polish learners of English. One study examined L1 phonetic drift, comparing learners of L2 English who were undergoing intensive L2 phonetic training with quasi-monolingual Polish speakers. The other study looked at L2 acquisition, comparing learners at two different levels of proficiency. Unlike most previous studies of Polish-English bilinguals, VOT data of both voiced and voiceless consonants were analyzed. In both experiments, an asymmetry was observed by which voiced stops were more susceptible to cross-language phonetic influence (CLI) than voiceless stops. These results build on evidence of a similar asymmetry observed in a number of other L1–L2 pairings. Predictions of competing phonological models are evaluated with regard to equivalence classification and phonetic CLI. It is shown that both traditional approaches to the phonological representation of voice contrasts fail to predict the observed asymmetry. An alternative theory, which predicts the asymmetry, is discussed.

Keywords: Polish-English bilinguals, phonetics and phonology, laryngeal features, VOT, Onset Prominence

1. Introduction

There is a large literature documenting experimental research into the phonetic realization of phonological categories in the speech of bilingual language users. Phonetic interaction between languages (cross-linguistic phonetic interaction; henceforth CLI) has been observed for a large range of speakers and situations, including the second language (L2) speech of members of immigrant communities (e.g., Flege, 1987; Flege, Schirro & McKay, 2003), L2 learners receiving formal instruction in both first language (L1) dominant environments (Herd, Walden,

https://doi.org/10.1075/lab.19092.sch | Published online: 9 October 2020
Linguistic Approaches to Bilingualism 12:2 (2022), pp. 103–132. ISSN 1879-9264 | E-ISSN 1879-9272

Knight & Alexander, 2015) and L2 immersion settings (Chang, 2012), as well as balanced or simultaneous bilinguals who are equally proficient in both languages (Sancier & Fowler, 1997). Notably, it has been shown that phonetic CLI can be bidirectional – resulting not only in a foreign accent in L2 (Piske, McKay & Flege, 2001), but also affecting the pronunciation of one's first language (L1 phonetic drift; Chang, 2019). Research observing bidirectional CLI has been valuable for both the formulation and empirical foundations of a number of influential models of bilingual speech production and perception, including Flege's Speech Learning Model (SLM; Flege, 1995), and Best and Tyler's Perceptual Assimilation Model (PAM; Best, 1995; Best & Tyler, 2007).

A primary goal of mainstream bilingual and multilingual speech research, particularly with respect to acquisition, has been to explore the factors responsible for phonetic CLI. These factors, for the most part, boil down to bilingual users' linguistic experience (e.g. Flege, 1995; Flege, Bohn & Jang, 1997; Flege, Yeni-Komshian & Liu, 1999; Moyer, 2011). In particular, researchers have explored the effects of age of onset of L2 learning, length of residence in the L2 country, and ratio of use between L1 and L2 (for a review see Piske et al., 2001). At the same time, each of these areas may be rendered more complex by factors such as genealogical connections between interacting languages, speakers' attitude (e.g. Moyer, 1999), any third or additional languages studied (Wrembel, 2015) or phonetic talent (Lewandowski, 2012). Clearly, language experience is a multifaceted phenomenon that can be difficult to quantify or generalize across all situations.

In exploring how aspects of linguistic experience may or may not contribute to phonetic CLI, researchers must accept a certain set of assumptions about the phonological systems of the interacting languages under study. That is, in most mainstream research, the linguistic status of the particular phonetic or phonological parameter being researched is taken as a given, while the studies themselves seek to identify factors responsible for phonetic interaction. By contrast, much less attention has been devoted to the question of what CLI can tell us about the phonologies of the interacting languages. Since phonology is still an active field of study with many unresolved controversies, CLI data from interacting languages may provide a basis for interpretation or reinterpretation of a given language's phonological system.

In this connection, an under-researched question may be formulated as follows: if two languages differ across two phonetic dimensions, but CLI is observed in only one, how can we explain the asymmetry? From the perspective of the SLM, we might consider this question in terms of *equivalence classification* (Flege, 1987). When bilingual language users perceive L1 and L2 phones as members of a single category, they miss acoustic differences between the sounds, and CLI should be expected. Consequently, if two phonetic dimensions behave asymmetrically in cross-language situations, we may assume equivalence classification only in the dimension in which more robust CLI effects are observed. This may in turn be assumed to reflect phonological equivalence across the two languages. Likewise, minimal CLI may be associated with less robust equivalence classification, which suggests cross-language differences in phonological representation that underlie separate cross-linguistic categories.

This paper will present acoustic data from late Polish-English bilinguals in which phonetic CLI shows asymmetrical behavior, observable to a larger degree in one member of a phonological contrast than in the other. The focus here will be on the laryngeal realization of stop consonants in word/utterance-initial position (Section 2), operationalized in terms of voice onset time (VOT; Lisker & Abramson, 1964). Polish is traditionally described as a 'voicing' language with prevoicing (negative VOT) in its voiced series and short-lag VOT in its voiceless series (Section 4). This is in opposition to English, with long-lag VOT associated with aspirated /ptk/ (fortis stops) and short-lag VOT associated with /bdg/ (lenis stops). Two experiments will be described. Study 1 (Section 5) investigates L1 phonetic drift. A group of B2-level Polish speakers of English, who had completed and were still undergoing intensive L2 phonetic training, showed signs of English-induced L1 drift in their realization of Polish /bdg/, but not in their pronunciation of Polish /ptk/. Study 2 (Section 6) shows asymmetrical success in the acquisition of L2 English stops – English-style aspiration appears to be mastered to a greater degree and earlier in the acquisition process than unvoiced lenis /bdg/.

Although VOT has been widely studied in bilingual speech research, voicing is an area of persistent debate for phonological theory (Wetzels & Mascaró, 2001; Beckman, Jessen & Ringen, 2013; Bennett & Rose, 2017; Schwartz, 2017). As might be expected, opposing phonological perspectives on voicing make different predictions with regard to both the amount of CLI between laryngeal systems, as well as whether CLI in the voiced and voiceless series of consonants should show symmetrical behavior. Since testing phonological predictions has not been a priority of bilingual speech research, many empirical findings regarding CLI remain without a clear interpretation in terms of phonological theory. A primary goal of this paper is therefore to consider ways to explain CLI from the perspective of refined phonological predictions (see Sections 3 and 7).

2. Laryngeal contrasts and bilingual speech – a brief review

2.1 Findings involving voiceless stops

Some of the most commonly cited experiments on CLI in laryngeal contrasts involve the voiceless series of stops. Flege (1987) compared monolingual speakers of American English and French with four different groups of bilingual speakers. He found that VOT values for /t/ produced by three of the groups fell in between those of the monolingual controls (77 ms for English, 33 ms for French) in both languages, and that the proximity to the monolingual norms was in line with the level of proficiency and amount of usage of each language. Another study that has been influential is Sancier and Fowler (1997), who looked at the speech of a single Brazilian Portuguese-English bilingual who divides her time between Brazil and the United States. It was observed that this speaker's realization of /p/ and /t/ depended on which country she had been spending time in. Time in Brazil tended to shorten the VOT in accordance with the norms of Brazilian Portuguese, while time in the USA led to longer VOT. Interestingly, listeners in the two countries behaved different with regard to their ratings of the speaker's accentedness. Brazilian listeners were sensitive to the VOT increases associated with time in the USA, while American listeners did not hear changes induced by the speaker's time spent in Brazil. Another study by Flege (1991) looked at voiceless stop production by two groups of Spanish-English bilinguals, as well as by monolingual controls. Speakers in this study also produced 'compromise' values for the VOT of voiceless stops that were in line with expectations due to age of learning.

The production of voiceless stops in CLI situations by L1 speakers of Polish, the L1 examined in the present study, has been investigated by Waniek-Klimczak (2011), Sypiańska (2013), and Wrembel (2015). In each of these studies, when the target items were L2 (or L3) aspirated stops, results were more or less what might be expected – Polish speakers produced longer VOT as a function of proficiency and frequency of use. In other words, L1 Polish speakers of English appeared to acquire a new category for L2 (or L3) English voiceless stops. Results of these studies were somewhat more equivocal with regard to productions of L1 Polish /ptk/, i.e. when the authors looked for effects of L2 (and L3) on L1. Waniek-Klimczak (2011) found that more proficient speakers of English produced longer VOT of Polish voiceless stops than Polish monolinguals. By contrast, the L1 Polish productions of /ptk/ observed by Sypiańska (2013) and Wrembel (2015) were in line with monolingual reference values from the literature, apparently unaffected by the speakers' other languages. Unfortunately, none of these studies looked at the realization of /bdg/.

At the same time, there are a number of studies in which equivalence classification between aspirated and short-lag /ptk/ apparently does not occur, and bilingual users establish separate categories for the voiceless set. Flege & Eefting (1987a) describe perception and production experiments of L1 Dutch speakers of English. In the perception study, Dutch listeners showed a small shift in the voiced-voiceless category boundary as a function of whether they were listening to English or Dutch. This finding suggests that L1 Dutch listeners were aware of categorical differences between English and Dutch voiceless stops. In production, L1 Dutch learners of English showed success in the acquisition of long-lag English VOT. In a study dealing with Spanish-English bilinguals, Magloire & Green (1999) found VOT values of voiceless stops that were largely equivalent to those of monolingual speakers. That study was careful in controlling for language mode (Grosjean, 2004), and provided evidence for separate categories for voiceless stops in the two languages. A similar result was obtained by Antoniou, Best, Tyler, and Kroos (2010) in a study of Greek-Australian English bilinguals.

As mentioned before, much of the research on VOT in bilingual speech has been chiefly concerned with factors contributing to CLI. For this reason, comparing CLI in voiced and voiceless stops has not been a priority, and there are fewer studies looking at both series than there are that look only at voiceless stops. Additionally, I have yet to come across a study that examines only the voiced series. The apparent cause of these asymmetrical priorities stems from the commonly accepted notion that English aspiration is a 'marked' feature, which renders English /ptk/ different from /ptk/ in a large number of other languages. Nevertheless, some studies have examined both series, and a review of this research reveals an interesting asymmetry by which CLI appears to be more prevalent in the voiced series than in the voiceless series.

2.2 CLI in voiced vs. voiceless stops

Two studies investigating the interactions between American English and Spanish have asymmetrical CLI between the voiced and voiceless series. Herd et al. (2015) studied L1 phonetic drift in American English by speakers in an L1-dominant environment. In their study, it was observed that L1 American English speakers learning Spanish did not shorten the VOT of English /ptk/ relative to monolinguals, but L2 Spanish influence was found in a greater likelihood of prevoiced /bdg/ in the participants' productions. That study also found some L1 drift effects in vowel quality. These results suggest equivalence classification between American English and Spanish for vowels and voiced stops, but not voiceless stops. Zampini (1998) found a parallel asymmetry in an L2 acquisition situation. American English learners of Spanish were more successful in acquiring L2 /ptk/ with short-lag VOT than prevoiced L2 /bdg/.

Chionidou and Nicolaidis (2015) studied the speech of simultaneous Greek-German bilingual children, including a comparison of the dominant language of the school environment. They found significant but modest (around 5 ms) differences between bilingual and monolingual speakers of the two languages in the VOT of /ptk/. By contrast, /bdg/ were subject to more dramatic cross-linguistic interaction, with /bdg/ without prevoicing in Greek, and /bdg/ with prevoicing in German, produced in about 30% of the items. In another study involving L1 Greek speakers, Antoniou, Best, Tyler, and Kroos (2011) found that for Greek-Australian English bilinguals, a phonetic code-switching task induced VOT shifts toward L1 monolingual norms for both series. However, the amount of prevoicing of English /b/ and /d/ induced by the Greek context appeared to be more dramatic than the VOT shortening effect on English /p/ and /t/. In initial position, code-switched English targets were often produced with prevoicing (Antoniou et al. 2011, p. 563, Table 1). The mean VOT in monolingual mode was −8.4 ms for /b/, 9.0 ms for /d/, while the mean VOT of code-switched English voiced targets was −37.8 ms for /b/ and −55 ms for /d/. This appears to be a qualitative shift by which speakers were more likely to produce prevoiced English tokens under the influence of Greek. By contrast, aspirated English targets were shortened by about 10 ms in the code-switching task, but those shortened English targets still had long VOTs of 68.6 ms and 81.8 ms for /p/ and /t/, respectively.

Other studies that have examined laryngeal interactions between English and Dutch, Czech, Tagalog, and Polish have observed the same asymmetry. In separate studies of L1 dominant Dutch-English bilinguals, both Simon (2009) and Bless (2015) found a large amount of prevoicing in L2 English voiced /bdg/, while the VOT of the voiceless stops approximated the norms for monolingual English. Sučková (2018) examined the speech of expatriate L1 English speakers living in Czechia. She observed that the VOT of their L1 voiceless stops was for the most part unaffected by short-lag Czech /ptk/, while their voiced stops showed a large amount of Czech-style prevoicing. Kang, George, and Soo (2016) looked at heritage speakers of Tagalog in Toronto.[1] In their study, nine speakers produced English VOTs that closely matched those of a monolingual control group in the case of voiceless stops, but produced prevoicing in around 20% of the voiced stops (compared to less than 5% by the control group). With regard to Tagalog stops, Kang et al. (2016) found dramatically less prevoicing in the heritage group (<50% of tokens) than in the control group (>80% of tokens), but voiceless stops that were not statistically different from native speaker controls. Finally, Newlin-

1. I am grateful to a LAB referee for bringing my attention to this paper.

Łukowicz (2014) examined the speech of heritage Polish speakers in New York city, and observed greater Polish interference for the voiced set than the voiceless set. Table 1 presents a summary of findings of asymmetrical CLI between two-series laryngeal systems.

Table 1. Summary of cases of asymmetrical CLI in two-series laryngeal contrasts

Language pairing (L1–L2)	Reference(s)	Relevant result
American English-Spanish	Zampini (1998); Herd et al. (2015)	More successful acquisition of short-lag /ptk/ than prevoiced /bdg/ in L2 Spanish; L1 phonetic drift in AE /bdg/ (and vowel quality) but not in /ptk/.
Dutch-British English	Simon (2009); Bless (2015)	Persistent Dutch-style prevoicing in L2 English; but VOT of /ptk/ approach monolingual norms.
Greek-German	Chionidou & Nicolaidis (2015)	Modest CLI in VOT of /ptk/; more dramatic effects for /bdg/.
Greek-Australian English	Antoniou et al. (2011)	Greater drift toward L1 norms for /bdg/ than /ptk/ in phonetic code-switching task.
(British and US) English-Czech	Sučková (2018)	Prevoicing in L1 /bdg/ of L1 English speakers in Czechia; VOT of /ptk/ unaffected by Czech short-lag VOT.
Tagalog-English	Kang et al. (2016)	Bidirectional CLI for English and Tagalog voiced stops, but not voiceless stops.
Polish-US English	Newlin-Łukowicz (2014)	Greater Polish interference in English voiced stops than voiceless stops.

In sum, there is an increasing amount of evidence from a variety of language pairs that voiced /bdg/ are more susceptible than voiceless /ptk/ to phonetic CLI. This is not to claim that CLI is unheard of for /ptk/, nor that this asymmetry will always be observed. Rather, the point is that in those studies that examine both voiced and voiceless stops, if any asymmetries are observed, it is the voiced series that shows more dramatic CLI effects. Asymmetrical behavior of the type described here has no direct explanation within the SLM or other models of bilingual speech. Equivalence classification is based on perceived cross-linguistic phonetic similarity. Similarity is something of a vague notion, and theories of bilingual speech do not often address the question of which cross-language phonetic differences induce CLI and which do not. Nevertheless, it appears that in CLI situations, unvoiced and prevoiced /bdg/ are perceptually more similar to each other than short-lag and aspirated /ptk/. To explain this finding in a satisfactory manner requires a reconsideration of the phonological representation of two-series laryngeal contrasts, an issue we address in the following section.

3. Laryngeal phonology and asymmetrical CLI

In the tradition of Chomsky & Halle (1968), laryngeal contrasts are represented by means of a binary feature [voice], in which both [+voice] and [–voice] specifications may be phonologically active. Under this approach, the phonetic realization of the contrast in terms of aspiration and prevoicing is a function of language-specific rules of phonetic implementation. The binary approach assumes that VOT is a phonetic detail that is not specified phonologically, reflecting the fact that English aspiration is predictable rather than phonemic.

More recently, many scholars have accepted an approach that encodes the VOT typology directly. That is, two-way contrasts are assumed not to be a question of phonetic implementation of a single binary feature, but reflect the presence or absence of privative features, the choice of which is a function of whether a system is one of aspiration or voicing. This approach, dubbed Laryngeal Realism (LR; Honeybone, 2005; Beckman et al., 2013), is claimed to reflect phonetic reality more closely than the binary approach. Under LR, short-lag VOT associated with Polish /ptk/ and English /bdg/ reflects the lack of any laryngeal specification – it is 'unmarked'. Meanwhile, prevoicing is claimed to signify the presence of a unary feature [voice], while aspiration is encoded by means of a feature [spread glottis] ([sg]). As its name might suggest, phonemic contrast is less important for LR than providing a transparent link between phonetics and phonological specifications.

An alternative perspective on laryngeal phonology is offered by the Onset Prominence framework (OP; e.g. Schwartz 2016, 2017), which reconciles the conflicting claims of the binary approach and LR, and at the same time is predictive of the asymmetrical CLI behavior described in this paper. The key postulate of OP representations for encoding two-series laryngeal contrasts is that stop consonants are not 'segments', but rather hierarchical structures comprised of the three discretely identifiable phases (closure, noise burst, and CV transition) in their production (cf. Aperture Theory; Steriade, 1993, and Q Theory; Shih & Inkelas, 2019). For thorough discussion, see Schwartz (2016, 2017). Thanks to this postulate, OP can encode both voicing and aspiration systems with a single privative feature for /ptk/, while leaving /bdg/ unspecified. Whether the voiceless series is aspirated or not depends on the representational level at which the feature is assigned. In other words, the phonetic realization of /ptk/ is determined by the timing of the feature within the internal structure of the stop.

OP representations for two-series laryngeal contrasts are shown in Figure 1. The pair of trees on the left represents aspiration systems. Aspirated stops assign a [fortis] feature at the Closure (C) level, and the specification 'trickles' down the structure, occupying the Noise node (N) to produce aspiration, as well as the Vocalic Onset (VO) node. In the pair of trees on the right we see voicing systems, in which unaspirated /ptk/ assigns [fortis] at the VO level, leaving Noise unaffected and yielding short-lag VOT. Note that in both systems, the voiced series lacks any laryngeal specification, a postulate that we will return to in Section 7.

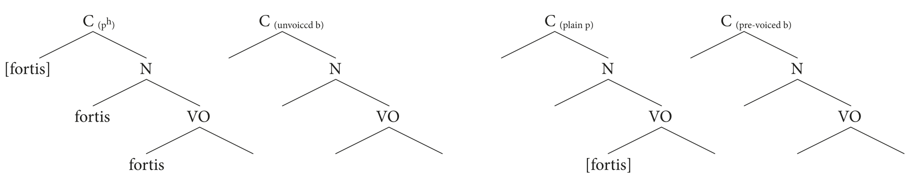

Figure 1. OP representations of aspiration (pair of trees on the left) and voicing systems

Table 2 summarizes the predictions of OP, the binary approach, and LR with regard to equivalence classification and Laryngeal CLI. Crucially, in both the binary approach and Laryngeal Realism, CLI is predicted to be symmetrical between the voiced series and the voiceless series. By contrast, the OP approach predicts asymmetrical behavior. Voiced stops have equivalent representations in voicing and aspiration languages, while short-lag and long-lag /ptk/ are represented differently, so CLI should be more prevalent in the voiced set. This is not to suggest that CLI should be completely absent in the voiceless set, which is clearly not the case. However, the prediction of the model is that CLI should be more robust in the voiced series.

4. The phonetics of laryngeal contrasts in Polish and English

Table 3 summarizes VOT findings in English and Polish. A few comments concerning this table are necessary. Firstly, there is evidence that Polish short-lag VOT values are somewhat longer than what is reported for other voicing languages, with mean values of around 30 ms for labials and coronals, and up to 50 ms for dorsals. Secondly, the English table conflates American (e.g. Lisker &

Table 2. Predictions of theoretical proposals for phonetic CLI

Theory	Representation (voiced vs. voiceless)	Predictions
Laryngeal Realism	[voice] vs. Ø in voicing systems; Ø vs. [sg] in aspiration systems	Minimal CLI; symmetrical behavior between /bdg/ and /ptk/
Binary	[+voice] vs. [–voice] in both types of system;	Dramatic CLI; symmetrical behavior between /bdg/ and /ptk/
OP	Ø vs. [fortis] at VO-level in voicing systems; Ø vs. [fortis] at C-level in aspiration systems	Asymmetrical behavior; more CLI for /bdg/ than /ptk/

Abramson, 1964) and British English (Docherty, 1992), which both show long-lag measures for fortis stops. Finally, the values in the table only cover short-lag measures for English /bdg/. Recently, however, prevoiced realizations of English voiced stops have become increasingly prevalent, particularly in accents found in the southern United States (e.g., Herd, 2017).

Table 3. Summary of VOT results for English and Polish

Series	Source: English VOT in ms (b/p, d/t, g/k)	Source: Polish VOT in ms (b/p, d/t, g/k)
/bdg/	– Lisker & Abramson (1964): 1, 5, 21 – Docherty (1992): 15, 21, 27 – Kim, Kim & Cho (2018): 15 (labial and coronal in trochaic words); 17 (labial and coronal in iambic words)	– Keating (1979): −88, −89, −66 – Malisz & Żygis 2015: −57, −46, −49 – Waniek-Klimczak (2011): no data – Sypiańska (2013): no data – Wrembel (2015): no data
/ptk/	– Lisker & Abramson (1964): 58, 70, 80 – Docherty (1992): 42, 64, 62 – Kim, Kim & Cho (2018): 71 (labial and coronal in trochaic words); 55 (labial and coronal in iambic words) • Keating (1979): 22, 28, 52	– Waniek-Klimczak (2011): 28, no /t/, 54 – Sypiańska (2013): 32, 35, 43 – Malisz & Żygis (2015): 18, 24, 35 – Wrembel (2015): 23, 30, 50

Beyond VOT, other phonetic correlates associated with laryngeal contrasts have also been documented. For example, voiceless consonants tend to show higher fundamental frequency (fo) at the onset of following vowels than voiced consonants (e.g., Hanson, 2009; Kirby & Ladd, 2016). Additionally, voiceless consonants have been found to clip CV transitions, with the effect of raising the first formant (F1) at vowel onset (Stevens & Klatt, 1974). Both of these effects, documented in Polish by Schwartz, Wojtkowiak & Brzoza (2019), are in line the OP representations in Figure 1, since it is at the Vocalic Onset level of structure, i.e. the beginning of the vowel, that voiceless stops in both types of two-series laryngeal system share a [fortis] specification.

5. Study 1: Asymmetrical L1 phonetic drift in the speech of Polish learners of English

5.1 Participants

The experiment compared two groups of speakers. The first group (students) was a set of 18 2nd year students of English at a Polish university. To gain admission to the English studies program at our university, students must pass the expanded *Matura* exam in English at the end of secondary school. Passing this exam may be equated with achieving a B1 (Threshold) level according to the Common European Frame of Reference (CEFR) for languages (Council of Europe, 2001). In their first year of the program, students complete 120 contact hours of speaking courses, 60 contact hours of grammar, 60 contact hours of writing, and 120 contact hours of pronunciation training. Additionally, students complete a 60 contact hour theoretical course on English phonetics and phonology. The students had successfully completed all of these courses, and were in the process of undergoing an additional 60 contact hours of practical courses devoted to each of these areas (speaking, grammar, writing, pronunciation). Overall, second year students in our program have at least a B2 (Vantage) level of proficiency in English, with both theoretical and practical training in phonology and phonetics. All of the students were female, and between the ages of 20 and 22.

The second group, which will be referred to as quasi-monolinguals, consisted of 15 L1 Polish speakers, recruited from elementary English courses at a private language school. This group did have some basic knowledge of English, since English is obligatory in the Polish school system. However, their level was quite low (A1–A2 according to the CEFR). They were capable of producing only simple utterances dealing with familiar situations, and their pronunciation is strongly marked by features of a Polish accent (substituting Polish vowels in English words, tapped /r/, final obstruent devoicing, regressive voice assimilation, etc.). Thus, to the extent that was feasible in an era when English is ubiquitous, this group can be considered functionally monolingual, an appropriate control group for comparison with the students. The quasi-monolingual group was made up of female speakers between the ages of 17 and 38 (with a median age of 25).

5.2 Materials and procedure

Experimental materials were comprised of forty-eight monosyllabic and disyllabic words in Polish (Appendix 1), each of them beginning with a stop consonant (24 voiced, 24 voiceless). The dataset was counterbalanced for place of articulation of the stop (16 labial, 16 coronal, 16 dorsal), and the following vowel was always a non-high vowel, as high vowels have been found to lengthen VOT (Maddieson, 1997). With a total of 33 speakers (18 Students and 15 Monolinguals), this yielded 1584 tokens (864 from the Students, 720 from the Monolinguals).

Recordings were made directly onto a laptop computer using a high quality head-mounted microphone and USB audio interface. Recordings of the Students were made in a sound treated recording booth at the university. Recordings of the quasi-monolinguals were made in a quiet classroom setting. Productions were elicited using Powerpoint slides, presented in a pseudo-randomized order. Each slide contained a single word. The first and last three slides in the presentation were fillers that were not analyzed. Speakers were instructed to read the word on the slide, after which the experimenter advanced the Powerpoint presentation to the next slide. The recording sessions were carried out by a native speaker of Polish, and instructions were given in Polish. Thus, this was clearly an L1 task, in which language mixing effects (Grosjean, 2004) were not expected.

5.3 Acoustic and statistical analysis

Annotation of the recordings was performed in Praat (Boersma and Weenink, 2017). VOT was marked according to the standard criteria in Praat text-grids, and durations were later extracted by script. For prevoiced items, VOT was labeled as the time window from the onset of periodicity during stop closure to the moment of stop release indicated by a noise burst in the waveform. Such measurements of course yielded positive numbers, so the VOT calculation from the script needed to be adjusted accordingly. For items without prevoicing, VOT was defined as the time window from stop release to the onset of periodicity in the waveform. In the case of /bdg/, in addition to VOT, the realizations were classified as either prevoiced or unvoiced.

Statistical analysis was carried out with the help of the SPSS Statistical Software (IBM Corporation, 2016). For VOT, a generalized linear mixed model was fitted with a group*voicing interaction term as the main predictor of interest. Treatment coding of these two variables allowed for a direct test of the significance of differences between the two predictors, while the choice of reference levels (quasi-monolinguals for group; voiceless for voicing) mediated between two opposing hypotheses, based on previous studies and the phonological discussion in Section 3, that voiceless stops will be either susceptible or resistant to CLI. The model included a by-speaker random slope for voicing, a by-item random slope for group, and by-speaker and by-item random intercepts. Additionally, a mixed-effects binary logistic regression model was run on /bdg/ items, which had been coded as prevoiced/unvoiced. Unvoiced /bdg/ (yes/no) was the dependent variable. Group was the main predictor, with quasi-monolinguals as the reference level, reflecting an expected default of voiced realizations. The logistic model included by-speaker and by-item random intercepts, and a by-item random slope for group.

5.4 Results

Mean VOT values (and standard deviations) for the two groups are given in Table 4. The VOT results are also shown in the line graphs in Figure 2. As can be seen in both the table and the figure, the students showed shorter negative VOT (less prevoicing) in /bdg/ than the quasi-monolinguals, but equivalent VOT measures in the voiceless series. Contrast estimates from the linear model revealed a significant group-based difference for the voiced series (Estimate = −27.86 ms; $SE = 3.866$; $t = -7.21$; $p < .001$), but not the voiceless series (Table 5: monolingual*voiceless). Full results of the linear model are in given Table 5.

Table 4. Mean VOT in milliseconds (and standard deviation) for L1 Polish stops

Series	Quasi-monolinguals	Students
voiced	−92.3 (27)	−64.2 (35)
voiceless	41.7 (16)	42.3 (17)

With regard to /bdg/ realization, the quasi-monolingual group produced only 2 unvoiced items (0.6% of their total), while the students produced 39 (9.1% of their total). The logistic regression model revealed that this effect was significant – the students were more likely to produce unvoiced /bdg/ items. These results are summarized in Table 6.

Table 5. Output of linear model for Polish stops – VOT results, Intercept is monolinguals*voiceless

Model term	*β*	SE	*t*	*p*	95% C.I. – Lower	95% C.I. – Upper
intercept	41.944	18.4024	2.279	.023	5.848	78.039
monolinguals*voiced	−133.561	4.6483	−28.733	<.001	−142.678	−124.443
students*voiceless	−0.945	3.864	−0.245	*ns.*	−8.524	6.634
students*voiced	−105.701	2.9664	−35.632	<.001	−111.519	−99.882

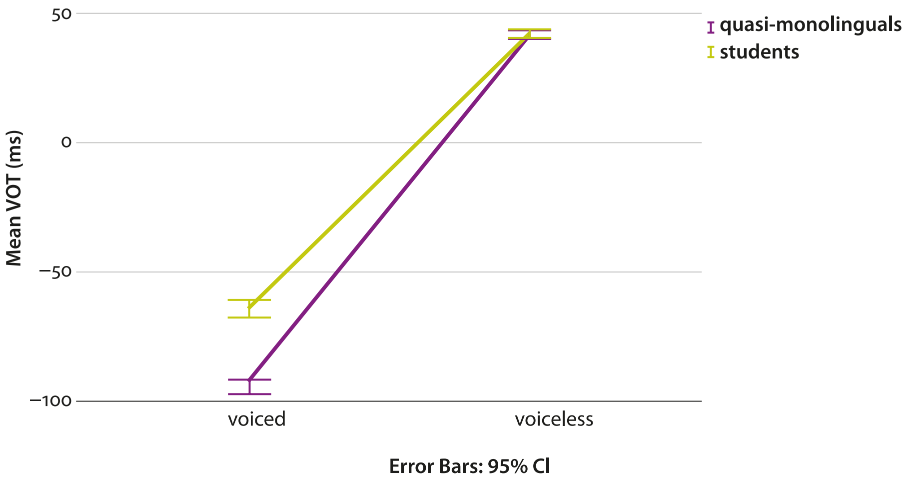

Figure 2. Line graph of mean VOT values for Polish stops. Error bars denote 95% confidence intervals

Table 6. Results of logistic regression model for Polish unvoiced /bdg/ realizations

Term	*β*	Exp(*β*)	SE	*t*	*p*
intercept	−5.357	0.005	1.089	−4.92	<.001
group:students	2.61	13.599	0.884	2.95	.003

Figures 3 and 4 present scatterplots of VOT values for voiceless and voiced stops in Polish, respectively. On the x-axis in the plots, speakers are arranged in order of their mean intra-speaker VOT (lowest values to highest). Speakers labeled “S” (green dots) are from the students group. Speakers labeled “M” (purple crosses) are from the quasi-monolingual group. In the case of the voiceless stops (Figure 3), "M" speakers and "S" speakers are interspersed more or less evenly throughout the range of intra-speaker mean VOT values. A strikingly different pattern is observed for voiced stops (Figure 4): the 10 highest mean VOT values are from the students group, while 8 out of the 9 lowest intra-speaker VOT means are from the quasi-monolingual group.

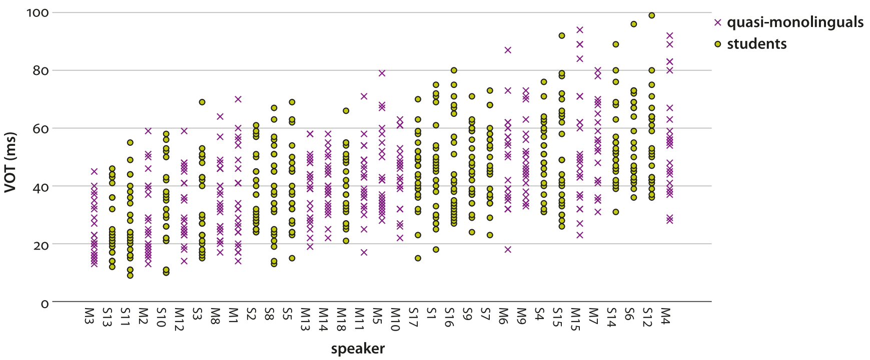

Figure 3. Scatter plot of VOT measures by speaker for Polish /ptk/

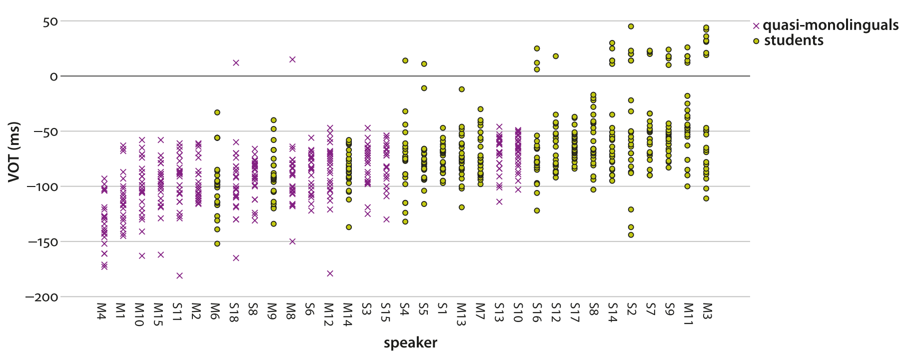

Figure 4. Scatter plot of VOT measures by speaker for Polish /bdg/

5.5 Discussion

The L1 drift study revealed group-based differences only in the case of the voiced set. The students, with B2-level English proficiency and explicit training in English pronunciation, produced less prevoicing in L1 Polish than the quasi-monolingual group, but they did not lengthen the VOT of voiceless stops. Thus, the results suggest that the Students experienced L2-induced phonetic drift only in the case of /bdg/ – exposure to unvoiced lenis stops in English appears to affect the likelihood and magnitude of prevoicing in Polish. The VOT of voiceless stops appears not to have been affected.

Assuming equivalence classification is a cause of CLI in L1 drift situations, the logical interpretation of these results is that /bdg/ are classified as the same phonetic category in both of the students' languages, while /ptk/ are perceived as different, and are therefore resistant to drift. The individual results shown in Figure 3 and 4 are for the most part compatible with this overall picture. The range of intra-speaker VOTs for voiced stops was distributed largely according to speaker group, while the range for voiceless stops was interspersed with speakers from both groups.

These findings may be interpreted in terms of what we have already discussed with regard to the phonetics of the laryngeal contrast in Polish, and how it might be encoded phonologically. If equivalence classification is what drives English-induced L1 drift for Polish /bdg/, we might expect the perceptual weight of prevoicing in L1 Polish to be relatively low, since bilingual speakers apparently disregard Polish prevoicing in establishing cross-linguistic correspondences. This interpretation finds support in three previous studies of the voicing perception in Polish (Keating, Mikos & Ganong, 1981; Schwartz & Arndt, 2018; Schwartz et al., 2019).

Issues relating to the perceptual weight of prevoicing in Polish, and bearing on the question of equivalence classification and cross-language perception of laryngeal contrasts, also play a role in determining the success of L2 phonological acquisition, the subject of Study 2.

6. Study 2: Asymmetrical achievement in L2 phonological acquisition

6.1 Participants

The L2 study compared two groups of Polish speakers of English. The first group was comprised of 1st year students (n=10; 3 male, 7 female, all between the ages of 18 and 20) of English at a Polish university, with a proficiency level of B1 according to the CEFR. This is the level that is demanded for admission to the English program. The second group (n=14; 6 male, 8 female, between the ages of 27 and 45 with a median age of 32) was comprised of professors and Ph.D. students in the same English program at the same university. These speakers had C2 proficiency in English, and had all been through the extensive phonetic training the Students were just starting. Crucially, the two groups represent L1 Polish speakers of English before and after two years of University-level instruction in English phonetics. This is in contrast to Study 1, which compared quasi-monolinguals with no phonetic training to a group of students who were undergoing training at the time. Due to these differences, the two studies were carried out separately and participant groups were not merged.

6.2 Materials, procedure and analysis

Experimental materials were comprised of a set of 54 English monosyllabic words starting with initial stops /p t k b d g/ (see Appendix 2). The list was counterbalanced for place of articulation of the stops, and the following vowel was non-high. The recording procedure was similar to that in the L1 drift study described earlier, except that all of the recordings were made in the recording studio at the Polish university, and the recording sessions and instructions were all carried out in English. Annotation procedures and statistical analyses were the same as in the previous study.[2]

6.3 Results

Mean VOT values (and standard deviations) for the two groups are given in Table 7. The VOT results are also shown in the line graphs in Figure 5 (error bars denote 95% confidence intervals). The professors/PhD students showed shorter prevoicing in /bdg/ than the 1st year students, but the VOT of voiceless stops did not differ significantly. Contrast estimates from the linear model revealed a significant group based difference for the voiced series (Estimate = −19.678 ms; $SE = 9.075$; $t = -2.17$; $p = .03$), but not the voiceless series (Table 8: students*voiceless). Full results of the linear model are given in Table 8 (Intercept: professors/PhD students*voiceless).

Table 7. Mean VOT in milliseconds (and standard deviation) for English stops for the two speaker groups

Series	Students	Professors/PhD students
voiced	−83.1 (57)	−62.6 (56)
voiceless	62.4 (35)	67.8 (25)

With regard to /bdg/ realization, 1st year students produced /bdg/ as unvoiced 14.9% of the time, while the professors/PhD students produced

2. The reference levels for the linear analysis were profs/PhD students for group, and voiceless for voicing. For the logistic analysis of /bdg/ realization the reference level was students for group. The motivation behind these choices was analogous to the L1 drift study.

Table 8. Output of linear model for English stops – VOT results

Model term	*β*	SE	*t*	*p*	95% C.I. – Lower	95% C.I. – Upper
intercept	67.821	32.322	2.098	.036	4.413	131.23
students*voiced	−150.105	9.5912	−15.65	<.001	−168.921	−131.289
students*voiceless	−6.406	9.0584	−0.707	.48	−24.177	11.364
profs/ PhD*voiced	−130.427	4.1365	−31.531	<.001	−138.542	−122.312

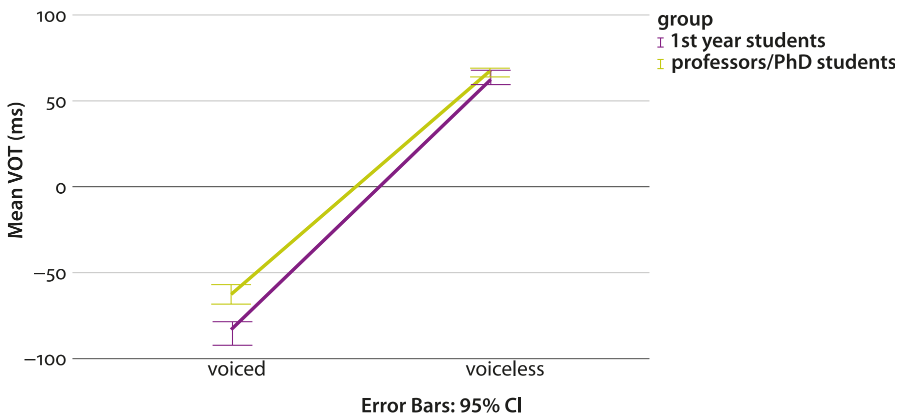

Figure 5. Line graph of mean VOT values for English stops. Error bars denote 95% confidence intervals

unvoiced /bdg/ at a rate of 27.5%. Although the professors/PhD students produced a greater percentage of unvoiced /bdg/ realizations, the mixed effects logistic regression model revealed a non-significant effect of group (Table 9).

Table 9. Results of logistic regression model for English unvoiced /bdg/ realizations

Term	*β*	Exp (*β*)	SE	*t*	*p*
intercept	−2.369	.094	1.54	−1.53	*ns.*
group:PhD students/profs	0.948	2.581	0.741	1.28	*ns.*

Figures 6 and 7 present scatterplots of VOT values for individual speakers for voiceless and voiced stops in English, respectively. On the x-axis in the plots, speakers are arranged in order of their mean intra-speaker VOT (lowest to highest). Speakers labeled with numbers are from the students' group (purple crosses).

Speakers labeled "A" (green dots) are from the professors/PhD group. Notice that in the case of the voiceless stops (Figure 6), speakers from the two groups are interspersed more or less evenly throughout the range of intra-speaker mean VOT values. For voiced stops (Figure 7), nine of the eleven highest /bdg/ mean VOT values are from the professors/PhD group, while roughly half of the unvoiced items produced by the students are attributable to a single speaker. Additionally, 4 out the 12 lowest intra-speaker VOT means come from the professors/PhD group.

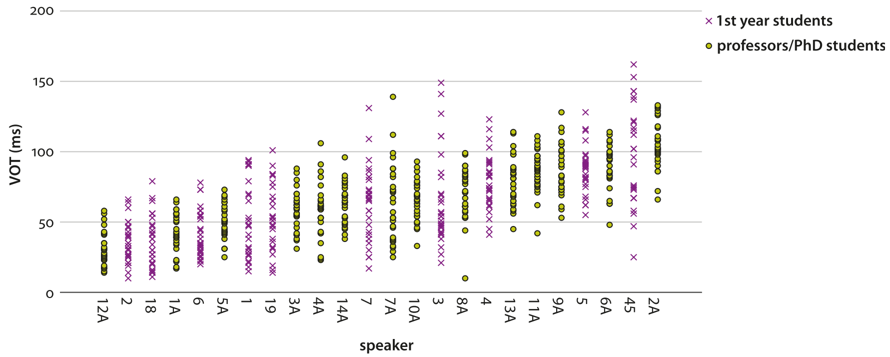

Figure 6. Scatter plot of VOT measures by speaker for English /ptk/

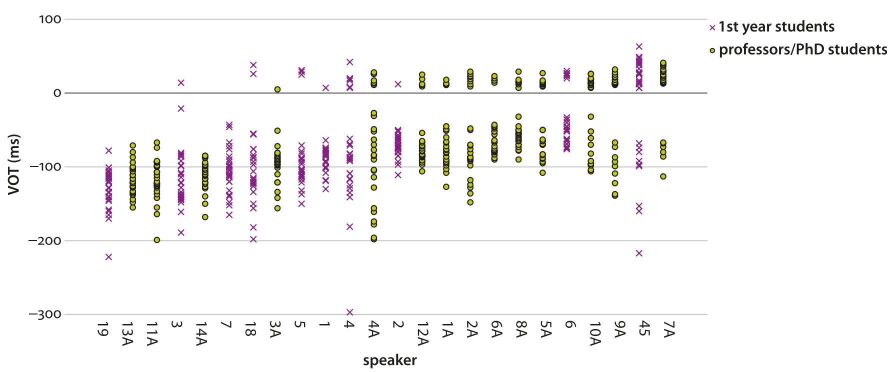

Figure 7. Scatter plot of VOT measures by speaker for English /bdg/

6.4 Discussion

In discussing the results of the L2 production study, we start with the realization of voiceless stops. For /ptk/, there was no significant difference between the groups in terms of VOT. The most likely interpretation of this result is speakers in the students' group, which has already achieved B1-level proficiency in English for admission to the university program, had 'acquired' the long-lag VOT associated with English voiceless stops. While we make no claims about whether this acquisition is on par with the norms of monolingual English,[3] it must be remembered that the teachers' pronunciation serves as a model for the students. Evidently, with regard to aspiration, learners acquire the norms of their teachers relatively quickly. This interpretation is compatible with Simon's (2009) suggestion that aspiration is 'salient' and relatively easily acquired.

In the case of voiced stops, however, group-based differences were observed in VOT values. The students produced longer negative VOT (more prevoicing) than the professors/PhD students. Thus, students' progress in suppressing L1-style prevoicing is delayed relative to their acquisition of long-lag VOT that is typical in English. At the same time, however, prevoicing was quite prevalent in the productions of the professors/PhD students, and the logistic regression analyses yielded non-significant results. In other words, although prevoicing was overall less prevalent for the professors/PhD students, statistically they were not significantly more likely to produce unvoiced tokens, despite the presence of a trend in this direction. The implication of this finding is that the suppression of prevoicing constitutes a persistent challenge for Polish speakers of English, even those with a very high level of proficiency.

Notably, for the group of professors/PhD students, it had been several years since they received explicit pronunciation training, which is part of the curriculum only for the first two years of the program. In this connection, a pilot study with 3rd year students and professors in the same university program (Schwartz & Dzierla, 2017) found /b/ realized as unvoiced about 33% of the time. In other words, 3rd year students who had just completed their two years of pronunciation training produced more native-like /bdg/ without prevoicing than the professors/PhD students in this study. Thus, it seems that with regard to prevoicing suppression, the more advanced group may have experienced something of a 'regression' in the years after their explicit pronunciation training had ceased (see Kartushina & Martin, 2019 for an analogous finding dealing with vowels; see also Chang, 2013, who observes partial reversal of L1 drift effects after L2 exposure had decreased). By contrast, apparently no such regression has occurred in the production of English voiceless stops with aspiration. Considering this result from the perspective of the SLM, it may be suggested that due to equivalence

3. The values here are more or less in line with what Docherty (1992) found for British English. Most studies of American English have yielded higher measures for /ptk/. The L2 speakers in this study typically learn British English, but have a great deal of exposure to American English from movies and television. Nevertheless, a comparison of British and American VOT values for /ptk/ is largely beyond the scope of this paper. They both fall in the category of long-lag, as do most of the items produced by the L2 speakers in this study.

classification, long-term acquisition of unvoiced /bdg/ in these speakers is hindered, but a similar effect is not observed for /ptk/. That is, long-lag VOT in /ptk/ appears to be learned earlier than unvoiced /bdg/, and this acquisition is more resistant to regression later on.

In sum, the results of the L2 production study, like those of the L1 drift study described earlier, show an asymmetry in CLI between voiced and voiceless stops. Acquisition of aspirated /ptk/ is more successful than the acquisition of unvoiced lenis /bdg/. This finding is consistent with the claim that equivalence classification is more likely for the voiced set than the voiceless set.

7. General discussion – voicing without [voice]

The experiments described in this paper add to a growing body of evidence that CLI between two-series laryngeal systems is asymmetrical, more prevalent in voiced stops than in voiceless stops. From the point of view of phonological theory, to the best of my knowledge only the Onset Prominence (Section 3) approach is predictive of this asymmetry. Crucially, under OP the voiced series is representationally equivalent in the two types of system, and is not specified for any laryngeal feature, while the voiceless series is not equivalent, facilitating new category formation in line with the SLM. The essence of the OP proposal is that there is no feature [voice] in either Polish or English.

The postulate that languages do not have a feature [voice] raises theoretical questions about the relationship between phonological specifications and the physical properties of speech. The most obvious question concerns the status of periodicity, which when it occurs in the production of obstruents is widely assumed to reflect the presence of a phonological specification. Indeed, I would claim that the vast majority of both phoneticians and phonologists take for granted that phonetically voiced obstruents are specified for a phonological feature [voice]. The representations shown in Figure 1 represent a significant departure from mainstream phonological thinking, and a few words in support of this perspective are necessary. For further discussion, see Schwartz (2017).

Conceptual arguments in favor of an approach without [voice] may be framed in terms of Traunmüller's (1994) Modulation Theory. In Modulation Theory, speech perception involves the demodulation of an acoustic *carrier signal*. Listeners extract content that is linguistic, including phonological specifications, from an acoustic signal that also encodes extra-linguistic information about the speaker, such as sex, age, and emotional state. In other words, Traunmüller's theory is built on the observation that listeners distinguish the phonological content of an utterance from the acoustic background upon which is transmitted, the carrier. However, since it carries extra-linguistic information, the carrier cannot be not silent. Rather, Traunmüller suggests it is a voiced, schwa-like vocoid that serves as an acoustic 'canvas' for phonological features. Viewers of paintings rarely notice the canvas. Similarly, in parsing the speech signal, listeners usually focus on the content, rather than the carrier. Since the carrier is voiced, it follows that demodulation cannot involve extracting a universal feature [voice] based solely on periodicity. Any phonological role for phonation must be epiphenomenal and language-specific. In other words, a phonological feature based on voicing may emerge (cf., Mielke, 2008) in individual languages, but phonation is not a universal primitive of phonological representation.

As a consequence of this claim, we must consider the phonological status of voicing during stop closure. In the OP/Modulation approach to laryngeal phonology, prevoicing is a phonetic effect without phonological relevance. Thus, the prediction is that for listeners of voicing languages, when prevoicing is absent it should not induce voiceless percepts. This prediction has been borne out in perception studies employing a number of different experimental paradigms. In Dutch, Flege & Eefting (1987a) observed a VOT boundary of 33.5 milliseconds between voiced and voiceless stops, while in Spanish, the same authors found a boundary of 23.1 milliseconds (Flege & Eefting, 1987b). If prevoicing reflected the presence of a feature [voice] in these languages, we would expect a category boundary of 0 milliseconds. In a discrimination task with Polish listeners, Schwartz & Arndt (2018) found that Polish listeners discriminate the laryngeal contrast highly accurately even in cases where prevoicing is absent. In that study, response times for unvoiced /bdg/ vs. /ptk/ pairs were no slower than pairs containing prevoicing. Likewise, Schwartz et al. (2019) carried out a phoneme monitoring task with Polish listeners. In the monitoring experiment, as in the discrimination study, the lack of prevoicing did not hinder listeners in identifying voiced stops. In short, there is evidence that in voicing languages, there is much more to the perception of /bdg/ than prevoicing, a fact that casts doubt on the common assumption that the contrast is encoded with a feature [voice].

Of course, in the absence of a feature [voice], we must explain what is responsible for the difference between voiced and unvoiced realizations of /bdg/. In short, prevoicing of /bdg/ reflects the emergence of the carrier signal in positions without a feature [fortis] that could provide a salient modulation, while a lack of voicing in /bdg/ is due simply to aerodynamic constraints on phonation. This claim makes a number of additional predictions that also appear to be borne out. First and foremost, it correctly predicts that languages should not show phonemic contrasts between short and long prevoicing the way they do between short-lag and long-lag positive VOT. Such a contrast would require listeners to distinguish between two types of phonation: one type encoding a phonological specification and the other type representing the carrier.

Another prediction concerns the realization of closure voicing, which due to aerodynamic constraints often ceases in the final stages of closure before stop release. Davidson (2016) refers to this as voicing 'bleed'. She observes that in intervocalic contexts periodicity carries over from the preceding vowel into stop closure, showing a constant decrease in amplitude regardless of whether phonation in fact ceases. While intervocalic 'bleed' is a phonetic effect that is independent of any [voice] specification, prevoicing in initial position shows an analogous decrease in amplitude, often leading to breaks before release, which lend support to the [voice]-less interpretation. Schwartz et al. (2019) found that breaks in voicing in Polish initial stops contribute to longer negative VOT measures. Assuming a feature [voice], this finding yields an undesirable theoretical contradiction. On the one hand, the breaks suggest a weakening of [voice], but on the other hand the longer negative VOT measures suggest longer segment duration of voiced stops, and longer segments are typically assumed to be a sign of phonological strength. Thus, positing a feature [voice] for Polish implies an interpretation that voicing breaks are simultaneously both a weakening and strengthening. The OP approach avoids such an undesirable contradictory claim.

Finally, the OP/Modulation perspective predicts that the realization of /bdg/ in aspiration languages should be subject to a great deal of variation. This has indeed been observed in North American English. Prevoicing in /bdg/ is quite common in a number of dialects in the Southern United States (Jacewicz, Fox & Lyle, 2009; Herd, 2017), yet speakers of these dialects do not differ from standard speakers in the VOT of /ptk/ (Hunnicutt & Morris, 2016; Herd, 2017). In other words, even when prevoicing is present, these are still 'aspiration' systems – prevoicing is just an alternative realization of the lenis stops. Likewise Swedish (Helgason & Ringen, 2008) and Turkish (Petrova, Plapp, Ringen & Szentgyorgyi, 2006) have been described as languages with aspiration and prevoicing. Phonological descriptions of such systems are simplified if it is assumed that they contain no active feature [voice] – the difference between Swedish-type and English-type systems need not be phonologically specified.

In conclusion, we may note that the asymmetry described in this paper lends support to a non-mainstream phonological framework, and points out the complementary relationship between phonological theory and bilingual speech research. New phonological theories make new predictions about CLI to help us better understand empirical observations from bilingual speech. At the same time, bilingual speech data can inform phonological theory, since patterns of CLI must be derivable from phonological representations, which may be in in need of refinement.

Funding

This research was supported by a grant from the Polish National Science Center (Narodowe Centrum Nauki), project number UMO-2016/21/B/HS2/00610, awarded to Geaffrey Schwartz.

Acknowledgments

Thanks to Jerzy Dzierla, Ewelina Wojtkowiak, Kamil Kaźmierski, participants of the studies, and the editors of *LAB*.

References

Antoniou, M., Best, C., Tyler, M., & Kroos, K. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, *38*, 640–653. https://doi.org/10.1016/j.wocn.2010.09.005

Antoniou, M., Best, C., Tyler, M., & Kroos, K. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: a study in phonetic code-switching. *Journal of Phonetics 39*, 558–570. https://doi.org/10.1016/j.wocn.2011.03.001

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, *49*, 259–284. https://doi.org/10.1017/S0022226712000424

Bennett, W., & Rose, S. (2017). Moro voicelessness dissimilation and binary [voice]. *Phonology*, *34*, 473–505. https://doi.org/10.1017/S0952675717000252

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. Munro & O.-S. Bohn (Eds.), *Second language speech learning – the role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Bless, M. (2015). Interference in Early Acquisition Dutch-English Bilinguals: A Phonetic Examination of Voice Onset Time in Dutch and English Bilabial Plosives. Master's Thesis, University of Amsterdam.

Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer (version 6.0.29) [Computer program]. http://www.praat.org/, accessed March 2019.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, *40*, 249–268. https://doi.org/10.1016/j.wocn.2011.10.007

Chang, C. B. (2013). A novelty effect in the phonetic drift of the native language. *Journal of Phonetics*, *41*, 520–533. https://doi.org/10.1016/j.wocn.2013.09.006

Chang, C. B. (2019). Phonetic drift. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 191–203). Oxford, UK: Oxford University Press.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Chionidou, A., & Nicolaidis, K. (2015). Voice onset time in bilingual Greek-German children. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK. http://www.icphs2015.info/pdfs/Papers/ICPHS0595.pdf

Council of Europe. (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge, U.K.: Press Syndicate of the University of Cambridge.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, *54*, 35–50. https://doi.org/10.1016/j.wocn.2015.09.003

Docherty, G. (1992). *The Timing of Voicing in British English Obstruents*. Berlin: de Gruyer. https://doi.org/10.1515/9783110872637

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. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, J. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*(1), 395–411. https://doi.org/10.1121/1.400473

Flege, J. (1995). Second language speech learning – theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J., & Eefting, W. (1987a). Cross-langauge switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*(3), 185–202. https://doi.org/10.1016/0167-6393(87)90025-2

Flege, J., & Eefting, W. (1987b). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*, 67–83. https://doi.org/10.1016/S0095-4470(19)30538-8

Flege, J., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470. https://doi.org/10.1006/jpho.1997.0052

Flege, J., Yeni-Komshian, G., & Liu, H. (1999). Age constraints on second language acquisition. *Journal of Memory and Language*, *41*, 78–104. https://doi.org/10.1006/jmla.1999.2638

Flege, J., Schirro, C., & MacKay, I. (2003). Interaction between native and second language phonetic subsystems. *Speech Communication*, *40*, 467–491. https://doi.org/10.1016/S0167-6393(02)00128-0

Grosjean, F. (2004). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*, 131–149. https://doi.org/10.1017/S136672899800025X

Hanson, H. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *Journal of the Acoustical Society of America*, *125*(1), 425–441. https://doi.org/10.1121/1.3021306

Helgason, P., & Ringen, C. (2008). Voicing and aspiration in Swedish stops. *Journal of Phonetics 36*(4), 607–628. https://doi.org/10.1016/j.wocn.2008.02.003

Herd, W. (2017). Southern stops: Phonation type differences in Mississippi. *Journal of the Acoustical Society of America 142*(4), 2680–2680. https://doi.org/10.1121/1.5014769

Hunnicutt, L., & Morris, P.A. (2016). Prevoicing and Aspiration in Southern American English. *University of Pennsylvania Working Papers in Linguistics 22*(1), Article 24. Available at: https://repository.upenn.edu/pwpl/vol22/iss1/24

Herd, W., Walden, R., Knight, W., & Alexander, S. (2015). Phonetic drift in a first language dominant environment. *Proceedings of Meetings on Acoustics 23*. https://doi.org/10.1121/2.0000100

Honeybone, P. (2005). Diachronic evidence in segmental phonology: The case of obstruent laryngeal specifications. In M. van Oostendorp & J. van de Weijer (Eds.), *The internal organization of phonological segments* (pp. 319–354). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110890402.317

IBM Corporation. (2016). *IBM SPSS Statistics for Windows*, Version 24.0. Armonk, NY: IBM Corp.

Jacewicz, E., Fox, R., & Lyle, S. (2009). Variation in stop consonant voicing in two regional varieties of American English. *Journal of the International Phonetic Association*, *39*, 313–334. https://doi.org/10.1017/S0025100309990156

Kang, Y., George, S., & Soo, R. (2016). Cross-language influence of stop voicing contrast in Heritage Tagalog. *Heritage Language Journal*, *13*(2), 184–218. https://doi.org/10.46538/hlj.13.2.6

Kartushina, N., & Martin, C. D. (2019). Third-language learning affects bilinguals' production in both their native languages: A longitudinal study of dynamic changes in L1, L2 and L3 vowel production. *Journal of Phonetics*, *77*, article 100920. https://doi.org/10.1016/j.wocn.2019.100920

Keating, P. (1979). A phonetic study of a voicing contrast in Polish. Ph.D. dissertation, Brown University.

Keating, P., Mikos, M., & Ganong, W. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America*, *70*, 1261–1271. https://doi.org/10.1121/1.387139

Kim, S., Kim, J., & Cho, T. (2018). Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics*, *71*, 65–80. https://doi.org/10.1016/j.wocn.2018.07.004

Kirby, J., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel f0: evidence from 'true voicing' languages. *Journal of the Acoustical Society of America*, *140*, 2400–2411. https://doi.org/10.1121/1.4962445

Lewandowski, N. (2012). Talent in non-native phonetic convergence. PhD dissertation, University of Stuttgart.

Lisker, L. & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word 20*, 384–422. https://doi.org/10.1080/00437956.1964.11659830

Maddieson, I. (1997). Phonetic universals. In W. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 619–639). Oxford: Blackwell.

Magloire, J., & Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica*, *56*, 158–185. https://doi.org/10.1159/000028449

Malisz, Z. & Żygis, M. (2015). Voicing in Polish: interactions with lexical stress and focus, *Proceedings of the 18th ICPhS in Glasgow*.

Mielke, J. (2008). *The emergence of distinctive features*. Oxford: Oxford University Press.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, *21*, 81–108. https://doi.org/10.1017/S0272263199001035

Moyer, A. (2011). An investigation of experience in L2 phonology: Does quality matter more than quantity? *The Canadian Modern Language Review*, *67*, 191–216. https://doi.org/10.3138/cmlr.67.2.191

Newlin-Łukowicz, L. (2014). From interference to transfer in langauge contact: Variation in voice onset time. *Language Variation and Change*, *26*, 359–385. https://doi.org/10.1017/S0954394514000167

Petrova, O., Plapp, R., Ringen, C. & Szentgyorgyi, S. (2006). Voice and aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish. *The Linguistic Review*, *23*, 1–35. https://doi.org/10.1515/TLR.2006.001

Piske, T., MacKay, I. & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics 29*, 191–215. https://doi.org/10.1006/jpho.2001.0134

Sancier, M. & Fowler, C. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*, 421–437. https://doi.org/10.1006/jpho.1997.0051

Schwartz, G. (2016). On the evolution of prosodic boundaries – parameter settings for Polish and English. *Lingua*, *171*, 37–74. https://doi.org/10.1016/j.lingua.2015.11.005

Schwartz, G. (2017). Formalizing modulation and the emergence of phonological heads. *Glossa – a journal of general linguistics*, *2*(1), 1–20.

Schwartz, G. & Arndt, D. (2018). Laryngeal Realism vs. Modulation Theory – evidence from VOT discrimination in Polish. *Language Sciences*, *69*, 98–112. https://doi.org/10.1016/j.langsci.2018.07.001

Schwartz, G. & Dzierla, J. (2017). Pre-voicing suppression in the speech of Polish learners of English. Paper presented at the 3rd Approaches to Phonetics and Phonology (APAP) conference. Lublin, Poland, June 2017.

Schwartz, G., Wojtkowiak, E., & Brzoza, B. (2019). Beyond VOT in the Polish laryngeal contrast. *Proceedings of the 19th International Congress of Phonetic Sciences*. Melbourne, Australia. https://icphs2019.org/icphs2019-fullpapers/pdf/full-paper_256.pdf

Shih, S., & Inkelas, S. (2019). Autosegmental aims in surface-optimizing phonology. *Linguistic Inquiry*, *50* (1), 137–196. https://doi.org/10.1162/ling_a_00304

Simon, E. (2009). Acquiring a new second language contrast: an analysis of the English laryngeal system of first language Dutch speakers. *Second Language Research*, *25* (3), 377–408. https://doi.org/10.1177/0267658309104580

Steriade, D. (1993). Closure, release, and nasal contours. In M. Huffman & R. Krakow (Eds.), *Nasals, nasalization, and the velum*. (pp. 401–470). San Diego: Academic Press.

Stevens, K., & Klatt, D. (1974). Role of formant transitions in the voiced-voiceless distinction for stops? *Journal of the Acoustical Society of America*, *55*, 653–659. https://doi.org/10.1121/1.1914578

Sučková, M. (2018). First language attrition and maintenance in the accents of native speakers of English living in the Czech Republic. Paper presented at Accents 2018. University of Łódź.

Sypiańska, J. (2013). Quality and quantity of language use and attrition of L1 Polish due to L2 Danish and L3 English. PhD dissertation, Adam Mickiewicz University in Poznan.

Traunmüller, H. (1994). Conventional, biological, and environmental factors in speech communication: A modulation theory. *Phonetica*, *51*. 170–183. https://doi.org/10.1159/000261968

Waniek-Klimczak, E. (2011). Aspiration in Polish: A sound change in progress?. In M. Pawlak, & J. Bielak (Eds.), *New Perspectives in Language, Discourse and Translation Studies* (pp. 3–11). Berlin: Springer. https://doi.org/10.1007/978-3-642-20083-0_1

Wetzels, W.L., & Mascaró, J. (2001). The typology of voicing and devoicing. *Language*, *77*, 207–244. https://doi.org/10.1353/lan.2001.0123

Wrembel, M. (2015). *In search of a new perspective – cross-linguistic influence in the acquisition of third language phonology*. Poznan: Wydawnictwo Naukowe UAM.

Zampini, M. (1998). The relationship between the production and perception of L2 Spanish stops. *Texas Papers in Foreign Language Education*, *3*(3), 85–100.

Appendix 1. Polish items for L1 drift study (Section 5)

/bdg/ initial	gloss	/ptk/ initial	gloss
bać	be afraid	*kac*	hangover
baz	bases	*kara*	punishment
bak	tank	*kawa*	coffee
bal	ball	*każ*	order (imp)
bar	bar	*keks*	biscuit
bas	bass	*koc*	blanket
bat	whip	*kos*	blackbird
bez	without	*kot*	cat
daj	give (imp)	*pan*	man
dam	I give	*para*	couple
dane	data	*pas*	belt
dar	gift	*pejs*	sideburn
data	date	*pet*	cigarette butt
demo	demo	*polo*	polo shirt
doba	day	*pop*	pop
dom	house	*tabu*	taboo
gad	reptile	*taca*	tray
gafa	gaffe	*tara*	washboard
gaj	grove	*targ*	market
gapa	oaf	*tas*	deck (of cards)
gasi	extinguishes	*ten*	this
gaz	gas	*test*	test
gen	gene	*toga*	toga
gest	gesture	*ton*	tone

Appendix 2. English items for L2 study (Section 6)

/bdg/ intial	/ptk/ initial
back	pace
bar	pass
beg	peck
bell	peg
ben	pen
bet	pet
buck	pop
bug	pub
bus	puck
dead	tech
debt	Ted
deck	tell
depp	ten
die	test
does	ton
done	tub
duck	tuck
Dutch	tug
get	come
gets	cub
guess	cup
guest	cut
gull	cap
gum	keg
gun	kelp
gut	coat
gap	kept

Address for correspondence

Geoffrey Schwartz
Adam Mickiewicz University
Al niepodleglosci 4
Poznan 61-874
Poland

geoff@wa.amu.edu.pl
https://orcid.org/0000-0002-0728-7820

Publication history

Date received: 12 October 2019
Date accepted: 21 August 2020
Published online: 9 October 2020