BEYOND VOT IN THE POLISH LARYNGEAL CONTRAST

Geoffrey Schwartz, Ewelina Wojtkowiak, Bartosz Brzoza

Faculty of English, Adam Mickiewicz University in Poznań
geoff@wa.amu.edu.pl, ewojtkowiak@wa.amu.edu.pl, bbrzoza@wa.amu.edu.pl

ABSTRACT

Acoustic and perceptual studies of the Polish laryngeal contrast in word-initial stops are presented. The production study revealed expected pre-voicing of /b,d,g/, yet also showed evidence that /p,t,k/ should be characterized as [fortis], with effects on f0 and F1 at vowel onset. In a phoneme monitoring experiment, Polish listeners recognized /p,b,t,d/ with mismatched laryngeal cues. Voiced targets were reliably identified when pre-voicing was absent from the stimuli. Voiceless targets showed more serious effects of the cue mismatches.

On the basis of these and other data, it is proposed that [voice] is absent from the representation of Polish /b,d,g/, which are phonologically unspecified, while /p,t,k/ are marked as [fortis]. This proposal is compatible with the familiar VOT typology under an aperture-based view of phonological structure. Aspirates align [fortis] with the closure phase of stops. Plain /p,t,k/ align [fortis] with the CV transition phase.

Keywords: phonetics-phonology interface laryngeal cues, Polish, representation, voicing

1. INTRODUCTION

Despite the widely accepted VOT-based typology of languages with two series of laryngeal contrasts [11], issues arise with regard to both phonological behavior and phonetic implementation. In so-called ‘voicing’ languages, we often encounter signs of phonologically active voicelessness [2],[12],[20], challenging the predictions of Laryngeal Realism (LR) [1],[7], which assumes that voiceless stops are unspecified for laryngeal features in voicing languages.

Phonetic findings bearing on this issue include more robust cross-linguistic interaction for /b,d,g/ than /p,t,k/ in L2 acquisition and L1 phonetic drift situations. Assuming, after [4], that the principle of *equivalence classification* guides such cross-linguistic phonetic interactions, it may be claimed that /b,d,g/, but not /p,t,k/, are phonologically equivalent across languages. Relevant findings, which have implications for L1 laryngeal representation, include the following:

- In [6], it was observed that American English speakers’ L1 /b,d,g/ were affected by L2 Spanish, but their /p,t,k/ were not.
- [14] and [21] found that Polish learners are more successful in their acquisition of aspirated /p,t,k/ in L2 English than they are at suppressing L1 pre-voicing, which induces errors such as [fejzbuk] ‘facebook’.
- In a study of Greek-English bilingual children, [3] observed a much greater degree of interaction between the two languages for /b,d,g/ than for /p,t,k/.
- [17] found that expatriate English speakers in Czechia produced high rates of L2-induced pre-voicing in their L1, but did not significantly shorten VOT of /p,t,k/.

These findings all suggest phonological equivalence between voicing and aspiration systems for voiced stops, but not voiceless stops.

Additional findings also raise questions about the the phonological status of closure voicing in voicing languages. In Dutch, [19] observed that pre-voicing was absent in approximately 25% of /b,d,g/ items. Meanwhile, [8] and [9] observed intervocalic closure voicing in dialects of Italian and Spanish, respectively. Intervocalic voicing is predicted by LR not to happen in voicing languages [1]. With regard to perception, [15] found that Polish listeners discriminate the laryngeal contrast with high accuracy when pre-voicing is absent from /b,d,g/ tokens.

These facts suggest that in voicing languages, /b,d,g/ may not be specified with a phonological feature [voice]. Rather, voicing may be claimed to be an element of an acoustic carrier signal as envisioned in [18], and formalized in the representational system described in [13]. The prediction of this system is that laryngeal cues at vowel onset, including f_0 and F1, should provide evidence of ‘fortisness’ of initial /p,t,k/ in voicing languages, while the perceptual weight of pre-voicing should be limited to some extent. This paper provides new acoustic and perceptual data from Polish that bear on these issues.

2. PRODUCTION STUDY

2.1. Method

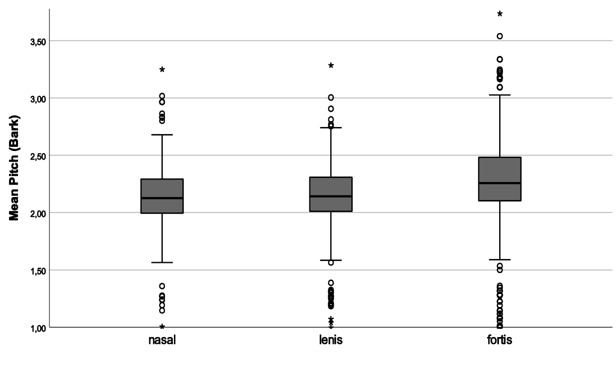
Fourteen native speakers of Polish, first year students of English (B1-level proficiency), all of them female, produced a list of citation form Polish words with initial /b,d,g/ and /p,t,k/. The following vowels were /a/, /e/ and /o/. Items were elicited using PowerPoint slides in a sound attenuated booth at a Polish university. Annotation was done by hand by the second author of this paper. Acoustic measures were extracted using Praat scripts, and included VOT, type of /b,d,g/ realization (fully voiced, partially voiced with a break in pre-voicing of at least 20 ms before stop release, or unvoiced), as well as mean f_0 (in Bark) and F1 (in Bark difference, F1-f_0) over the first 25% of the following vowel. To provide a baseline for the f0 analysis (cf. [5]), nasal-initial items were also recorded. Generalized linear mixed models were run in SPSS, with speaker as a random factor, and Voicing, Voicing*Vowel, and Voicing*Place as fixed factors. A total of 1032 items, including nasal onsets, were analyzed for the vowel onset measures (f0 and F1), and 797 items (lenis-fortis) for VOT.

2.2. Results

Mean VOT was 43 ms (*SD*=18) for the fortis items and -89 ms (*SD*=27) for the lenis items. Of the lenis items, 88% were fully voiced, 11.2% were partially voiced, and 0.8% were unvoiced. Partially voiced items showed significantly longer negative VOT (*M*=-109.6 ms, *SD*=22.4) than fully voiced (*M*=-87.1 ms, *SD*=24.3) realizations of /b,d,g/, p<.001.

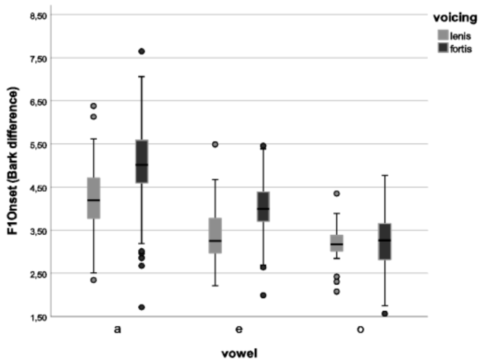
With regard to f0, using nasal-initial items as a baseline it was observed that fortis items raised f0 (Intercept:nasal = 2.13); *B*=0.117, *S.E.*=0.02, *t*=5.39, *p*<.001) but lenis items had no effect (*B*=.004, *S.E.*=0.02, *t*=0.17, *p*=.863). These results are shown in Figure 1.

Figure 1: Effects of consonant type on f0.



For F1 at vowel onset, fortis items showed significantly higher F1 values (*p*<.001) than lenis items for /a/ and /e/, but not /o/ (*p*=.814). These results are summarized in Figure 2.

Figure 2: Effects of voicing on F1.



3. PERCEPTION STUDY

Our perception study consisted of a phoneme monitoring task aimed at providing evidence about the relative perceptual weight of VOT, f_0, and F1 in the laryngeal contrast in Polish.

3.1. Participants

38 native speakers of Polish participated in the experiment. They were all first year students of English at a Polish university between the ages of 18 and 24. 29 of the participants were female, 8 were male, and one chose not to respond the initial question about gender.

3.2. Stimuli

Stimuli for the phoneme monitoring experiment were made from recordings of two native speakers of Polish, one male and one female, producing the pairs *dam-tam* 'give-there' and *pas-bas* 'belt-bass' in citation form. A number of filler items were obtained from recordings of 8 additional speakers. From the original recordings, acoustic manipulations were made in Praat, yielding three types of stimuli, summarized in Table 1.

Table 1: Stimulus types for monitoring task

Type	Mismatched cue	Matched cue
1	VOT	Pitch, F1Onset
2	Pitch	VOT, F1Onset
3	F1Onset	VOT, Pitch

The stimuli were designed to pit one laryngeal cue to one series against two laryngeal cues to the other. For example, a Type 1 token of /t/ had the pre-voicing of that speaker's /d/ pasted into the closure portion of the signal, while the f0 and F1 cues matched the original /t/. A Type 1 token of /d/ simply had its pre-voicing silenced. For Type 2, the original target productions of lenis and fortis items, respectively, had their pitch at vowel onset raised or lowered by 10%. In Type 3 tokens, the mismatched F1 Onset was obtained by taking the opposite series as the base. Type 3 lenis tokens were obtained from original fortis ones, with their f_0 at vowel onset lowered and pre-voicing added to match the lenis item. Type 3 fortis items were made by silencing the pre-voicing and raising the f_0 of the lenis items. For each trial in the phoneme monitoring task, the target stimulus was grouped into single sound files with three other filler words, creating a sequence of four words with a 150 ms interval between each of the words. The position of the target item in the sequence (1^{st}, 2^{nd}, 3^{rd}, 4^{th}) was counterbalanced. The total number of experimental trials was 24 (3 types of stimuli*4 target words* 2 speaker voices), with 24 additional filler items. With 38 participants this yielded a total of 912 total responses for analysis.

3.3 Procedure

The experiment was carried out in E-Prime in a psycholinguistic laboratory at a Polish university. The experimental stations were equipped with high-quality headphones. In each trial of the monitoring task, an orthographic symbol matching the initial sound of the target item (/p t b d/), or a sound appearing in one of the fillers, was displayed on the screen. After 500 ms, the recording of the sequence of four words was played. The order of the trials was randomized. Listeners were instructed to respond as quickly as possible by pressing the SPACE key when they heard the target phoneme in the sequence of four words. If they did not hear the target phoneme in the sequence of four words, listeners were instructed to respond by pressing one of the SHIFT keys. Participants were told to use their dominant hand on the space bar for positive responses, and their off hand on the corresponding SHIFT key for negative responses. None of the participants reported being ambidextrous, and none reported any hearing deficiencies.

3.4. Analysis

Statistical analysis concentrated on two dependent variables. The first was binary, involving whether or not the target phoneme was actually identified (SPACE or SHIFT response). The second was response time (RT), a continuous variable calculated from the onset of the vowel following the target phoneme. RT analysis was based only on positive responses, those in which listeners identified the target phoneme. For both binary and linear analyses Stimulus Type*Fortis-Lenis was included as the main fixed factor of interest, while participant was included as a random factor.

3.4. Results

Figure 3 shows the proportion of positive responses as a function of stimulus type, sorted for lenis and fortis targets. When the target was lenis, the mismatched cue did not have a significant effect on the likelihood of a missed response. In other words, voiced items without pre-voicing, voiced items with a fortis-like pitch pattern, and voiced items with fortis-like F1 onset were all recognized with near-ceiling performance (>96%). Fortis targets showed more variable responses. Pre-voiced fortis targets (Type 1) were recognized 31.2% percent of the time, significantly less frequently than f_0 or F1 mismatches (92% and 94.9%). In a binary logistic regression analysis, the VOT-mismatched fortis items were the only ones that behaved differently with regard to positive responses.

Figure 3: Proportion of responses in which target phoneme was recognized

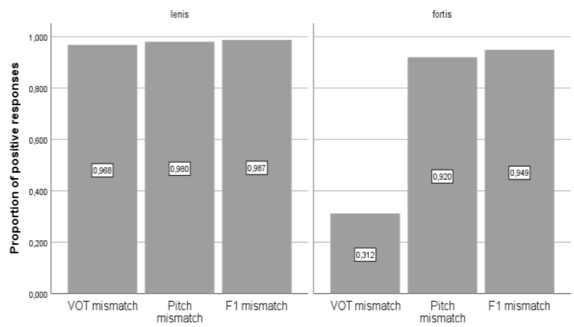
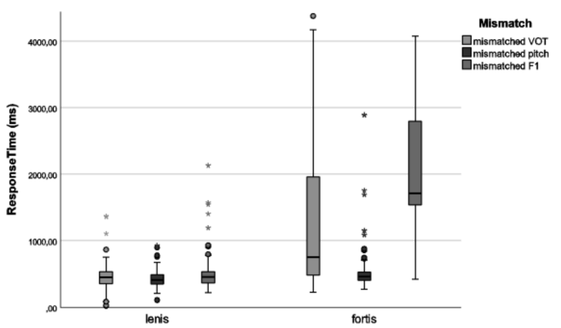


Figure 4: Mean RT for positive responses



For the analysis of response time, we consider only those items in which the target phoneme was recognized. In the case of the lenis targets, there were no significant effects of cue mismatches. The lenis items without pre-voicing, the ones with fortis-like pitch raising, and the ones with fortis-like F1 onsets were all recognized with the same response latency (p=.089). For the fortis targets, the F1 mismatches induced the slowest responses, followed by VOT mismatches, and finally pitch mismatches. Pairwise comparisons revealed significant differences in each case (p<.001).

4. DISCUSSION

Our studies provide perspective on the realization of the word-initial laryngeal contrast in Polish, and the relative weight of three laryngeal cues: VOT, f_0, and F1. In what follows, we will reflect on the implications of our data for the phonological representation of Polish laryngeal categories.

In the domain of production, we may note that the VOT results are largely compatible with other descriptions [10]. However, we also provide previously undocumented (to our knowledge) data on whether pre-voicing in /b,d,g/ is constant or interrupted. Interestingly, but not surprisingly, partial pre-voicing was associated with greater negative VOT, which is difficult to reconcile with the traditional view that the feature [voice] is active in Polish. Positing the feature [voice], one could argue that the items with interrupted pre-voicing exemplify both a weakening of the feature [voice] on the one hand, and a segmental strengthening on the other hand, since longer negative VOT is associated with increased segment duration. This apparent contradiction is avoided if we assume that the Polish /b,d,g/ are not in fact specified for the feature [voice].

The f_0 and F1 results constitute effects that, as far as we know, had not been previously documented for Polish. In the case of f_0, including nasal items enables us to suggest that voiceless items raise f0 in Polish, while voiced items have no effect, which is compatible with the view that Polish /p,t,k/ may be described as [fortis].

The perception results may also be taken as evidence in support of this phonological outlook. Lenis target items were reliably and quickly identified in the phoneme monitoring experiment regardless of cue mismatches. Thus, pre-voicing was by no means necessary for the recognition of /b,d,g/, nor did it accelerate processing speed relative to the other cues.

The monitoring results for the fortis target items are somewhat more difficult to interpret. The overall results, by which fortis targets with mismatched cues were recognized much less consistently than lenis ones, suggests that listeners expected fortis items, to a greater extent than lenis items, to contain acoustic evidence of a phonological feature. This would explain why the consequences of cue mismatches were more serious for the fortis items. Furthermore, pre-voiced items with fortis-like f_0 and F1 cues were recognized as voiceless in 31.2% of cases, so it appears that listeners may sometimes simply ignore pre-voicing. At the same time, however, it was these pre-voiced 'fortis' items that were recognized least reliably. Since these items lacked silent closure, a crucial perceptual cue to voicelessness, this result is expected under the assumption of phonological 'fortisness'. When a primary cue was absent, recognition performance suffered. Note also that the failure to identify fortis stops in a monitoring task does not necessarily imply these items were heard as lenis. The other finding of note for the fortis targets was that F1 mismatches affected listeners more than pitch mismatches, suggesting that the F1 cue carries more perceptual weight than the f0 cue.

To summarize the phonological interpretation of our results, it is proposed that Polish /p,t,k/ are specified for a phonological feature [fortis], while pre-voiced /b,d,g/ lack phonological specification – pre-voicing reflects the emergence of an acoustic carrier signal [18].

Under this proposal, the question that remains is how to reconcile a [voice]-less representational system with the widely-accepted VOT typology for languages with two series of consonants. This reconciliation is possible if we adopt an aperture-based view of phonological structure [13] [16]. Stated briefly, in aspiration systems, the [fortis] feature is assigned to the closure phase in the representation of a stop, inducing long-lag VOT. In voicing systems such as Polish, the [fortis] feature aligns with the CV transition phase following stop release. The consequence of this is that VOT remains short, promoting the role of vowel-based laryngeal cues such as f_0 and F1.

5. FINAL REMARKS

A large body of phonetic and phonological evidence suggests that the LR proposal is not entirely adequate in its view of the phonetics-phonology interface. We have presented new data from Polish that presents additional challenges to LR, and at the same time points to an alternative proposal.

6. ACKNOWLEDGEMENT

This research was supported by a grant from the Polish National Science Centre (Narodowe Centrum Nauki), grant # UMO-2016/21/B/HS2/00610.

7. REFERENCES

[1] Beckman, J., Essen, M., Ringen, C. 2013. Evidence for laryngeal features: Aspirating vs. true-voice languages. *Journal of Linguistics* 49(2), 259–284.

[2] Bennett, W., Rose, S. 2017. Moro voicelessness dissimilation and binary [voice]. *Phonology* 34, 473-505.

[3] Chionidou, A., Nicolaidis, K. 2015. Voice Onset Time in Greek-German Bilingual Children. *Proceedings of ICPhS, Glasgow.*

[4] Flege, J. 1987. The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification. *JPhon* 15, 47-65.

[5] Hanson, H. 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *J. Acoust. Soc. Am.* 125 (1), 425-441.

[6] Herd, W., Walden, R., Knight, W., Alexander, S. (2015). Phonetic drift in a first language dominant environment. *Proceedings of Meetings on Acoustics* 23.

[7] Honeybone, P. 2005. Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. In van Oostendorp, M, van de Weijer, J. (eds.), *The internal organization of phonological segments*, 317–352. Berlin: Mouton de Gruyter.

[8] Hualde, J., Nadeu, M. 2011. Lenition and phonemic overlap in Rome Italian. *Phonetica* 68, 215-242.

[9] Hualde, J., Simonet, M., Nadeu, M. 2011. Consonant lenition and phonological recategorization. *Laboratory Phonology* 2, 301-329.

[10] Keating, P. 1979. A phonetic study of a voicing contrast in Polish. Ph.D. dissertation, Brown University.

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

[12] Rubach, J. 1996. Non-syllabic analysis of voice assimilation in Polish. *Linguistic Inquiry* 27, 69–110.

[13] Schwartz, G. 2017. Formalizing modulation and the emergence of phonological heads. *Glossa: a journal of general linguistics*, 2(1), 81.

[14] Schwartz, G. Dzierla, J. 2017. Pre-voicing suppression in the speech of Polish learners of English Paper presented at APAP, Lublin.

[15] Schwartz, G., Arndt, D. 2018. Laryngeal Realism vs. Modulation Theory – evidence from VOT discrimination in Polish. *Language Sciences* 69, 98-112.

[16] Steriade, D. 1993. Closure, release, and nasal contour. In Huffman, M., Krakow, R. (eds.), *Nasals, nasalization, and the velum*, 401–470. San Diego, CA: Academic Press.

[17] 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.

[18] Traunmüller, H. 1994. Conventional, biological, and environmental factors in speech communication: A modulation theory. *Phonetica* 51, 170–183.

[19] van Alphen, P., Smits, R. 2004. Acoustic and perceptual of the voicing distinction in Dutch initial plosives: the role of pre-voicing. *JPhon* 32, 455-491.

[20] Wetzels, W. L., Mascaro, J. 2001. The typology of voicing and devoicing. *Language* 77, 207-244..

[21] Zając, M. Phonetic convergence in the speech of Polish learners of Enlish. Ph.D. dissertation, University of Łódź