FORTIS STOPS IN POLISH – EVIDENCE FROM ACOUSTIC MEASURES OF VOICE QUALITY

Geoffrey Schwartz, Maral Asiaee, Ewelina Wojtkowiak

Adam Mickiewicz University, Poznań, Poland
geoff@amu.edu.pl, marasi@amu.edu.pl, ewelina.wojtkowiak@amu.edu.pl

ABSTRACT

An acoustic study of the effects of the Polish laryngeal contrast on the phonation of vowels following voiced and voiceless obstruents is presented. Data from 15 native speakers of Polish reveal voicing-based phonation differences in H1*-H2*, H2*-H4*, H4*–H2kHz*, and H2kHz*–H5kHz, providing evidence that phonation properties may help distinguish between Polish laryngeal categories in onset position. Implications for the phonological encoding of the laryngeal contrast are discussed. It is suggested that Polish voiceless stops may be categorized as [fortis], postulated to be an abstract category with a wide range of possible phonetic effects.

Keywords: laryngeal contrasts, phonetics-phonology interface, voice quality, Polish.

1. INTRODUCTION

It is well known that the phonetic realization of two-series laryngeal contrasts is phonetically much more complex than one might infer from phonological descriptions based on features such as [voice] or [spread glottis]. Phonological voicing has been shown to be cued by a large number of acoustic properties, including voice onset time (VOT; [1]), the duration and height of neighboring vowels [2], [3], f_0 in the vicinity of the consonant [4], F1 transition duration [5], noise burst duration or frication duration, as well as the amplitude of stop release bursts, frication, or aspiration. Of these parameters, the phonological literature has focused largely on VOT [6]. On the basis of VOT, it has often been claimed that two-series systems may be divided into two types, commonly referred to as aspiration and true-voice languages. Further, in the approach referred to as laryngeal realism (LR; e.g. [7], [8]) it is suggested that these two types differ with regard to the phonological feature specification that appears, [spread glottis] in aspiration systems as opposed to [voice] in true-voice systems.

If voicing and aspiration systems were truly phonologically distinct, we would expect phonetic distinctions in additional parameters aside from VOT. However, phonetic studies of parameters beyond VOT often reveal parallel behavior between voicing and aspiration systems. The voicing effect on f_0, by which pitch is higher in the vicinity of voiceless obstruents, is an example. In American English, [9] found that voiceless onsets raise f_0, while voiced obstruents have no effect, relative to the baseline produced by nasal onsets. Similar findings have been obtained for voicing languages [10], [11] including French, Italian, and Polish. F1 transition duration, in which clipped transitions after fortis consonants induce higher F1 measurements in non-high vowels, has been shown to behave in a parallel fashion in English [5] and Polish [11]. In short, the current state of research into the phonetics of two-series laryngeal systems has not identified a single guiding phonetic feature upon which phonological categories may be constructed.

In this connection, recent research has begun to identify additional phonetic correlates, based on phonation, which may play a role in maintaining contrast in two-series laryngeal systems (e.g. [12], [13]). Most of this research has come from varieties of English and arose out the observation that voiceless obstruents, most commonly the plosive /t/, frequently undergo glottal reinforcement or even glottal replacement. This glottalization is typically associated with voiceless obstruents in English, while it is for the most part absent in the case of voiced obstruents (but see [14]). Acoustically, glottalization is often described in terms of acoustic measures of voice quality, including a number of measures quantifying spectral tilt. Since glottalization weakens the first harmonic of the vocal wave, glottalized consonants typically show more gradual dropoffs in spectral tilt and lower values for acoustic measures comparing the amplitude differences between lower and higher harmonics. Notably, [15] found perceptual differences between glottalization induced by voiceless consonants and creaky voice that often appears phrase-finally.

The research connecting phonation type with laryngeal contrast discussed above focuses on post-vocalic or coda position in syllables. Thus, for English, experimental findings have linked aspects of phonation with the perception of vowel duration [13] which in this language has been found to be a primary cue to the contrast [16]. At the same time, given the overall phonetic complexity of the laryngeal contrasts, we suggest that it is worth investigating the effects of phonological voicing in pre-vocalic or onset positions as well.

Polish is a language in which the laryngeal contrast in the coda position is said to be neutralized, completely or incompletely. In the initial position, the familiar VOT contrast between pre-voiced and short-lag stops is attested. At the same time, the non-VOT cues such as f_0 and F1 transitions, housed on vowels following the consonants, have been shown to play an important role in maintaining laryngeal contrasts [11], [17]. In light of the discussion above, we raise the question of whether acoustic measures of voice quality may also constitute part of the phonetic anatomy of the contrast in Polish between voiced and voiceless consonants. A working hypothesis, on the basis of the findings [11], is that voiceless stops in Polish may be described as [fortis], which may be reflected in smaller values for spectral tilt measures gleaned from neighboring vowels, even if the effect is not dramatic enough to be described as glottalization. In what follows, we present a preliminary acoustic study investigating this question.

2. METHOD

2.1. Participants and recordings

15 monolingual Polish speakers took part in a production study. They were all females, aged 17-38 (median age: 25). While they had some history of learning foreign languages in school, they claimed not to be fluent in any of them.

The materials consisted of a word list of forty-eight monosyllabic and disyllabic words in Polish, each of them beginning with a plosive (24 voiced, 24 voiceless). The dataset was counterbalanced for the place of articulation of the stop (16 labial, 16 coronal, 16 dorsal), and the following vowel was always non-high. The participants were recorded in a quiet room in a language school where they were enrolled in an elementary English course. They were recorded directly onto a laptop, using a head-mounted microphone and a USB interface. The items were shown to the participants on PowerPoint slides.

2.2 Segmentation and extraction of acoustic parameters

The acoustic annotation was performed manually using Praat [18] on three separate tiers. Tier 1 contained the consonantal onsets (VOT: positive for voiceless stops, negative for voiced stops, measured according to standard criteria). Tier 2 categorised the onsets into voiceless or voiced. Tier 3 contained the measurements of the vowel only.

Two Praat scripts were used to extract the VOT values from tier 1 and F1 (Bark difference: F1-f_0), and f_0 (in Hz) averages from the first 20% of the vowel from tier 3. VoiceSauce [19] was used to extract voice quality measures from the vocalic part in each word in tier 3. After obtaining all data frames at each 1ms intervals, impossible or incorrect 0 values were omitted.

Four spectral tilt measures, H1*-H2*, H2*-H4*, H4*–H2kHz*, and H2kHz*–H5kHz), were chosen for analysis. These measures allow for the quantification of spectral tilt over a wide range fo the acoustic spectrum [20]. Since in this study we are only looking at phonologically conditioned variation within the confines of modal voice, measures of periodicity such as Cepstral Peak Prominence or Harmonics to Noise Ratio, which quantify the presence of noise in the vocal wave and are more effective in distinguishing between modal and non-modal voices [20], are not analysed here.

2.3. Statistical analysis

All statistical analyses were performed using R version 4.0.4 [21], while visualisations of the data were made using JASP [22]. Since the data were not normally distributed (p<.05), the Mann-Whitney U test was performed to find out whether spectral tilt values in vowels differ after voiced/voiceless obstruents.

3. RESULTS FOR TRADITIONAL LARYNGEAL PARAMETERS

We start by reporting results for acoustic parameters traditionally associated with the laryngeal contrast.

As far as VOT is concerned, the results are summarised in Table 1.

Table 1: mean VOT measures of Polish monolinguals

Series	VOT (ms)	Std Dev
voiceless	41.70ms	16
voiced	-92.05ms	26

Table 2 presents the average values of F1 (in Bark) and f_0 (in Hz) after voiceless- and voiced-initial onsets.

Table 2: mean F1 and f_0 measures of Polish monolinguals

Parameter	Onset	Average
F1- f_0	voiceless	4.9 Bark
	voiced	4.4 Bark
f_0	voiceless	195 Hz
	voiced	186.6 Hz

The results from the traditional acoustic parameters associated with the laryngeal contrast may be summed up as follows. In line with previous studies, VOT in Polish voiceless stops is in a range approaching what [23] would call ‘slightly aspirated’, while prevoicing in voiced stops is very common. With regard to correlates deriving from the vowel following the stops, the effects of underlying voicing were detectable in both the f_0 (difference between voiceless and voiced onsets of 8.4Hz; p=.019) and F1 (difference between voiceless and voiced onsets of 0.5 Bark; p=.003) parameters.

4. RESULTS OF VOICE QUALITY MEASURES

Figures 1-4 show raincloud plots of the four spectral tilt measures analysed here. From the plots it is clear that the voicing induced differences are relatively subtle and small in magnitude, as might be expected for within-modal-voice variation, as opposed to modal-non-modal contrasts.

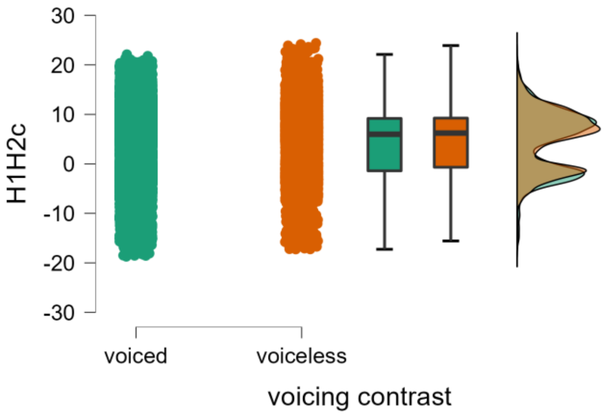

Figure 1: The difference between voiced and voiceless onsets in the H1*-H2* measurement.

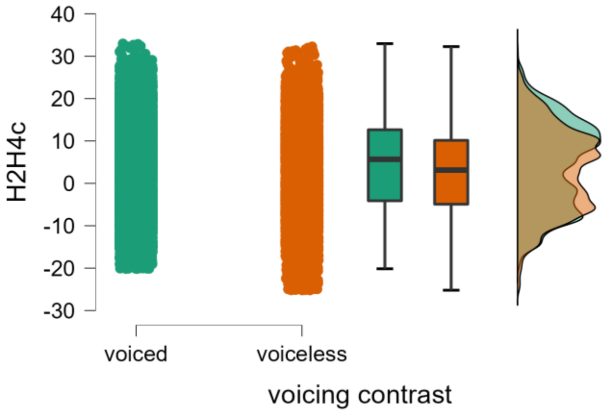

Figure 2: The difference between voiced and voiceless onsets in the H2*-H4* measurement.

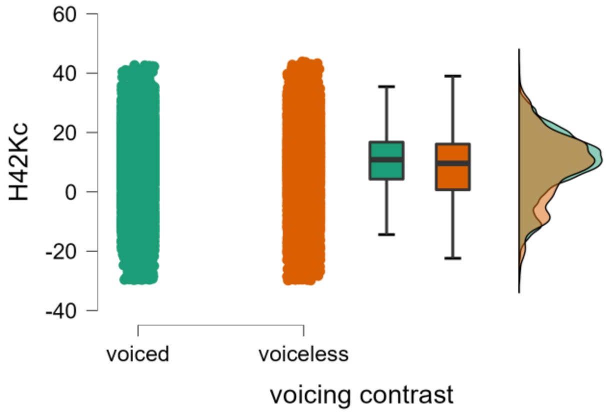

Figure 3: The difference between voiced and voiceless onsets in the H4*-H2kHz* measurement.

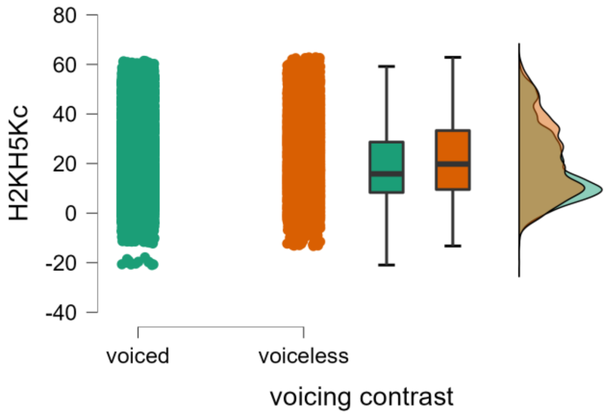

Figure 4: The difference between voiced and voiceless onsets in the H2kHz*-H5kHz* measurement.

The median and Interquartile range (IQR) of voice quality parameters are presented in Table 3.

Table 3: Descriptive data exhibiting median and (IQR) of spectral tilt measures

Spectral tilt measures	Voiceless	Voiced
H1*-H2*	6.22 (9.93)	5.98 (10.6)
H2*-H4*	3.14 (15.1)	5.68 (16.8)
H4*–H2kHz*	9.60 (15.4)	10.8 (12.5)
H2kHz*–H5kHz	19.8 (23.8)	15.8 (20.4)

Results from the Mann-Whitney U test revealed that the median of all parameters depicting the harmonic source spectral shape were statistically different (p<.001) between the vowels following the voiced and voiceless obstruents.

5. DISCUSSION

The present study investigated the effects of obstruent voicing on voice quality in Polish. Our working hypothesis was that voiceless consonants in Polish should be characterized as ‘fortis’, and as such should induce a slightly stiffer voice quality on the following vowel than voiced consonants. We presented results for four spectral tilt measures, H1*-H2*, H2*-H4*, H4*–H2kHz*, and H2kHz*–H5kHz. In all four measures, we observes significant differences resulting from underlying voicing. Thus, overall, the results provide evidence for the claim that voice quality is indeed an acoustic correlate of voicing in Polish.

With regard to our working hypothesis, results from two of the measures were compatible with the 'voiceless as fortis' hypothesis, while results for the other two measures showed a difference in the opposite direction. For the H1*-H2* measure, which compares the difference in amplitude between the first and second harmonics of the vowel wave, voiced consonants induced smaller measures, indicating a stronger H2 relative to H1. Likewise, the H2kHz*–H5kHz measure was slightly lower after voiced consonants, indicating a stronger high-frequency element in the spectrum that is expected in stiffer voice qualities. By contrast, the H2*-H4* and H4*–H2kHz* measures both showed signs of stiffer phonation after voiceless obstruents. The H2*-H4* and H4*-H2kHz results provide additional support for the 'voiceless as fortis' hypothesis, which originally arose out of the f_0 and F1 results described by [11].

Overall, the voice quality results suggest a situation in which the bottom and the top of the spectrum are indicative of slacker phonation, while the middle of the spectrum is indicative of stiffer phonation. In this situation, a comparison of the relative merits of the different acoustic measures of voice quality is warranted The H1*-H2* measure describes amplitude of harmonics at the very bottom of the acoustic spectrum, typically in the area below 300Hz. Perceptual sensitivity to amplitude differences in this frequency range is relatively limited, as is visible in equal loudness curves [24]. H2*-H4*, by contrast, compares harmonics in a higher frequency range that shows slightly greater auditory sensitivity. Meanwhile H2kHz, the harmonic nearest 2000 Hz, is in the most sensitive frequency range when it comes to amplitude differences. For this reason, it may be hypothesized that the differences in the H2*-H4* and H4*–H2kHz* measures observed here may be more perceptually robust than H1*-H2* in characterizing the effects of voiceless consonants on voice quality. It is also possible that since Polish voiceless stops have been described as 'slightly aspirated', as shown above, they contributed to a more breathy result in some measures but not others.

We also note that the H4*–H2kHz* measure shows an acoustic resemblance to measures of spectral balance that have been shown to distinguish stressed from unstressed syllables [25], [26]. Spectral balance measures typically divide signal up into two bands, 0-1 kHz and 1-5 kHz. More prominent (or stressed) syllables typically have higher relative amplitude in the higher band where perceptual sensitivity is greater. It is also worth noting that the comparison inherent within the H2kHz*–H5kHz measure, between harmonics closest to 2000 Hz and 5000 Hz, falls entirely within the upper band of spectral balance measures. With regard to the split results observed here, in future studies it is worth investigating the hypothesis that some measures of phonation are manipulated for phonological purposes, while others are more crucial for speaker identification.

Overall, our results provide preliminary evidence of a connection between 'fortisness' in Polish stop production and perceptual prominence of the voice quality that appears on the vowel following the stop. In other words, there may be a phonetic connection between the phonatory effects elicited by stress and those elicited by voiceless consonants. In this connection, we note the presence of onset-sensitive stress systems in the world's languages. In a few such systems, a pattern is attested in which voiceless onsets attract stress, but voiced onsets do not. The opposite pattern is not found [27]. We suggest that in certain instances, voicing-induced differences in phonation of the type observed here may evolve into phonological prominence.

Returning to the phonological issue of laryngeal contrasts, we suggest that their phonetic complexity points to an approach to the phonetics-phonology relationship in which phonological specifications are abstract in nature and do not highlight any one phonetic property. Instead of a specification such as [-voice], which makes direct reference to the presence or absence of periodicity in the signal, we would opt in favour of an abstract feature such as [fortis], cued by a wide range of phonetic properties, but not tied to any one property in particular. Notably, this feature has not only consonant-internal acoustic correlates such as VOT and burst amplitude, but also encodes phonetic properties observed on neighbouring vowels, including raised f_0, clipped F1 transitions, and, as we have shown stiffer voice quality as reflected in higher amplitude harmonics in perceptually prominent areas of the acoustic spectrum.

6. ACKNOWLEDGEMENTS

This research was supported by a grant from the Polish National Science Centre (Narodowe Centrum Nauki), project number 2021/41/B/HS2/00239.

7. REFERENCES

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

[2] Chen, M. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22(3), 129-159.

[3] Esposito, A. 2002. On vowel height and consonantal voicing effects: Data from Italian. *Phonetica* 59, 197−231.

[4] Ohde, R. N. 1984. Fundamental frequency as an acoustic correlate of stop consonant voicing, *J. Acoust. Soc. Am.* 75, 224–230.

[5] Stevens, K., Klatt, D. 1974. Role of formant transitions in the voiced-voiceless distinction for stops, The *J. Acoust. Soc. Am.* 55, 653–659.

[6] Beckman, Jill, Petur Helgason, Bob McMurray & Catherine Ringen. 2011. Rate effects on Swedish VOT: Evidence for phonological over specification. *Journal of Linguistics* 39, 39-49.

[7] Honeybone, P. 2005. Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. In: van Oostendorp, M., van der Weijer, J. (eds), *The internal organization of phonological systems.*Mouton de Gruyter, 319–354.

[8] Beckman, Jill, Michael Jessen & Catherine Ringen. 2013. Empirical evidence for laryngealfeatures: Aspirating vs. true voice languages. *Journal of Linguistics* 49, 259–284.

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

[10] Kirby, J., Ladd D. R. 2016. Effects of obstruent voicing on vowel f0: evidence from 'true voicing' languages", *J. Acoust. Soc. Am.* 140, 2400–2411.

[11] Schwartz, G., Wojtkowiak, E., Brzoza, B. 2019. Beyond VOT in the Polish laryngeal contrast, *Proc. 19th ICPhS*, Melbourne, 2383-2387.

[12] Penney, J,, Cox, F., Miles, K., Palethorpe, S. 2018. Glottalisation as a cue to consonant voicing in Australian English. *J Phon* 66, 161-184.

[13] Sanker, C. 2019. Influence on coda stop features in perceived vowel duration. *J Phon* 75, 43-56.

[14] Koops, C., & Niedzielski, N. 2009. /t, d/-glottalization and vowel variation in Houston AAE. Paper presented at New Ways of Analyzing Variation (NWAV) 38, University of Ottawa, Ottawa, Canada.

[15] Garellek, M. 2015. Perception of glottalization and phrase-final creak. *J. Acoust. Soc. Am.*, 137, 822-831.

[16] Port, R. F., Dalby, J. 1982. Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics* 32, 141–152

[17] Schwartz, G., Arndt, D. 2018. Laryngeal realism vs. modulation theory – evidence from VOT discrimination in Polish, *Language Sciences* 69, 98-112.

[18] Boersma, P. 2001. Praat, a system for doing phonetics by computer, *Glot International* 5, 341-345.

[19] Shue, Y.-L., Keating, P., Vicenik, C. 2009. "VOICESAUCE: A program for voice analysis.," *J. Acoust. Soc. Am.*, 126, 2221- 2009.

[20] Garellek, M. 2019. The phonetics of voice. *Routledge Handbook of Phonetics*, Routledge, 75-106.

[21] R Core Team. 2021. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

[22] JASP Team. 2023. JASP (Version 0.17.1) [computer software].

[23] Cho, T. and P. Ladefoged. "Variation and universals in VOT: evidence from 18 languages". *JPhon* 27, 207-229, 1999.

[24] Robinson, D. W., Dadson, R. S. 1957. Threshold of Hearing and Equal-Loudness Relations for Pure Tones, and the Loudness Function, . *J. Acoust. Soc. Am.* 29, 1284.

[25] Plag, G. K., Schramm, M. 2011. Acoustic correlates of primary and secondary stress in North American English. *J Phon* 39, 362-374.

[26] Sluijter, A., van Heuven, V.J. 1996. Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proc. ICSLP96*, 630-633.

[27] Gordon, M. 2005. A perceptually driven account of onset-sensitive stress, *Natural Language and Linguistic Theory* 23, 595-653.