Laboratory Phonology
Journal of the Association for Laboratory Phonology

Schwartz, G., Wojtkowiak, E., Asiaee, M., & Kaźmierski, K. (2026). Beyond VOT – phonation as a correlate of the laryngeal contrast in Polish. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 17(1), pp. 1–39. DOI: https://doi.org/10.16995/labphon.25171

Open Library of Humanities

Beyond VOT – phonation as a correlate of the laryngeal contrast in Polish

Geoffrey Schwartz*, Experimental Phonology and Phonetics Laboratory, Faculty of English, Adam Mickiewicz University in Poznań, Poland, geoff@amu.edu.pl

Ewelina Wojtkowiak, Experimental Phonology and Phonetics Laboratory, Faculty of English, Adam Mickiewicz University in Poznań, Poland

Maral Asiaee, Experimental Phonology and Phonetics Laboratory, Faculty of English, Adam Mickiewicz University in Poznań, Poland

Kamil Kaźmierski, Experimental Phonology and Phonetics Laboratory, Faculty of English, Adam Mickiewicz University in Poznań, Poland

*Corresponding author.

This article presents an acoustic experiment describing the effects of word-initial stop consonant voicing on the realization of the following vowel in Polish. We provide data from 30 speakers and include a wide range of acoustic measures of voice quality, fundamental frequency, and the first formant of the vowel. Trajectories estimated with Generalized Additive Mixed Models and the results of a Principal Components Analysis reveal that word-initial voiceless stops in Polish induce slightly breathier modal phonation than voiced stops, particularly over the early portion of vowel duration. We also observe higher f0 and F1 frequencies at vowel onset after voiceless stops. Taken together, our results are suggestive of an active phonetic effect of voicelessness on the acoustics of the following vowel. Implications of these results for current models of laryngeal phonology are discussed.

Laboratory Phonology: Journal of the Association for Laboratory Phonology is a peer-reviewed open access journal published by the Open Library of Humanities. © 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/. Articles published by the Open Library of Humanities may also contain other CC-licensed material, or material used or reproduced under an exception or limitation to copyright. Such material is not subject to the article's license. It is either marked with its own license, or marked that it has been reproduced with permission or under fair-use.
OPEN ACCESS

1. Introduction

The phonetic realization of laryngeal contrasts in languages with two series of stop consonants is probably the most commonly described domain for investigations into the relationship between phonology and phonetics. This fact is likely due to a single phonetic parameter, voice onset time (VOT; Lisker & Abramson, 1964; Cho et al., 2019), which is described in countless studies of a large number of languages. Indeed, Ladd and Schmid (2018, p. 229) refer to VOT as a "runaway success" for describing voice contrasts, a somewhat unusual (and perhaps slightly sarcastic) characterization for an objectively measurable physical phenomenon. The ubiquity of VOT stems from two factors. First, it is a relatively simple parameter to observe and measure on acoustic displays. Second, VOT appears to show rather categorical behavior with regard to two-series laryngeal contrasts. Most frequently, it is claimed that three categories of VOT are found across two-series languages: voicing lead (pre-voicing, or negative VOT), short-lag VOT, and long-lag VOT. Individual systems typically make use of two of these three categories in constructing their voicing contrast. So-called aspiration languages (see e.g., Beckman et al., 2013), which include English and most of Germanic, contrast short-lag lenis stops with long-lag fortis stops, although pre-voicing is known to appear in Swedish (e.g., Beckman et al., 2011) and some varieties of English (e.g., Jacewicz et al., 2009). So called true-voice or voicing languages (e.g., Slavic and Romance languages) contrast pre-voiced and short-lag stops.

The categoricity of VOT has been interpreted as evidence in support of a theoretical perspective that has come to be known as Laryngeal Realism (LR; Iverson & Salmons, 1995; Avery & Idsardi, 2001; Honeybone, 2005; Beckman et al., 2013; Raimy, 2021). Within LR, the strongest claims related to VOT are made by Beckman et al. (2013), for whom pre-voicing is an instantiation of a feature [voice] (cf. [slack] in Avery & Idsardi's [2001] model), long-lag VOT is assumed to reflect a feature [spread glottis] ([spread]), while short-lag VOT indicates a lack of laryngeal specification. Thus, short-lag stops are attributed to a putative default category that is both phonetically and phonologically inactive. While the earliest works within LR focus on how differences between voicing and aspiration languages are reflected in phonological phenomena such as assimilation, many working today within LR assume that VOT offers a transparent perspective on the phonetics-phonology interface, by which in two-series systems, initial stops reveal a what-you-see-is-what-you-get relationship between phonetics and underlying phonological specification. According to this view, the representation of the contrast based on aspiration (as in, e.g., English) is phonologically distinct from that of contrasts based on voicing (as in, e.g., French).

Underlying the discussion above is a fundamental issue concerning whether and how phonetic evidence may indicate the presence of an active phonological specification. For LR, phonologically active status should give rise to observable phonetic effects, while a lack of such

effects should indicate unmarked status. For the most part, we agree with this perspective, yet given the phonetic complexity of voice contrasts, we feel that it is important to search for these effects across a range of laryngeal cues. This, we believe, is a necessary endeavor, since some oft-cited work within LR (e.g., Beckman et al., 2011, 2013) bases its phonological claims primarily on VOT data.

While VOT suggests phonological interpretations that are relatively straightforward, the picture becomes more complicated when we consider additional phonetic correlates of laryngeal contrasts. Available evidence from one such correlate, consonant-induced f0 contours (Cf0; Kirby & Ladd, 2016) reveals largely parallel behavior in voicing languages (e.g., Kirby & Ladd [2016] for French and Italian; Pinget & Quené [2023] for Dutch; Schwartz et al. [2019] for Polish; Gao & Arai [2019] for Tokyo Japanese) and aspiration languages (e.g., Hanson [2009] for English). These experiments have found that, at the onset of the vowel following the obstruent, voiceless consonants induce f0 raising, while voiced consonants have no effect, relative to an f0 baseline produced by nasals. Kirby and Ladd (2018) suggest that this effect is due to active targeting of voicelessness by speakers, even in voicing languages in which, according to LR, unaspirated voiceless stops are phonologically unspecified. In short, comparisons of VOT with other laryngeal cues may give rise to conflicting interpretations of the phonological representations underlying two-series systems: VOT suggests distinct representations for voicing and aspiration languages. Cf0 in the studies cited above suggests parallel representations.

In examining how different phonetic cues suggest conflicting phonological interpretations, it is worth noting the distinct loci of the cues in question. VOT, of course, focuses our attention on stop closure and release. Meanwhile, a number of other correlates are housed on the vowel preceding or following voiced and voiceless obstruents (for discussion, see e.g., Kingston & Diehl, 1994). These include Cf0 mentioned above, as well as F1 onset frequency (Stevens & Klatt, 1974). An additional vowel-based correlate to laryngeal contrasts is voice quality. Broadly speaking, a given laryngeal category may be expected to induce breathier or stiffer phonation relative to other members of a contrast. These differences may be observed in acoustic measures of spectral tilt, such as H1*-H2*, or H1*-A1* (for more details and description of additional measures, see Section 2.4), based on amplitude differences between harmonics in the vocal wave. Breathier phonation is generally associated with a drop-off in amplitude of higher harmonics, yielding higher positive values for various spectral tilt measures (see e.g., Garellek, 2019). Voice quality may also be characterized acoustically with measures of periodicity such as cepstral peak prominence (CPP) and harmonics-to-noise (HNR) ratios (again, see Section 2.4 for details and description of additional measures), which quantify the amount of noise in the vocal wave. Higher values for periodicity measures are generally associated with modal voice relative to both breathy and creaky phonation (see e.g., Garellek, 2019).

Given the role phonation appears to play in evolutionary connections between voice contrasts and those based on tone or register (see e.g., Brunelle & Kirby, 2016; further discussion will follow in Section 3), we expect oppositions traditionally described as voiced vs. voiceless to have effects on the phonatory properties observed on neighboring vowels. For two-series languages, such effects have been documented in English, where textbook descriptions of allophonic effects on voice quality (e.g., Cruttenden, 2001), including aspiration and glottalization, have given rise to experimental (e.g., Kong et al., , 2012; Penney et al., 2018; Sanker, 2019) and corpus studies (Hejná et al., 2021) documenting phonation as a cue to consonant voicing. For example, Kong et al. (2012) include data from H1*-H2* as a function of initial voicing in English and Japanese, and find breathier phonation (higher values) after voiceless consonants. For other two-series systems, particularly those found in voicing languages, we know of very little published data on how underlying voicing may affect voice quality. Documenting such effects can give us further insight into the phonetic complexity of laryngeal contrasts. This insight, in turn, should be valuable for our understanding the nature of the phonological representations underlying the oppositions and the types of diachronic changes they may undergo.

This paper will describe an acoustic study of the two-way laryngeal contrast of phrase-medial word-initial stops in Polish, a language known to have a "standard" voiced-voiceless distinction in pre-vocalic positions. VOT in Polish has been thoroughly described (see, among others, Keating, 1980; Waniek-Klimczak, 2011; Sypiańska, 2013; Malisz & Żygis, 2015; Schwartz, 2022; Wojtkowiak & Schwartz, 2022), so our focus here is on correlates housed on the vowel following the initial stop, including a large set of acoustic measures of phonation, Cf0, as well as F1. While our primary empirical goal is to document the effects of underlying voicing in initial stop consonants on the phonetic realization of the following vowel, our theoretical aim is to consider the question of how laryngeal contrasts are encoded phonologically and predictions that phonological representations may make for the phonetic implementation of the contrast. In particular, we consider two recent proposals (Raimy, 2021; Schwartz, 2022), which would have us expect voicing-induced phonation differences on vowels but differ in the scope of their predictions, as well as their likely interpretation of the data we present.

While a more thorough presentation is provided in Section 3, the predictions of the Raimy (2021) and Schwartz (2022) proposals may be summed up briefly as follows. Raimy's account is based on the Avery and Idsardi (2001) model, in which phonological features are organized into *dimensions*, and two additional levels of representation are posited. Dimensions are linked with a *phonology-phonetics* level of representation through a process called *completion*. Another process, referred to as *enhancement*, may alter specifications between the phonology-phonetics level and an additional level of representation: *categorical phonetics*. Raimy's approach would have us expect phonetic effects of laryngeal marking induced by both voiced and voiceless obstruents. In the case of voiced obstruents, effects would be due to a feature [slack] that *completes* the Glottal

Tension dimension. Meanwhile, effects of voiceless obstruents would be due to a postulate that they are *enhanced* at the level of categorical phonetics with a feature [spread] (for more details, see Section 3). With regard to the phonetic realization of these effects, we may glean some predictions from Raimy's approach, although they are not stated explicitly. One prediction we might expect is a lowered f0 on the vowel following voiced obstruents (Kingston & Diehl, 1994), since the lesser laryngeal tension dictated by [slack] should reduce the frequency of vocal fold vibration. Meanwhile, Raimy's approach would presumably foresee raised f0 after voiceless obstruents due to [spread] *enhancement*, in parallel with Cf0 results from languages where [spread] is specified phonologically (e.g., Hanson, 2009). With regard to phonation, it is somewhat more difficult to make inferences from Raimy's account. In principle, both [slack] on voiced onsets and [spread] on voiceless onsets might be expected to induce some degree of breathiness (higher values for spectral tilt measures, lower values for periodicity measures). However, since [spread] is a function of phonetic *enhancement* that should be most perceptually robust after stop release, while [slack] is claimed to be a phonological specification responsible for pre-voicing during stop closure, perhaps we can interpret the Raimy proposal as being predictive of more robust breathiness effects after voiceless onsets than voiced onsets.

Schwartz's proposal, couched within the Onset Prominence representational framework (OP; see Schwartz, 2022, for an OP account of two-series laryngeal phonology), makes slightly different predictions. Onset Prominence is a non-segmental framework in which manner of articulation is specified structurally, derived from the hierarchy of phonetic events inherent to a stop-vowel sequence. This approach encodes the CV transition phase as a structural node within the primary building block of the model[1] (see Section 3 for more details). In accordance with the structural properties of obstruents within OP, Schwartz predicts that effects of onset laryngeal specifications in voicing languages should be expected in the early portion of the vowel following the consonant and that these effects should diminish later in the vowel. Raimy's proposal makes no explicit predictions about the time course of phonetic effects on the following vowel–presumably they would be attributed to implementation. Schwartz's other prediction is that phonetic effects on the following vowel should be induced by voiceless onsets but not voiced ones. This is because the latter lack laryngeal specification in the OP account (see Section 3), while the former are specified with a feature [fortis]. At the same time, the Schwartz proposal makes no explicit predictions about what the phonetic effects induced by [fortis] at vowel onset might be. A summary of the predictions gleaned from two phonological accounts is given in **Table 1**.

[1] Other models that posit a structural position for CV transitions include Aperture Theory (Steriade, 1993) and Q Theory (Shih & Inkelas, 2019). Onset Prominence differs from these models in positing a hierarchical rather than linear relationship among structural positions.

Prediction	Raimy (2021)	Schwartz (2022)
Effects of both voiced and voiceless onsets?	Yes	No. Effects only after voiceless onsets
Time course of effects	No explicit predictions	Effects in the first half of vowel
Phonetic shape of effects	F0 lowered after voiced onsets, raised after voiceless onsets. Perhaps breathier phonation after voiceless onsets	No explicit predictions

Table 1: Phonetic predictions of Raimy and Schwartz models of Polish laryngeal phonology.

The differences between these proposals bear on persistent issues in laryngeal phonology, including markedness, the phonetic underpinnings of phonological feature specifications, the status of the segment, and the use of multiple levels of representation in modeling the phonetics-phonology interface. We will revisit these issues in Section 3.

Returning to our experimental study, we state our research questions below.

- RQ1: To what degree do phonation, along with f0 and F1, contribute to the realization of the two-series laryngeal contrast in Polish?
- RQ2: How can data from these features help us evaluate phonological proposals for encoding the two-series laryngeal systems?
    - RQ2a: Are phonetic effects of onset laryngeal specification on the following vowel induced by voiced consonants, voiceless consonants, or both?
    - RQ2b: What is the time course of any effects we may observe?

We will present data from a wide variety of acoustic measures of phonation, as well as Cf0 and F1, to offer what we believe to be the most comprehensive look to date at the phonetics of the word-initial laryngeal contrast in Polish. Before proceeding, we review previous literature on the phonetics of the Polish word-initial voice contrast.

1.1. Voicing in initial stops in Polish

With regard to the phonetic realization of voicing contrasts in Polish, a large number of studies over the past several decades (Keating, 1980; Keating et al., 1981; Sypiańska, 2013; Malisz & Żygis, 2015; Wrembel, 2015; Schwartz, 2022; Wojtkowiak & Schwartz, 2022, among others) have described VOT. Over these years, the prevalence of pre-voicing and VOT values for voiceless plosives reported in these studies have been relatively stable. Back in 1981, Keating et al. reported on word-list data from Polish. They observed mean VOT values of Polish voiceless stops

in the range between 22–53 ms and pre-voicing in nearly all voiced plosives. The later papers cited above show similar results, with somewhat shorter values (means between 15–30 ms) for /p t k/ reported in studies based on larger utterances (Sypiańska, 2013; Malisz & Żygis, 2015; Wojtkowiak & Schwartz, 2022) than those based on words produced in isolation (Wrembel, 2015; Schwartz, 2022). In those studies that looked at voiced plosives, pre-voicing is nearly always present. The mean VOT for voiceless plosives appears to be slightly longer than what is described for short-lag lenis stops in aspiration languages such as English. Raimy (2021, p. 86) refers to this as "mid-lag" VOT, which constitutes one basis for a phonological analysis to be considered in Section 3.

Looking beyond VOT, we are aware of only a few studies that describe vowel-based acoustic correlates of the Polish voicing contrast. Schwartz et al. (2019) provide preliminary data on Cf0 and F1, based on words produced in isolation by 14 speakers. For f0, they observed that voiceless plosives induced higher f0 in the first 25% of the following vowel, while voiced plosives produced f0 values that were indistinguishable from a nasal-based reference point. Ting et al. (2025) describe a corpus-based study of both Cf0 and f0 effects derived from vowel quality in 20 languages, one of which is Polish. In their Appendix D, the researchers provide trajectories estimated by Generative Additive Mixed Models (GAMMs). The Cf0 trajectories for Polish (as well as a couple of other Slavic languages) do not include nasals but, nevertheless, suggest a raising effect after voiceless obstruents (see Ting et al., 2025, Appendix D) and a relatively neutral trajectory after voiced obstruents. These results are compatible with other recent studies describing voicing languages, including French and Italian (Kirby & Ladd, 2016), as well as Dutch (Pinget & Quené, 2023). Finally, the study by Schwartz et al. (2019) also found higher F1 frequency at vowel onset after voiceless obstruents than voiced ones (Stevens & Klatt, 1974; Lisker, 1975). It is assumed that this effect occurs when the typically rising F1 transition from consonant to vowel is obscured after voiceless onsets. In the Schwartz et al. (2019) data on Polish, the effect is most reliable for low and mid vowels, since high vowels with a low F1 produce a less robust transition.

The perceptual implications of these effects are described in Schwartz and Arndt (2018) and Schwartz et al. (2019). In the former, Polish listeners showed very high discrimination accuracy of the contrast, even when pre-voicing was absent from the voiced series, suggesting an important perceptual role of vowel-based acoustic correlates such as Cf0 and F1 onset. In the latter study, a phoneme monitoring task used cue mismatches among VOT, Cf0 and F1 to test the perceptual weight of each cue. The results revealed no perceptual advantage for VOT over the vowel-based cues. The results of those studies are compatible with Keating et al.'s (1981) earlier conclusion, based on VOT continuum effects, that VOT in Polish bears less perceptual weight than it does in English.

Returning to production, Schwartz et al.'s (2019) Cf0 results were replicated in a separate production study in Wojtkowiak and Schwartz (2022), based on 20 speakers reading meaningful Polish sentences. With regard to acoustic measures of phonation, Schwartz et al. (2023) provide preliminary data on four measures of spectral tilt from 15 speakers producing words in isolation, observing significant but small differences as a function of voicing of the preceding consonant. The present article builds on this previous research, providing data from a larger number of speakers (30), a much larger inventory of acoustic measures of phonation, and a more detailed look at the time course of the effects of onset voicing on the following vowel.

2. Acoustic Study

2.1. Participants

The participants of our study were 30 native speakers of Polish. The group consisted of 20 females and 10 males, aged 18–29, with a median age of 24. Participants signed informed consent forms and were compensated for their time, receiving gift certificates valid at local retail businesses.

The participants included both university students and young adults recruited from the local community. Those that were students were studying a range of academic disciplines, but those disciplines did not include linguistics, phonetics, or foreign-language-related subjects. While all participants reported having had some exposure to foreign languages, primarily English or German, as part of compulsory school education, they claimed not to be fluent in, nor able to actively speak, any language other than Polish. Despite classroom exposure to foreign languages, they reported no practical speaking ability and did not use any other language in daily communication. Additionally, none of the participants had received any training in phonetics or foreign language pronunciation. Accordingly, we would characterize these speakers as being functionally monolingual.

Evidence in support of the monolingual status of our speakers can be found in Wojtkowiak (2022), which compares VOT in this group of speakers to several groups of students specializing in English and receiving explicit training in English pronunciation. That study found robust differences in the probability of pre-voicing as a function of group (more proficient users of English showed less pre-voicing in L1 Polish), indicating L1 phonetic drift (e.g., Chang, 2010) in the speech of the students of English, but not in the speech of the monolinguals.

2.2. Materials

Our dataset for the analysis of the voicing contrast was comprised of disyllabic Polish words beginning with the plosives /p, t, k, b, d, g/.[2] The target words were consistently two syllables

[2] Unlike some studies of Cf0, which used nasals as a reference point for examining effects of obstruent voicing, in this study we do not include a nasal reference point. The reason for this is that we are also examining the effects of the onset on acoustic measures of phonation, based on harmonic amplitudes, which nasality is known to affect (Chen, 1997; Styler, 2017).

long to ensure uniform lexical stress on the penultimate syllable. The plosives were followed by the vowels /a/, /ɛ/, or /ɔ/, and always preceded by /i̯/. Our study focuses on data from a subset of recordings made for a larger project looking at the effects of prosodic position on segmental phonetics in Polish. The analysis is based on a set of target words that were embedded in meaningful sentences, carefully controlled for prosodic position. In each case, the target word is preceded by eight syllables, followed by either seven or eight syllables. While the original collection of sentences placed target words in either utterance-initial, phrase-initial, or phrase-medial position (see details of the distinct prosodic positions in Wojtkowiak & Schwartz, 2022), and both accented (focused) and unaccented productions were elicited, the present analysis is limited to the unaccented (i.e., not under focus), phrase-medial condition, in an attempt to control for various prosodic factors that might influence phonation. While we acknowledge that an analysis of only phrase-medial, unaccented position may limit the applicability of our study to other prosodic environments, we consider unaccented and phrase-medial as the default with regard to prosodic structure. In our view, this allows for a clearer focus on the phonetic effects of segmental phonological specification without possible confounds that may be induced by prosodic variation.

In (1) we present an example of one of the test sentences containing a target word to give readers an idea of the structure of the utterances that were read. The entire collection of sentences is given in the Appendix. In (1), the target word *dowód* [dɔvut] 'proof' is italicized.

(1) Example of a sentence with target word *dowód* (see Appendix for the entire set)
Na rozprawie przedstawimy *dowód* winy oskarżonego
'At the trial we will present *proof* of the accused's guilt'

In each case, the target word acts as an object (usually a noun in the accusative case or, alternatively, an infinitive verb) to the finite verb that directly precedes it. This structure was designed to ensure a phrase-medial reading of the target word. To establish whether unaccented, phrase-medial prosodic positions were successfully elicited, the recordings were checked by two phonetically trained native speakers of Polish to establish whether the intended prosodic positions were indeed produced. Cohen's Kappa coefficient was calculated over the larger corpus ($k = 0.97$) and items with interrater discrepancies were removed. Of the intended 540 items (6 plosives * 3 vowel contexts * 30 speakers), 29 items were removed due either to inter-rater discrepancies or errors on the part of the speaker, leaving a total of 511 total items.

2.3. Recording procedure and annotation

As mentioned above, the data to be described here come from recordings made in a project devoted to the effects of prosodic position on segmental phonetics. For this reason, meaningful Polish sentences were created to elicit prosodic position in a controlled manner. Prior to the recording session, participants familiarized themselves with the sentence list, and they were instructed to

practice reading them before recording began. The experimenter also gave instructions designed to ensure that the speakers produced the intended prosodic structures. The experiment took place in a sound-treated recording studio at a Polish university. Using SpeechRecorder (Draxler & Jänsch, 2015), which randomizes the order of elicited items, the target sentences were presented on slides on a monitor inside the booth, and speakers read directly off the slides. Recordings were made directly onto a laptop computer (44.1 kHz sampling rate; 16-bit resolution) using a high-quality head-mounted microphone (SHURE SM35-XRL) connected to a USB audio interface (Edirol UA-25EX).

Recordings were annotated manually into Praat textgrids (Boersma & Weenink, 2025). The analysis here is based on a tier housing stressed vowels in the target words. In this way, our focus is on phonation in the vowel following either a voiced or voiceless plosive. The segmentation of the vowels was based on onset and offset of full formant structure, or formant columns, as described by Machač and Skarnitzl (2009). During the annotation process, care was taken to verify that the stressed vowels in the target words were produced with impressionistically modal phonation. As it happens, it was not necessary to exclude any non-modal items. In our recordings, creaky voice was sometimes found phrase-finally—and quite regularly on word-initial vowels—but it did not appear in the target words analyzed here.

Figures 1 and **2** present examples of boundaries established during the annotation procedure with a close-up view of the word-initial stops of the target words, the preceding vowel, and the following vowel that constitutes the empirical focus of the present analysis.

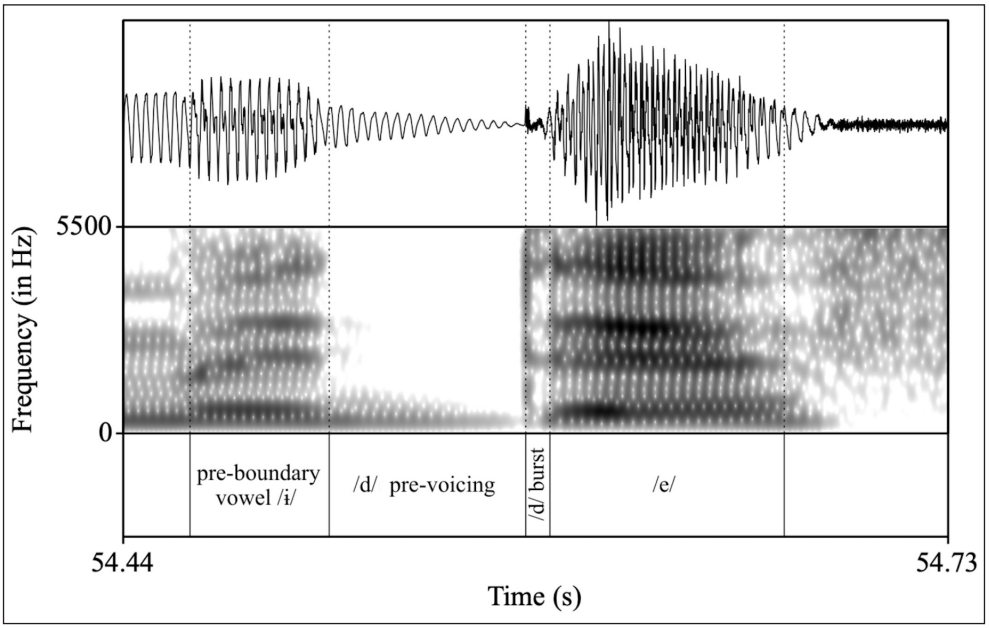

Figure 1: Close-up view of an annotation of *usuniemy defekt* 'we'll fix the glitch', including the word-final vowel in *usuniemy*, the word-initial /d/ and the following vowel.

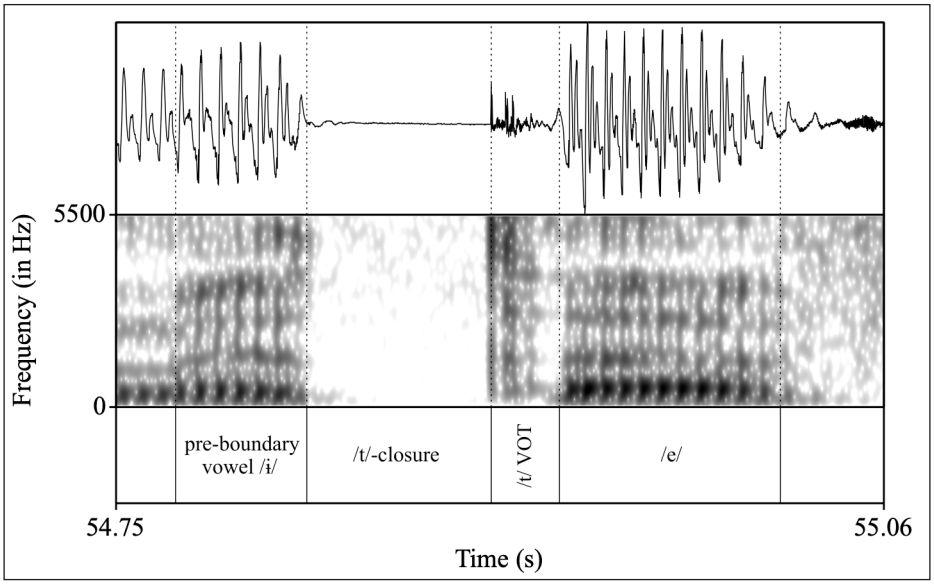

Figure 2: Close-up view of an annotation of *stosujemy teflon* 'we use teflon', including the word-final vowel in *stosujemy*, the word-initial /t/ and the following vowel.

2.4. Acoustic parameters

A total of 27 acoustic parameters, including fundamental frequency, harmonic amplitudes, spectral tilt, spectral noise, formant frequencies and energy related parameters were extracted using VoiceSauce (Shue et al., 2011). These parameters were estimated at 1 ms intervals, with pitch range settings of 100–500 Hz for female and 75–300 Hz for male speakers (see e.g., Johnson & Babel, 2023).

Fundamental frequency is of interest due to previous research investigating the effects of onset voicing, mentioned in Section 1. F0 is measured using the STRAIGHT algorithm (Kawahara et al., 1998). VoiceSauce then uses the f0 estimate to identify the harmonics within the spectrum and determine their corresponding amplitudes. To reduce variability in harmonic magnitudes computed with a fixed time window, VoiceSauce computes harmonic spectra pitch-synchronously over windows spanning three pitch periods. By default, the first four formants and their corresponding bandwidths are found using the Snack Sound Toolkit (Sjölander, 2004). The frequency of the first formant (F1) at vowel onset is of particular interest considering earlier findings on the effects of consonant voicing, mentioned in Section 1. The harmonics closest to the formant frequencies are then identified, and their amplitudes are assigned as the amplitudes of the corresponding formants. To account for the filtering effect of vocal tract, harmonic amplitudes are corrected using the formant frequencies and their corresponding bandwidths in an algorithm described by Iseli and Alwan (2004). The spectral tilt measures returned by VoiceSauce (H1*-H2*, H2*-H4*, H4*-H2K*, H2K*-H5K, H1*-A1*, H1*-A2*, H1*-A3*) reflect amplitude differences, while the asterisks indicate that the relevant harmonics have been corrected for vocal tract filtering.

In addition to spectral tilt measures, VoiceSauce also employs cepstral analysis to quantify periodicity and noise-related characteristics through various parameters. Specifically, harmonics-to-noise ratios (HNRs) are computed across distinct frequency bands (0–500 Hz, 0–1500 Hz, 0–2500 Hz, and 0–3500 Hz) using de Krom's (1993) algorithm. Cepstral peak prominence (CPP) is computed following the methodology outlined by Hillenbrand et al. (1994). This measure quantifies the prominence of the cepstral peak by determining the amplitude difference between the peak itself and the corresponding value on a regression line fitted to the cepstrum at the same quefrency. The subharmonic-to-harmonic ratio (SHR) is another periodicity measure, which calculates the relative amplitude of subharmonics to harmonics. It serves as an indicator of spectral noise linked to phonatory irregularities, including period doubling. In VoiceSauce, SHR is computed following the algorithm proposed by Sun (2002). Root mean square Energy is a spectral noise measure computed over a window spanning five pitch periods and reflects the overall amplitude. The strength of excitation (SoE) quantifies the relative amplitude of the impulse-like excitation occurring at the moments of significant excitation during voiced speech. Its implementation in VoiceSauce is based on the algorithm described by Murty and Yegnanarayana (2008).

2.5. Data visualization and analysis

Initial data cleaning was done according to guidelines described in Johnson and Babel (2023). Individual data points were removed if VoiceSauce returned a value of zero for f0, formants 1–4, CPP, and H5k Hz. Data points in which Energy exceeded three standard deviations above the grand mean were also removed. For SHR, zeros were replaced by 0.0001 in accordance with Johnson and Babel. The original VoiceSauce analysis yielded 40,223 data points, of which 5,166 were removed, leaving a total of 35,057 observations. Following removal of individual data points, entire tokens were removed if fewer than 10 data points remained per token. If an entire token was removed, then the corresponding item within its voiced-voiceless pair was also removed to maintain balance in the dataset with regard to the laryngeal contrast. For example, if a /da/-initial word was removed, then the corresponding /ta/-initial word from the same speaker was also removed. Token removal left us with 462 items for analysis. All acoustic parameters were centered and standardized within speakers using z-scoring.

VoiceSauce analyses return a very large number of acoustic measures that are often correlated with each other. As noted by Keating et al. (2023, p. 378), the sheer volume of measures may be somewhat overwhelming, leaving researchers with the dilemma as to whether to reduce the number of acoustic dimensions in the dataset. A commonly used dimension-reduction strategy is principal components analysis (PCA), which we will use in our study, as well. At the same time, one of the main goals of the present study is descriptive: We are interested in the effects of onset voicing on the acoustics of the following vowel over a range of acoustic measures, as well as the

time course of those effects. Therefore, before proceeding to the PCA results, we will present visualizations of raw data for a selection of parameters based on a segmentation of each vowel into 10 intervals.

As we shall see, the visualizations suggest that the most robust effects of consonant voicing occur in the early part of the vowel (intervals 1–4), after which the trajectories tend to come together. For this reason, we resegmented the vowels into thirds and performed the PCA on the initial 33.3% of the vowel's duration. To identify the most important acoustic parameters associated with the voicing contrast, we employ the following three-step process.

1. PCA is conducted on by-token means of the first one-third of each vowel.
2. The components returned by PCA are used as predictors in a binary logistic regression model to see which components contribute to the odds of the initial consonant being voiced or voiceless.
3. We examine the correlations of the original acoustic parameters with the components that come out as significant at step 2.

Additionally, to further examine the time course of effects induced by the voicing specification of the preceding plosive on the phonation of the following vowel, we ran a GAMM analysis fitted with the mgcv package (Wood, 2011) on all of the acoustic measures and interpreted with the itsadug package (van Rij et al., 2025) in R (R Core Team, 2025). These measures served as dependent variables in the models, which, like the first set of visualizations to be shown, were performed over the 10-interval segmentation. Each model contained onset voicing (voiced-voiceless) as a categorical predictor and also included the quality of the following vowel (/a ɛ ɔ/) as a control variable. A smooth term for interval, as well as a by-voicing difference smooth for interval, were also included. The inclusion of the difference smooth enabled significance testing of the effect of voicing through model comparison (e.g., Sóskuthy, 2017). Sóskuthy (2021), through Type I/Type II error simulations, compares several methods of significance testing in the GAMM context. Significance testing through the inclusion of a difference smooth and model comparison results in acceptable rates of Type I errors in contrast to the visual method of assessing whether the confidence band of the difference smooth does not include 0, without a specific hypothesis regarding the cutoff for minimal interval of lack of said overlap. Additionally, the models include by-speaker and by-word random smooths, which may be thought of as the dynamic counterparts of random intercepts and slopes in standard mixed models. Finally, to pre-empt issues with autocorrelated errors, each GAMM model included an error model with ρ derived from a model without an error term, in a way suggested by Baayen et al. (2018). We used the results of the PCA to decide on a smaller selection of acoustic measures to report on from the GAMM analysis. Results of the GAMMs for all the measures are provided in the supplementary materials.

In sum, our analysis is intended to provide a detailed picture of the effects of voicing in Polish plosives on the acoustics of the following vowel. For this reason, we are most interested in effects that are concentrated in earlier portions of the vowel, as these may be attributable to the preceding consonant. The data and scripts necessary to reproduce our analysis are available at https://osf.io/zv9w6/?view_only=0ad6979a8d22478fb518470e93fb2074.

2.6. Results

While our focus here is on acoustic correlates of obstruent voicing housed on the vowel following stop release, VOT results for these recordings are reported in Wojtkowiak and Schwartz (2022). The mean VOT for voiceless plosives, averaged across places of articulation and vowel qualities, was 29 ms (SD = 12) for voiceless plosives (/p/: M = 21.6 ms, SD 6.5; /t/: M = 25.9 ms, SD = 8.3; /k/: M = 42.1 ms, SD = 10.3), and –72 ms (SD = 22) for voiced plosives (/b/: M = –83.7, SD = 26.5; /d/: M = –65.3 ms, SD = 19.9; /g/: M = –67.9, SD = 17.6).[3] Those results are compatible with other descriptions of VOT available in the literature on Polish.

2.6.1 Visualizations of raw data for vowel-based voicing correlates

For the vowel-based cues, we start by showing visualizations of the raw data based on a segmentation of the vowel into 10 intervals and Generative Additive Models with vowel interval and initial consonant voicing as predictors. These visualizations were performed to gain an overview of the contrast before carrying out more rigorous statistical analyses. Inspection of the initial visualizations, both the ones shown here and those provided in the supplementary materials, informed the statistical analyses described in sections 2.6.2 and 2.6.3.

We show raw data from Cf0 and F1, and from the following voice quality measures: H1*-H2*, H2*-H4*, H4*-H2K*, H2K*-H5K, as well as two measures of periodicity: CPP and HNR05. The spectral tilt measures are included in Kreiman et al.'s (2014) psychoacoustic model of voice quality, supplemented with CPP as suggested by Garellek (2019), while HNR05 emerged as a heavily weighted parameter in Keating et al.'s (2023) cross-language study aimed at establishing a common acoustic space for voice quality distinctions. Confidence bands in the initial visualizations represent extrapolations of 95% confidence intervals through successive sections of the vowel. Visualizations for all measures extracted using VoiceSauce can be found in the supplementary materials.

Figure 3 shows Cf0 as a function of consonant voicing. The confidence bands suggest that f0 is higher after voiceless consonants, but shows a neutral contour after voiced consonants

[3] The VOT study described in Wojtkowiak and Schwartz (2022) examined only 20 speakers, but included a wider range of prosodic positions (including phrase-initial and utterance-initial) in both accented (under focus) and unaccented conditions. It was found that VOT values were relatively stable, showing only a small degree of variation as a function of prosodic position and accent.

(cf. findings from a number of other studies of voicing languages: e.g., Kirby & Ladd, 2016, for French and Italian; Pinget & Quené, 2023, for Dutch; Gao & Arai, 2019, for Tokyo Japanese).

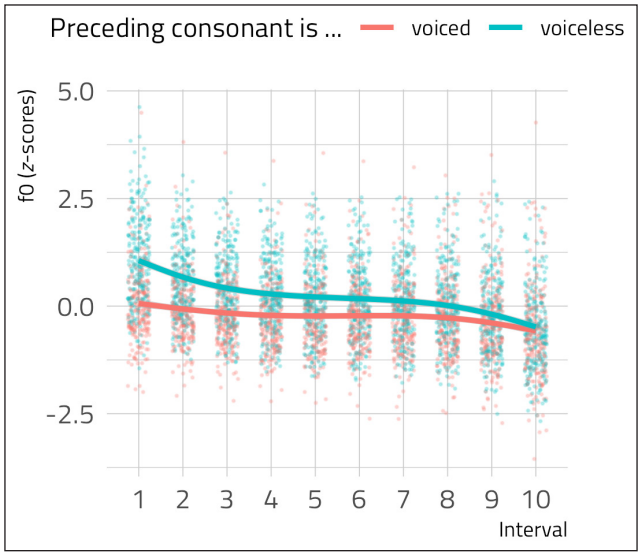

Figure 3: Cf0 after voiced (orange) and voiceless (aqua) consonants. Confidence bands extrapolate 95% confidence intervals over each of the 10 sections of the vowel.

Figure 4 shows the onset-induced trajectories for the first formant (F1). We can observe slightly higher values early in the vowel after voiceless consonants than voiced consonants. F1 transition frequencies are a known correlate of consonant voicing, and what we see in **Figure 3** is compatible with other findings (e.g., Stevens & Klatt, 1974).

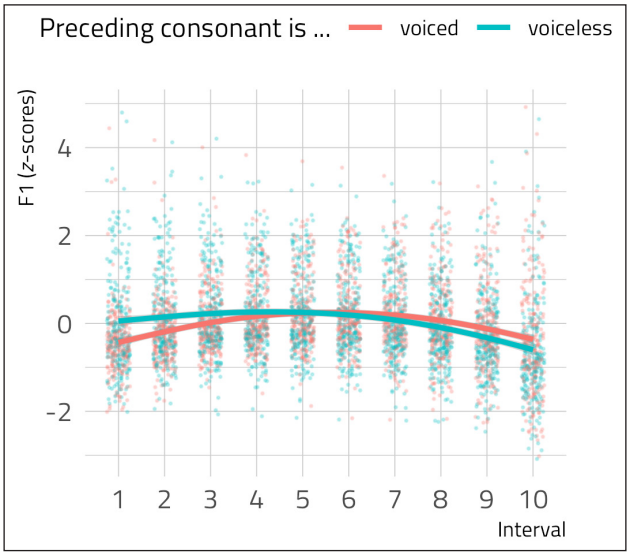

Figure 4: F1 after voiced (orange) and voiceless (aqua) consonants. Confidence bands extrapolate 95% confidence intervals over each of the 10 sections of the vowel.

Figure 5 presents raw data from four spectral tilt measures included by Kreiman et al. (2014) in their psychoacoustic model of voice quality. The trajectories for H1*-H2* and H2*-H4* appear to indicate that voiceless consonants induce breathier phonation (higher values). This difference appears to be minimized for H4*-H2K*, while H2K*-H5K appears to induce higher values for voiced consonants early in the vowel. On the whole, the visualizations suggest that breathiness induced by voiceless onsets is concentrated in the lower portions of the spectrum.

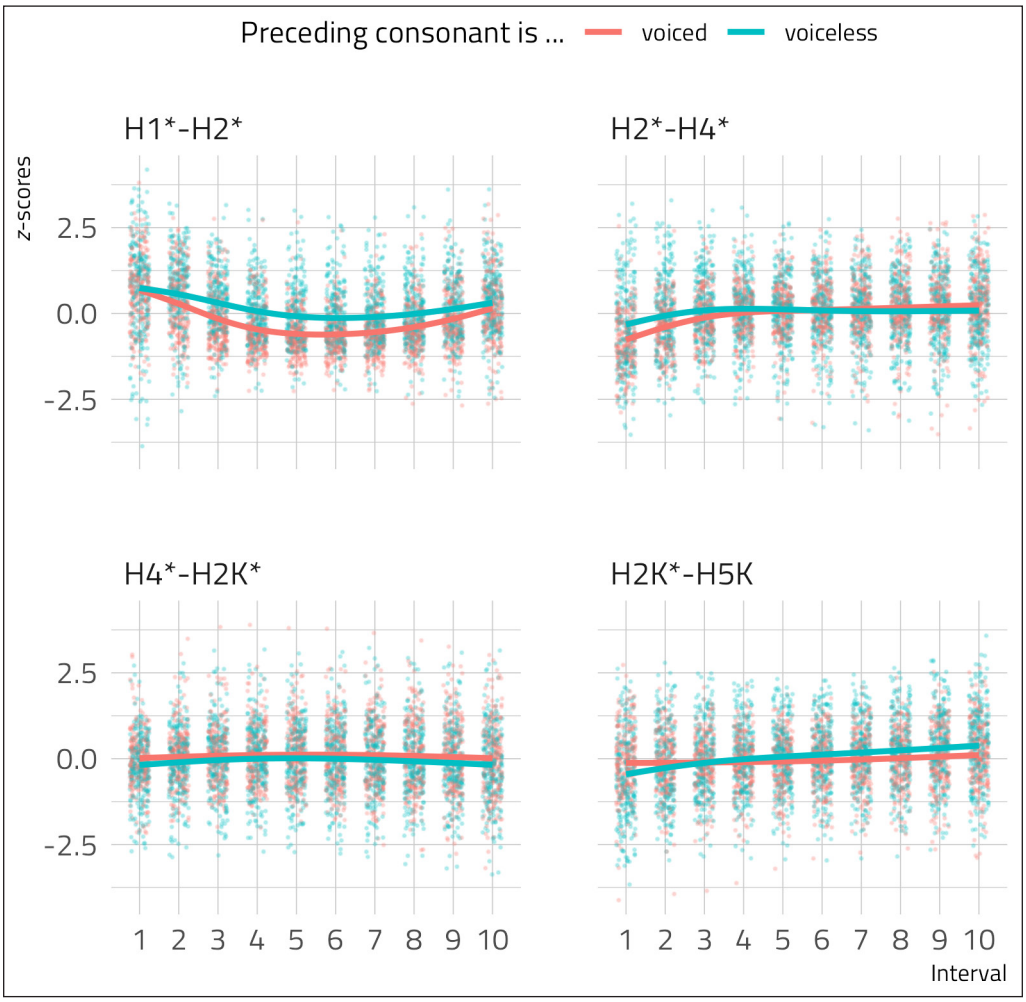

Figure 5: Spectral tilt after voiced (orange) and voiceless (aqua) consonants. Confidence bands extrapolate 95% confidence intervals over each of the 10 sections of the vowel. Top left: H1*-H2*, top right: H2*-H4*, bottom left: H4*-H2K*, bottom right: H2k*-H5K.

Figure 6 presents raw data from two periodicity measures: CPP and HNR05. CPP is included by Garellek (2019) as a supplement to the four spectral tilt measures used by Kreiman et al. (2014). Meanwhile, HNR05 is suggested as one of the most informative parameters by Keating et al. (2023). These two measures, while closely related, differ in the frequency range from which

they are calculated. CPP is taken from the entire harmonic spectrum, while HNR05 derives from the frequency band between 0–500 Hz. Both voiced and voiceless onsets appear to lower the periodicity measures, which increase later in the vowel, while overall the periodicity measures are higher after voiced consonants. Interestingly, the trajectory for HNR05 suggests that the general increase in periodicity from consonant to vowel is delayed somewhat after voiceless onsets, such that only after Interval 3 does the value increase sharply. Such a delay is not visible in the trajectory for CPP.

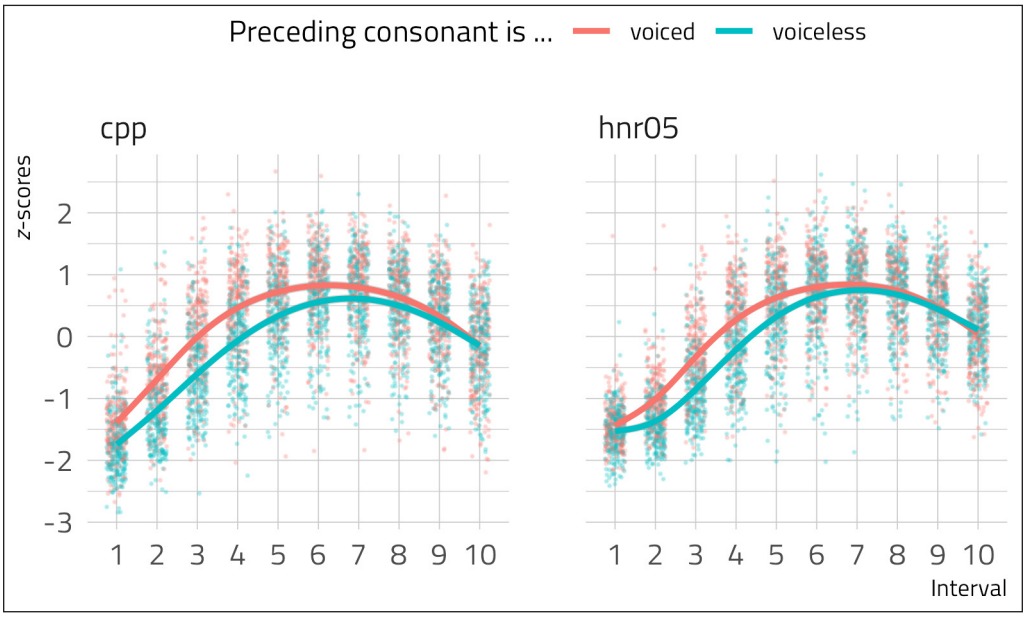

Figure 6: CPP (left) and HNR05 (right) after voiced (orange) and voiceless (aqua) consonants. Confidence bands extrapolate 95% confidence intervals over each of the 10 sections of the vowel.

Summing up our visual investigation of eight vowel-based correlates of the voicing contrast, we may formulate the following tentative conclusions. First of all, it appears that onset voicing influences a range of vowel-based cues, including Cf0 and F1, as well as voice quality. Second, apparent voicing-induced differences are concentrated relatively early in the vowel, while the trajectories tend to converge later in the vowel. Finally, with regard to voice quality, it appears that voiceless consonants induce breathier modal phonation than voiced consonants, reflected in acoustic measures encompassing the lower part of the harmonic spectrum. In what follows, we present the results of statistical analyses to determine the reliability of these conclusions.

2.6.2 Principal Components Analysis (PCA)

In an attempt reduce the acoustic dimensionality of the data extracted using VoiceSauce, we performed a PCA on 16 acoustic parameters (HNR05, HNR15, HNR25, HNR35, CPP, SHR, Energy, H1*-H2*, H2*-H4*, H4*-H2k*, H2k*-H5k, H1*-A1*, H1*-A2*, H1*-A3*, f0, and F1; SoE

and individual harmonic amplitudes were not included). On the basis of the visual inspection of the raw data, we decided to resegment the vowels into three intervals and run the PCA on the mean values for the acoustic parameters over the first third of the vowel. The resegmentation focused attention on effects induced by the initial consonants. Inspection of the scree plot (see supplementary materials) produced by the PCA revealed three relevant components. Component 1 accounts for 28.5% of the variance, Component 2 for 22.2%, Component 3 for 10.3% and Component 4 for 10%. The drop going from Component 3 to Component 4 is much less steep compared to the much steeper drop from Component 2 to Component 3, indicating that the cutoff should be placed after Component 3.

To model voicing as a function of the PCA components, we ran a mixed-effects binary logistic regression model with onset voicing as the response variable, Components 1, 2, and 3 as predictor variables, and a by-speaker random intercept (voicing ~ dim_1 + dim_2 + dim_3 + (1 | speaker)) fitted with the lme4 package (Bates et al., 2015) in R (R Core Team, 2025). Significance testing was performed with the "joint_tests()" function in the emmeans package (Lenth, 2025). The model revealed that estimates for all three components were significant. Component 1 scores show a positive association with the odds that initial consonant is voiced ($b = .74$, $p < .001$). Component 2 also shows a positive association with the odds that the onset was voiced ($b = .333$, $p < .001$). Component 3 shows a negative association with the odds that the onset was voiced ($b = -.70$, $p < .001$). Note that the association strengths for Components 1 and 3 are greater (their estimates are further from 0) than that for Component 2. These effects are summarized visually in the partial-effect plots shown in **Figure 7**.

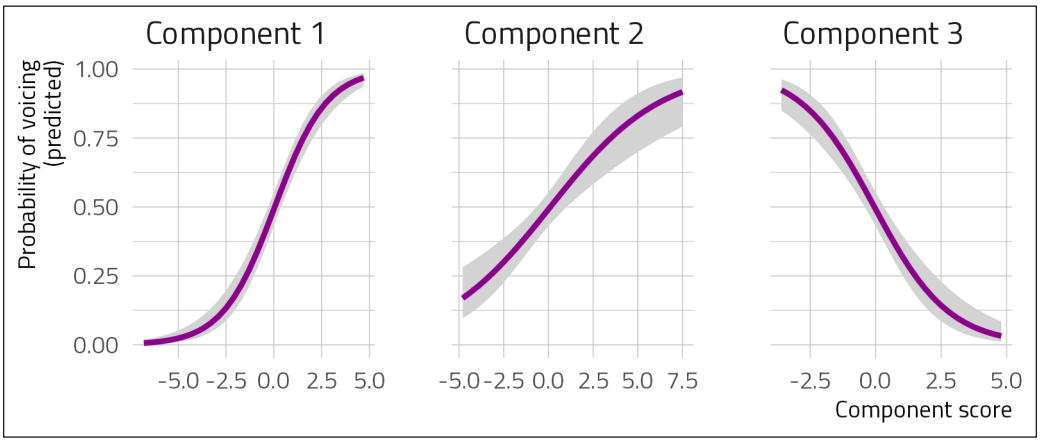

Figure 7: Effects of Components 1, 2, and 3 on the probability that the onset consonant is voiced. Effects estimated with the "effects" package (Fox & Weisberg, 2019).

Table 2 provides correlation coefficients for each acoustic parameter within each component. In the case of Component 1 and Component 2, a positive value is associated with an increase in the odds that the consonant is voiced, while in Component 3, a positive value is associated

with a decrease in the odds that the onset is voiced. In accordance with guidelines outlined in Tabachnick and Fidell (2013) and described by Johnson and Babel (2023), component-parameter correlations with absolute values of more than 0.63 are "very good," and correlations with an absolute value higher than 0.71 are "excellent." In **Table 2**, all correlations that are either "very good" or "excellent" are shown in bold. For phonation measures in Component 1, these include HNR25 with a positive correlation (higher values are associated with higher probability that the onset is voiced), H1*-A1* and H1*-A2* with a negative correlation (lower values are associated with higher probability that the onset is voiced). F1 (higher values for voiceless consonants) also appears to have a good correlation with Component 1. In Component 2, we see three

Parameter (estimated effect on the odds of onset being "voiced")	**Comp1 (.735)**	**Comp2 (.325)**	**Comp3 (–.703)**
HNR05	.52***	***.74*****	ns
HNR15	.62***	***.68*****	–.1*
HNR25	***.67*****	***.63*****	.21***
HNR35	.62***	.6***	.33***
CPP	.55***	.43***	–.14**
SHR	.31***	ns	–.13**
Energy	.3***	ns	–.14**
H1*-H2*	ns	ns	.42***
H2*-H4*	–.57***	.33***	.18***
H4*-H2k*	–.22***	.35***	***–.78*****
H2k*-H5k	.58***	–.41***	.55***
H1*-A1*	***–.75*****	.44***	.33***
H1*-A2*	***–.64*****	.52***	.3***
H1*-A3*	–.6***	***.65*****	ns
f0	–.26***	–.21***	.32***
F1	***–.68*****	.51***	ns

Table 2: Correlation strengths between individual acoustic parameters and PCA components (***: $p < .001$; **: $p < .01$; *: $p < .05$; ns: non-significant).

HNR measures, as well as H1*-A3*, showing strong positive correlations with voicing (higher values more likely to be associated with voiced consonants). Finally, for Component 3, which is negatively correlated with the probability that the onset was voiced, H4*-H2K* shows a strong negative correlation (i.e., higher values more likely to be associated with voiced consonants).[4]

For the most part, inspection of the phonetic makeup of the individual Components in the PCA is compatible with the tentative conclusions suggested by the visual inspection of the raw data. The association of voiceless onsets with breathier phonation is seen quite clearly in Component 1, most robustly in H1*-A1* and HNR25. Additionally, the PCA results appear to be compatible with the claim that the breathiness effects of voiceless consonants disappear at the higher end of the acoustic spectrum. This perhaps explains the fact that the H2K*-H5K, a spectral tilt measure based on higher-frequency harmonic amplitudes, behaves differently from the other spectral tilt measures across components 1 and 2.

2.6.3 GAMM analyses and significance testing

In the PCA analysis, components 1 and 3 had the strongest association with the voicing contrast, as we saw in **Figure 7**. To further examine the effects of onset voicing over the time course of the vowel, we ran GAMMs according to the description provided in Section 2.5. To test the significance of the GAMM trajectories, we used the *compareML()* function. The *p*-values were adjusted using the Holm-Bonferroni correction for multiple comparisons among all the measures. As shown in **Table 2**, the measures that contributed most to these components were HNR25, H1*-A1*, H1*-A2*, and F1 in Component 1, and H4*-H2K* in Component 3. These tests revealed significant effects of consonant voicing for each of these measures, with the exception of H4*-H2K*, and are summarized in **Table 3**.

Parameter	Adjusted R^2	Deviance Explained	p-value corrected
H1*-A1*	0.27	0.31	$<.001$
H1*-A2*	0.18	0.23	.004
H4*-H2K*	0.24	0.27	.908
HNR25	0.58	0.60	$<.001$
F1	0.39	0.42	.001

Table 3: Summary of GAMM results for most heavily weighted parameters in the PCA (see **Table 2**).

[4] A *Laboratory Phonology* reviewer of this paper asked about differences between measures that emerged in our PCA analysis and measures found in PCA analyses of other studies. Perhaps the differences stem from the fact that many of those studies compare categorically different voice qualities, while in our study we are looking at differences within the confines of modal voice.

Visualizations of the raw data for H4*-H2K* and F1 have been shown in **Figures 4** and **5**, respectively. To illustrate effects of voicing in the other three parameters that emerged from the PCA, **Figures 8–10** show prediction and difference smooths from the GAMMs. In these visualizations, the difference smooths plot the difference in the measures between voiced onsets and voiceless onsets. When the curve is above the line, the value for the measure is higher after voiced consonants, and when it is below the line, the value is higher after voiceless consonants. For HNR25 (**Figure 8**), we observe lower values for voiceless consonants early in the vowel, as we saw for HNR05 in **Figure 6**. For the other two spectral tilt measures in **Table 3**, we observe higher values after voiceless consonants, an effect which is more robust for H1*-A1* (**Figure 9**) than H1*-A2* (**Figure 10**). GAMM visualizations (both prediction and difference smooths) for all the acoustic measures may be found in the supplementary materials.

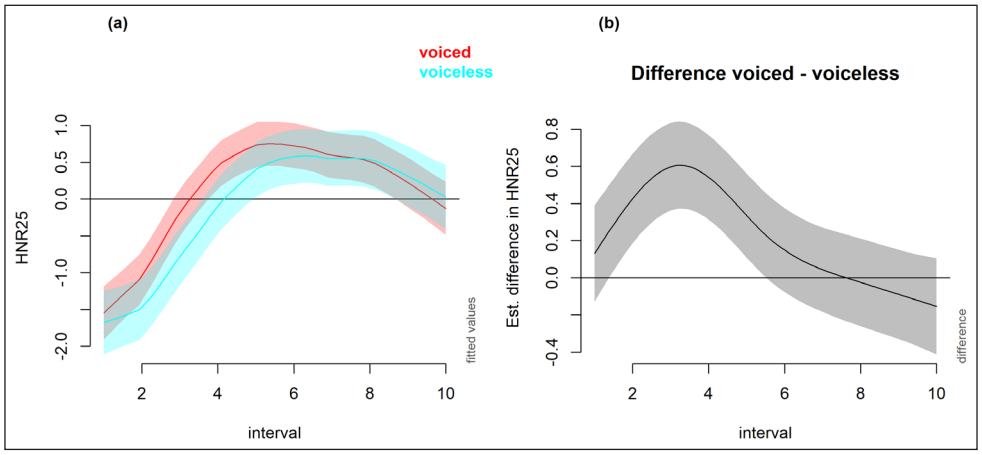

Figure 8: GAMM prediction smooths (a) and difference smooth (b) for HNR25. Portions where the difference smooth is above the zero line indicate higher values after voiced onsets.

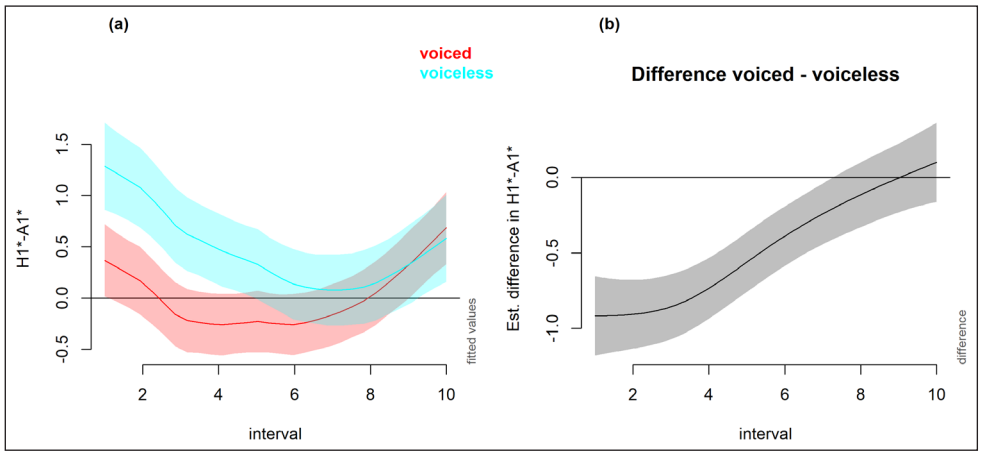

Figure 9: GAMM prediction smooths (a) and difference smooth (b) for H1*-A1*. Portions where the difference smooth is below the zero line indicate lower values after voiced onsets.

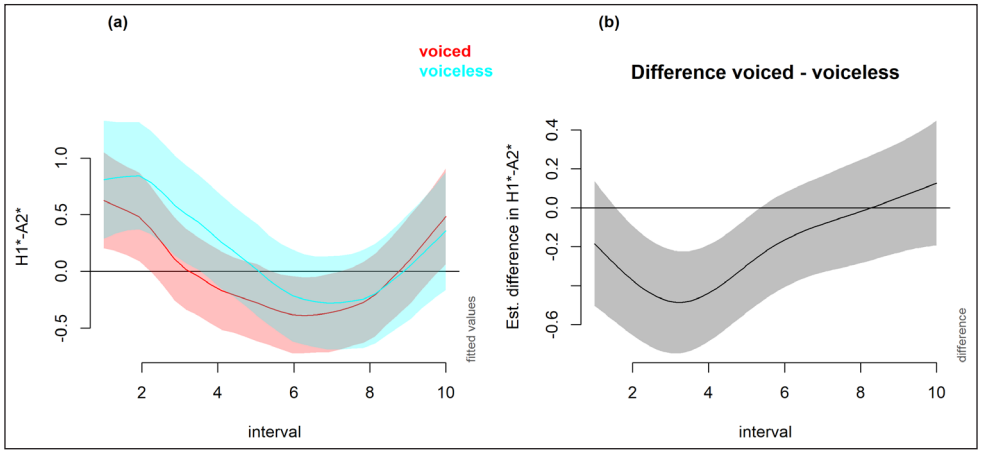

Figure 10: GAMM prediction smooths (a) and difference smooth (b) for H1*-A2*. Portions where the difference smooth is below the zero line indicate lower values after voiced onsets.

2.7 Discussion

The most important findings of our study are twofold. First, it is clear that the word-initial stop voicing contrast in Polish has acoustic effects on phonation in the following vowel, at least in the unfocused phrase-medial context that we examined. Thus, phonation may be claimed to be an additional acoustic correlate of the phonological categories traditionally described as voiceless and voiced in this language. That the observed acoustic effects are due to the onset consonant can be seen in the raw visualizations (**Figures 4–6**), GAMM prediction and difference smooths (**Figures 8–10**), as well as in Component 1 of the PCA carried out over the first third of vowel duration. Second, there is evidence that voiceless onsets in Polish produce breathier phonation, as reflected in higher spectral tilt and lower periodicity values than in voiced onsets. This breathiness effect is concentrated in the lower harmonics of the acoustic spectrum. The low-frequency concentration of the effect is reflected in the fact that the H2K*-H5K spectral tilt measure, as well as H4*-H2K*, both incorporating higher frequency elements, appear to behave differently from other spectral tilt measures. Finally, we must clarify that what we call a "breathiness" effect is observed within the confines of modal phonation. In other words, for most measures, voiceless onsets in Polish induce modal phonation that is acoustically closer to breathy voice than what is observed after voiced onsets.[5]

For all of the periodicity measures, both in **Figure 6** (CPP and HNR05) and the remaining ones in the supplementary materials, the trajectories over the course of the vowel suggest a voicing-independent pattern of increasing periodicity and decreasing spectral tilt in the first half of the vowel, effects that reverse direction in the second half of the vowel. Thus, it appears that

[5] In principle, we could consider using the term *slack* to describe this phonation. We refrain from doing this, however, to avoid confusion with one of the phonological accounts of Polish voicing considered in Section 3.

consonants on either side of the vowel diminish the overall acoustic robustness of phonation. This acoustic robustness reaches its peak around vowel midpoint. Such a pattern may be expected if we consider the aerodynamic effects of consonant and vowel articulation on phonation. One question that remains is whether the voicing-induced divergences of the trajectories in the early part of the vowel are a function of voiced consonants enhancing the slope and magnitude of the contours, voiceless consonants diminishing them, or both.

A possible answer to this question can be seen in the visualization of HNR05 in **Figure 6**, as well as the visualizations and smooths for the other HNR parameters in the supplementary materials. After voiceless consonants, the increase in periodicity associated with the vowel is delayed until the second or third interval, while in the case of voiced consonants no such delay is present. This aspect of the trajectories is compatible with an interpretation that voiceless onsets have an active effect on phonation in Polish, but voiced onsets do not. At the same time, it must be noted that in the CPP trajectory (**Figure 6**), a delay in the increase in periodicity after voiceless onsets is not observed, while it is for HNR. This discrepancy may be due to the fact that HNR measures are calculated in lower bands of the acoustic spectrum up to a maximum of 3500 Hz, while CPP is computed over the entire spectrum. Since the breathiness effect appears to be concentrated lower in the spectrum, the differences between the CPP trajectory and those of the HNR measures are not unexpected. Further support for the interpretation that voiceless onsets have an active effect on the following vowel in Polish can be gleaned from an examination of trajectories from Cf0 and F1 in **Figure 3** and **Figure 4**, respectively.

In sum, we believe that our data are compatible with a claim that voiceless onsets show active phonetic effects on the following vowel in Polish, while voiced onsets do not. Such an interpretation would be compatible with documented Cf0 effects for a number of other voicing languages, including French and Italian (Kirby & Ladd, 2016), Dutch (Pinget & Quené, 2023), and Tokyo Japanese (Gao & Arai, 2019). We would characterize this effect as a breathier version of modal voice, with breathiness concentrated in the lower portion of the spectrum. If this interpretation is correct, it would appear to leave VOT as the sole phonetic property indicative of an active [voice] feature in Polish. To investigate the implications this finding, and more generally how our data may be interpreted phonologically, in what follows we shall take a closer look at two recent phonological proposals (Raimy, 2021; Schwartz, 2022) concerning the laryngeal contrast in Polish.

3. General Discussion – implications for phonological representation

One of the main goals of our study is to provide data bearing on the phonological representation of the voicing contrast in Polish. The phonological literature on two-series laryngeal systems presents a rather complicated picture, in which we may encounter competing claims as to which feature specifications to employ (Iverson & Salmons, 1995; Avery & Idsardi, 2001), whether

feature specifications are binary or privative (Rubach, 1996; Wetzels & Mascaró, 2001; Bennett & Rose, 2017), whether one member of the laryngeal contrast is unmarked for laryngeal specification (e.g., Honeybone, 2005), and whether we need multiple levels of representation to explain a given system's behavior (e.g., Raimy, 2021). These questions focus attention on which (if any) phonetic aspects of laryngeal realization should be encoded in phonological systems, with wider theoretical implications for the relationship between phonetics and phonology.

Debate about these issues has centered on whether voicing languages and aspiration languages show equivalent phonological representations of the laryngeal contrast. In practice, this question has boiled down to whether to treat the VOT typology as a function of phonological specification. One influential early proposal is that of Keating (1984), who posits binary [voice] in phonological representation for both system types and an additional level of phonetic representation to capture the three basic VOT categories. Later work (e.g., Harris, 1994; Iverson & Salmons, 1995; Honeybone, 2005) adopted the feature [spread glottis] (or element |H|) to represent aspiration and long VOT, [voice] for negative VOT and characterize stops with short-lag VOT as unmarked. These proposals, which, as mentioned in Section 1, are often referred to under the heading of Laryngeal Realism (LR; see Avery & Idsardi, 2001; Beckman et al., 2013; Raimy, 2021), assume that voicing and aspiration systems are phonologically distinct. In addition, the representational differences posited by LR are said to be reflected in distinct patterns of assimilation as a function of the laryngeal system (e.g., regressive voicing is apparently not attested in languages without [voice]). Finally, LR has come to adopt privative feature specifications, in which plain voiceless stops with short-lag VOT are assumed to be phonologically unspecified, regardless of the symbol used to transcribe them. For LR, the typological frequency of plain unaspirated voiceless stops (see e.g., Maddieson, 1984) is attributable to their supposed unmarked status phonologically, which is said to reflect phonetic naturalness and a lack of laryngeal adjustment (e.g., Westbury & Keating, 1986). Under this view, the term "unmarked" has at least three distinct meanings, including "phonetically natural," "typologically frequent," and "phonologically unspecified" (see Hume, 2011, for discussion of the many uses of the term "markedness" in phonological theory).

At first glance, the insights of LR would seem to be a welcome development for our understanding of laryngeal phonology and its relationship with phonetics. In short, a binary approach must stipulate VOT and assimilation patterns, while LR appears to explain them. However, one of LR's fundamental claims, that short-lag stops in voicing languages are unspecified for laryngeal features, has proven controversial (e.g., Rubach, 1996; Wetzels & Mascaró, 2001; Bennett & Rose, 2017). In Polish, there is phonological evidence that voicelessness may spread both regressively and progressively to neighboring segments. Rubach (1996) attributes these facts to a binary [voice] specification in which voiceless obstruents are phonologically active (see also Cyran [2014, 2017], who posits a privative specification for voicelessness in two dialect regions of Polish). In this connection, two recent proposals, described in Raimy (2021) and

Schwartz (2022), respectively, seek to reconcile the insights of LR in accounting for VOT patterns with the evidence for active voicelessness in Polish. In what follows, we will compare those two proposals and their compatibility with the data presented in this paper.

Raimy (2021) analyzes Polish laryngeal phonology from the perspective of the model developed by Avery and Idsardi (2001). That system may be considered a representative of LR in that it posits that voicing and aspiration systems are phonologically distinct. It is also realist in that it employs features whose labels imply phonetically explicit descriptions of laryngeal activity. In the Avery and Idsardi model, phonological features are organized into hierarchical *dimensions*, which house antagonistic feature pairings (e.g., [spread] vs. [constricted]). Voicing languages are phonologically specified with the *dimension* of Glottal Tension (GT), which is then *completed* with a feature [slack]. *Completion* in Polish yields negative VOT in stops. By contrast, aspiration systems are phonologically specified on the *dimension* of Glottal Width, *completed* by means of a feature [spread]. For Avery and Idsardi, *dimensions* are the domain of truly phonological features, which may be underspecified, while *completion* yields an additional level of representation, referred to by Raimy (p. 86) as the *phonology-phonetics* level. The Avery and Idsardi model also features an additional mechanism, *enhancement*, to define relations between *phonology-phonetics* and an additional level of *categorical phonetics*. For Raimy, Polish voiceless stops are *enhanced* with the feature [spread]. Notably, this approach appears to explain both the complex patterns of assimilation in the language (see Rubach, 1996; Cyran, 2014), as well as the fact that Polish voiceless stops may exhibit what Raimy (p. 86), citing data from Keating (1984), refers to as "mid-lag" VOT.

A summary of Raimy's (2021) proposal for Polish compared with English is given in **Table 4**, which shows derivational steps from phonology to surface phonetics. Note that Raimy describes English /b d g/ as "voiceless unaspirated" (rather than "lenis"). Additionally, the "Larynx-gt" and "Larynx-gw" markers indicate phonological underspecification, indicated by a use of lower-case letters in "gt" and "gw," which may be altered at later stages of the derivation.

Schwartz (2022) analyses Polish laryngeal phonology (and its interaction with English in the context of bilingual and L2 speech) from the perspective of the Onset Prominence representational environment (OP; e.g., Schwartz, 2016). A key aspect of OP is that manner of articulation is a structural property, such that stop consonants are represented as the top three layers of a four-layer hierarchical tree structure. The layers of the OP hierarchy are derived from the closure, release, CV transition, and vowel steady-state phases (*Closure*, *Noise*, *Vocalic Onset*, and *Vocalic Target* nodes within OP representations) inherent in the production of stop-vowel sequence. Crucially, a complete stop-vowel CV unit, rather than skeletal positions for individual segments, represents the fundamental building block of the representations, which unite segmental phonology and prosodic constituents into a single ontological whole, rather than connecting them by means of association lines. OP representations for two-series laryngeal systems are shown in **Figure 11**.

	Polish voiced	**Polish voiceless**	**English voiceless unaspirated (lenis)**	**English aspirated**
Phonology dimensions	Glottal Tension (GT)	Larynx-gt	Larynx-gw	Glottal Width (GW)
Phonology-phonetics completion	[slack]	Larynx-gt	Larynx-gw	[spread]
Categorical phonetics enhancement	[slack]	[spread]	Larynx-gw	[spread]
Surface phonetics	Pre-voiced	Mid-lag VOT	Short-lag VOT	Long-lag VOT

Table 4: Polish and English word-initial laryngeal contrasts (after Raimy, 2021, p. 86).

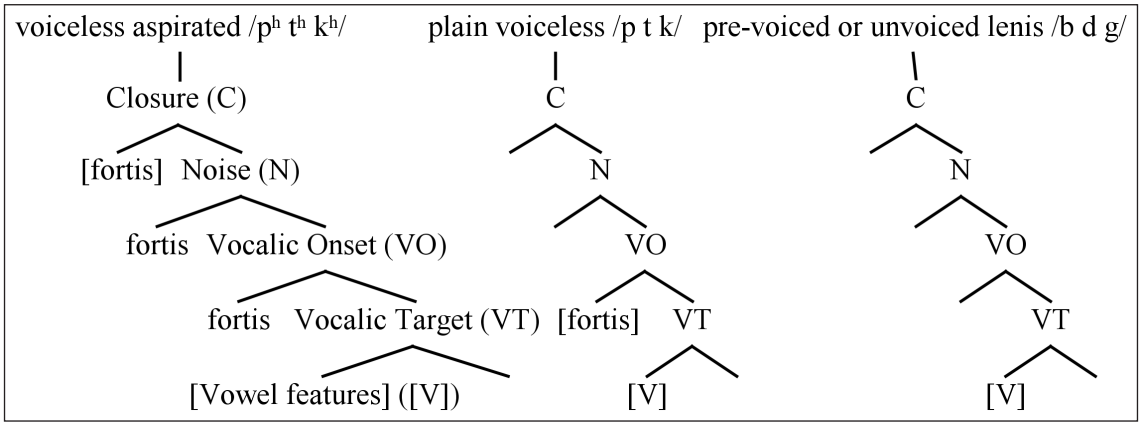

Figure 11: Onset Prominence representations for two-series laryngeal systems (after Schwartz, 2022). Left-hand tree: voiceless aspirated stops. Center tree: plain voiceless stops. Right-hand tree: voiced/lenis stops in both aspiration and voicing systems.

According to the OP approach (Schwartz, 2022, p. 111), the difference between aspiration languages and voicing languages is based on the hierarchical level of assignment of a single laryngeal feature, labeled [fortis]. When [fortis] is assigned at the *Closure* level (C), the feature specification occupies the top three nodes, including the Noise node (N), and VOT effects are expected. When it is assigned to VO, the Noise node lacks laryngeal specification, and short-lag VOT is expected. Notably, in both aspiration and voicing systems, the *VO* node of the structure serves as a docking point for the [fortis] feature,[6] thus predicting parallel effects of voicelessness

[6] The assignment of a feature at a higher level is associated with a "trickling" (Schwartz, 2016, p. 45) mechanism, by which the specification also occupies lower-level nodes. This is why we see the [fortis] feature both with (assigned) and without (trickled) brackets in **Figure 4**.

on the phonetics of the following vowel. At the same time, no predictions are made as to the appearance of pre-voicing: Pre-voiced and unvoiced /b d g/ have equivalent representations. In other words, unlike in LR, pre-voicing in OP need not reflect the presence of an underlying phonological feature.

Returning to the data presented in this paper, it may be said that the results are broadly supportive of both of these theoretical approaches. Both proposals predict active phonetic effects resulting from consonant voicelessness, thus challenging the claim, made explicit in some versions of LR (e.g., Beckman et al., 2013), that voiceless stops are phonologically inert. We found effects of voicelessness across a range of voice quality measures. These effects were indicative of breathier (yet still modal) phonation. For Raimy (2021), the breathiness effect would presumably be due to the *enhancement* mechanism by which a feature [spread] is added at the *categorical phonetic* level of representation. At the same time, since it is fundamentally a skeleton-based model, Raimy's proposal makes no explicit predictions as to the time course of the *enhancement* effects that we observed in our data. Presumably, for Raimy, the timing of the acoustic correlates of [spread] at vowel onset would be an inherent aspect of the implementation of *enhancement*. For Schwartz, the localization of the effects of voiceless consonants falls out directly from the [fortis] specification on the VO node of structure, which aligns with the CV transition portion of the signal in stop-vowel sequences. At the same time, Schwartz's proposal refrains from stating just what those [fortis]-like effects might be. In essence, [fortis] is a purely abstract feature specification whose phonetic realization is not explicitly defined. In sum, Raimy is more explicit about *what* the phonetic effects of voicelessness should be (breathier phonation resulting from [spread]), while Schwartz is more explicit about *where* those effects may be found (at vowel onset), but both predict that effects should be present.

Nevertheless, the two proposals diverge in a number of ways. The first difference is in their treatment of the voiced series of stops. In Raimy's proposal (2021), voiced stops in Polish are specified for a phonological feature [slack] on the Glottal Tension *dimension*, which is responsible for the presence of negative VOT. In the OP proposal, the voiced series is unspecified for any laryngeal feature, so pre-voicing and regressive voicing assimilation are not the result of a phonological specification—they are due to purely phonetic effects (see Cyran, 2014, for similar postulates for southern and western varieties of Polish).[7] The data described in this paper do not reveal strong evidence for a [slack] or [voice] feature in Polish beyond VOT, whose perceptual

[7] The OP account of laryngeal neutralization in Polish, including both final devoicing (see Schwartz et al., 2021) and regressive (de)voicing, derives from the OP postulate that the locus of laryngeal contrast in voicing languages such as Polish is the VO node. In non-prevocalic positions, this node may be lost and neutralization may occur. In the case of regressive voicing in clusters, voicing is then attributable to voicing *bleed* (Davidson, 2016) that persists after the preceding vowel. Thus, phonetically, voicing in Polish clusters is not regressive (see Schwartz, 2019), while phonologically the process does not imply spreading, but rather the production of a neutralized "allophone" lacking the VO node. By contrast, traditional spreading accounts require us to assume that a phonological process moves in the opposite direction (right to left) to the observable phonetic process (left to right).

weight in voicing languages appears to be somewhat less robust than in aspiration languages (see Keating et al., 1981; Schwartz, 2022). That is, voiced onsets seem to have only negligible, if any, effects on the acoustics of the following vowel. To explain the lack of effect of voiced consonants on the following vowel, Raimy could perhaps suggest that the [slack] specification on voiced stops does not need *enhancement*, since it already bears phonological status. This suggestion certainly has a plausible phonetic explanation—gestures responsible for pre-voicing may be timed earlier so as to align with stop closure rather than the following vowel. At the same time, such an interpretation would entail an implicit claim that phonetic *enhancement* targets only phonologically unspecified segments, but is unnecessary for marked objects. It is unclear whether this claim would stand up to further empirical testing.[8]

There is also reason to question the conventional wisdom that [voice] (or [slack]) is phonologically active in Polish, despite the presence of pre-voicing and regressive voicing assimilation in the language. Beyond the lesser perceptual weight of VOT mentioned earlier, the status of [voice] may be challenged by findings from L2 and bilingual speech research, in which perceptual *equivalence classification* (Flege, 1987) gives rise to cross-language phonetic interaction. In a number of studies of interactions between voicing and aspiration systems (for an overview, see Schwartz, 2022), we observe more robust interaction in the voiced series than in the voiceless series. Such findings suggest that Polish-style pre-voiced stops and English-style unvoiced lenis stops are perceptually more similar to each other, with a common phonological representation (unspecified for [voice]/[slack]), than short-lag and long-lag fortis stops. Bilinguals and L2 learners are more likely to keep /p t k/ distinct across languages than /b d g/, suggesting distinct representations for the former, but phonological equivalence for the latter. In our data, the presence of pre-voicing was the only phonetic cue suggesting active [voice]—nothing was found on the vowel-based cues. This fact in itself, of course, does not rule out the feature's phonological status. However, when considered in light of data from both VOT perception and L2/bilingual speech, our findings are easier to interpret if we assume that voiced stops in Polish are phonologically unspecified.

Independent of the status of [voice]/[slack], a more fundamental difference between the Raimy and Schwartz proposals is that the former posits several levels of representation, while representations in the latter are said to map directly to speech (see Schwartz, 2016, p. 40). Perhaps somewhat ironically, this direct mapping does not imply phonetically precise feature specifications. As pointed out earlier, OP is explicit about [fortis] effects appearing on vowel onset, but says nothing about what those effects might be—the fact that they involve f0 raising and breathier phonation, rather than f0 lowering and stiffer phonation, is essentially arbitrary, even if a phonetic explanation for them is readily available. Raimy's use of [slack] and [spread]

[8] For instance, Raimy's approach would presumably have us expect enhancement of the schwa vowel, which is often assumed to be phonologically unspecified.

feature specifications, by contrast, makes explicit claims about laryngeal activity. The question that remains is whether such phonetically precise specifications, and increasingly detailed levels of representation are, in fact, desirable.

In this connection, consider Ladd (2011), who argues against a level of categorical phonetic representation assumed in many mainstream approaches. In his discussion, Ladd discusses Cho and Ladefoged's (1999) VOT data from 18 languages. In Figure 9 (p. 223) of that paper, Cho and Ladefoged group their VOT results into four categories using boxes above the relevant bars. Ladd (2011) questions whether those boxes indeed capture distinct VOT categories that would be identifiable if the corpus included a larger number of languages and/or tokens. Likewise, Raimy's (2021) proposal apparently assumes a categorical distinction between short-lag VOT (without *enhancement*) in some true-voice languages and mid-lag VOT in Polish (with *enhancement*). Based on what we know about Polish VOT (see references cited in Section 1), by which, in addition to mid-lag stops, larger data sets also reveal plenty of truly short-lag stops, Raimy would presumably have to claim that the language has both non-[spread]-*enhanced* (short-lag) and [spread]-*enhanced* (mid-lag) stops, begging an explanation of the factors that determine when [spread]-*enhancement* occurs and when it does not. A more plausible characterization of Polish voiceless stops, it seems to us, is that they are indeed short-lag, but with a VOT distribution that is skewed towards longer values than is found for unvoiced lenis stops. The OP approach is compatible with such a characterization: A [fortis] specification on the VO node with a bare Noise node (see **Figure 6**), encodes their short-lag status and also explains the results of the vowel-based cues presented in this paper, while the skew towards longer VOT values is a product of statistical noise inherent to phonetic data.

From a wider perspective on two-series systems, including their potential diachronic development into tone and register contrasts, the theoretical difficulties associated with phonetically precise phonological features comes into sharper focus. Brunelle and Kirby (2016, p. 193) observe that register systems, which apparently developed historically from contrasts in voicing, are based on a set of acoustic properties that typically co-occur. Low register, deriving from voiced stops, typically involves lower f0, breathy phonation, lowered F1, and longer VOT. Meanwhile, High register is associated with higher f0, stiffer phonation, raised F1 and shorter VOT.

Considering possible links between register and standard Polish-style voicing systems from which they may have evolved, we can see a number of contradictions. **Table 5** provides a selection of Polish laryngeal correlates. An additional column shows how the realization of Polish voiceless stops would correspond to the related cue in register systems. The f0 and F1 effects suggest a link between Polish voiceless stops and High register, while the phonation effects show a link between Polish voiceless stops and Low register. Meanwhile, it is not entirely clear how to characterize the VOT link. Brunelle and Kirby (2016) associated plain stops and short-lag VOT with High register, but do not include mid-lag VOT.

Phonetic property	Polish realization of /p t k/	Register categories associated with cue (after Brunelle & Kirby, 2016)
F0 on following vowel	Higher	High
Phonation on following vowel	Breathier	Low
F1 on following vowel	Raised	High
VOT	Short-to-mid lag	High or unclear

Table 5: Selected laryngeal cues for Polish voiceless stops, and links to register systems.

The fact that the phonetic properties associated with the Polish contrast show mismatches with those of register contrasts argues against positing a large amount of phonetic detail in the phonology. The reason for this is that if phonology includes increasing amounts of phonetic detail in successive levels of representation, it is not entirely clear how phonetic divergences between laryngeal contrasts might develop diachronically. In other words, if a phonological system indeed includes mechanisms such as completion and enhancement, encoded with feature specifications based on explicitly defined laryngeal activity, we might expect less dramatic cross-language mismatches of the type shown in **Table 5**, since the grammar would give explicit phonetic instructions to speakers, diminishing the potential for diachronic instability.[9] By contrast, purely abstract laryngeal specifications such as [fortis] leave the analyst with a clearer narrative for describing mismatches and sound change. Changes leading to mismatches, such as those shown in **Table 5**, may develop due to perceptual reinterpretation (Ohala, 1981; Blevins, 2004) of a rather ambiguous, but quite numerous, set of acoustic correlates.

4. Final remarks

This article has presented data on how the Polish laryngeal contrast may affect the phonetic realization of vowels following voiced and voiceless stops. Over a range of acoustic measures of voice quality, as well as f0 and vowel height (F1), active effects of consonant voicelessness were observed. In particular, voiceless obstruents appear to induce modal phonation on the following vowel that is acoustically closer to breathy voice than what is observed after voiced consonants. Similarly, effects of voicelessness on f0 and F1 (both higher after voiceless consonants), compatible with previous findings in the literature, were also observed. Meanwhile, voiced onsets appeared to have negligible, if any, effects on the phonation, f0, and F1 of the following vowel.

[9] More recent work on the development of register (e.g., Brunelle et al., 2020, 2022) in Southeast Asian languages suggests a variety of avenues by which registrogenesis can take place, so the point in the main text about diminished potential for diachronic change should not be taken as suggestion that the LR approach renders such change impossible.

The implications of these findings for the phonological representation of two-series voice contrasts were also considered. Our results present a challenge to theoretical perspectives that assume phonologically inactive voiceless obstruents in true-voice languages. Additionally, our results suggest that pre-voicing may be the only reliable correlate of a [voice] or [slack] feature in Polish. Since the locus of the effects we observed is at the onset of the vowel following stop release, we suggest that theories of phonological representation should have ways of encoding the time course of such effects. At the same time, we suggest that laryngeal representations need not be defined in terms of phonetically explicit phonological features.

Appendix. List of experimental items and the sentences they were contained in

Target word	Vowel	Sentence	English translation
bawić [bavit͡ɕ] -play/ to entertain' (inf.)	[a]	Na tym szkoleniu uczymy **bawić** publiczność dowcipami.	At this training we're learning how to entertain the audience with some jokes.
bochen [bɔxɛn] - 'a loaf' (nom. sg.)	[ɔ]	Od pokoleń wypiekamy **bochen** wiejskiego chleba na mące.	For generations, we've been baking a loaf of rustic bread.
befsztyk [bɛfʂtɨk] - 'a beefstake' (nom. sg.)	[ɛ]	Tradycyjnie zamawiamy **befsztyk** wieprzowy z ziemniakami.	As per usual we order a beefsteak with potatoes.
dawać [davat͡ɕ] - 'to give' (inf.)	[a]	My najbardziej uwielbiamy **dawać** książki z okazji urodzin.	We love giving people books for birthdays the most.
dowód [dɔvut] - 'proof' (nom. sg.)	[ɔ]	Na rozprawie przedstawimy **dowód** winy oskarżonego.	During the trail we'll present the proof that incriminates the perpetrator.
defekt [dɛfɛkt] - 'a glitch' (nom. sg.)	[ɛ]	Na przeglądzie usuniemy **defekt** przewodu paliwowego	During the inspection will fix the glitch in the fuel line.
gafy [gafɨ] - 'blunders' (nom. pl.)	[a]	Nieświadomie popełniamy **gafy** podczas spotkań towarzyskich.	Unbeknownst to us we make blunders during social occasions.
gofry [gɔfrɨ] - 'waffles' (nom. pl.)	[ɔ]	W naszym punkcie sprzedajemy **gofry** z polewą czekoladową.	Here we sell waffles with chocolate sauce.
geny [gɛnɨ] - 'a gene' (nom. pl.)	[ɛ]	W naszym ośrodku badamy **geny** ludzi chorych na raka.	In this center we study the genes of cancer patients.
Pawła [pavwa] - 'Paul' (gen. sg.)	[a]	Chyba najbardziej lubimy **Pawła** Sowę za jego dowcipy.	It's likely that our favorite thing about Paul Sowa is his sense of humor.

(Contd.)

powód [pɔvut] - 'a cause' (nom. sg.)	[ɔ]	W krótkim czasie ustalimy **powód** złego stanu gospodarki.	In a very short time we'll establish the cause of the poor state of the economy.
pewność [pɛvnɔɕt͡ɕ] - 'certainty' (nom. sg.)	[ɛ]	W przybliżeniu oceniamy **pewność** istnienia życia w kosmosie.	We're roughly assessing the certainty of extraterrestrial life.
taflę [taflɛw̃] 'a tile' (acc. sg.)	[a]	W przerwie meczu polerujemy **taflę** lodu całego lodowiska	During the break we polish the tile of the entire ice rink.
towar [tɔvar] - 'goods' (nom. sg.)	[ɔ]	Na bieżąco kupujemy **towar** z hurtowni po niskich cenach.	We buy cheap goods from wholesalers on regular basis.
teflon [tɛflɔn] - 'teflon' (nom. sg.)	[ɛ]	Przy produkcji stosujemy **teflon** jako powłokę patelni.	In the proces of production we use teflon as the pan's coating.
kawior [kaviɔr] - 'caviar' (nom. Sg.)	[a]	Podczas bankietów jadamy **kawior** z mintaja jako przystawkę.	At banquets we eat pollock fish caviar as a starter.
Kowno [kɔvnɔ] - 'Kaunas' (nom. Sg.)	[ɔ]	Według programu zwiedzamy **Kowno** zaraz po meczu w Wilnie.	According to the schedule we're sightseeing Kaunas right after the game in Vilnius.
kefir [kɛfir] - 'kefir' (nom. sg.)	[ɛ]	Każdego ranka pijemy **kefir** smakowy do śniadania.	Every day we drink flavoured kefir for breakfast.

Acknowledgements

This research was supported by a grant from the Polish National Science Centre (Narodowe Centrum Nauki), project number UMO-2021/41/B/HS2/00239. We are grateful to the editors of *Laboratory Phonology*, two anonymous *Laboratory Phonology* reviewers, James Kirby, Adam Olender, Joanna Maciąg, Zuzanna Cal, and participants at the 20th International Congress of Phonetic Sciences in Prague. Any remaining errors are our own responsibility.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

Geoffrey Schwartz: Funding acquisition, conceptualization, supervision, writing and editing

Ewelina Wojtkowiak: Speech data collection and annotation, writing and editing

Maral Asiaee: VoiceSauce analysis, writing and editing

Kamil Kaźmierski: Data visualization and analysis, writing and editing

References

Avery, P., & Idsardi, W. (2001). Laryngeal dimensions, completion and enhancement. In T. Hall (Ed.), *Distinctive Feature Theory* (pp. 41–70). De Gruyter Mouton. https://doi.org/10.1515/9783110886672.41

Baayen, H., van Rij, J., de Cat, C., & Wood, S. (2018). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. In D. Speelman, H. Heylen, & D. Geeraerts (Eds.), *Mixed-effects regression models in linguistics* (pp. 49–69). Springer.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics*, *39*(1), 39–49. https://doi.org/10.1016/j.wocn.2010.11.001

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, *49*(2), 259–284. https://doi.org/10.1017/S0022226712000424

Bennett, W., & Rose, S. (2017). Moro voicelessness dissimilation and binary [voice]. *Phonology*, *34*(3), 473–505. https://doi.org/10.1017/S0952675717000252

Blevins, J. (2004). *Evolutionary Phonology: The emergence of sound patterns*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486357

Boersma, P., & Weenink, D. (2025). *Praat: Doing phonetics by computer* (Version 6.4.37) [Computer software]. https://www.praat.org/

Brunelle, M., Brown, J., & Thị Thu Hà, P. (2022). Northern Raglai voicing and its relation to Southern Raglai register: Evidence for early stages of registrogenesis. *Phonetica*, *79*(2), 151–188. https://doi.org/10.1515/phon-2022-2019

Brunelle, M., & Kirby, J. (2016). Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass*, *10*(4), 191–207. https://doi.org/10.1111/lnc3.12182

Brunelle, M., Tấn, T., Kirby, J., & Giang, Đ. (2020). Transphonologization of voicing in Chru: Studies in production and perception. *Laboratory Phonology*, *11*(1), Article 15. https://doi.org/10.5334/labphon.278

Chang, C. (2010). *First language phonetic drift during second language acquisition* [Doctoral dissertation, University of California, Berkeley].

Chen, M. (1997). Acoustic correlates of English and French nasalized vowels. *Journal of the Acoustical Society of America*, *102*(4), 2360–2370. https://doi.org/10.1121/1.419620

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

Cho, T., Whalen, D., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, *71*, 52–65. https://doi.org/10.1016/j.wocn.2018.11.002

Cruttenden, A. (2001). *Gimson's pronunciation of English* (6th ed.). Arnold.

Cyran, E. (2014). *Between phonology and phonetics: Polish voicing*. De Gruyter Mouton. https://doi.org/10.1515/9781614515135

Cyran, E. (2017). 'Voice' languages with no [voice]? Some consequences of laryngeal relativism. *Acta Linguistica Academica*, *64*(4), 477–511. https://doi.org/10.1556/2062.2017.64.4.1

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

de Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research*, *36*(2), 254–266. https://doi.org/10.1044/jshr.3602.254

Draxler, C., & Jänsch, K. (2015). *SpeechRecorder* (Version 2.12.16) [Computer software]. http://www.bas.uni-muenchen.de/Bas/software/speechrecorder/

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). SAGE.

Gao, J., & Arai, T. (2019). Plosive (de-)voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery. *Journal of Phonetics*, *77*, Article 100932. https://doi.org/10.1016/j.wocn.2019.100932

Garellek, M. (2019). The phonetics of voice. In W. F. Katz & P. E. Assmann (Eds.), *The Routledge handbook of phonetics* (pp. 75–106). Routledge. https://doi.org/10.4324/9780429056253-5

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

Harris, J. (1994). *English sound structure*. Blackwell.

Hejná, M., Kaźmierski, K., & Guo, W. (2021). Even Americans pre-aspirate. *English World-Wide*, *42*(2), 200–226. https://doi.org/10.1075/eww.00065.hej

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, *37*(4), 769–778. https://doi.org/10.1044/jshr.3704.769

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. 317–352). De Gruyter Mouton. https://doi.org/10.1515/9783110890402.317

Hume, E. (2011). Markedness. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*. Wiley-Blackwell. https://doi.org/10.1002/9781444335262.wbctp0004

Iseli, M., & Alwan, A. (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. In *Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 669–672). IEEE. https://doi.org/10.1109/ICASSP.2004.1326074

Iverson, G., & Salmons, J. (1995). Aspiration and laryngeal representations in Germanic. *Phonology*, *12*(3), 369–396. https://doi.org/10.1017/s0952675700002566

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*(3), 313–334. https://doi.org/10.1017/S0025100309990156

Johnson, K., & Babel, M. (2023). The structure of acoustic voice variation in bilingual speech. *Journal of the Acoustical Society of America*, *153*(5), 3221. https://doi.org/10.1121/10.0019659

Kawahara, H., de Cheveigné, A., & Patterson, R. D. (1998). An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: Revised TEMPO in the STRAIGHT suite. In *Proceedings of the 5th International Conference on Spoken Language Processing* (pp. 1–4). https://doi.org/10.21437/icslp.1998-555

Keating, P. (1980). *A phonetic study of a voicing contrast in Polish* [Doctoral dissertation, Brown University].

Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286–319. https://doi.org/10.2307/413642

Keating, P., Kuang, J., Garellek, M., & Esposito, C. (2023). A cross-language acoustic space for vocalic phonation distinctions. *Language*, *99*(2), 351–389. https://doi.org/10.1353/lan.2023.a900090

Keating, P., Mikos, M. J., & Ganong, W. F. (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*(5), 1261–1271. https://doi.org/10.1121/1.387139

Kingston, J., & Diehl, R. (1994). Phonetic knowledge. *Language*, *70*(3), 419–454. https://doi.org/10.1353/lan.1994.0023

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*(4), 2400–2411. https://doi.org/10.1121/1.4962445

Kirby, J., & Ladd, D. R. (2018). Effects of obstruent voicing on vowel F0: Implications for laryngeal realism. *Yearbook of the Poznań Linguistic Meeting*, *4*, 213–235. https://doi.org/10.2478/yplm-2018-0009

Kong, E. J., Beckman, M., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics*, *40*(6), 725–744. https://doi.org/10.1016/j.wocn.2012.07.002

Kreiman, J., Gerratt, B. R., Garellek, M., Samlan, R., & Zhang, Z. (2014). Toward a unified theory of voice production and perception. *Loquens*, *1*(1), Article e009. https://doi.org/10.3989/loquens.2014.009

Ladd, D. R. (2011). Phonetics in phonology. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The handbook of phonological theory* (2nd ed., pp. 348–373). Blackwell.

Ladd, D. R., & Schmid, S. (2018). Obstruent voicing effects on f0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops. *Journal of Phonetics*, *71*, 229–248. https://doi.org/10.1016/j.wocn.2018.09.003

Lenth, R. (2025). *emmeans: Estimated marginal means, aka least-squares means* (R package version 1.11.2-8) [Computer software]. https://CRAN.R-project.org/package = emmeans

Lisker, L. (1975). Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America*, *57*(6), 1547–1551. https://doi.org/10.1121/1.380602

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Machač, P., & Skarnitzl, R. (2009). *Principles of phonetic segmentation*. Epocha.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.

Murty, K. S. R., & Yegnanarayana, B. (2008). Epoch extraction from speech signals. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(8), 1602–1613. https://doi.org/10.1109/TASL.2008.2004526

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178–203). Chicago Linguistic Society.

Penney, J., Cox, F., Miles, K., & Palethorpe, S. (2018). Glottalisation as a cue to coda consonant voicing in Australian English. *Journal of Phonetics*, *66*, 161–184. https://doi.org/10.1016/j.wocn.2017.10.001

Pinget, A. F., & Quené, H. (2023). Effects of obstruent voicing on vowel fundamental frequency in Dutch. *Journal of the Acoustical Society of America*, *154*(4), 2124–2136. https://doi.org/10.1121/10.0021070

R Core Team. (2025). *R: A language and environment for statistical computing* [Computer program]. R Foundation for Statistical Computing. http://www.R-project.org

Raimy, E. (2021). Privativity and ternary phonological behavior. In S. Bendjaballah, A. Tifrit, & L. Voeltzel (Eds.), *Perspectives on element theory* (pp. 65–110). De Gruyter. https://doi.org/10.1515/9783110691948-004

Rubach, J. (1996). Nonsyllabic analysis of voice assimilation in Polish. *Linguistic Inquiry*, *27*(1), 69–110. http://www.jstor.org/stable/4178926

Sanker, C. (2019). Effects of coda voicing on vowel phonation. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 3323–3327). Australasian Speech Science and Technology Association.

Schwartz, G. (2016). On the evolution of prosodic boundaries: Parameter settings for Polish and English. *Lingua*, *171*, 37–73. https://doi.org/10.1016/j.lingua.2015.11.005

Schwartz, G. (2019). *There is no [voice] in Polish.* Adam Mickiewicz University in Poznań. https://www.researchgate.net/publication/333193352_There_is_no_voice_in_Polish

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

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., Asiaee, M., & Wojtkowiak, E. (2023). Fortis stops in Polish: Evidence from acoustic measures of voice quality. In *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 2259–2263). International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2023/full_papers/434.pdf

Schwartz, G., Kaźmierski, K., & Wojtkowiak, E. (2021). Perspectives on final laryngeal neutralization: New evidence from Polish. *Phonology*, *38*(4), 693–717. https://doi.org/10.1017/S0952675721000373

Schwartz, G., Wojtkowiak, E., & Brzoza, B. (2019). Beyond VOT in the Polish laryngeal contrast. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2383–2387). Australasian Speech Science and Technology Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_2432.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

Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011). VOICESAUCE: A program for voice analysis. In *Proceedings of the 17thth International Congress of Phonetic Sciences* (pp. 1846–1849). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Shue/Shue.pdf

Sjölander, K. (2004). *The Snack sound toolkit* [Computer software]. KTH Stockholm.

Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction.* arXiv. https://arxiv.org/pdf/1703.05339

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, Article 101017. https://doi.org/10.1016/j.wocn.2020.101017

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

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced–voiceless distinction for stops. *Journal of the Acoustical Society of America*, *55*(3), 653–659. https://doi.org/10.1121/1.1914578

Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *Journal of the Acoustical Society of America*, *142*(4), 2469–2482. https://doi.org/10.1121/1.5008854

Sun, X. (2002). Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (Vol. 1, pp. 333–336). IEEE. https://doi.org/10.1109/icassp.2002.5743722

Sypiańska, J. (2013). *Quality and quantity of language use and attrition of L1 Polish due to L2 Danish and L3 English* [Doctoral dissertation, Adam Mickiewicz University in Poznań].

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Ting, C., Sonderegger, M., Clayards, M., & McAuliffe, M. (2025). The crosslinguistic distribution of vowel and consonant intrinsic F0 effects. *Language*, *101*(1). https://doi.org/10.1353/lan.2025.a955269

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2025). *itsadug: Interpreting time series and autocorrelated data using GAMMs* (R package version 2.5) [Computer software]. CRAN. r-project.org

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). Springer. https://doi.org/10.1007/978-3-642-20083-0_1

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, *22*(1), 145–166. https://doi.org/10.1017/S0022226700010598

Wetzels, W. L., & Mascaró, J. (2001). The typology of voicing and devoicing. *Language*, *77*(2), 207–244. https://doi.org/10.1353/lan.2001.0123

Wojtkowiak, E. (2022). The effects of intensive phonetic training on the acquisition of English stops. In V. G. Sardegna & A. Jarosz (Eds.), *Theoretical and practical developments in English speech assessment, research, and training* (pp. 261–277). Springer. https://doi.org/10.1007/978-3-030-98218-8_15

Wojtkowiak, E., & Schwartz, G. (2022). The acoustics of plosives and the formation of prosodic structure in Polish. *Folia Linguistica*, *56*(2), 273–324. https://doi.org/10.1515/flin-2022-2020

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wrembel, M. (2015). *In search of a new perspective: Cross-linguistic influence in the acquisition of third language phonology*. Wydawnictwo Naukowe UAM. http://hdl.handle.net/10593/14647