The phonology of vowel VISC-osity – acoustic evidence and representational implications

GEOFFREY SCHWARTZ (ID)

Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

While vowel inherent spectral change (VISC) is becoming increasingly prominent in phonetic descriptions of varieties of English, there is little data on whether and how formant dynamics may shape the vowel systems of other languages. This paper provides cross-language acoustic comparisons of the degree of VISC in Polish and British English. In one study, British English showed a greater concentration of F1 movement earlier in the time course of vowels than Polish. In a second study, proficient Polish speakers of English showed a greater concentration of F1 movement earlier in the time course of vowels while speaking their L2 English than while speaking their L1 Polish. These findings are in line with a proposal formulated in the Onset Prominence (OP) framework, by which the relative degree of formant dynamics in the two languages is attributable to consonantal or vocalic affiliation of the Vocalic Onset (VO) node of structure. Consonantal VO affiliation in English contributes to a greater degree of VISC concentrated earlier in vowel duration. Further empirical patterns associated with the OP representational settings in the two languages are also discussed, along with implications for other languages.

CORRESPONDING AUTHOR:
Geoffrey Schwartz

Faculty of English, Adam Mickiewicz University in Poznań, Ul. Grunwaldzka 6, 60-780 Poznań, Poland

geoff@amu.edu.pl

1 INTRODUCTION

Vowel inherent spectral change (VISC; see e.g. Nearey & Assmann 1986; Morrison & Assmann 2013) has become an increasingly prominent area of phonetic investigation. Data on vowel-internal formant movement has been shown to have significant implications for speech perception (e.g. Strange 1989), socio-phonetics and language change (e.g. Fox & Jacewicz 2009), as well as second language acquisition (e.g. Rogers et al. 2013; Schwartz & Kaźmierski 2020). At the same time, however, the study of VISC has yet to break into the mainstream of linguistics. On the basis of textbook descriptions, one might conclude that it is sufficient to characterize vowels in terms of a two-dimensional chart populated by a finite set of phonetic symbols, occasionally supplemented by additional features such as length or nasalization. Nevertheless, on many occasions it has been shown that the dynamic properties of vowels, as much as their location on a two dimensional chart, play an important role in the behaviour of vowel systems. Vowels that are in close proximity in static F1–F2 space may be distinguished by the direction and magnitude of their formant movement (Hillenbrand 2013; Chládková et al. 2016), while perceptual experiments have shown that listeners attend to this type of information in vowel identification (Strange et al 1983; Jenkins & Strange 1999).

While VISC is becoming an increasingly prominent element of experimental descriptions of varieties of English (Fox & Jacewicz 2009; Williams & Escudero 2014; Elvin et al. 2016), data from other languages, as well as systematic cross-language comparisons, are rare (but see Williams et al. 2015). It is therefore difficult to say anything about the degree to which VISC is a language-specific or universal phenomenon. In many varieties of English, there is strong evidence that formant movement is an integral aspect of the vowel system, even for nominal monophthongs. This evidence can be gleaned from both experimental phonetic studies of both production (Williams & Escudero 2014; Williams et al. 2015) and perception (Strange et al. 1983; Jenkins & Strange 1999; Chládková et al. 2016), as well as textbooks of English pronunciation, which note that many of the so-called monophthongs in the language are in fact diphthongized to a significant degree (e.g. Cruttenden 2001; Collins & Mees 2009). From the point of view of phonology, this work implies that the perceptual identity of a given English vowel, its phonological representation, if you will, is intimately connected with its dynamic formant patterns, at least in the dialects described in those studies.

With regard to other languages, one may encounter impressionistic comments referring to vowels that are pure in quality, but it is difficult to find descriptions of acoustic data bearing on this issue. Existing studies suggest that the degree of formant movement in Dutch (Williams et al. 2015) and German (Strange & Bohn 1998) is of systematically lesser magnitude than in Southern British English and American English, respectively. With regard to perception, Schwartz et al. (2016) found that Polish learners of English do not show 'dynamic specification' effects (Strange 1989) in L1 Polish vowel identification, but they do in L2 English. Apparently, vowel perception in Polish is weighted more heavily toward static formant targets than it is in English. The present paper contributes to the relatively sparse literature on cross-language differences in vowel formant dynamics, presenting two acoustic studies comparing CVC monosyllables in Polish and British English (Section 3). The first study compares productions from a corpus of Southern British English with recordings of Polish made for the purposes of this study. The second study compares productions by proficient Polish users of English in their L1 and their L2.

To the extent that the English productions differ systemically from the Polish items across both studies, it raises a question about whether the cross-language differences constitute an inherent aspect of the two phonological systems. In other words, are the language-specific differences in VISC phonological in nature, or are they simply phonetic? The role of sub-segmental phonetic details in the structure of phonological grammars has been a point of disagreement in phonological theory. In the perspective assumed here, phonological representations and behaviour are determined largely on the basis of phonetic considerations (cf. Donegan & Stampe 1979; Hayes et al. 2004). However, while phonetics research often concentrates on physical properties that are gradient, speech perception imposes a categorical element on the phonetics-phonology relationship; ambiguities in the acoustic signal are parsed according to a restricted number of phonological categories (Ohala 1981). One such ambiguity is found in the early portion of vowels following consonant articulations in consonant-vowel sequences. Strictly speaking, this portion of the signal is vocalic, characterized by robust formant structure

normally associated with vowels. At the same time, listeners use acoustic cues contained in formant transitions for the identification of the preceding consonant (e.g. Wright 2004).

This ambiguity bears directly on an important theoretical question. How is the consonant-vowel distinction encoded at the interface between phonetics and phonology? This is not only a matter of identifying phonetic correlates of consonants and vowels, such as formant structure, obstruent noise, or silence associated with stop closures. It is fundamentally a question of parsing and segmentation. A particular portion of the speech signal, the CV transition, is perceptually ambiguous with regard to consonant-vowel distinction. Individual phonological systems, in defining the relationship between consonant and vowel categories and the speech signal, must resolve the ambiguity in some way. If phonology is to resolve the ambiguity, the ambiguity must be encoded in the representational system. In other words, if phonological systems indeed make categorical distinctions between consonants and vowels, which is a relatively uncontroversial claim, then it is phonology, not phonetics, that must define what a consonant is, what a vowel is, and how these units are mapped to the speech signal.

The Onset Prominence framework (OP; Schwartz 2010 *et seq.*), encodes the parsing of CV transitions in terms of the Vocalic Onset (VO) structural node. The VO node is derived directly from the initial portion of vowels (see Schwartz 2016a, 2017) in CV sequences. The key question is whether this portion of the signal maps onto the consonantal or vocalic 'segment' in the phonological string.[1] When the VO node is affiliated with consonants, CV transitions may extend further into a vowel's duration, opening the door to a phonological reinterpretation of the formant trajectories as vowel-inherent properties. When VO is built into vowel representations, formant targets are reached earlier in vowel duration, leading to impressionistically purer vowel quality. The two languages examined in the current study, Polish and British English, differ with regard to the affiliation of VO. In Polish VO is contained in the representation of vowels, while in English it is built into consonants. These opposing specifications predict cross-language differences in the degree and time course of formant movement in the two languages. English is expected to exhibit greater VISC earlier in the vowel than Polish. The phonological status of the cross-language differences described here is reinforced in earlier work (Schwartz 2016a), in which it is shown that implications of VO affiliation pervade the phonological systems of Polish and English to explain a wide range of seemingly unrelated oppositions between the two languages (4.1), and provide insight into phonetic patterns found in additional languages (4.2).

2 VISC – PHONETIC AND PHONOLOGICAL BACKGROUND

2.1 PREVIOUS RESEARCH ON VISC PRODUCTION AND PERCEPTION

Research on VISC dates back to acoustic experiments in second half of the 20th century. One of the first studies was by Peterson & Lehiste (1960), who examined acoustic aspects of American English vowels. They described, in addition to aspects of vowel-inherent duration and pitch, vowel-based differences in formant trajectories. Later research (e.g. Nearey & Assmann 1986) documented these observations in more detail. In particular, it was observed that so-called tense vowels tend to show movement toward the periphery of the acoustic vowel space, while lax vowels are characterized by movement toward the centre (see e.g. Nearey 2013: 52–54). Nearey & Assmann (1986) coined the term *Vowel Inherent Spectral Change* (VISC), which was hypothesized to be a truly intrinsic aspect of the vowels of North American English. To test this hypothesis, researchers typically employed discriminant analyses, establishing VISC as a significant predictor of vowel identity.

VISC may be related to the effects of neighbouring consonants. Early studies of vowel production described *target undershoot* in CVC contexts (e.g. Lindblom 1963 for Swedish; Stevens & House 1963 for American English), by which canonical formant targets associated with vowels produced in isolation are not reached. The target undershoot problem posed a new challenge for speech perception researchers, who asked how vowel identification could remain constant in the face of consonant-induced acoustic variability. This issue was the focus of a series of experimental studies carried out with North American listeners in the 1970s and 1980s (for a review, see Strange 1989). A common finding was that consonant-induced co-

---

1   Two other logical possibilities present themselves. This portion of the signal can map to both consonants and vowels, or the CV transition can be ignored by the phonological system, and map to neither.

articulation did not hinder vowel identification. Indeed, in some cases co-articulated vowels were identified more accurately than vowels produced in isolation. These findings led to the formulation of a hypothesis that static formant targets in a two-dimensional space were insufficient for describing the perceptual identity of American English vowels. Rather, in the 'dynamic specification' approach (Strange 1989), formant trajectories over the duration of the vowel also provide listeners with crucial cues for vowel perception.

As it happens, target undershoot does not necessarily imply formant movement. Targets may differ under the influence of neighbouring consonants, but if the transition to and from those undershot targets is sufficiently rapid, you may have pure vowels regardless of their location in F1–F2 space. For this reason, examining VISC means looking not only at the extent to which vowel formants diverge from canonical targets, but also the duration of CV and VC transitions relative to overall vowel duration. The role of CV and VC transitions in vowel identification by North American listeners was the focus of a series of experiments in which naturally produced stimuli were altered by silencing various parts of a vowel's duration. In one such stimulus condition, referred to as the Silent Center condition (SC; e.g. Strange et al. 1983), the central quasi-steady-state portion of the vowel is silenced, leaving listeners to identify vowels on the basis of CV and VC transitions. Silent Center tokens are compared for perception accuracy with tokens in which central portion of the vowel is included, or others in which only the CV or VC transitions are included, or unmodified tokens. A consistent finding in these experiments was that the SC tokens were identified most accurately of all the modified stimuli, with error rates often not significantly higher than unmodified tokens (Jenkins & Strange 1999). Other stimulus types, including those constructed from a vowel 'nucleus', induced higher error rates.

More recently, descriptions of VISC have appeared in sociolinguistic studies of English dialectal variation (Fox & Jacewicz 2009; Williams & Escudero 2014; Elvin et al. 2016), as well as studies on English as an L2 (Jin & Liu 2013; Rogers et al. 2013). However, cross-language comparisons are somewhat difficult to find. In one study, Williams et al. (2015) compared the production of vowels in Southern British English and Dutch, and found that spectral change was a better predictor of vowel identity in the former than in the latter. With regard to perception, Schwartz et al. (2016) employed the Silent Center paradigm with L1 Polish learners of English both in their L1 and L2. While the SC items were identified most accurately in L2 English, Polish listeners showed no dynamic specification effects in L1 perception, with constant identification accuracy regardless of the portion of the vowel they heard. Taken together, these studies suggest that vowel inherent spectral change plays a more significant role in the vowel system of British English than it does in Dutch or Polish. The present study is intended to contribute to the dearth of cross-language studies, describing production data from Polish and English.

2.2 THE PHONOLOGICAL ORIGINS OF VISC

In the tradition of generative phonology, as advanced in the *Sound Pattern of English* (SPE) by Chomsky & Halle (1968) and later work that it inspired, VISC would be seen as a matter of phonetic implementation, rather than something that is a systemic aspect of phonological grammars. In what follows, a different perspective is offered, one in which phonological representations encode a parsing ambiguity familiar from speech perception research. This is not to claim that *all* ambiguities from speech perception must be captured in phonological representations. However, the particular ambiguity to be discussed here directly concerns the mapping between speech and the perceived string of phonological units, which any adequate theory of the phonetics-phonology relationship must deal with.

The development of VISC may be thought of in terms of a listener-oriented view of phonology (Ohala 1981; Blevins 2004), in which perceptual ambiguities in the acoustic signal lead to reinterpretation of phonological specifications. Speech perception research informs us that listeners rely on vowel formant transitions in the identification of consonant place of articulation (e.g. Wright 2004). Assuming that the transitions occupy approximately the first and last 25% of a vowel's duration, there are consequences when they are produced more slowly extending further into the vowel. In such cases, listeners may be expected to reinterpret the movement as a feature inherent to the vowel itself. In other words, canonical vowel representations may be reorganized to encode movement that was originally a product of formant transitions from neighbouring consonants. While different consonant places of articulation of course produce

different formant trajectories, it may be hypothesized that patterns associated with the most common place of articulation, typically coronal, may be extended analogically over the course of diachronic development.

If the origins of VISC may be found in listener-induced reinterpretation of consonant-vowel transitions, there are implications for any theory of phonological representation in which consonant and vowel segments are universal entities. Vowel formant transitions constitute an inherent perceptual ambiguity. Phonetically, they are vocalic, but listeners may use them for consonant identification. Each phonological system must therefore interpret CV transitions in terms of its consonant-vowel distinction. In other words, are the transitions built into the representation of consonants, vowels, neither, or both? This is essentially a parsing problem. A single phonetic entity (formant transitions) may be associated with more than one phonological object (consonant and vowel segments), and individual phonological systems must choose from between those objects in determining the principles underlying the phonetics-phonology interface.

2.3 VISC AND SEGMENTAL STRUCTURE IN PHONOLOGICAL REPRESENTATION

One account of how this parsing problem may be resolved may be found within the Onset Prominence representational environment (Schwartz 2010 *et seq.*), in which 'segmental' representations emerge from a representational hierarchy encoding a stop-vowel CV sequence. Each 'segment' in the OP model has internal structure that is derived from the various phonetic phases in the articulation of the CV. Although this basic idea is not unique to OP, having been explored elsewhere in Aperture Theory (e.g. Steriade 1993) and Q Theory (Shih & Inkelas 2019), those theories have focussed primarily on the phonological behaviour of individual entities such as contour segments. By contrast, OP's hierarchical perspective offers an account of parsing ambiguities arising from transitions *between* segments.

Sample OP representations are given in the trees in (1), which depict two potential parses of a stop-vowel sequence.[2] Each level of the representational hierarchy corresponds to a more or less discretely identifiable phonetic event in the production of the stop-vowel CV: Closure (C) which produces silence or near silence, Noise (N) encoding aperiodic release bursts and aspiration/frication, Vocalic Onset (VO) encoding CV transitions, and Vocalic Target (VT) capturing relatively stable vowel quality. For more discussion, see Schwartz (2016a) and Schwartz (2017).

(1)   VO parameters in the OP framework. Consonantal VO affiliation (trees a. and b.), vocalic VO affiliation (trees c. and d.).
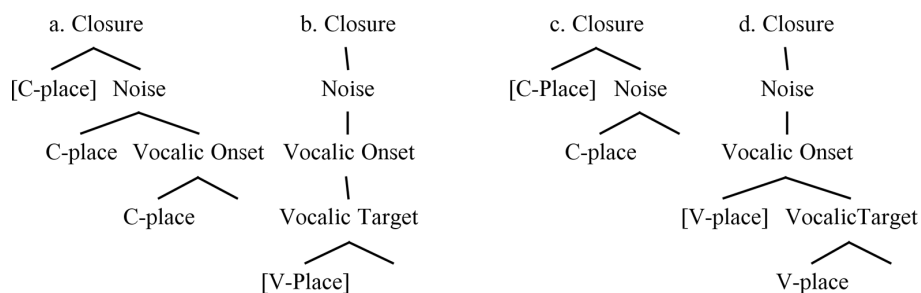
In (1), stops are shown in trees a. and c., and occupy the higher levels (Closure and Noise) of the OP hierarchy. Vowels (in b. and d.) are found at the bottom of the hierarchy. In the OP framework there is no segmental 'skeleton'; OP trees directly encode manner of articulation as prosodic structure. Place and laryngeal features are assigned at various levels of the hierarchy, and 'trickle' down the structure (Schwartz 2016a: 45). For example, in tree (1a) the stop's place feature is assigned at the Closure level, but also occupies the Noise and Vocalic Onset nodes. Assigned place is indicated in square brackets; 'trickled' place appears without brackets. Laryngeal features are not shown in (1), but will be considered in 4.2. The 'trickling' mechanism encodes the causal relationship between various supra-laryngeal articulations and

---

2   Another possibility is that the VO node is built into both consonant and vowel representations. This may be posited for Eastern Arrernte, which supports rich place contrasts cued by formant transitions (Tabain et al. 2004), but has also been analysed as having only 'onsetless' syllables (Breen & Pensalfini 1999). Consonantal VO encodes the role of formant transitions, while vocalic VO affiliation ensures that vowel-initial syllables are prosodically well-formed (Schwartz 2013a).

their acoustic consequences. For instance, a coronal specification assigned to the Closure level in a plosive will determine the burst spectrum at the Noise level, and will also affect CV formant transitions at the VO level, which may be represented as 'trickled' feature specifications.

Trickling is crucial to the OP account of the emergence of VISC. It encodes the phonetic causality by which stop place affects the spectrum of noise bursts (e.g. Stevens & Blumstein 1981) and formant transitions (e.g. Delattre et al. 1955). However, in the OP environment the trickling mechanism is subject to a formal restriction: it is blocked by the assignment of a feature at a lower level (Schwartz 2016a: 45). Thus, in tree (1a) the C-place feature occupies both Noise and VO since the V-place is assigned at VT, while in (1c) C-place trickles only as far as the Noise node, after which it is blocked by the VO-level V-place specification. When C-place specifications are allowed to trickle onto VO (1a), it creates a configuration conducive to the development of VISC – the stop's place specification impinges on the vowel's structural space, and the vowel's place features are anchored later in the vowel (1b). Diachronically, the movement toward the VT-level docking point for the vowel features may be phonologized to become a 'vowel-inherent' feature. By contrast, when VO-level vowel specifications block trickling of C-place features, CV transitions are more rapid and vowel quality is more stable.[3]

The trickling mechanism is also crucial in distinguishing OP from other theories of segment-internal phonological structure mentioned earlier. In Aperture Theory (Steriade 1993), consonants are decomposed into separate root nodes encoding closure (A_0), noise (A_f) and release (A_{max}). In Q Theory (Shih & Inkelas 2019), segments are comprised of three 'sub-segments' that are claimed to be linked to Articulatory Phonology's (Browman & Goldstein 1989) 'onset transition', 'target', and 'release transition' phases of segmental structure (see Gafos 2002). What sets apart OP from these other models is the hierarchical nature of the CV unit that comprises its building block. To account for the acoustic effects of closure localization on later phases in the structure of a stop, both Aperture Theory and Q Theory would require a linear spreading mechanism, which would in turn necessitate stipulations to explain the direction of movement. In OP, trickling is an automatic by-product of the hierarchical organization of the model's segment-internal structures, and directly encodes the left-to-right directionality of acoustic effects in CV sequences.[4] An additional advantage of the OP perspective over the other models of segment-internal structure is that those models do not make any predictions about cross-language differences in the implementation of CV formant transitions, which are the empirical focus of this paper. For OP, the relevant predictions come from the ambiguous status of the VO node, as shown in (1). When VO is parsed as part of the consonant, more VISC is predicted. To the best of my knowledge, neither Aperture Theory nor Q Theory encode parsing ambiguities of this type.

The experimental studies in this paper describe systemic differences between Polish and English with regard to vowel formant dynamics, which suggest that English may be seen as a system with consonantal VO affiliation, while Polish shows vocalic VO affiliation. The representational parameters also make predictions for other phonetic features, including the status of CV formant transitions for consonant perception – English places greater perceptual weight on transitions than Polish (Schwartz & Aperliński 2014; Aperliński & Schwartz 2015), and the likelihood of vowel-initial words to undergo linking processes – Polish does not resyllabify C#V sequences (Rubach & Booij 1990). These issues will be revisited in Section 4.

One additional point about the predictions of these representations must be made at this time. The OP hierarchy is built from a stop-vowel sequence in which CV transitions are built into the representations. The representations as shown make no predictions about VC transitions and the behaviour of 'coda' consonants. As a consequence of the fact that the representational configurations in (1) are associated with CV formant transitions and not VC transitions, the basic empirical prediction that English should show more VISC than Polish may be refined. Rather, it might be hypothesized that English should show more dramatic formant movement relatively early in the time course of the vowel, since VISC is seen to arise from extended CV transitions.

---

3   The 'stable' quality predicted by vocalic VO affiliation does not imply a further prediction that there are no co-articulatory effects of consonants on vowel quality in languages with Vocalic VO affiliation. These effects exist, but are reflected in altered static target locations in F1–F2 space, and do not necessarily induce more formant movement.

4   Regressive effects in the OP model are due to an additional mechanism, submersion, that is not directly relevant to this paper. For details, see Schwartz (2016a).

At the same time, perceptual considerations suggest that CV transitions should play a greater role in phonological organization than VC transitions. The initial portion of vowels after an onset consonant is associated with a perceptual boost in which auditory sensitivity is increased, while VC transitions are typically associated with a period of lessened auditory sensitivity (e.g. Wright 2004). As a result, formant movement later in the vowel may be assumed to be less perceptually robust.

2.4 RESEARCH HYPOTHESIS

Before proceeding to a description of the cross-language acoustic studies, it is necessary to state the main research hypothesis. The hypothesis involves two basic predictions. First, consonant-based VO affiliation in English predicts that English should show a greater degree of formant movement in its vowel system than Polish, which associates the VO node with the representation of vowels. The second prediction is that the greater formant movement in English should be concentrated in the first half of the vowel's duration, since consonantal VO specification involves an encroachment by the CV formant transitions into the structural space of the vowel.

3 ACOUSTIC PHONETIC EXPERIMENTS

This section will describe two cross-language acoustic studies of vowel formant dynamics in Polish and British English. The first study compares data from a corpus of Southern British English with our own recordings of L1 Polish. This study will be referred to as the L1 Comparison Study. The second examines the speech of proficient Polish speakers of English in both their L1 and L2. The second experiment will be referred to as the L1–L2 Study. While VISC in English has been described in a number of published works, available phonetic descriptions of vowels in standard Polish (Dukiewicz & Sawicka 1995) make no mention of VISC or diphthongization, and its vowels are impressionistically pure in quality. Thus, the working hypothesis is that vowel inherent spectral change is less prevalent in Polish than in English, as was outlined in 2.3.

The analysis in both studies is based on citation form productions (see Procedure) of four different Polish-English vowel pairs in that may be thought to correspond in terms of their position in two-dimensional vowel charts. The English vowels employed in the study are /iː/ /ɪ/ /e/ and /æ/, corresponding to Wells' (1982) keywords FLEECE, KIT, DRESS, and TRAP, respectively. The Polish vowels used in the study were /i/ /ɨ/[5] /ɛ/ and /a/, spelled <i> <y> <e> and <a>, respectively. According to Sobkowiak (2008), the FLEECE vowel is slightly more peripheral than Polish /i/, the KIT vowel slightly further forward than Polish <y>, the DRESS vowel is higher than Polish <e>, and the TRAP vowel is further forward than Polish /a/.[6] The four vowel pairs chosen for analysis were selected to provide approximate pairings of 'similar' vowels in the two languages, facilitating cross-language comparison of VISC. Other vowels were left out of the study because their Polish-English correspondences are not as clear. For example, Polish lacks contrasting high rounded vowels (cf. English GOOSE vs. FOOT), while Polish /ɔ/ is typically higher than the British English LOT vowel, and lower than the THOUGHT vowel.

3.1 CROSS-LANGUAGE COMPARISON OF L1 POLISH AND L1 ENGLISH (L1 COMPARISON STUDY)

3.1.1. Procedure

The L1 Comparison data are taken from two sources. For Polish, citation form recordings were made of 24 L1 Polish speakers (17 female, 7 male) producing the target vowels /i ɨ ɛ a/ in CVC words (see Appendix) in Polish. These speakers were all first year students starting an English-language program at a Polish university. Although these speakers had intermediate level of proficiency in English, B1 according to the Common European Frame of Reference, the

---

5   The status of this vowel requires additional comment. Although traditionally transcribed with the symbol /ɨ/, this vowel is more of a slightly retracted and lowered front vowel, rather than a high central vowel as its transcription suggests. In this sense, it may be considered analogous to the KIT vowel in British English. In addition, the phonological status of this vowel as a phoneme in the language has been the subject of debate due to its restricted distribution. It is banned from word-initial position.

6   Recent work has shown that the TRAP vowel is undergoing a process of retraction in younger generations of RP speakers (Hawkins & Midgley 2005), rendering it closer to Polish /a/.

students had yet to receive any phonetic training in English. In addition, the recording session was carried out entirely in Polish. This may be assumed to have prevented language mixing effects (Grosjean 2004) as it has in other studies of bilingual speech (e.g Antoniou et al. 2010). Thus, it may be suggested that the data are reasonably characteristic of Polish as a whole.

Recordings of Polish were made in a sound treated room at the English department of a Polish university. Items were presented one at a time on a monitor located within the recording booth, using Speech Recorder (Draxler & Jänsch 2015). The order of the slides was randomized. Speakers produced included four repetitions of each item. The coda consonant was always /t/, while the onset consonants were lenis stops /bdg/, counterbalanced for place of articulation (see Appendix). The dorsal context before /i/ was excluded due to Polish phonotactic restrictions. A total of 1056 L1 Polish items was recorded (24 speakers * 3 vowels * 3 onsets * 4 repetitions + 24 speakers*/i/ * 2 onsets * 4 repetitions). During manual annotation in the Praat program (Boersma & Weenink 2017), 85 items were excluded due to speech errors or irregularities of formant tracking in Praat, leaving a total of 971 Polish items for analysis.

Unfortunately, it was not feasible to collect a representational sample of vowels from L1 English speakers, since in the Polish city in which this research was carried out it was impossible to find a homogeneous group of native speakers of English. For this reason, acoustic measurements for British English native speakers were gathered from recordings in the UCL Speaker Database (Markham & Hazan 2002), made at University College, London. The corpus is said to contain 45 "speakers of British English with a fairly neutral accent or mild South-Eastern English accent" (Markham & Hazan 2002:1). The speech materials include word lists, sentence lists, reading passages and unscripted speech. The recordings were made at the Department of Phonetics and Linguistics, UCL, in an anechoic chamber.

For the present study, only recordings of adults (18 female,15 male) reading the word lists were used, in order to ensure experimental conditions that are comparable with our Polish speakers, who also read word lists. The lists in the UCL Database contain a number of monosyllabic words (CVC shape). Those containing the vowels of interest, FLEECE, KIT, DRESS, and TRAP, were extracted. In choosing words for analysis, consonantal context was considered. Unfortunately, the word list was not perfectly counterbalanced in this respect. Twelve words were selected (see Appendix). In selecting these words, aspirated stops were avoided, as these have been shown to affect CV formant transitions (Stevens and Klatt 1974). Given the limitations of the dataset in the UCL database, it was necessary to include both fortis (7) and lenis codas (5). With 33 speakers, the twelve words created a corpus of 396 L1 English vowel tokens.

3.1.2 Analysis

The recordings were annotated manually in Praat (Boersma & Weenink 2017), with F2 onset and offset determining vowel boundary location. A script was used to segment each vowel into four vowel-internal intervals (0–25%, 25–50%, 50–75%, and 75–100% of vowel duration). As is common practice in VISC research, the first and fourth intervals were excluded from interval-based analyses in an attempt to minimize the effects of neighbouring consonants on formant trajectories (cf. Fox & Jacewicz, 2009; Williams & Escudero, 2014). The measures extracted by the script included F1 and F2 Slopes (in Bark/100ms units), and mean Bark normalized F1 and F2 values (F1–f0; F3–F2; Syrdal & Gopal 1986) for the 2nd and 3rd intervals. The mean Bark normalized F1 and F2 measures were not analysed statistically – they are presented simply to provide a general overview of the vowels' positions in F1–F2 space, as well as the movement between the 2nd and 3rd intervals (25–50%; 50–75%).

The absolute values of the formant slope measures for each interval (F1-2nd, F2-2nd, F1-3rd, F2-3rd) were calculated in order to quantify the distance of the formant slopes from zero, regardless of whether they were positive or negative. This was done in order to facilitate interpretation of the statistical results to be described in what follows. The absolute values of the formant slope measures served as dependent variables in a series of generalized linear mixed models performed in SPSS (IBM corporation 2013). Separate models were fitted for each of these four measures. Predictor variables included Language, Vowel Pair, Onset place (labial, coronal dorsal), Coda place (labial, coronal, dorsal), with Vowel Duration as a continuous predictor. The models also included by-speaker random slopes and intercepts. Results will be reported as contrast estimates in a set of models with a Language*Vowel*Onset interaction term as

predictor variable, in order to quantify the language-induced differences in formant slopes for each vowel pairing and onset consonant. Coda place and Vowel Duration were included as control variables in the models to quantify variation they may have induced, but will not be discussed in detail.[7]

3.1.3 Results

Figure 1 provides an overview of mean Bark normalized F1 and F2 values in the 2nd and 3rd interval of each vowel, sorted for language. The numbers in the vowel space denote the vowel interval in which the mean was recorded. An overview of the degree of formant movement in the two languages can be gleaned by examining the distance between the 2nd and 3rd interval means. Visual inspection of the figure reveals that for all four vowel pairings, English appears to show a larger excursion between the two intervals, a difference that appears to hold primarily in the height dimension. In the analyses to follow, this movement is quantified as slopes of individual formants in each interval on a vowel-by-vowel basis, summarized in *Figures 2* and *3*
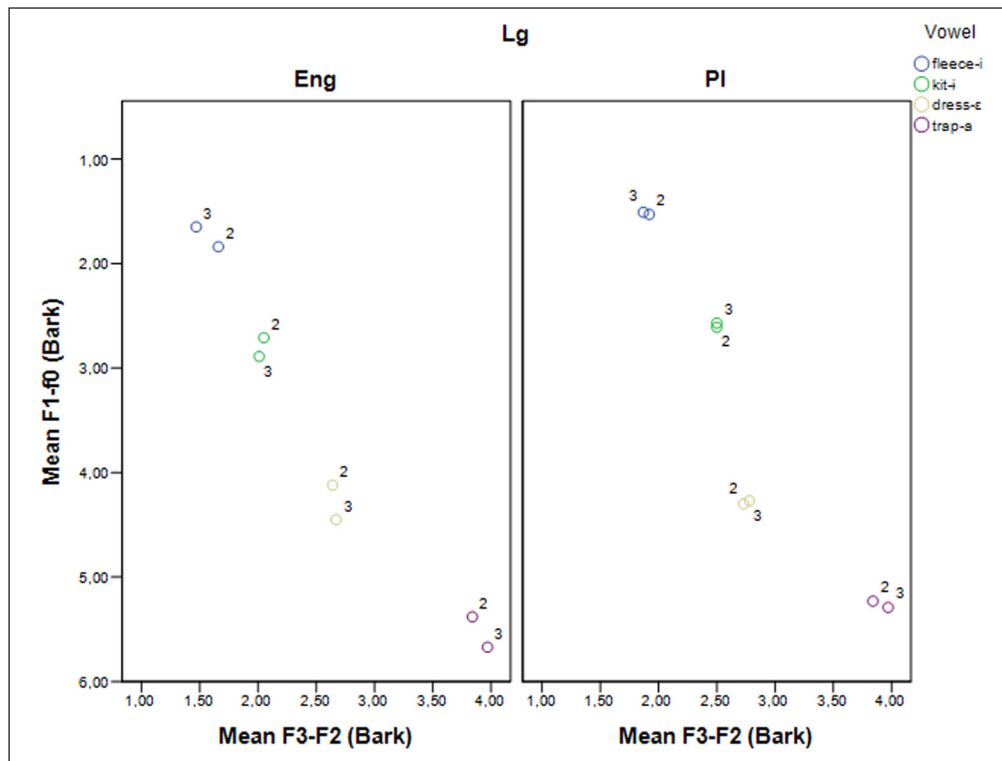

Figure 1 Mean F1-f0 and F3-F2 values for 2nd and 3rd interval sorted for Language in the L1 comparison study.
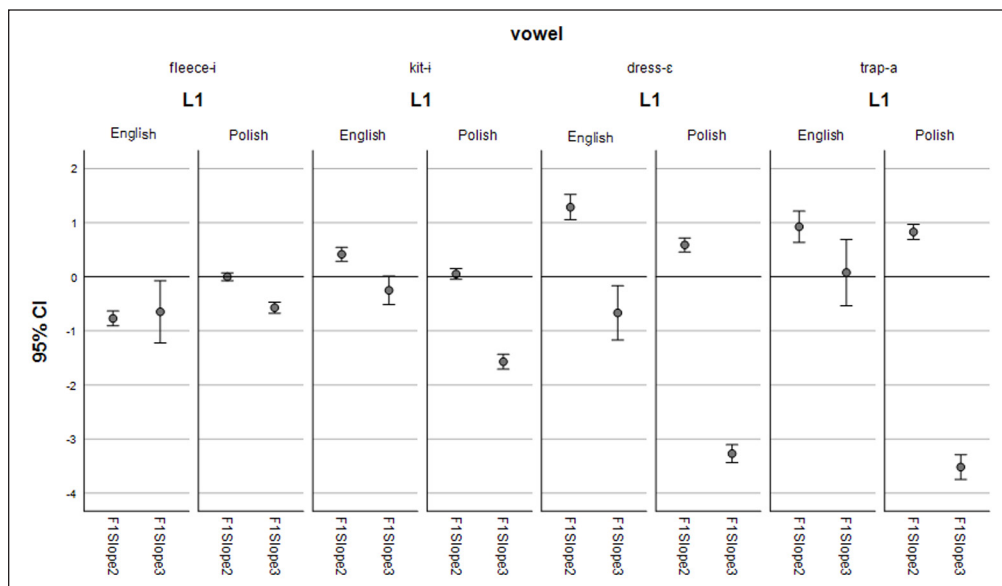

Figure 2 Error bars denoting 95% confidence intervals for F1 slopes by vowel and language. Reference line at zero.

---

7   An additional comparison of Polish and L1 English, using recordings from online pronunciation dictionaries of English, is described in Appendix B. Except for two of the English target words, *sheep* and *stick*, all codas were coronal. Coda place was investigated in a set of linear models with a Language * Vowel * Coda interaction term as predictor. Effects of coda place are shown in Table 12 and Figure 9 in Appendix C.
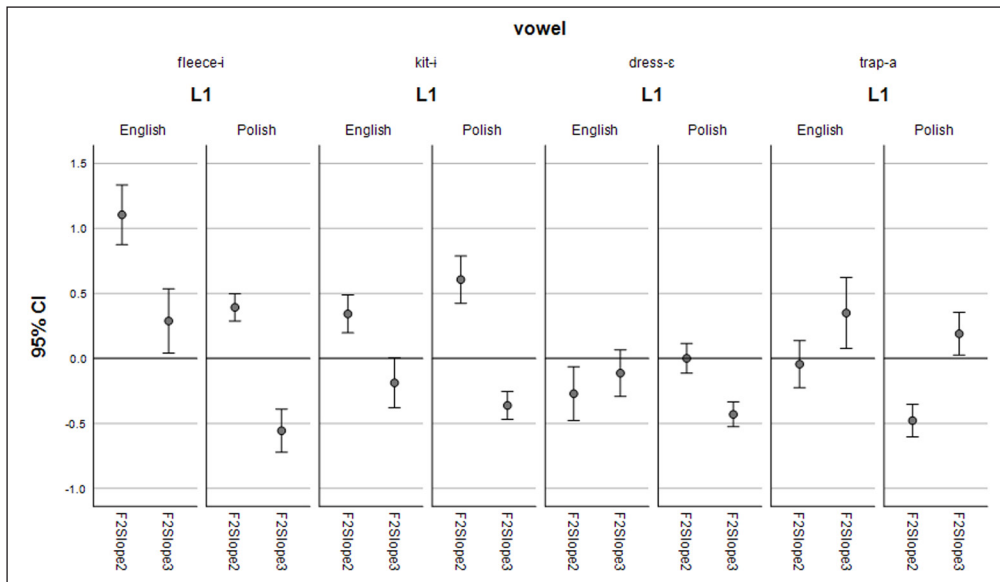
Figure 3 Error bars denoting 95% confidence intervals for F2 slopes by vowel and language. Reference line at zero.

Tables 1–4 present the estimated cross-language contrasts of the linear models for each interval-based measure, sorted for vowel pairing and onset consonant. The estimates represent the difference between English and Polish predicted by the models. When the estimate is positive, it means that English has a steeper formant trajectory. When it is negative, Polish has a steeper a steeper trajectory. Recall that the statistical models were run on the absolute values of the slope measures. To see whether a given formant was rising or falling, see *Figures 1* and *2*.

VOWEL	ONSET	CONSTRAST ESTIMATE (ENGLISH – POLISH)	STD. ERROR	t	p
trap-ɑ	labial	0.158	0.165	0.96	.337
	coronal	1.221	0.153	7.99	<.001
dress-ɛ	labial	0.09	0.161	0.562	.574
	coronal	–0.003	0.139	–0.024	.981
	dorsal	1.469	0.133	11.02	<.001
kit-ɨ	labial	0.196	0.135	1.45	.148
	coronal	0.543	0.125	4.341	<.001
fleece-i	coronal	0.517	0.127	4.07	<.001

Table 1 Contrast estimates for 2nd interval F1 slope.

VOWEL	ONSET	CONSTRAST ESTIMATE (ENGLISH – POLISH)	STD. ERROR	t	p
trap-ɑ	labial	0.319	0.189	1.692	.091
	coronal	–0.13	0.177	–0.731	.465
dress-ɛ	labial	0.28	0.186	1.505	.132
	coronal	0.683	0.160	4.273	<.001
	dorsal	0.545	0.153	3.565	<.001
kit-ɨ	labial	0.235	0.155	1.513	.131
	coronal	0.657	0.143	4.60	<.001
fleece-i	coronal	0.862	0.145	5.96	<.001

Table 2 Contrast estimates for 2nd interval F2 slope.

3.1.4 Discussion

The results of the L1 comparison study reveal that effects of Language are largely dependent on vowel interval. Formant movement in English is more dramatic in the second interval (25–50% of vowel duration), while in Polish it is more dramatic in the third (50–75%). English exhibited greater F1 movement than Polish in the 2nd interval in four of eight vowel-onset combinations, three of which with coronal onsets, representing all four vowel pairings (*Table 1*). Steeper 2nd

VOWEL	ONSET	CONSTRAST ESTIMATE (ENGLISH – POLISH)	STD. ERROR	t	p
trap-ɑ	labial	−1.801	0.344	−5.24	<.001
	coronal	−2.601	0.317	−8.21	<.001
dress-ɛ	labial	−1.327	0.334	−3.98	<.001
	coronal	−1.62	0.293	−5.53	<.001
	dorsal	−2.285	0.281	−8.13	<.001
kit-ɨ	labial	−0.321	0.285	−1.13	.26
	coronal	−0.233	0.266	−0.88	.381
fleece-i	coronal	0.243	0.271	0.89	.37

Table 3 Contrast estimates for 3rd interval F1 slope.

VOWEL	ONSET	CONSTRAST ESTIMATE (ENGLISH – POLISH)	STD. ERROR	t	p
trap-ɑ	labial	0.279	0.224	1.243	.214
	coronal	−0.068	0.208	−0.325	.745
dress-ɛ	labial	0.205	0.219	0.933	.351
	coronal	0.249	0.19	1.316	.188
	dorsal	−0.088	0.181	−0.485	.628
kit-ɨ	labial	0.422	0.184	2.292	.022
	coronal	0.423	0.17	2.485	.013
fleece-i	coronal	0.059	0.173	0.339	.735

Table 4 Contrast estimates for 3rd interval F2 slope.

interval F2 movement in English was also observed in four of the vowel-onset combinations (**Table 2**). In the third interval, by contrast, Polish showed greater F1 movement for the non-high vowels, regardless of onset (**Table 3**). This effect was clearly attributable to greater negative slopes in the Polish items (**Figure 2**, left panel), suggesting the F1 transition to the final coda consonant was well underway in Polish, but less so in English. With regard to F2 in the third interval, effects were observed only for the KIT-/ɨ/ pair (**Table 4**).

Taken together, the results of the L1 comparison study are compatible with the phonological proposal from Section 2. In English, the greater movement in the second interval points to VISC as the result of phonologization of extended CV transitions, encoded in the OP framework as consonantal VO affiliation as shown in (1).

3.2 L1-L2 STUDY

We turn now to a study of L1 Polish learners of English speaking both in their native language and in their L2. The goal of the second study is to see if the cross-language effects observed in the L1 Comparison are found within a single group of speakers in their L1 and L2. A positive finding in this regard would strengthen the claim that patterns of formant movement are a systemic element of the phonological systems of the two languages, as predicted by the OP representational model.

3.2.1 Participants

Twelve L1 Polish speakers of English took part in the experiment. All of the participants were female. At the time of recording they were students in their third year of an English language specialization at a Polish university. All of the participants were L1-dominant Polish native speakers with C1 level proficiency in English according to the Common European Frame of Reference for Languages. Admission to our English programme requires B1–B2 level proficiency. In the first two years of the programme, students receive intensive language instruction, theoretical courses in linguistics, including a theoretical course on English phonetics and phonology, as well as cultural studies courses, all conducted in English. Most relevant for our purposes, the programme also includes two years of intensive instruction in English pronunciation using a Southern British English model. Pronunciation instruction includes extensive drilling, listening, as well as acoustic comparisons of student productions with native recordings. Pronunciation is

also an important aspect of the evaluation procedure of the oral component. It is therefore safe to assume that third year students who 'survive' the first two years of the English programme have achieved at least a C1 level of proficiency in English, with only negligible traces of a Polish accent. Nevertheless, these students function in an L1 dominant environment, and use Polish more than English in their everyday lives.

3.2.2 Materials and procedure

The materials and procedure of the L1–L2 study, including acoustic measures and statistical procedures, closely match those of the Polish part of the L1 study described earlier, with an additional recording session devoted to L2 English. The only difference in the materials is that in this case it was possible to gather a counterbalanced data set in English with regard to consonantal context. The vowels examined in this study were produced in three different consonantal contexts: /b_t/, /d_t/, and /g_t/. In total, 1056 items were recorded, including 576 English items (12 speakers*4 vowels*4 repetitions*3 onsets) and 480 Polish items (the same number, but without /gɨt/). From these, 80 items were eliminated due to speaker errors or formant tracking irregularities, leaving 976 total items for analysis.

Recordings were made in two sessions separated by at least one week. The purpose of separating the recording sessions was to prevent language mixing effects (Grosjean 2004). In the first session, Polish was recorded, and the recordings were conducted in Polish by a native speaker. In the 2nd session, the experiment was carried out in English by either a native speaker of English or a C2-level Polish speaker of English. The participants were seated in a sound treated booth equipped with a computer monitor, while experimental items were elicited one at a time on slides using Speech Recorder, which also randomized the order of presentation. Recordings were made directly onto a laptop computer with a head-mounted Shure SM35-XRL microphone connected to a Roland UA-25 USB Audio Interface.

3.2.3 Results

Figure 4 provides an overview of mean Bark normalized F1 and F2 values in the 2nd and 3rd interval of each vowel, sorted for language. The numbers in the vowel space denote the mean for the corresponding interval. An overview of the degree of formant movement in the two languages can be gleaned by examining the distance between the 2nd and 3rd interval means. Visual inspection of the Figure suggests that English shows greater movement in the non-high vowels.
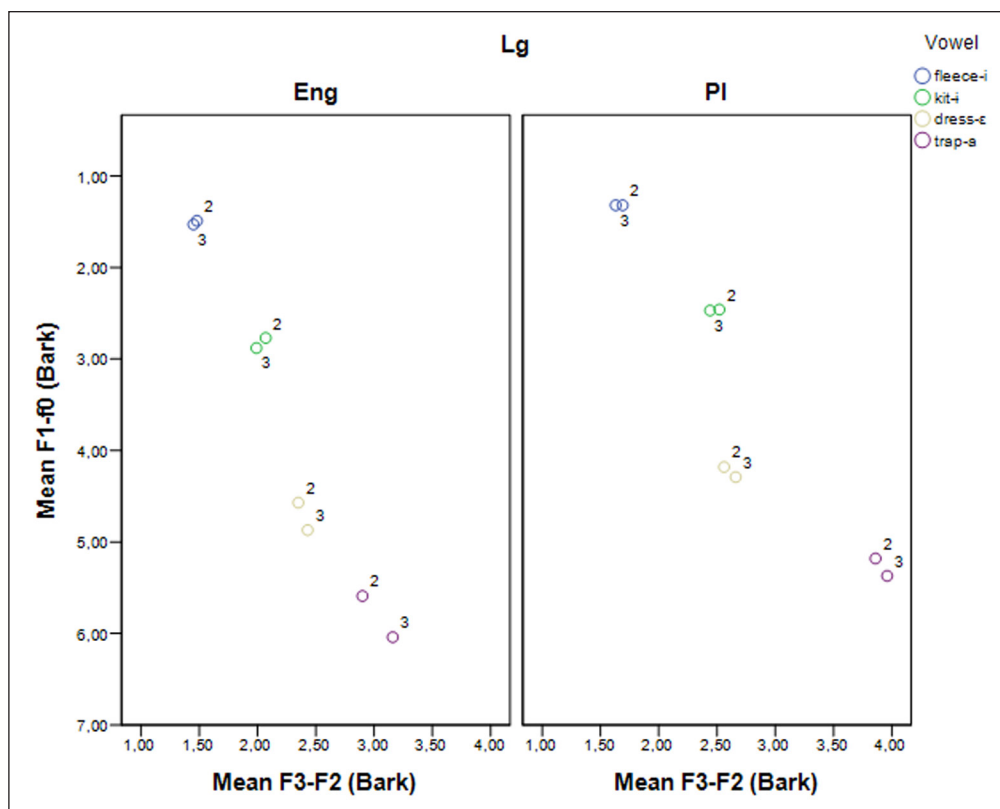
Figures 5 and *6* summarize formant slope measures for the second and third intervals, respectively. The figures suggest that more consistent effects of Language occur for the first formant.
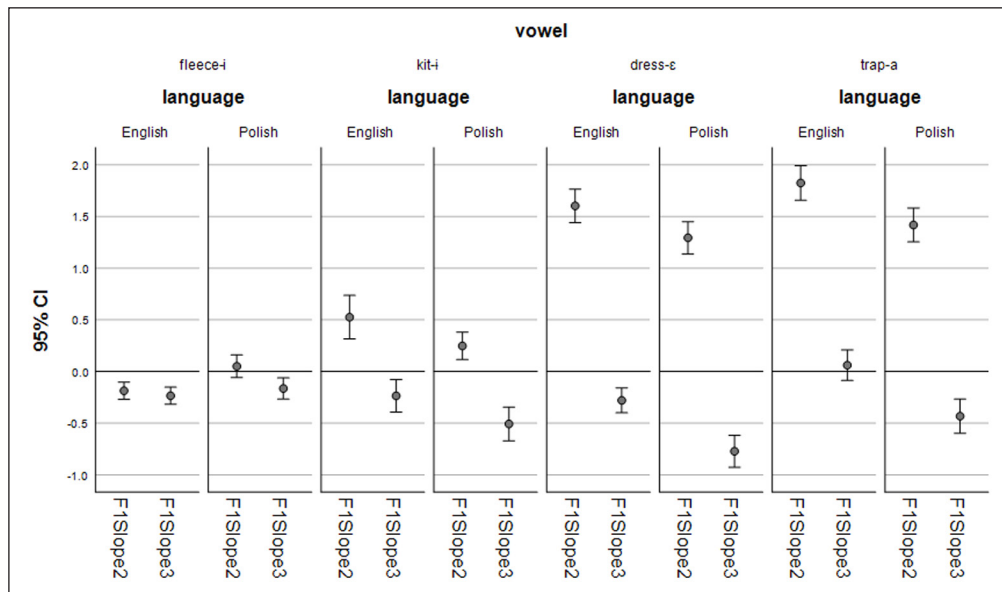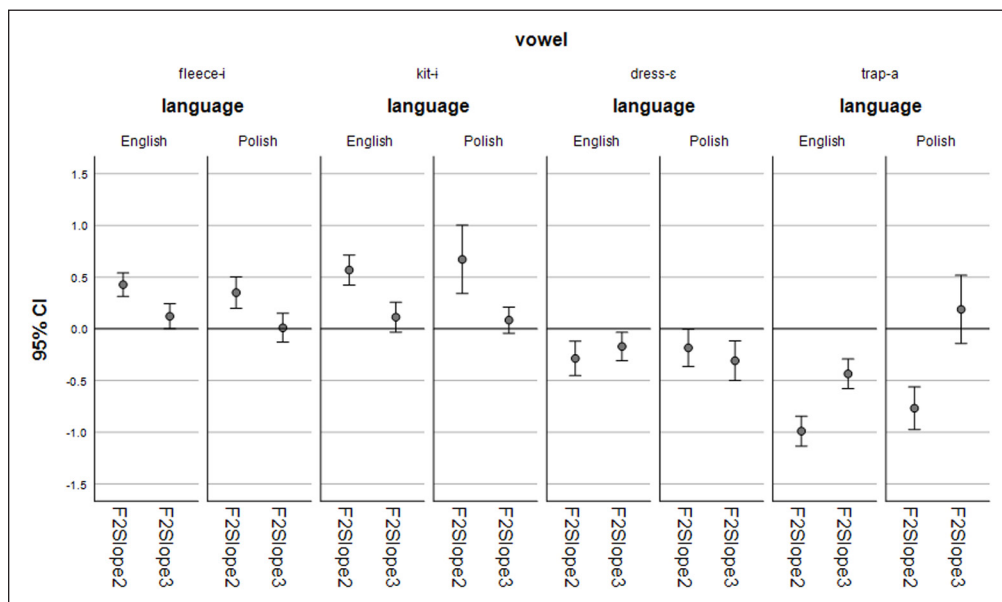
Statistical analyses were nearly identical to those of the L1 comparison study, except that coda place was excluded as a control variable in L1–L2 study, since all codas were coronal (see Appendix). *Tables 5–8* present the estimated cross-language contrasts of the linear models for each interval-based measure, sorted for vowel pairing and onset consonant. The estimates represent the difference between English and Polish predicted by the model. When the estimate is positive, it means that English has a steeper formant trajectory. When it is negative, Polish has a steeper a steeper trajectory. Recall that the statistical models were run on the absolute values of the slope measures. To see whether a given formant was rising or falling, see *Figures 5* and *6*.

3.2.4 Discussion

The results of the L1–L2 study may be summarized as follows, the English items showed greater formant movement than the Polish items in the 2nd interval of the vowel, while Polish showed greater movement in the 3rd interval. These differences were found primarily in the first formant of the non-high vowel pairings. As with the L1 comparison study, we observed steeper F1 slopes in the second interval for English (*Table 5*), and steeper F1 slopes in the third interval for Polish (*Table 7*). The effects of Language on the F2 trajectories (*Tables 6* and *8*) appear to be much less consistent.

VOWEL	ONSET	CONSTRAST ESTIMATE (ENGLISH – POLISH)	STD. ERROR	*t*	*p*
trap-ɑ	dorsal	0.376	0.163	2.303	<.001
	coronal	0.915	0.132	6.96	<.001
	labial	0.355	0.142	2.498	<.001
dress-ɛ	dorsal	0.553	0.149	3.706	<.001
	coronal	0.578	0.124	4.652	<.001
	labial	0.057	0.13	0.44	.66
kit-ɨ	coronal	0.613	0.13	4.707	<.001
	labial	0.378	0.138	2.732	.006
fleece-i	dorsal	0.246	0.172	1.428	.154
	coronal	0.366	0.137	2.668	.008
	labial	0.275	0.139	1.983	.048

Table 5 Contrast estimates for 2nd interval F1 slope, L1–L2 study.

VOWEL	ONSET	CONSTRAST ESTIMATE (ENGLISH – POLISH)	STD. ERROR	*t*	*p*
trap-ɑ	dorsal	0.438	0.194	2.26	.024
	coronal	0.08	0.156	0.511	.61
	labial	0.165	0.169	0.976	.329
dress-ɛ	dorsal	0.163	0.177	0.921	.357
	coronal	0.282	0.147	1.909	.057
	labial	–0.004	0.155	–0.024	.981
kit-ɨ	coronal	0.179	0.155	1.155	.249
	labial	–0.298	0.164	–1.815	.07
fleece-i	dorsal	0.316	0.205	1.542	.124
	coronal	0.057	0.163	0.351	.726
	labial	0.416	0.165	2.522	.012

Table 6 Contrast estimates for 2nd interval F2 slope, L1–L2 study.

VOWEL	ONSET	CONSTRAST ESTIMATE (ENGLISH – POLISH)	STD. ERROR	*t*	*p*
trap-ɑ	dorsal	–1.731	0.478	–3.624	<.001
	coronal	–2.58	0.43	–6.003	<.001
	labial	–2.037	0.445	–4.573	<.001
dress-ɛ	dorsal	–1.035	0.455	–2.274	.023
	coronal	–1.557	0.419	–3.718	<.001
	labial	–0.854	0.426	–2.006	.045
kit-ɨ	coronal	0.074	0.426	0.173	.863
	labial	–0.234	0.438	–0.535	.593
fleece-i	dorsal	–0.879	0.504	–1.743	.082
	coronal	–0.806	0.448	–1.8	.072
	labial	–0.666	0.447	–1.491	.136

Table 7 Contrast estimates for 3rd interval F1 slope, L1–L2 study.

In Section 2, it was suggested that the origins of VISC may be found in extended CV formant transitions, in accordance with the phonological representations in (1). A capsule summary of the results of both studies is presented in *Table 9*, which allows us identify language-specific differences in the temporal coordination between consonants and vowels in CVC syllables. English appears to be characterized by slower CV transitions that appear to extend to vowel midpoint or even beyond, while in Polish these transitions are more rapid. This was reflected in the fact that English showed greater F1 movement in the second interval (25–50%) in both

VOWEL	ONSET	CONSTRAST ESTIMATE (ENGLISH – POLISH)	STD. ERROR	t	p
trap-ɑ	dorsal	0.408	0.193	2.114	.035
	coronal	−0.249	0.156	−1.595	.111
	labial	−0.312	0.168	−1.854	.064
dress-ɛ	dorsal	−0.098	0.177	−0.554	.58
	coronal	−0.179	0.147	−1.219	.223
	labial	0.061	0.154	0.395	.693
kit-ɨ	coronal	0.218	0.154	1.415	.157
	labial	0.045	0.164	0.273	.785
fleece-i	dorsal	0.122	0.203	0.603	.546
	coronal	0.068	0.162	0.418	.676
	labial	0.439	0.164	2.681	.007

Table 8 Contrast estimates for 3rd interval F2 slope, L1–L2 study.

VOWEL PAIRING	COMMON RESULTS OF BOTH STUDIES
Fleece-i	2nd interval: steeper F1 fall for English
Kit-ɨ	2nd interval: steeper F1 rise for English
Dress-ɛ	2nd interval: steeper F1 rise for English; 3rd interval: steeper F1 fall for Polish
Trap-ɑ	2nd interval: steeper F1 rise for English; 3rd interval: steeper F1 fall for Polish

Table 9 Capsule summary of results for both acoustic comparisons.

studies. In the third interval, Polish showed more F1 movement in the non-high vowel pairs, which suggests that the VC transitions in the Polish items were already underway in the third quarter of the vowel.

Considering the proposed phonological connection between CV transitions and VISC, it is necessary to comment on the fact that phonologization of VISC was observed in F1 rather than F2, despite the established role of the latter as a cue to consonant place (Wright 2004).[8] All three places of articulation studied here (labial, coronal, velar) are associated with a low F1 locus, inducing a rising trajectory in all but the highest of vowels (Delattre et al. 1955; Stevens 1998). By contrast, F2 trajectories are more variable as a function of place (Delattre et al. 1955; Stevens 1998). In other words, onset consonants affect trajectories of both F1 and F2, but the F1 effect is more consistent, and therefore is a better candidate for phonologization. An additional contributing factor is that listeners are more sensitive to small acoustic differences in the F1 frequency range (<1000 Hz) than the F2 frequency range (1000–2500 Hz), so F1 should be more conducive to phonologization than F2. This is reflected in the typology of vowel systems (e.g. Ladefoged & Maddieson 1996), which typically show a larger number of F1 categories than F2 categories.

Returning to the representations in (1), an important cross-language difference consists in the relative location of the vowel's target specification, which is later in English (the VT node) than in Polish (the VO node). By way of illustration, *Figure 7* presents an annotated waveform and spectrogram display of an L1 British English speaker producing the word *that* and an L1 Polish speaker producing the Polish word *dat* 'date (gen.pl.)'. These items were taken from the L1 Comparison study. Of particular interest is the trajectory of the first formant. In the case of low vowels in a context between coronal consonants, the F1 maximum may be assumed to represent a 'target' value for the formant. In comparing the two spectrograms, notice the position in the vowel at which the F1 maximum is reached. These time points are marked in the top tier of annotation. In the English token, this point falls quite close to the end of the vowel. Over the course of the vowel, the F1 shows a steady rise. Conversely, in the Polish item, the F1 maximum is reached quite early in the vowel, after which we observe a large portion with a flat F1 slope.

8   I am grateful to a *Glossa* reviewer to bringing this question to my attention.
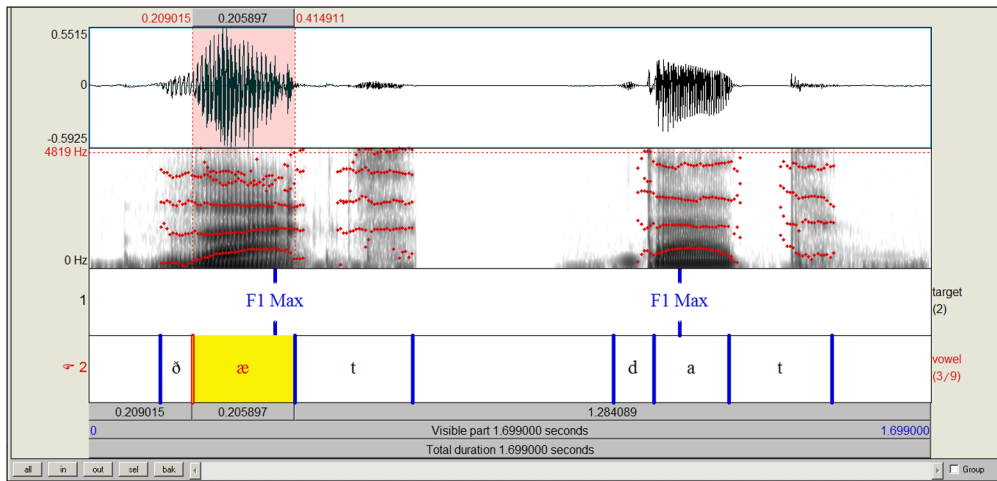
One aspect of the illustration in *Figure 7* requires further comment. In the spectrograms, the effects of the final fortis coda on the vowel are visible. In the English item, we can see pre-glottalization and pre-fortis clipping, which are absent in Polish. Pre-fortis clipping leads to shorter vowels, which of course affected annotation boundaries and formant trajectories in the two studies. This explains the effects by which Polish showed steeper negative F1 slopes in the 3rd interval. It may be suggested that in Polish, the VC transition to the coda is already underway in the 3rd interval, contributing to the F1 drop. Conversely, English pre-fortis clipping truncates the VC formant transitions. As a consequence, the 'target' formant values are housed later in the vowel, and 3rd interval slopes are flatter.

4 WIDER IMPLICATIONS OF VO AFFILIATION

On the whole, the results of the acoustic studies are compatible with the phonological perspective presented in 2.3. However, in order for the phonological predictions to be meaningful, they must be shown to be related to independent aspects of the relevant phonological systems. In this section, we consider the wider implications of the representational parameters shown in (1).

4.1 POLAR OPPOSITIONS IN ENGLISH AND POLISH

VO affiliation, as shown in (1), is predictive of a number of additional phonetic characteristics in Polish and English. These predictions fall out from two claims inherent to the different parses of CV sequences. The first claim is that VO specification in consonants constitutes an encroachment by the consonantal representation into the structural space of the vowel. The second claim is that VO specification in vowels results in a built-in consonantal element (Schwartz 2013a) in the absence of onset consonants.

The first prediction concerns the relative perceptual weight of formant transitions and stop release bursts in the perception of stop place of articulation. Since the CV transition encoded by the VO node is part of the 'consonant' in English, but part of the 'vowel' in Polish, the relative perceptual weight of CV transitions for consonant identification is predicted to be greater in English than in Polish. This hypothesis was tested in experimental studies described in Schwartz & Aperliński (2014) and Aperliński & Schwartz (2015). Those studies found that English listeners performed better than Polish listeners in identifying items with their release bursts removed, forcing listeners to rely on the CV transition. Additionally, for items with conflicting cues (e.g. dorsal burst vs. coronal transition) Polish listeners were less likely to use the CV transition for identification than English listeners.

As a corollary to the relative weight of CV transitions in English as opposed to Polish, we may consider the fact that consonants in English are much more susceptible to lenition than consonants in Polish. If English listeners are more likely to identify consonants on the basis of formant transitions, speakers may spare the articulatory effort required to produce robust release bursts. In Polish, where consonant weakening is generally not attested (Polish casual speech tends to elide consonants, rather than weaken them), listeners apparently attend to aperiodic noise to a greater extent. The enhanced capacity of Polish listeners to rely on noise spectra for consonant identification may also be reflected in the typologically rare sibilant

contrasts in the language, for which noise spectrum is an important cue (Nowak 2006a; Żygis & Padgett 2010).

An additional prediction falling out from (1) is related to the second claim mentioned above, and concerns the behaviour of word-initial vowels. In the Vocalic VO system posited for Polish, the VO node may be seen to act as a built-in consonantal element. That is, in systems with vocalic VO affiliation initial vowels are not, strictly-speaking, onsetless (cf. Schwartz 2013a).[9] Rather, the VO node ensures that vowel-initial syllables in Polish are prosodically well-formed according to a minimality constraint on OP constituents (Schwartz 2013a), given here in (2).

(2)   Minimal Constituent (MC) – A well-formed prosodic constituent contains active nodes both above and below the VT level

Since they satisfy the MC constraint, vowel-initial words in Polish are resistant to *sandhi* linking processes, and 'resyllabification' of C#V sequences is not expected (Rubach & Booij 1990). By contrast, English is well known for processes such as C#V liaison (*find out-fine doubt*), and linking [r] by which word-initial vowels acquire an 'onset', and are no longer truly word-initial.

In Polish, vocalic VO affiliation is associated with a greater likelihood for vowel glottalization, which serves as a '*sandhi*-blocker' to maintain the prosodic integrity of the vowel-initial word. While vowel glottalization is common in English, it is far more frequent in phrase-initial position, where it serves as a boundary marker (Dilley et al. 1996; Garellek 2012).[10] In Polish, by contrast, vowel glottalization is relatively common in phrase-medial position (Schwartz 2013b, Malisz et al. 2013), and can even occur word-medially. In a cross-language study of vowel glottalization, Schwartz (2016b) looked at glottalization rates in the speech of B2-level Polish learners of English both in their L1 and their L2. He found higher rates for L1 Polish than L2 English, despite the fact that L2 speech is sometimes claimed to exhibit a *word integrity constraint* (Cebrian 2000). Inter-language word integrity would have us expect higher rates of glottalization in L2 English, since glottalization preserves word boundaries.

Schwartz (2016a) shows how VO affiliation gave rise to additional phonological oppositions between the two languages. These include phonemic vowel length, and phonological (as opposed to phonetic) vowel reduction, and many of the unusual aspects of Polish phonotactics (Schwartz 2016a: 54–56). While details of the representational mechanisms underlying the Polish-English oppositions described in Schwartz (2016a) are beyond the scope of this paper with its focus on VISC, it is important to note that the oppositions pervade the sound systems of the two languages, from prosodic organization to sub-segmental phonetic details, as summarized in *Table 10*.

4.2 BEYOND ENGLISH AND POLISH

With regard to the presence of VISC and the behaviour of word-initial vowels, other languages do not present such neat polar oppositions as those that are found in Polish and English and presented in *Table 10*. Nevertheless, it will be shown that OP is able to explain these complications, offering something of a typology of VO affiliation and its interaction with place and laryngeal features.

Three additional languages, French, Russian, and German, for which descriptions of VISC and/or the behaviour of vowel-initial words are available, will be considered.[11] We shall see that complexities arise primarily in consonantal VO systems, in which VO-level place and laryngeal features may compete for perceptual primacy. In other words, since CV transitions

---

9   While at first glance, this proposal might seem to be an unwarranted phonological abstraction, a phonetic explanation for the emergence of 'empty onsets' may be suggested. An initial vowel that follows a pause, and a stop-vowel sequence both produce a dramatic and rapid rise in the amplitude envelope of the periodic portion of the acoustic wave. Since stop release bursts are often very low in amplitude, they are often not heard by listeners, so there is a clear auditory link between stop-vowel and silence-vowel sequences, provided there is no significant aspiration or affrication of the stop.

10   In recent years, evidence has emerged to suggest that vowel glottalization is increasing in frequency in English, particularly in ethnically diverse urban areas (Britain & Fox 2008, Davidson & Erker 2014).

11   Eastern Arrernte is an interesting case that will not be discussed due to space restrictions. The basic postulate, as mentioned in Footnote 2, is that VO is contained is incorporated into both consonants and vowels (Schwartz 2013a).

PHONOLOGICAL PROPERTY	ENGLISH	POLISH
Vowel quality and VISC (this paper)	VISC in 1st half of vowel	Flatter formant trajectories in first half of vowel
Linking of initial vowels, and resyllabification of C#V (Cruttenden 2001; Rubach & Booij 1990)	Yes	No
Consonant perception (Nowak 2006a; Schwartz &Aperliński 2014; Aperliński & Schwartz 2015; Walley & Carrell 1983)	Greater weight of CV transitions	Lesser relative weight of CV transitions; greater weight of noise spectra
Intervocalic consonant lenition (Cruttenden 2001; Dukiewicz & Sawicka 1995)	Yes	No
Release of coda stops (Cruttenden 2001; Dukiewicz & Sawicka 1995)	Optional	Obligatory (except for homorganic clusters)
Fronting of vowel in coronal contexts (Ladefoged 1999; Nowak 2006b)	Yes	No (except in context of palatals)
Syllable prominence and vowel reduction (Malisz & Wagner 2012; Nowak 2006b, Rojczyk 2019)	Very strong syllable prominence and phonological reduction (to schwa)	Relatively weak syllable prominence and some phonetic reduction
Sonority violations in onset clusters (many descriptions)	Only sibilant-stop clusters	Largely unrestricted

Table 10 VO-induced oppositions in Polish and English (after Schwartz 2016a).

are strictly speaking part of the vowel, yet provide cues for both primary and secondary place of articulation in consonants, as well as laryngeal features, phonological systems must sort out to what extent these features may coexist on the VO node. In the OP model, the possibility for features to share the VO node relates to the question of the representational level at which a given feature is assigned.

An overview of five languages is provided in (3), which presents OP representations of stop-vowel CV sequences. Consonant place, laryngeal, and vowel place features are indicated in these structures, labelled [place] (and [place2]), [lar], and [V], respectively. These features 'trickle' onto empty nodes below them, and trickling ceases when another feature is assigned at a lower level, as was discussed in 2.3. Note also that these representations show entire CV sequences as single units, rather than individual consonant and vowel structures. In looking at the structure of the entire OP hierarchy for a given language, instead of at individual consonant and vowel 'segments', we have a better view of the consonant-vowel interactions that are crucial the OP accounts of these languages. These structures will be unpacked into individual 'segments' in the discussion that follows.

(3)   The OP hierarchy and feature specifications in five languages
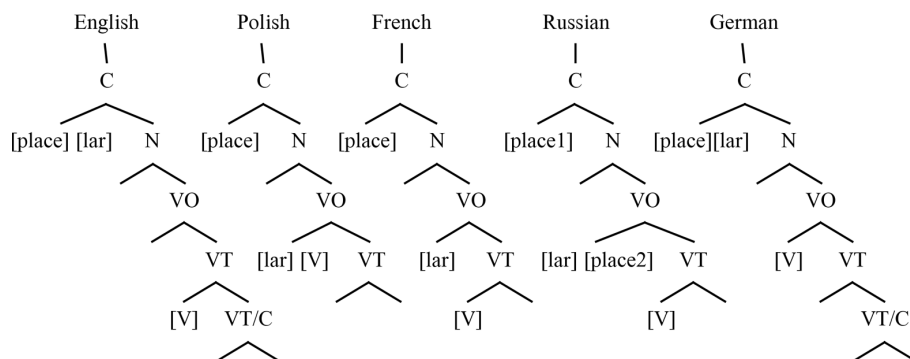

4.2.1 English, Polish and the place of laryngeal features in the OP environment

Before discussing the additional languages, it is necessary to enrich the representations with laryngeal features for English and Polish, as is shown in the two leftmost trees in (3). In English, both place and laryngeal features are assigned to the Closure node, while vowel features are assigned at the VT level. The place and laryngeal specifications in English 'trickle' down the structure to occupy both the Noise node and the VO node, while [V] is assigned at the VT level.

Trickling of the laryngeal feature yields aspiration (see Schwartz 2017). Trickling of the [place] feature, combined with VT-level [V] assignment, results in VISC. VT-level [V] assignment also facilitates linking of initial vowels, which violate the MC constraint in (2). Note also that English has an additional node below VT, capturing the familiar requirement for either a long vowel or coda consonant in stressed monosyllables (see Schwartz 2016a). The structure for English is taken apart into 'segments' in (4).

(4)   OP representations for a stop-vowel sequence in English, extracted into individual 'segments'
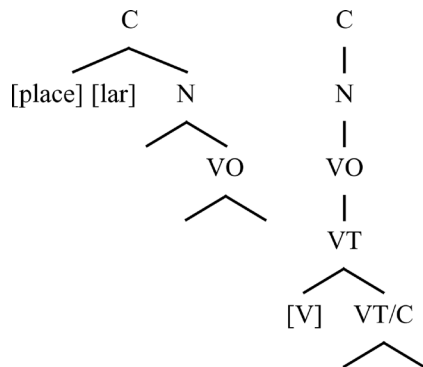

In Polish, VISC is prevented by the presence of non-place features assigned at the VO level – trickling of the Closure-assigned place feature can proceed only as far as Noise. In (1), it was shown that VO-specified vowels blocked trickling of consonant place. VO specification in Polish vowels ensures prosodic well-formedness of vowel-initial words (satisfaction of the MC constraint), and prevents linking processes of the kind observed in English.

With regard to Polish, the behaviour of laryngeal specifications requires some additional discussion. In the OP account of laryngeal phonology (Schwartz 2017), laryngeal features in 'voicing' languages such as Polish are assigned at the VO level. Thus, we see in (3) that the VO level in Polish contains both [lar] and [V] specifications. The former is traditionally associated with consonants while the latter encodes vocalic features. Both features may be assumed to block tricking of consonant place, preventing the development of VISC.

The presence of both [V] and [lar] features on VO in Polish raises an important question for the OP model. How can we reconcile VO-level laryngeal specification with the configuration in (1), in which the VO node was shown as absent from the representation of stops? The representations in (3), which show the OP hierarchy as a single prosodic unit containing both a stop and a vowel, rather than a string of two 'segments', provide perspective on this question. Stated briefly, the 'absence' of the VO node from Polish stops in (1) does not mean that the VO level is absent from the OP hierarchy. Rather, it means that when the entire CV hierarchy is unpacked into individual consonants and vowels, in the representation of the consonant the VO node may be unary. Unary nodes in the OP framework are latent placeholders that may house melodic features, but are not realized on their own (Schwartz 2016a: 43). In (5), the Polish tree from (3) is shown as two individual 'segmental' structures. The laryngeal feature is housed on a unary VO node in the stop (the tree on the left), and is realized when the stop combines with the following vowel (the tree on the right) into a CV unit.

(5)   OP representations for a stop-vowel sequence in Polish, extracted into individual 'segments'
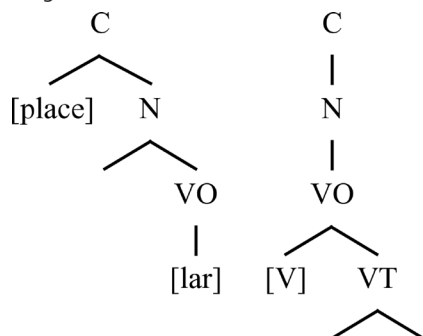
What happens when no vowel follows a Polish obstruent with a unary VO node (i.e. when it is not an 'onset')? The VO node may be lost, resulting in laryngeal neutralization. This is shown in (6).

(6)   OP variants of Polish stops
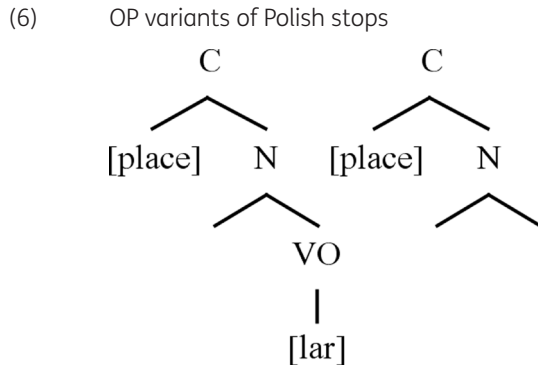
The tree on the left shows a canonical stop representation for Polish, one that appears in 'onsets'. On the right, the VO node and the [lar] specification are absent, yielding a neutralized variant with no laryngeal feature.

The two structures in (6) are instructive in illustrating the OP perspective on laryngeal neutralization (see Wojtkowiak & Schwartz 2018). The idea is that processes like final devoicing and regressive voicing/devoicing are not the result of synchronic 'delinking' or 'spreading' mechanisms. Rather, it is assumed that Polish speakers have both neutralized and non-neutralized variants of obstruents in their mental inventory of speech sounds. In pre-vocalic positions, the variant with unary VO always surfaces, and the laryngeal specification is a property of the entire 'onset' structure (cf. Kehrein & Golston 2004), as is shown in (3). In other positions, either variant may be chosen, and neutralization occurs but is not obligatory. In this connection, it is worth considering phonetic findings concerning laryngeal neutralization in Polish. Strycharczuk (2012) found that in pre-sonorant *sandhi* contexts in Polish, voicing showed a bimodal distribution, suggesting that neutralization was categorical but optional, rather than gradient. The two variants of Polish stops shown in (6) actually predict this finding. Speakers may choose either neutralized or non-neutralized variants to pronounce in non-prevocalic positions. These are categorical options.

Although the interaction between VO specification and laryngeal phonology in Polish appears to be something of a digression in our discussion, it is important in that it illustrates OP relationships between prosodic and segmental units. Stated briefly, the prosodic trees in (3) constitute the building block from which all 'segmental' representations emerge. It is only in the context of the prosodic unit that we can reliably see a laryngeal contrast in Polish. Only in the prosodic unit do we see that the VO node is the focal point for both vowel features and laryngeal specifications, leaving no room on VO for consonantal place features, and preventing the development of VISC in Polish. In English, place and laryngeal features trickle in parallel onto the VO node, and consonant place has more robust effects on the time course of vowel formants.
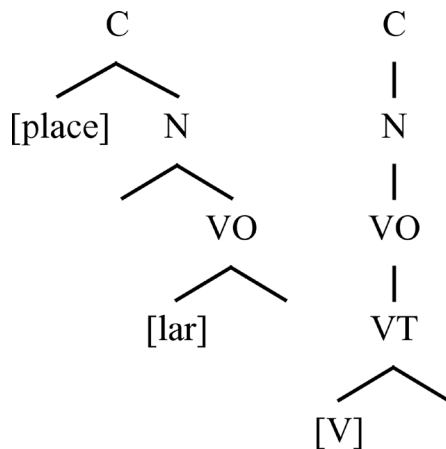
4.2.2 French: Linking but no VISC

Now we turn our attention to French. With regard to VISC, French is similar to Polish, characterized by impressionistically pure vowels.[12] Unlike Polish, however, French is well known for linking processes, such as *enchaînement* and *liaison*, affecting word-initial vowels. The behaviour of vowel-initial words in French suggests that initial vowels are prosodically ill-formed, meaning French initial vowels are not specified at the VO level (except in *h-aspire* words), and do not satisfy the MC constraint in (2). With regard to linking processes, French therefore resembles English.

---

12   French vowels have been shown to be susceptible to co-articulatory effects of neighbouring consonants (Strange et al 2007; Strange et al. 2009). However, these effects do not lead to an increase in formant movement, but rather cause changes in static vowel target locations in F1–F2 space.

In OP, the behaviour of French may be explained as follows. VO-level [lar] specification prevents VISC by blocking trickling of place specifications, while VT-level [V] specification ensures linking processes. The French version of the hierarchy from (3) is taken apart into the 'segmental' sequence in (7).

(7)   OP representations for a stop-vowel sequence in French, extracted into individual 'segments'

```
            C                        C
           / \                       |
     [place]  N                      N
             / \                      |
            VO  VO                    VO
           / \                        |
      [lar]   VT                      VT
             / \
           [V]
```

The main representational difference proposed in (3) between French and Polish is the location of [V] features. In the former they are assigned at the VT level, while in the latter they are found on VO. This difference may in fact be reflected in the perceptual makeup of laryngeal contrasts in the two languages, despite the fact that they both are counted as 'voicing' languages according to a VOT-based typology (Lisker & Abramson 1964). In French, the F1 onset (Stevens & Klatt 1974) is not a heavily weighted perceptual cue to consonant voicing (Serniclaes 1987; Boulakia 1990; Hazan & Boulakia 1993), while in Polish there is evidence that it plays a fairly significant role (Schwartz & Arndt 2018; Schwartz et al. 2019). Since in Polish, VO is shared by [lar] and [V], listeners may put more weight on secondary laryngeal cues. In other words, laryngeal features in Polish are at a perceptual disadvantage by virtue of the fact that they share the VO node with vowel features. As a result, it may be hypothesized that Polish listeners compensate for this by developing greater sensitivity than French listeners to formant-based cues to voicing. In French, the laryngeal feature is the lone occupant of the VO node, and is robustly cued by fundamental frequency (Kirby & Ladd 2016) in addition to VOT. There is little need for listeners to attend to the formant cues.
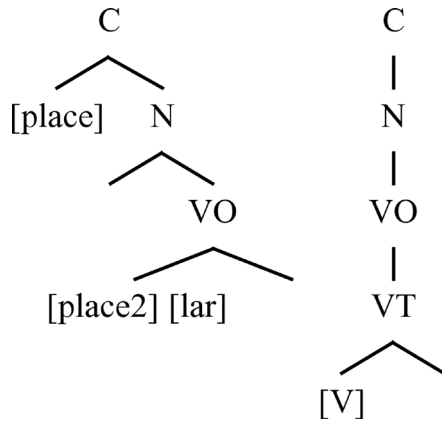
4.2.3. Russian: VISC induced by secondary articulation

Russian, like French, is a voicing language that is known to link vowel-initial words (Knyazev 2006), but unlike French it is characterized by a relatively high degree of VISC (Kuznetsov 2001). Formant movement in Russian vowels may be attributed to secondary articulations on consonants, palatalization and velarization, which are most robustly cued by formant transitions (Kochetov 2001, 2006; Padgett 2001).[13]

In the OP environment, the VO node is the natural docking site for secondary place features. The Russian version of the OP hierarchy from (3) is taken apart into 'segmental' structures in (8). As we can see, both laryngeal features and secondary place features ([place2]) are assigned to VO in Russian. It is the latter that is responsible for the significant degree of formant movement. Thus, like in English, Russian VISC is attributable to a consonantal place specification on VO. The difference is that in English, VO is occupied by a trickled primary place feature, while in Russian it is occupied by an assigned secondary place feature.

---

13   Secondary palatalization is commonly described in Polish as a post-lexical phonetic effect. Polish is therefore clearly distinct from Russian, in which palatalization is phonemic. It is also phonetically distinct, since Russian palatalized consonants are often affricated, while post-lexical palatalized consonants in Polish are not.
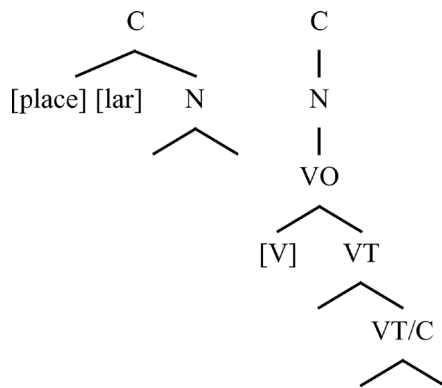
(8)   OP representations for a stop-vowel sequence in Russian, extracted into individual
        'segments'

```
         C                    C
        / \                   |
  [place]   N                 N
           / \                |
          VO                  VO
         /  \                 |
[place2] [lar]                VT
                             / \
                            [V]
```

4.2.4. German: No linking and no VISC in a system with diphthongs

The final language we will examine here is German, shown in (3) to have vocalic VO affiliation, which is reflected in the lack of linking processes and prevalence of word-initial glottalization, or *harter Einsatz* on initial vowels (e.g. Wiese 1996). The German version of the hierarchy from (3) is extracted into individual 'segmental' structures in (9).

(9)   OP representations for a stop-vowel sequence in German, extracted into individual
        'segments'

```
            C                 C
           / \                |
  [place] [lar]  N            N
               / \            |
              VO              VO
             /  \            / \
           [V]    VT
                 / \
                VT/C
               / \
```

In German, like in Polish, glottalization prevents linking and reinforces the prosodic integrity of a VO-specified initial vowel. With regard to VISC, Strange & Bohn (1998) compared formant dynamics in German and American English, and found that German monophthongs are pure in quality relative to English monophthongs. Thus, in both formant movement and the behaviour of word-initial vowels, German appears to be similar to Polish.

At the same time, there are of course many parallels between German and English vowel systems, including length/tenseness contrasts in high vowels and the presence of diphthongs. These parallels reflect a shared prosodic feature – the requirement that monosyllabic words contain either a long vowel or a coda consonant – and are independent of the status of VO in the two systems. This is shown in the structures in (3) in which both German and English show an additional node below the VT level.

4.2.5 Summary

In this section, we have examined how VO specification in the OP framework interacts with other features in the phonologies of various languages to govern two seemingly unrelated properties: the propensity for linking of vowel-initial syllables and the development of vowel inherent spectral change. In the OP system, VISC is a product of consonant place features appearing on the VO node, while linking occurs when initial vowels lack VO specification, which renders them prosodically ill-formed. The effects for each language discussed are summarized in *Table 11*.

LANGUAGE	VISC	LINKING
English	Yes, place features trickle onto VO	Yes, initial vowels lack VO
Polish	No, VO houses [V] and [lar] features	No, initial vowels contain VO
French	No, VO houses [lar] features	Yes, initial vowels lack VO
Russian	Yes, VO houses secondary place	Yes, initial vowels lack VO
German	No, VO houses [V] features	No, initial vowels contain VO

Table 11 Summary of VO-related effects on VISC and Linking.

The representations in (3), which are predictive of these effects in five different languages, offer a promising perspective in which systemic cross-language differences in non-contrastive phonetic properties may be explained. Admittedly, five languages, all of them from the Indo-European family, is hardly a representative sample. Much empirical work remains to be done in documenting cross-language differences in phonetic properties such as VISC. The contribution of OP is that it offers representational tools in which hypotheses for empirical study may be formulated. By contrast, models that treat vowels as single 'segments' in a linear string of phonological units make no predictions about the hypothesis investigated in this paper. Without a phonological perspective on the interactions between vowels and neighbouring consonants, cross-language patterns of formant dynamics can be described as physical phenomena, but not explained as a systemic element of linguistic structure.

5 CONCLUSION

This paper has presented cross-language acoustic comparisons of vowel inherent spectral change in British English and Polish. The results suggest that British English is characterized by a greater degree of formant dynamics, concentrated earlier in the time course of vowels, than Polish. The paper also suggests that explaining the origins of this difference is a task for phonological theory. Onset Prominence representations, in which structural ambiguities may be used to derive language specific differences in certain non-contrastive phonetic details, provide an insightful perspective on this issue. In the Onset Prominence representational environment, VISC is a product of consonant place specifications housed on the Vocalic Onset node of structure, which is ambiguous with regard to the consonant-vowel distinction. It was shown that the cross-language differences in VISC are predictive of other aspects of Polish and English phonology. It was also shown that in additional languages, the interaction between VO and other phonological features may govern a wider set of language-specific phonetic details.

The crucial aspect of the OP framework, which allows the model to form phonological predictions about phonetic details such as VISC, is that the 'segment' is an emergent, rather than primitive entity. This is not to say the segment is absent from the OP model. Rather, the claim is that unlike models based on a segmental string, OP offers a story about how the mapping between segments and the speech signal reflects perceptual ambiguities. We know very well that speakers of different languages differ in their parsing of the speech signal. By enriching phonological representations to encode language-specific differences in parsing and segmentation of the signal, OP provides a truly phonological perspective on the phonetics-phonology relationship, and facilitates the formulation of new hypotheses for experimental phonetic study.

APPENDIX A – WORDLISTS FROM ACOUSTIC EXPERIMENTS

L1 Polish items from both studies

> *bit* 'beat'
> *byt* 'state of being'
> *bet* 'baby's blanket'
> *bat* 'whip'
> *dit* 'type of poetry'
> *dyt* (nonce word)
> *det* (nonce word)
> *dat* 'date' (gen. pl.)
> *git* 'super!'
> *get* (nonce word)
> *gat* (nonce word)

Words analyzed from UCL corpus

> FLEECE: *seat, sheep, seen*
> KIT: *dish, fish, stick,*
> DRESS: *get, said, bed*
> TRAP: *that, man, bad*

L2 items from cross-language study

> Labial onsets: *beat, bit, bet, bat*
> Coronal onsets: *neat, knit, net, stat*
> Dorsal onsets: *skeet, skit, get, scat*

APPENDIX B – ADDITIONAL L1 COMPARISON

To better control for consonantal context in the L1 comparison study, an additional comparison was carried out of the Polish items and L1 productions of the English words from the L1–L2 study, taken from online dictionaries of English with embedded sound files. Unfortunately, the sound quality was adequate for acoustic analysis only in the case of 5 online dictionaries (dictionary. cambridge.org, collinsdictionary.dom, oxfordlearnersdictionaries.com, macmillandictionary. com, howjsay.com). This yielded 60 total items (12 target words * 5 dictionaries). A generalized linear mixed-effects model was run, with absolute values of formant slopes as dependent variable, a Lg*Vowel*Onset interaction term as predictor variable, Vowel duration included as a control variable, and by-speaker/dictionary random slopes and intercepts. The results of this comparison are given in **Table 10**, and summarized graphically in **Figure 8**. The basic patterns observed in the main L1 comparison study may also be found here. Greater F1 movement in English in the 2nd interval was found for all vowel pairs except for FLEECE-i. In this pair, a trend in this direction for 2nd interval F1 slope could be observed for labial and coronal onsets. Greater F1 movement in Polish in the 3rd interval was found in the non-high vowels.

Table 12 Contrast estimates (Eng-Pol) of online dictionary recordings with L1 Polish data.

FORMANT-INTERVAL	VOWEL	ONSET	ESTIMATE	STD. ERROR	t	p
F1–2nd	trap-a	labial	0.857	0.297	2.885	0.004
		coronal	1.516	0.355	4.273	<.001
		dorsal	0.76	0.278	2.737	0.006
	dress-ɛ	labial	0.741	0.282	2.631	0.009
		coronal	1.042	0.33	3.163	0.002
		dorsal	1.482	0.289	5.129	<.001
	kit-ɨ	labial	0.666	0.294	2.264	0.024
		coronal	1.039	0.315	3.301	0.001
	fleece-i	labial	0.56	0.319	1.754	0.08
		coronal	0.54	0.316	1.708	0.088
		dorsal	0.163	0.318	0.514	0.607
F2–2nd	**VOWEL**	**ONSET**	**ESTIMATE**	**STD. ERROR**	**t**	**p**
	trap-a	labial	–0.1	0.302	–0.33	0.742
		coronal	–0.478	0.374	–1.277	0.202
		dorsal	0.227	0.274	0.828	0.408
	dress-ɛ	labial	–0.207	0.283	–0.73	0.466
		coronal	–0.009	0.343	–0.026	0.979
		dorsal	0.654	0.292	2.244	0.025
	kit-ɨ	labial	0.036	0.3	0.119	0.906
		coronal	0.357	0.326	1.095	0.274
	fleece-i	labial	0.491	0.327	1.499	0.134
		coronal	0.078	0.326	0.24	0.811
		dorsal	–0.052	0.326	–0.159	0.874

(Contd.)

| F1–3rd	VOWEL	ONSET	ESTIMATE	STD. ERROR	*t*	*p*
	trap-a	labial	–2.002	0.579	–3.455	0.001
		coronal	–2.975	0.666	–4.47	<.001
		dorsal	–2.27	0.548	–4.141	<.001
	dress-ɛ	labial	–2.752	0.559	–4.928	<.001
		coronal	–3.252	0.627	–5.185	<.001
		dorsal	–1.737	0.568	–3.06	0.002
	kit-ɨ	labial	–0.944	0.577	–1.637	0.102
		coronal	–0.402	0.607	–0.662	0.508
	fleece-i	labial	–0.017	0.608	–0.028	0.978
		coronal	0.08	0.607	0.131	0.895
		dorsal	–0.358	0.607	–0.589	0.556
F2–3rd	**VOWEL**	**ONSET**	**ESTIMATE**	**STD. ERROR**	***t***	***p***
	trap-a	labial	–0.419	0.336	–1.246	0.213
		coronal	–0.183	0.418	–0.438	0.661
		dorsal	0.028	0.304	0.092	0.927
	dress-ɛ	labial	0.12	0.315	0.382	0.703
		coronal	0.05	0.383	0.131	0.896
		dorsal	0.316	0.324	0.975	0.33
	kit-ɨ	labial	0.19	0.334	0.569	0.57
		coronal	–0.286	0.363	–0.787	0.432
	fleece-i	labial	0.916	0.365	2.512	0.012
		coronal	–0.867	0.363	–2.386	0.017
		dorsal	–0.62	0.364	–1.704	0.089
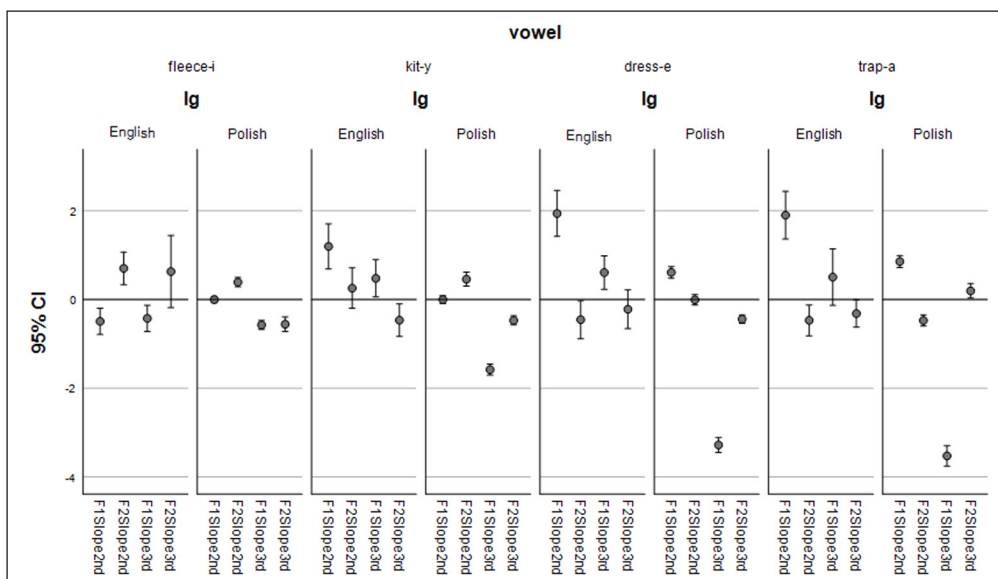
Figure 8 95% confidence intervals of formant slope measures in comparison between L1 Polish data and online dictionary recordings.

APPENDIX C – CODA PLACE EFFECTS FROM L1 COMPARISON STUDY

FORMANT-INTERVAL	VOWEL	CODA CONTRAST	ESTIMATE	STD.ERROR	t	p
F1-2nd	fleece-i	labial-coronal	–0.101	0.16	–0.63	.527
F1-2nd	kit-ɨ	coronal-dorsal	0.114	0.099	1.15	.249
F2-2nd	fleece-i	labial – coronal	–0.461	0.176	–2.62	.009
F2-2nd	kit-ɨ	coronal – dorsal	–0.086	0.108	–0.79	.428
F2-3rd	fleece-i	labial-coronal	0.215	0.282	0.76	.447
F1-3rd	kit-ɨ	coronal – dorsal	–0.412	0.174	–2.37	.018
F1-3rd	fleece-i	labial-coronal	–0.024	0.198	–0.12	.903
F2-3rd	kit-ɨ	coronal – dorsal	0.28	0.122	2.29	.022

Table 13 Effects of coda place on formant slope measures.

Figure 9 95% C.I. error bars for all four formant slope measures in vowel items with non-uniform codas.
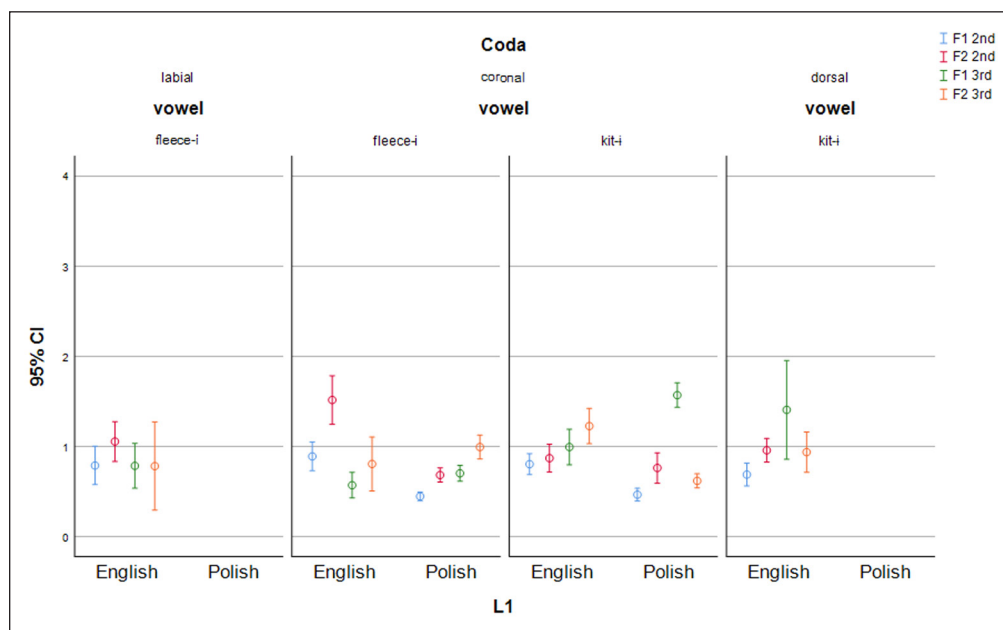
FUNDING INFORMATION

ACKNOWLEDGEMENTS

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATION

Geoffrey Schwartz [ID] *orcid.org/0000-0002-0728-7820*
Faculty of English, Adam Mickiewicz University in Poznań, Ul. Grunwaldzka 6, 60-780 Poznań, Poland

REFERENCES

Aperliński, Grzegorz & Geoffrey Schwartz. 2015. Release bursts vs. formant transitions in Polish stop place perception. In: The Scottish Consortium for ICPhS (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow, online.

Blevins, Juliette. 2004. *Evolutionary Phonology – the Emergence of Sound Patterns*. Cambridge: Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9780511486357*

Boersma, Paul & David Weenink. 2017. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.36. *http://www.praat.org/*

Boulakia, Georges. 1990. Use of spectral cues for initial-stop voicing perception by French-English bilinguals. *Journal of the Acoustical Society of America* 88(S1). DOI: *https://doi.org/10.1121/1.2029175*

Breen, Gavan & Rob Pensalfini. 1999. Arrernte – a language with no syllable onsets. *Linguistic Inquiry* 30(1). 1–25. DOI: *https://doi.org/10.1162/002438999553940*

Britain, David & Sue Fox. 2008. Vernacular universals and the regularization of hiatus resolution. *Essex Research Reports in Linguistics* 57. 1–42.

Browman, Catherine & Louis Goldstein. 1989. Articulatory Gestures as Phonological Units. *Phonology* 6. 201–251. DOI: *https://doi.org/10.1017/S0952675700001019*

Cebrian, Juli. 2000. Transferability and productivity of L1 rules in Catalan-English interlanguage. *Studies in Second Language Acquisition* 22. 1–26. DOI: *https://doi.org/10.1017/S0272263100001017*

Chládková, Kateřina, Silke Hamann, Daniel Williams & Sam Hellmuth. 2016. F2 slope as a Perceptual Cue for the Front-Back Contrast in Standard Southern British English. *Language and Speech*, 1–22. DOI: *https://doi.org/10.1177/0023830916650991*

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Collins, Beverly & Inger Mees. 2009. *Practical Phonetics and Phonology, A Resource Book for Students*. London: Routledge.

Cruttenden, Alan. 2001. *Gimson's Pronunciation of English,* 6th edition. London: Arnold.

Davidson, Lisa & Danny Erker. 2014. Hiatus resolution in American English: the case against glide insertion. *Language* 90(2). 482–514. DOI: *https://doi.org/10.1353/lan.2014.0028*

Delattre, Pierre, Alvin Libermann & Franklin Cooper. 1955. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America* 27(4).769–773. DOI: *https://doi.org/10.1121/1.1908024*

Dilley Laura, Stephanie Shattuck-Hufnagel & M. Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24. 423–44. DOI: *https://doi.org/10.1006/jpho.1996.0023*

Donegan, Patricia & David Stampe. 1979. The study of Natural Phonology. In Daniel A. Dinnsen (ed.), *Current Approaches to Phonological Theory*, 126–173. Bloomington, IN: Indiana University Press.

Draxler, Christoph & Klaus Jänsch. 2015. SpeechRecorder v. 2.X.X. [Software]. Available from: *http://www.bas.uni-muenchen.de/Bas/software/speechrecorder/*

Dukiewicz, Leokadia & Irena Sawicka. 1995. *Gramatyka Współczesnego Języka Polskiego – Fonetyka i Fonologia* [Grammar of Modern Polish – Phonetics and Phonology] PAN: Instytut Języka Polskiego.

Elvin, Jaydene, Daniel Williams & Paola Escudero. 2016. Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *Journal of the Acoustical Society of America* 140(1). 576–581. DOI: *https://doi.org/10.1121/1.4952387*

Fox, Robert & Ewa Jacewicz. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America* 126. 2603–2618. DOI: *https://doi.org/10.1121/1.3212921*

Gafos, Adamantios I. 2002. A Grammar of Gestural Coordination. *Natural Language & Linguistic Theory* 20. 269–337. DOI: *https://doi.org/10.1023/A:1014942312445*

Garellek, Marc. 2012. Word-initial glottalization and voice quality strengthening. *UCLA Working Papers in Phonetics* 111. 92–122.

Grosjean, François. 2004. Studying bilinguals: Methodological and conceptual issues. In Tej Bhatia & William Ritchie (eds.), *The Handbook of Bilingualism,* 32–63. Oxford: Blackwell Publishing. DOI: *https://doi.org/10.1002/9780470756997.ch2*

Hawkins, Sarah & Jonathan Midgley. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35(2). 183–199. DOI: *https://doi.org/10.1017/S0025100305002124*

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.). 2004. *Phonetically Based Phonology*. Cambridge: Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9780511486401*

Hazan, Valerie & Georges Boulakia. 1993. Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech* 36. 17–38. DOI: *https://doi.org/10.1177/002383099303600102*

Hillenbrand, James. 2013. Static and dynamic approaches to vowel perception. In Geoffrey Morrison & Peter Assmann (eds.), *Vowel Inherent Spectral Change*, 9–30. Berlin: Springer. DOI: *https://doi.org/10.1007/978-3-642-14209-3_2*

IBM Corporation. 2013. *IBM SPSS Statistics for Windows, Version 22.0.* Armonk, NY: IBM Corp.

Jenkins, James & Winifred Strange. 1999. Perception of dynamic information for vowels in syllable onsets and offsets. *Perception and Psychophysics* 61. 1200–1210. DOI: *https://doi.org/10.3758/BF03207623*

Jin, Su-Hyun & Chang Liu. 2013. The vowel inherent spectral change of English vowels spoken by native and non-native speakers. *Journal of the Acoustical Society of America* 133(5). 363–369. DOI: *https://doi.org/10.1121/1.4798620*

Kehrein, Wolfgang & Chris Golston. 2004. A prosodic theory of laryngeal contrasts. *Phonology* 21. 325–357. DOI: *https://doi.org/10.1017/S0952675704000302*

Kirby, James & D. Robert Ladd. 2016. Effects of obstruent voicing on vowel F0: evidence from true voicing languages. *Journal of the Acoustical Society of America* 140(1). 2400–2411. DOI: *https://doi.org/10.1121/1.4962445*

Knyazev, Sergej. 2006. *Struktura fonetičeskogo slova v russkom jazyke: Sinkhronija i Diakhronija* (The Structure of Phonetic Words in Russian: Synchrony and Diachrony). Moscow: Maks-Press.

Kochetov, Alexei. 2001. *Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization.* Ph.D. dissertation, University of Toronto.

Kochetov, Alexei. 2006. Testing Licensing by Cue: A case of Russian palatalized coronals. *Phonetica* 63. 113–148. DOI: *https://doi.org/10.1159/000095305*

Kuznetsov, Vladimir. 2001. Spectral dynamics and the classification of Russian vowels. *XI Session of the Russian Acoustical Society*, 439–443.

Ladefoged, Peter. 1999. American English. *Handbook of the International Phonetic Association*, 41–44. Cambridge: Cambridge University Press.

Ladefoged, Peter & Ian Maddison. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35. 1773–1781. DOI: *https://doi.org/10.1121/1.1918816*

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

Malisz, Zofia & Petra Wagner. 2012. Acoustic-phonetic realization of Polish syllable prominence: a corpus study. *Speech and Language Technology* 14/15. 105–114.

Malisz, Zofia, Marzena Żygis & Berndt Pompino-Marschall. 2013. Rhythmic structure effects on glottalisation: A study of different speech styles in Polish and German. *Laboratory Phonology* 4. 119–158. DOI: *https://doi.org/10.1515/lp-2013-0006*

Markham, Duncan & Valerie Hazan. 2002. The UCL Speaker Database. *Speech Hearing and Language: UCL Work in Progress* 14. 1–17.

Morrison, Geoffrey & Peter Assmann. 2013. *Vowel Inherent Spectral Change*. Berlin: Springer. DOI: *https://doi.org/10.1007/978-3-642-14209-3*

Nearey, Terrance. 2013. Vowel inherent spectral change in the vowels of North American English. In Geoffrey Morrison & Peter Assmann, (eds.), *Vowel Inherent Spectral Change*, 49–85. Berlin: Springer. DOI: *https://doi.org/10.1007/978-3-642-14209-3_4*

Nearey, Terrance & Peter Assmann. 1986. Modeling the role of vowel inherent spectral change in vowel identification. *Journal of the Acoustical Society of America* 80. 1297–1308. DOI: *https://doi.org/10.1121/1.394433*

Nowak, Paweł. 2006a. The role of vowel transitions and noise in the perception of Polish sibilants. *Journal of Phonetics* 34. 139–152. DOI: *https://doi.org/10.1016/j.wocn.2005.03.001*

Nowak, Paweł. 2006b. Vowel reduction in Polish. Ph.D. dissertation. University of California at Berkeley.

Ohala, John. 1981. The listener as a source of sound change. In C. S. Masek et al. (eds.), *Papers from the Parasession on Language and Behavior*, 178-203. Chicago: Chicago Linguistic Society.

Padgett, Jaye. 2001. Contrast dispersion and Russian palatalization. In Elisabeth Hume & Keith Johnson (eds.), *The role of speech perception in Phonology*, 187–218. San Diego: Academic Press.

Peterson, Gordon & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32(6). 693–703. DOI: *https://doi.org/10.1121/1.1908183*

Rogers, Catherine, Merete Glasbrenner, Teresa DeMasi & Michelle Bianchi. 2013. Vowel inherent spectral change and the second language learner. In Geoffrey Morrison & Peter Assmann (eds.), *Vowel Inherent Spectral Change*, 231–259. Berlin: Springer. DOI: *https://doi.org/10.1007/978-3-642-14209-3_10*

Rojczyk, Arkadiusz. 2019. Quality and duration of unstressed vowels in Polish. *Lingua* 217. 80–89. DOI: *https://doi.org/10.1016/j.lingua.2018.10.012*

Rubach, Jerzy & Geert Booij. 1990. Syllable structure assignment in Polish. *Phonology* 7. 121–158. DOI: *https://doi.org/10.1017/S0952675700001135*

Schwartz, Geoffrey. 2010. Phonology in the speech signal – unifying cue and prosodic licensing. *Poznań Studies in Contemporary Linguistics* 46(4). 499–518. DOI: *https://doi.org/10.2478/v10010-010-0025-3*

Schwartz, Geoffrey. 2013a. A representational parameter for onsetless syllables. *Journal of Linguistics* 49(3). 613–646. DOI: *https://doi.org/10.1017/S0022226712000436*

Schwartz, Geoffrey. 2013b. Vowel hiatus at Polish word boundaries – phonetic realization and phonological implications. *Poznań Studies in Contemporary Linguistics* 49. 557–585. DOI: *https://doi.org/10.1515/psicl-2013-0021*

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

Schwartz, Geoffrey. 2016b. Word boundaries in L2 speech – evidence from Polish learners of English. *Second Language Research* 32. 397–426. DOI: *https://doi.org/10.1177/0267658316634423*

Schwartz, Geoffrey. 2017. Formalizing modulation and the emergence of phonological heads. *Glossa: a journal of general linguistics* 2(1). 81. DOI: *https://doi.org/10.5334/gjgl.465*

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

Schwartz, Geoffrey & Grzegorz Aperliński. 2014. The phonology of CV transitions. In Jolanta Szpyra-Kozłowska & Eugeniusz Cyran (eds.), *Crossing Phonetics-Phonology lines*, 277–298. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Schwartz, Geoffrey, Grzegorz Aperliński, Mateusz Jekiel, and Kamil Malarski. 2016. Spectral Dynamics in L1 and L2 Vowel Perception. Research in Language 14(1). 61–77. DOI: *https://doi.org/10.1515/rela-2016-0004*

Schwartz, Geoffrey & Kamil Kaźmierski. 2020. Vowel dynamics in the acquisition of L2 English – an acoustic study of Polish learners. *Language Acquisition – a Journal of Developmental Linguistics* 27(3). 227–254. DOI: *https://doi.org/10.1080/10489223.2019.1707204*

Schwartz, Geoffrey, Ewelina Wojtkowiak & Bartosz Brzoza. 2019. Beyond VOT in the Polish laryngeal contrast. *Proceedings of ICPhS,* Melbourne.

Serniclaes, Willy. 1987. *Etude experimentale de la perception du trait du voisement des occlusive du francais*. Ph.D. dissertation, Free University of Brussels.

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

Sobkowiak, Włodimierz. 2008. *English Phonetics for Poles,* 3rd Edition. Poznań: Wydawnictwo Poznańskie.

Steriade, Donca. 1993. Closure, release and nasal contours. Huffman, Marie & Loren Trigo (eds.), *Phonetics and Phonology 5: Nasals, nasalization and the velum*, 401–470. DOI: *https://doi.org/10.1016/B978-0-12-360380-7.50018-1*

Stevens, Kenneth. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Stevens, Kenneth N., & Arthur House. 1963. Perturbation of vowel articulations by consonant context. *Journal of the Acoustical Society of America* 85. 2135–2153.

Stevens, Kenneth N., & Blumstein, Sheila E. 1981. The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (eds.), *Perspectives on the study of speech*, 1–38. Hillsdale, NJ: Erlbaum.

Stevens, Kenneth N., & Dennis Klatt. 1974. Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America* 55. 653–659. DOI: *https://doi.org/10.1121/1.1914578*

Strange, Winifred. 1989. Evolving theories of vowel perception. *Journal of the Acoustical Society of America* 85. 2081–2087. DOI: *https://doi.org/10.1121/1.397860*

Strange Winifred, James Jenkins & Thomas Johnson. 1983. Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America* 34. 695–705. DOI: *https://doi.org/10.1121/1.389855*

Strange, Winifred & Ocke-Schwen Bohn. 1998. Dynamic specification of coarticulated German vowels: perceptual and acoustical studies. *Journal of the Acoustical Society of America* 104. 488–504. DOI: *https://doi.org/10.1121/1.423299*

Strange, W., A. Weber, E. Levy, V. Shafiro, M. Hisagi & K. Nishi. 2007. Acoustic variability within and across German, French, and American English vowels: phonetic context effects. *Journal of the Acoustical Society of America* 122. 1111–1129. DOI: *https://doi.org/10.1121/1.2749716*

Strange, W., E. Levy & F. Law. 2009. Cross-language categorization of French and German vowels by naïve American listeners. *Journal of the Acoustical Society of America* 126. 1461–1476. DOI: *https://doi.org/10.1121/1.3179666*

Strycharczuk, Patrycja. 2012. *Phonetics-phonology interactions in pre-sonorant voicing*. PhD dissertation, University of Manchester.

Syrdal, Anne & H. Gopal. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79(4). 1086–1100. DOI: *https://doi.org/10.1121/1.393381*

Tabain, Marija, Gavan Breen & Andrew Butcher. 2004. VC vs. CV syllables: A comparison of Aboriginal languages with English. *Journal of the International Phonetic Association* 34. 175–200. DOI: *https://doi.org/10.1017/S0025100304001719*

Walley, A. C. & T. D. Carrell. 1983. Onset spectral and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America* 73. 1011–1022. DOI: *https://doi.org/10.1121/1.389149*

Wells, John. 1982. *Accents of English: an introduction*. Cambridge: Cambridge University Press.

Wiese, Richard. 1996. *The phonology of German*. Oxford: Oxford University Press.

Williams, Daniel & Paola Escudero. 2014. A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *Journal of the Acoustical Society of America* 136(5). 2751–2761. DOI: *https://doi.org/10.1121/1.4896471*

Williams Daniel, Jan-Willem van Leussen & Paola Escudero. 2015. Beyond North American English: modelling vowel inherent spectral change in British English and Dutch. In: The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow, online.

Wojtkowiak, Ewelina & Geoffrey Schwartz. 2018. Sandhi-voicing in dialectal Polish: Prosodic implications. *Studies in Polish Linguistics* 13(2). 123–143. DOI: *https://doi.org/10.4467/23005920SPL.18.006.8745*

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically Based Phonology*, 34–57. Cambridge: Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9780511486401.002*

Żygis, Marzena & Jaye Padgett. 2010. A perceptual study of Polish sibilants, and its implications for historical sound change. *Journal of Phonetics* 38(2). 207–22. DOI: *https://doi.org/10.1016/j.wocn.2009.10.003*