

Refining representations for L2 phonology

Second Language Research

2020, Vol. 36(4) 691–707

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0267658319852383

journals.sagepub.com/home/slr**Geoffrey Schwartz** 

Adam Mickiewicz University, Poznan, Poland

Abstract

This article discusses the implications of phonological representation for the study of L2 speech acquisition. It is argued, on the basis of empirical findings from diverse phenomena in L2 phonology, that refined representations in which ‘segments’ have internal prosodic structure offer a more insightful view of cross-linguistic phonetic interaction than traditional phonological models. These refinements may be implemented in the Onset Prominence representational environment, in which diverse structural parses affect sub-segmental phonetic properties, transitions between segments, and the formation of prosodic boundaries.

Keywords

L2 speech, Onset Prominence, phonetics and phonology, phonological representation

1 Introduction

For a majority of studies on second language (L2) phonological acquisition, it is segments and segmental contrasts that constitute the linguistic focus of experimental work. Researchers have investigated the effects of cross-language differences in segmental inventories, as well as the realization of phonological categories that are presumed to be similar across languages. Many studies are devoted to the acquisition of new L2 vowel contrasts by speakers whose first language (L1) lacks the opposition in question. For example, Escudero and Boersma (2004) is a frequently cited study examining Spanish L1 speakers’ perception of the English /i:/–/ɪ/ contrast (keywords FLEECE–KIT; Wells, 1982). Another popular area is the implementation of laryngeal contrasts, measured in terms of the phonetic parameter of voice onset time (VOT; Lisker and Abramson, 1964). Studies typically compare VOT of stops in speakers whose L1 and L2 differ in this parameter, such as French and American English (Flege, 1987) or Greek and Australian English (Antoniou et al., 2010).

Corresponding author:

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

Email: geoff@wa.amu.edu.pl

Research into the perception and production of L2 segments has been instrumental in the formation of the more influential theoretical models of L2 speech acquisition. Flege's Speech Learning Model (SLM; see Flege, 1995), Best's Perceptual Assimilation Model (PAM; Best, 1995; Best and Tyler, 2007), and Escudero's (2005) Second Language Linguistic Perception (L2LP) model facilitate the formulation of important predictions concerning learners' success or lack of success in the acquisition of L2 phonology. For example, the SLM maintains that sufficient acoustic distance between an L2 sound and its nearest L1 neighbor is conducive to the formation of a new perceptual category, paving the way to successful acquisition. The PAM and L2LP models focus on L2 contrasts, predicting that when an L2 contrast is phonetically similar to an L1 contrast, learners will be successful in its discrimination. While the PAM and SLM may be seen as strictly phonetic models in that their attention is concentrated on L2 sounds and the contrasts between them, the L2LP offers an enhanced environment in which other factors such as lexical competition and meaning may influence the acquisition of L2 speech.

In each of the models mentioned above, the notion of 'similarity' is a crucial concept that constitutes the basis for experimental hypotheses. Difficulties in acquisition of L2 segments and contrasts are predicted on the basis of how similar they are to those found in the L1. Interestingly, similarity has been seen as both a help and a hindrance. On the one hand, the presence of parallel contrasts in L1 and L2 has been found to facilitate discrimination of the L2 contrast (Best and Tyler, 2007). On the other hand, 'equivalence classification' (Flege, 1987) between corresponding L1 and L2 sounds may be an obstacle to the acquisition of phonetically faithful L2 items. When an L1 and L2 sound are classified as the same, learners often miss subtle acoustic differences between them.

A lingering issue concerns how similarity is to be defined. Although research has looked at similarity from a range of perspectives (Bohn, 2002; Fleischhacker, 2005; Mielke, 2012), a number of contentious points remain unsolved. One such problem is whether cross-linguistic similarity should be defined in terms of phonetics or phonology. For example, Chang (2015) provides a literature review that brings together findings to suggest that phonological similarity 'wins out' over phonetic similarity. In one case he discusses, French /y/ is closer acoustically (in its second formant frequency) to English /i:/ than /u/, yet its production by L1 English learners of French suggests equivalence classification with English /u:/ (Flege, 1987). Chang claims that this is due to the phonology of the two vowel systems rather than acoustic details of vowel production.¹ In the systems of French and English, the segments /u:/ and /u/ are said to correspond with each other at an abstract phonemic level, constituting the basis of cross-language similarity for users of both languages.

By making the phonetics–phonology comparison, Chang accepts a perspective on the phonetics–phonology relationship in which the segment is a crucial element. However, the role of the segment in phonology, and the phonetics–phonology relationship in particular, has long been questioned. In a variety of theories, ranging from traditional generative phonology (Chomsky and Halle, 1968), to phonetically-oriented models such as Articulatory Phonology (Browman and Goldstein, 1989) and Exemplar Theory (Bybee, 2001; Johnson, 1997), as well as more abstract representational theories based on syllables or other prosodic constituents, the status of segments as universal primitives in phonological systems is secondary. In Chomsky and Halle (1968), for example,

segmental symbols are simply shorthand for matrices of binary features. In Articulatory Phonology, the units of interest are articulatory gestures, and segments are merely clusters of gestures with a given temporal organization. In Exemplar Theory, segments emerge from clouds of input exemplars of entire words and phrases: they are not primitive. Likewise, in Government Phonology (e.g. Kaye et al., 1990), prosodic positions like onset and nucleus, rather than phonemes, are the main building blocks of phonological structure.

In such a theoretical environment, it appears that phonological segments are at once negated and taken for granted. That is, segments are widely accepted as the appropriate descriptive units for defining cross-language phonological similarity, yet there is no theoretical consensus on the mechanisms involved in their emergence, or their representation in relation to both smaller and larger entities. This article will argue that these issues have major implications for our understanding of L2 phonological acquisition, particularly with regard to the question of cross-language phonological similarity.

Over the course of the arguments to be laid out in what follows, I will summarize empirical findings for which traditional representations based on the segment fall short of explanatory adequacy. I discuss diverse aspects of L2 speech, including word boundaries, vowel quality, and laryngeal contrasts. In each case, it will be shown that a perspective in which ‘segments’ emerge with internal structure (Inkelas and Shih, 2017; Schwartz, 2010; Steriade, 1993) offers an insightful view of cross-language phonetic differences that determine the nature of cross-linguistic interaction (CLI). Crucially, these differences arise from ambiguities in the mapping between phonological representations and the perceived string of phonological segments. Before addressing these empirical areas, it is necessary to provide a sketch of a representational system that encodes these ambiguities: the Onset Prominence representational framework (OP; for the most comprehensive introductions, see Schwartz, 2016a, 2017).

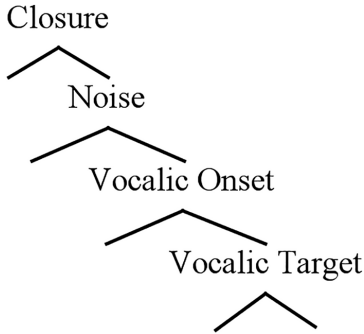
II A new representational perspective for L2 speech

To paint a picture of how refined phonological representations can provide a useful window for looking at L2 speech acquisition, it is necessary to present some basic considerations relating to how various categories of speech sounds relate to the acoustic signal, and the perception of the signal as a string of phonological units. The speech signal is inherently ambiguous with respect to these issues (Hockett, 1955; Liberman et al., 1967; Ohala, 1981). The primary goal of the OP framework is to represent these ambiguities as phonological parameters by which languages may differ, in order to facilitate the formulation of hypotheses concerning CLI in the course of L2 acquisition and other language contact situations.

Onset Prominence representations are built from a hierarchy of phonetic events associated with a stop-vowel CV sequence, shown in (1). The stop-vowel CV, the most commonly encountered ‘syllable’ type across languages, enjoys privileged status in the OP model. It is the only universal primitive in the framework, the building block from which all other representations derive. Each layer in (1) is derived from a specific phonetic entity associated with the realization of the stop-vowel sequence, which is treated as a single unit. The top node (Closure; C) encodes the closure phase of stop consonants. The

Noise node (N) encodes the release bursts/aspiration of stops, as well as frication in fricatives. The Vocalic Onset (VO) node reflects periodicity and formant structure associated with approximant consonants, as well as CV formant transitions (cf. A_{\max} in Aperture Theory; Steriade, 1993). At the bottom of the hierarchy is the Vocalic Target (VT) node, which houses formant frequencies that define vowel quality.

(1) The OP hierarchy



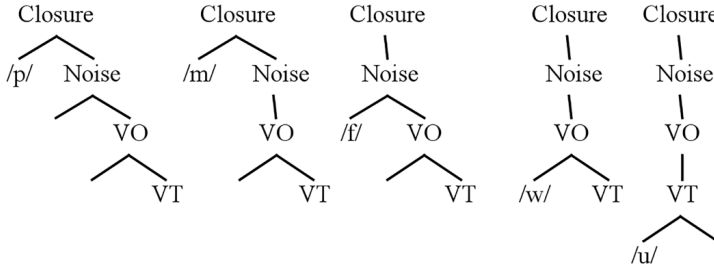
A crucial aspect of the CV primitive is its hierarchical structure. For the OP model, a CV does not represent a linear string. A linear string may be extracted from the OP hierarchy, and I assume that it is on a language specific basis (see below). However, we shall see that in the parsing of the OP hierarchy into individual consonants and vowels, there are ambiguities that play a crucial role in defining L1–L2 phonological relationships. The primary motivation of the hierarchical rather than linear arrangement of OP representations is to encode causal relations between articulation and acoustics. Assuming for example that the place of articulation of a stop is aligned with its Closure node, the hierarchical structure ensures that the acoustic effects of this place specification will also be present in the noise spectrum at closure release (the Noise node) and in CV formant transitions at vowel onset (the VO node). In a linear configuration this causal relationship is not inherent: some sort of spreading process would need to be stipulated in order to express the articulatory-acoustic link.² The need for hierarchical rather than linear representations will become apparent in the application of the OP model to issues such as vowel inherent spectral change (Section 4) and laryngeal phonology (Section 5).

‘Segments’ in the OP environment are extracted from the structure in (1), encoding manner of articulation in terms of the binary nodes present in an individual representation. In other words, each ‘segment’ is its own tree structure that includes parts of the basic hierarchy from (1). This is shown in (2). Binary nodes correspond to entities that are present in the phonetic realization of a given manner category, while unary nodes indicate that a given phonetic event is absent. Fricatives – the central tree in (2) – lack full closure, so their Closure node is unary. Nasals lack noise bursts, so their Noise node is unary. Approximants have unary Closure and Noise nodes, since they lack both these properties.

The segmental symbols in (2) are simply shorthand for place and laryngeal specifications. These specifications are assumed to also occupy lower-level nodes in a given

structure through a process of ‘trickling’ (Schwartz, 2016a: 45), which reflects the fact that the acoustic consequences of a given articulation (e.g. constriction location) are audible in the noise spectra of release bursts and in CV formant transitions. Trickling is not shown in (2) in the interest of clarity.

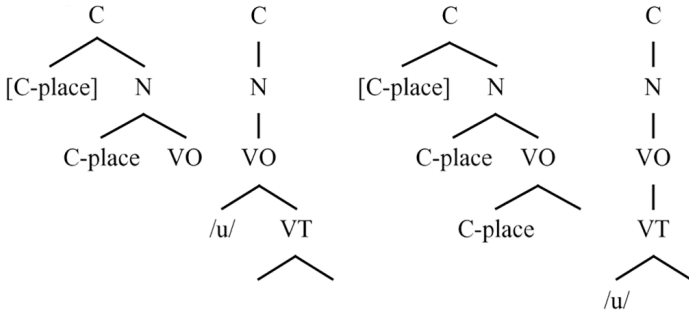
(2) OP manner categories



The OP representational system is characterized by ambiguities in the parsing mechanisms that can derive segments from the CV hierarchy shown in (1). The most important such ambiguity for the purposes of our discussion is the status of the VO node. Acoustically this node reflects a portion of the signal that is vocalic, with periodicity and formant structure. However, CV transitions are also crucial for the perception of consonants (Wright, 2004). Thus, there are (at least) two ways in which a phonological system may parse the OP hierarchy into a consonant–vowel sequence.

This is shown in (3), in which we see two different possibilities for representing a stop–vowel sequence. In these structures, the [C-Place] annotations reflect a generic place of articulation specification for consonants (e.g. [Labial], [Coronal], [Dorsal]), which are assigned at the Closure node and trickle down the structure. The /u/ symbol is shorthand for features specifying vowel quality. In the leftmost pair of trees in (3) the VO node is part of the vowel’s representation, while in the rightmost structures the VO is built into the representation of the consonant. The trickling of the consonant’s place specification is shown in the representations of the stop. Place features occupy the VO node in the system on the right, but not in the system on the left, where it is blocked by the place specification associated with the vowel.

(3) Parses of the OP hierarchy according to VO affiliation



The parses shown in (3) produce phonological parameters with consequences for a number of diverse phonological and phonetic phenomena. These include the prosodic behavior of vowel-initial syllables with respect to linking processes at word boundaries, the relative susceptibility of vowels to diphthongization, the relative weight of formant transitions vs. noise spectra in consonant place perception, and the susceptibility of consonants to lenition (Schwartz, 2016a). In what follows, we will examine implications of these representational parameters for selected empirical areas of L2 speech research.

III Word boundaries in L2 speech

Although word boundaries have not received too much attention in L2 speech research (see Zsiga, 2011), existing studies have produced some hypotheses about their behavior in second language speech. Cebrian (2000) studied the speech of L1 Catalan learners of English, and observed an asymmetry with regard to the phonetics of word boundaries. Catalan learners transferred L1 final obstruent devoicing into L2, but did not transfer a regressive voicing assimilation process that takes place across word boundaries (*vas petit* ‘small glass’ vs. *vaz gran* ‘large glass’). In other words, Cebrian found that a word-internal process (final devoicing) was a source of interference, while a process spanning a boundary (regressive voicing) was not. To explain this asymmetry, Cebrian (2000: 19) proposed ‘an inter-language prosodic constraint that restricts the application of rules to the level of the phonological word’.

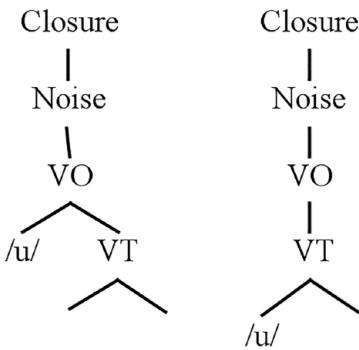
Lleo and Vogel (2004) investigated the acquisition of L2 German by L1 Spanish speakers. Their study included two phonological processes with implications for Cebrian’s proposal of L2 word integrity. One was a *sandhi* spirantization process, which takes place between vowels across word boundaries in L1 Spanish (*cuatro* [χ]atos ‘four cats’), while the other was *harter Einsatz*, or glottalization of initial vowels, which is common in L1 German (Wiese, 1996). Lleo and Vogel investigated the extent to which Spanish learners of German suppressed the L1 spirantization and produced L2 glottalization, and thereby produced correct realizations of initial stops and vowels in German. They found that Spanish learners suppressed the spirantization process at a high rate (around 80%), yet their acquisition of *harter Einsatz* was much less successful (under 50%). This finding casts doubt on Cebrian’s hypothesis. Inter-language word integrity would have us expect successful acquisition of glottalization, since it provides a natural auditory reinforcement of the boundary.

To interpret their results, Lleo and Vogel propose a typology of phonological behavior of segments at word boundaries. In so-called ‘demarcating’ languages, which is how they characterize German, we may observe processes that preserve or reinforce word boundaries (or even syllable boundaries). *Harter Einsatz* is a classic example of this type of boundary reinforcement. Polish is another example of a demarcating language, in which vowel glottalization is quite common (Schwartz, 2016b). In ‘grouping’ languages such as Spanish, on the other hand, we typically observe processes that weaken segments at word boundaries. With regard to word-initial vowels, grouping languages show a tendency to link them with the preceding word, as is observed in French *liaison* and *enchaînement*. English is also known for this type of linking process, including intrusive

/r/, glide-like realization of V#V sequences, and resyllabification of C#V sequences (*find out ~ fine doubt*). English therefore would also be classified as a grouping language.

For all intents and purposes, the grouping–demarcating distinction suggests that the prosodic structure of a ‘word’ is different in different languages. What remains unexplained, however, is why languages differ in this regard. In the OP framework, the ‘grouping’ vs. ‘demarcating’ distinction falls out from the phonological parameters presented in (3). This is illustrated here with respect to the behavior of word-initial vowels. The vowel structures derived from the representations in (3) are shown again in (4), which shows a VO-specified vowel alongside a vowel from a system in which the VO node is claimed by the consonant representations.

(4) Vowels with and without VO specification



In demarcating languages, vowels are specified with the VO node, and vowel-initial syllables are well-formed entities that can stand alone as prosodic constituents. Prosodic well-formedness in the OP model is described in terms of a Minimal Constituent constraint (Schwartz, 2013) in (5).

(5) Minimal Constituent (MC): A minimal prosodic constituent must contain a binary VT node and at least one additional binary higher-level node. (Schwartz, 2013)

When an initial vowel satisfies the MC constraint, there is no phonological impetus for the vowel to be linked with the preceding word.³ That is, a well-formed initial vowel constitutes a built-in prosodic boundary, which is frequently reinforced by glottalization. In other words, glottalization blocks *sandhi* linking processes.

From the representations in (4), we see that vowel glottalization may take one of two phonological forms. It is either the preservation of an element that is present in the input representation, or the insertion of a feature that is absent from the input. The first option clearly obtains in demarcating languages like German: vowel glottalization is a way of preserving the prosodic integrity of a vowel specified with the VO node. In other words, a boundary is built into the representation of vowel-initial lexical items in German, so *harter Einsatz* does not insert anything. It merely preserves the boundary, preventing linking with the preceding word. This *sandhi*-blocking effect of glottalization is also common in Czech

(Palková, 1997) and Polish (Schwartz, 2016b). By contrast, the acquisition of German vowel glottalization by L1 Spanish speakers entails the second option: learners must add a boundary. In traditional phonological models, phonemic contrast determines the nature of input representations. Since glottal stops are not phonemic in these languages, the mainstream view must treat all cases of glottalization as segmental ‘insertions’. This strategy, unfortunately, would require us to stipulate the ‘grouping vs. demarcating’ typology. In OP, by contrast, the typology falls out directly from the representations.

While at first glance this may seem like a minor issue dealing with an allophonic process in an L2, it may have farther reaching consequences for other aspects of L2 pronunciation. Consider final obstruent devoicing, which is a well-known source of interference for L2 learners, and a salient feature of Polish-accented English (Gonet and Pietroń, 2004; Scheuer, 2002). At the same time, however, final devoicing is a universal phonetic tendency that is also frequently observed in English (e.g. Collins and Mees, 2008), despite the fact that the lenis–fortis contrast is maintained. One strategy that L1 English speakers use for maintaining the contrast in C#V contexts is linking the final consonant to the following initial vowel (e.g. Giegerich, 1992). That is, final consonants in English are often resyllabified to become onsets: they are no longer ‘final’. In a study of Polish learners of English, Schwartz et al. (2014) found a robust correlation between students’ productions of the final voice contrast in C#V contexts and their suppression of vowel glottalization, which is common in their L1 Polish. Vowel glottalization reinforces the ‘final’ position of the preceding consonant, preserving the context for devoicing. Successful acquisition involves suppressing the L1 glottalization. Suppression of L1-induced glottalization facilitates C#V linking and the accurate pronunciation of ‘final’ voiced obstruents in L2 English.

This connection between vowel glottalization and final devoicing is predicted in the OP framework, since there is a direct representational link between the prosodic context and the segmental phonetics. The initial vowel in L1 English is linked because it is not specified with the VO node. Interference for Polish learners of English consists of the substitution of a VO-specified initial vowel found in L1. For mainstream models based on the segment, the prosodic conditioning of these processes must be stipulated. In the traditional view, linking and final devoicing would enter into an opaque relationship: linking eliminates the context for final devoicing.

IV Vowel dynamics in L2 speech

It has been observed many times that vowels that are spelled or transcribed with the same symbol in different languages are frequently quite distinct in their phonetic quality (see e.g. Bohn, 2017). These differences are often a function of formant trajectories, indicating changes in vowel quality over the time course of a vowel. The dynamic properties of vowel quality have been described in a number of varieties of English, unified under the heading vowel inherent spectral change (VISC; Morrison and Assmann, 2013; Nearey and Assmann, 1986). In what follows we shall consider the phonological origins of VISC in English, its representation, as well as implications for L2 learners.

One source of VISC is co-articulation with neighboring consonants. Awareness of the relationship between vowel formants and neighboring consonants goes back at least as

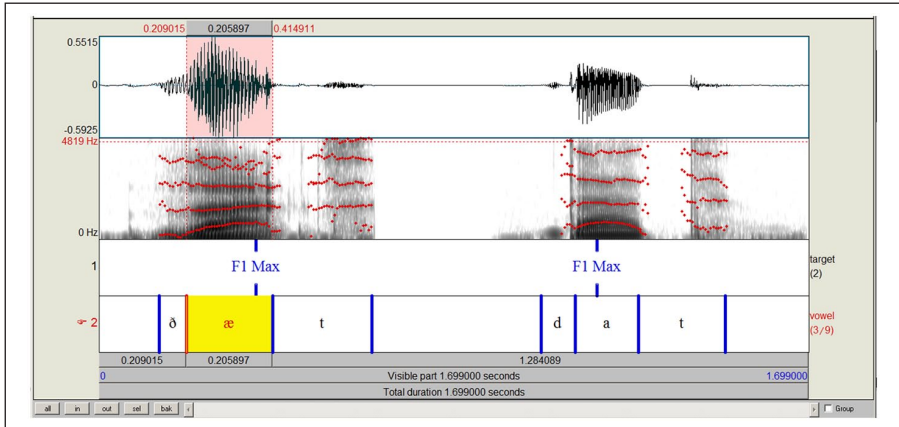


Figure 1. Waveform/spectrogram display of *that*, produced by an L1 British English speaker, and *dat* ‘date’ (genitive plural), produced by an L1 Polish speaker.

far as a study by Stevens and House (1963), who observed that in CVC contexts in American English, vowel formants differ from those produced in isolation, a phenomenon they described as ‘target undershoot’. Later, Hillenbrand et al. (2001) observed that despite significant effects of neighboring consonants on vowel formant trajectories, English vowels are classified most successfully on the basis of formant samples from two points in a vowel’s duration, allowing for a characterization of formant movement. In vowel perception, the role of formant dynamics was studied in an experimental paradigm in which naturally produced stimuli were altered by silencing various parts of a vowel’s duration. In one such stimulus condition, referred to as Silent Center (SC; Strange et al., 1983), the quasi-steady-state portion of the vowel is silenced, leaving listeners to identify vowels on the basis of CV and VC transitions. A consistent finding in these experiments was that the SC tokens were identified more accurately than items containing the central portion of the vowel (Jenkins and Strange, 1999; Strange, 1989) encompassing its putative ‘target’. These findings suggest a significant role for formant movement in English vowel perception, in which CV and VC transitions are important cues to vowel identity. Crucially, it appears that L1 English listeners delay their identification decisions until late in the vowel, after they have computed the course of formant movement.

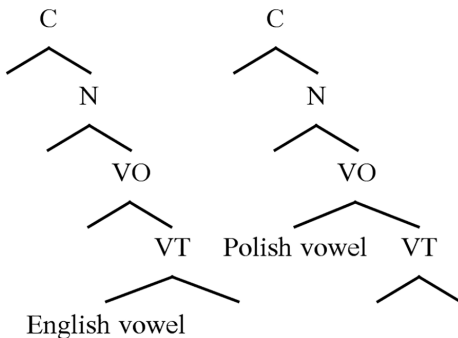
At the same time, there is no reason to assume that formant dynamics play the same role in all other languages that they do in English, and indeed cross-linguistic studies have found that VISIC is less prevalent in both Dutch (Williams et al., 2015) and Polish (Schwartz et al., 2016) than it is in English. These differences create a testing ground for L2 speech research. For a visualization of the challenge facing a speaker of an L1 with relatively steady vowel quality, consider Figure 1, which 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 (genitive plural)’. Of particular interest is the trajectory of the first formant (F1). 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.

The differences visible in these spectrograms suggest an explanation for well-documented difficulties in the acquisition of English vowels by L1 Polish speakers. The TRAP vowel represented in this spectrogram is often confused with Polish /ɛ/ (Sobkowiak, 2008). If one takes recordings of British English TRAP in a CV context, and listens only to the first 20%–30% of the vowel, the auditory impression resembles /ɛ/, since the F1 transition from the onset consonant to the F1 maximum is not yet complete. It therefore appears that Polish listeners’ confusion with this vowel may be a function of which portion of the vowel they are attending to for perception. Poles apparently base their identification on formant cues housed earlier in the vowel, while English listeners delay perceptual decisions.

The distinct parses of the VO node shown in the representations in (3) provide a natural expression of the cross-linguistic difference shown in the spectrograms in Figure 1. This is shown in the structures in (6), in which we see side by side versions of the OP hierarchy annotated with proposed temporal anchor points of vowel formant targets in the two languages. These anchor points are labeled ‘English vowel’ and ‘Polish vowel’, respectively. In English CV sequences, the VO node that encodes the early portion of the vowel is claimed by the representation of the onset consonants. As a consequence, listeners associate vowel quality based on formant values later in the time course of the vowel (under the VT node), opening the door to a greater degree of VISC. In Polish, VO is built into vowel representations, so phonological specifications defining vowel quality are aligned with an earlier portion of a vowel’s duration, and the Polish vowel is more pure in quality.

(6) Vowel ‘target’ alignment in English (left) and Polish (right)



To conclude this section, I will digress briefly by considering practical aspects of L2 pronunciation teaching and learning. For non-phonologists it may not be entirely clear

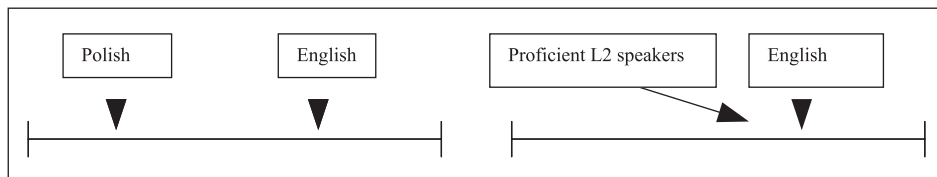


Figure 2. Visualization of temporal target locations for L1 Polish and English (left), and for proficient L2 speakers (right).

how the representational issues discussed in this article might be applied to L2 pronunciation learning. Clearly it is difficult to imagine a textbook of L2 English pronunciation that would include trees such as those in (6). Thus, it is important to think about how these phonological issues might be presented in a manner that is more user-friendly for non-specialists. Figure 2 attempts to do that with respect to the discussion of VISC. The figure provides a schematic visualization of the proposed effects of the cross-linguistic representational differences in (6) on the time course of vowel production. The timelines represent normalized vowel duration, while the arrows represent the endpoints of the articulatory approach to a target vowel location in Polish and English as an L1 on the left, and near target-like L2 English productions on the right. This schematic representation, which may serve as a visual aid for Polish learners (and instructors) of English, falls out directly from the OP model, since what is phonetically a vowel contains internal structure comprised of two separate representational nodes (VO and VT).

V Laryngeal contrasts in L2 speech

The reflection of CLI in measures of Voice Onset Time (VOT) is probably the single most frequently encountered type of phonetic experiment in the L2 speech literature. Starting with Flege (1987) or even earlier, many experiments have examined the production of the fortis series of stops /ptk/, in order to document CLI between languages like English or German featuring aspirated stops and languages in which unaspirated /ptk/ are produced with short-lag VOT. By contrast, far fewer studies have looked at the realization of the lenis series /bdg/. As an illustration of this state of affairs, consider Zampini's (2008) review chapter on experimental research in L2 speech production. This chapter contains a five page summary of research on the production of L2 stop consonants that cites over a dozen published studies. L2 speakers' production of initial /bdg/ is not discussed in this section.

A number of studies that have examined *both* /ptk/ and /bdg/ in the same experimental context have observed an interesting asymmetry. Lenis stops appear to be more susceptible than fortis stops to CLI.⁴ For example, Zajac (2015) found that Polish learners successfully converge with long-lag VOT of L2 English fortis stops, but fail to suppress L1 pre-voicing to produce unvoiced lenis stops in English. This result was replicated for Polish learners of English by Schwartz and Dzierla (2017). Other examples of this type of asymmetry have been found in situations of L2-induced L1 phonetic drift (Chang,

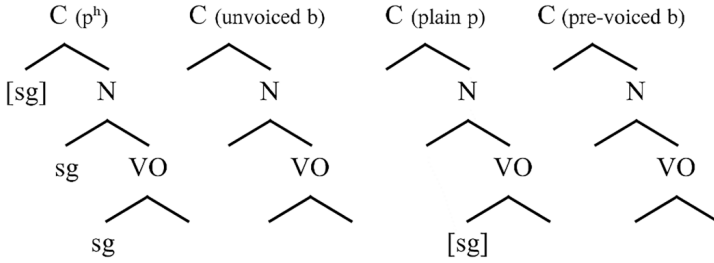
2012). For example, Herd et al. (2015) found that /bdg/ is produced with pre-voicing by US English speakers with a high level of proficiency in L2 Spanish, which has pre-voiced stops. However, L2 Spanish did not induce shortening of the VOT of /ptk/ in these speakers' L1 English. Along these same lines, Schwartz and Wojtkowiak (2017) found that Polish learners of English produced less pre-voicing than Polish monolinguals, but the VOT of their /ptk/ was not longer. Taken together, these findings all suggest that /bdg/ are subject to 'equivalence classification' (Flege, 1987) and CLI, but /ptk/ are not. In what follows, we will consider implications of this asymmetry for the phonological representation of laryngeal contrasts.

In the literature on the laryngeal phonology of languages with two series of stop consonants, two theoretical approaches dominate. In one approach based on the binary feature specifications of Chomsky and Halle (1968), languages employ a binary feature [voice], and VOT is a question of phonetic implementation, rather than phonological representation. The binary perspective is argued for by scholars such as Rubach (1996), Wetzels and Mascaró (2000), and Bennett and Rose (2017). In the other approach, Laryngeal Realism (Beckman et al., 2013), aspiration is due to a unary feature [spread glottis] ([sg]) and pre-voicing reflects a unary feature [voice], while short-lag VOT is unspecified. The main argument in favor of this approach is that it links unaspirated voiceless stops, typologically the most common, with unmarked phonological status. Incorporating this argument, proponents of Laryngeal Realism claim compatibility with the wider literature in which typologically common and phonologically unspecified are synonymous with 'unmarked'.

If we consider equivalence classification, which according to the SLM is the primary source of phonetic CLI, neither of the two phonological perspectives described above is capable of explaining observed asymmetries between /ptk/ and /bdg/. In both Laryngeal Realism and the binary approach, the representations of voicing contrasts predict symmetrical behavior for /ptk/ and /bdg/. In the binary approach voicing and aspiration languages have identical representations for both series ([+voice] and [-voice]), so we would expect equivalence classification for both series. Laryngeal Realism posits representational differences, and thus no equivalence classification, for both types of language ([voice] vs. \emptyset in voicing languages; \emptyset vs. [sg] in aspiration languages).

The OP approach to laryngeal phonology (for a thorough introduction, see Schwartz, 2017) predicts the asymmetry between /bdg/ and /ptk/. Relevant representations are shown in (7). Importantly, the VOT difference between voicing and aspiration systems receives natural expression in the internal structure of stop consonants. Unlike place features, which of course dock onto Closure since constriction location is what defines place, laryngeal specifications may be expected to exhibit flexibility with regard to the hierarchical level at which they can appear. In other words, VOT is a function of the timing of laryngeal gestures that are for the most part independent from supra-laryngeal articulation. In the structures in (7), a single laryngeal feature [sg] specifies only the fortis series of obstruents in both systems, while the lenis set is always unspecified (more on this postulate momentarily). In aspiration systems, [sg] is assigned at the Closure level, and trickles down to occupy the Noise node, which encodes aspiration. Unaspirated /ptk/ are specified for [sg] at the VO level, Noise is unaffected, and VOT remains short.⁵

(7) Two-series laryngeal systems in the OP framework



With the difference between plain and aspirated /ptk/ represented in terms of the level of [sg] specification, there is no need to posit a [voice] feature to represent /bdg/ with pre-voicing. As a result, the phonological equivalence between pre-voiced and unvoiced /bdg/, suggested by the CLI asymmetries discussed above, is captured in (7).

What are the implications of the representations in (7) for the phonological behavior and phonetic realization of laryngeal contrasts? The primary implication concerns the phonological status of voicing during stop closure. In the OP representations in (7), pre-voicing is simply a phonetic effect without phonological relevance. There is no feature [voice]. What then is responsible for the difference between voiced and unvoiced realizations of /bdg/? In essence, pre-voicing may simply be a contrast-enhancing mechanism in systems without aspiration. It does not reflect the presence of a phonological feature specification. This postulate explains the fact that languages do not show phonemic contrasts between short and long pre-voicing the way they do between short- and long-lag positive VOT.

Another prediction of the OP representations is that since voicing is phonologically irrelevant, the phonetic realization of /bdg/ in aspiration languages should be subject to a great deal of variation. This type of variation has been documented in North American English. Pre-voicing in /bdg/ has been observed in a number of dialects in the Southern United States (Herd, 2017; Jacewicz et al., 2009). Interestingly, it appears that speakers in the South who produced pre-voicing exhibit VOT values for /ptk/ that do not differ from speakers who do not pre-voice (Herd, 2017; Hunnicutt and Morris, 2016). In other words, pre-voicing dialects still count as ‘aspiration’ systems: the realization of /ptk/ is apparently more stable than that of /bdg/. The situation in pre-voicing dialects of American English may be said to resemble that of Swedish, which has been described as a language with both pre-voicing and aspiration (Helgason and Ringen, 2008). Phonological descriptions of such systems are simplified if it is assumed that they contain no active feature [voice].

At first glance, eliminating the feature [voice] might seem like a radical proposal. However, there are conceptual arguments to be made for such a break from phonological tradition. This outlook may be thought of in terms of Trau Müller’s (1994) Modulation Theory, in which speech perception entails the demodulation of an acoustic carrier. What is an acoustic carrier? According to Modulation Theory, the ideal carrier is a vocoid, produced by a combination of phonation and the resonance properties associated with neutral configurations of the articulators. In other words, it is a voiced, schwa-like vowel that serves an acoustic ‘canvas’ for phonological features. Viewers of paintings rarely notice the canvas. Similarly, in parsing the speech signal, listeners extract the linguistic content, i.e. phonological features, from the carrier. The carrier itself is the voice, the quality of which also transmits non-linguistic

speaker-specific information such as age and emotional state. Since the carrier is voiced, it follows that the demodulation cannot involve extracting a feature [voice] that is based solely on periodicity. Any phonological role for phonation must be epiphenomenal, and adopted on a language-specific basis.

In sum, our discussion of laryngeal phonology suggests that L2 speech research should devote more attention to documenting whether /ptk/ and /bdg/ behave symmetrically with regard to cross-linguistic phonetic influence. At present, evidence is mounting that they do not. Explaining why this is the case is a task for phonological theory.

VI Final remarks

Recent years have witnessed the dramatic development of technologies and empirical methods that have contributed to the study of L2 phonological acquisition. However, despite these advances, most research into second language phonology is still centered on traditional phonological models based on segmental transcription. This perspective, which Ladd (2011) calls ‘phone idealization’, may hamper progress for both phonological theory and the study of L2 speech. In this article I have argued for the need to refine models of phonological representation. In this way we may challenge theories of second language speech, which crucially rely on the notion of similarity, to be more explicit about defining what is similar to what. OP offers a promising path forward, splitting the segment apart to render an insightful view of both cross-language phonetic comparison and L2 phonological acquisition.

Declaration of Conflicting Interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research has been supported by grants from the Polish National Science Centre (Narodowe Centrum Nauki), project number UMO-2014/15/B/HS2/00452, ‘Vowel dynamics for Polish learners of English’, project number UMO-2016/21/B/HS2/00610, ‘Modulation in laryngeal features: Evidence from Polish learners of English’.

Notes

1. Considering the 3rd formant for measures of acoustic similarity could reveal that French /y/ is more similar to English /u:/ than /i:/. The rounded vowels both have a relatively low 3rd formant, while /i:/ does not.
2. The relationship described here describes acoustic effects that are progressive in direction. Regressive effects are derived by other mechanisms in the OP model that are not relevant for the present article.
3. The MC requirement conflates two constraints, ONSET and NUC, which in mainstream theory require that syllables contain both a nucleus and a consonantal onset.
4. Antoniou et al. (2010), who found no evidence of CLI between Greek and Australian English when controlling for language mode effects (Grosjean, 1998), note that language context was not controlled for in many experiments describing CLI in the VOT of /ptk/.
5. Comparing these representations with the ones in (3), one might ask about voicing systems in which the VO node is included in the representation of vowels. In such systems, the [sg] feature

and features associated with the vowel are both assigned at the VO level. The representation of stops contains a unary VO that houses the laryngeal feature. For details, see Schwartz (2017).

ORCID iD

Geoffrey Schwartz  <https://orcid.org/0000-0002-0728-7820>

References

- Antoniou M, Best C, Tyler M, and Kroos C (2010) Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics* 38: 640–53.
- Beckman J, Essen M, and Ringen C (2013) Evidence for laryngeal features: Aspirating vs. true-voice languages. *Journal of Linguistics* 49: 259–84.
- Bennett W and Rose S (2017) Moro voicelessness dissimilation and binary [voice]. *Phonology* 34: 473–505.
- Best C (1995) A direct realist view of cross-language speech perception. In: Strange W (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 171–204.
- Best C and Tyler M (2007) Nonnative and second-language speech perception: Commonalities and complementarities. In: Munro M and Bohn O-S (eds) *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: John Benjamins, pp. 13–34.
- Bohn O-S (2002) On phonetic similarity. In: Burmeister P, Piske T, and Rohde A (eds) *An integrated view of language development: Papers in honor of Henning Wode*. Trier: Wissenschaftlicher Verlag, pp. 191–216.
- Bohn O-S (2017) Cross-language and second language speech perception. In: Fernandez E and Cairns H (eds) *The handbook of psycholinguistics*. New York: Wiley, pp. 213–39.
- Browman C and Goldstein L (1989) Articulatory gestures as phonological units. *Phonology* 6: 201–51.
- Bybee J (2001) *Phonology and language use*. Cambridge: Cambridge University Press.
- Cebrian J (2000) Transferability and productivity of L1 rules in Catalan–English interlanguage. *Studies in Second Language Acquisition* 22: 1–26.
- Chang C (2012) Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40: 249–68.
- Chang C (2015) Determining cross-linguistic phonological similarity between segments: The primacy of abstract aspects of similarity. In: Raimy E and Cairns C (eds) *The segment in phonetics and phonology*. Chichester: John Wiley and Sons, pp. 199–217.
- Chomsky N and Halle M (1968) *The sound pattern of English*. New York: Harper and Row.
- Collins B and Mees I (2008) *Practical phonetics and phonology: A resource book for students*. 2nd edition. New York: Routledge.
- Escudero P (2005) *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht: LOT.
- Escudero P and Boersma P (2004) Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–85.
- Flege J (1987) The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65.
- Flege J (1995) Second language speech learning: Theory, findings, and problems. In: Strange W (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 233–76.
- Fleischhacker H (2005) Similarity in phonology: Evidence from reduplication and loanword adaptation. Unpublished PhD dissertation, UCLA, Los Angeles, CA, USA.

- Giegerich H (1992) *English phonology: An introduction*. Cambridge: Cambridge University Press.
- Gonet W and Pietróń G (2004) *The Polish tongue in the English ear*. In: Waniek-Klimczak E and Sobkowiak W (eds) *Materiały z konferencji 'Dydaktyka fonetyki języka obcego'* [Proceedings of 'foreign language pronunciation didactics' conference]. Konin: Wydawnictwo PWSZ w Koninie, pp. 55–65.
- Grosjean F (1998) Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1: 131–49.
- Helgason P and Ringen C (2008) Voicing and aspiration in Swedish stops. *Journal of Phonetics* 36: 607–28.
- Herd W (2017) Southern stops: Phonation type differences in Mississippi. *Journal of the Acoustical Society of America* 142: 2680–80.
- Herd W, Walden R, Knight W, and Alexander S (2015) Phonetic drift in a first language dominant environment. *Proceedings of Meetings on Acoustics* 23. Electronic publication. Published online: 25 October 2015 (doi: 10.1121/2.0000100).
- Hillenbrand J, Clark M, and Nearey T (2001) Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America* 109: 748–63.
- Hockett C (1955) *A manual of phonology*. Baltimore, MD: Waverly Press.
- Hunnicut L and Morris PA (2016) Prevoicing and Aspiration in Southern American English. *University of Pennsylvania Working Papers in Linguistics* 22. Available at: <https://repository.upenn.edu/pwpl/vol22/iss1/24> (accessed May 2019).
- Inkelas S and Shih S (2017) Looking into segments. *Proceedings of Annual Meetings on Phonology 2016*. Linguistic Society of America.
- Jacewicz E, Fox RA, and Lyle S (2009) Variation in stop consonant voicing in two regional varieties of American English. *Journal of the International Phonetic Association* 39(3): 313–334.
- Jenkins J and Strange W (1999) Perception of dynamic information for vowels in syllable onsets and offsets. *Perception and Psychophysics* 61: 1200–10.
- Johnson K (1997) Speech perception without speaker normalization: An exemplar model. In: Johnson K and Mullenix J (eds) *Talker variability in speech processing*. New York: Academic Press, pp. 3–26.
- Kaye J, Lowenstamm J, and Vergnaud J-R (1990) Constituent structure and government in phonology. *Phonology* 7: 193–231.
- Ladd DR (2011) Phonetics in phonology. In: Goldsmith J, Riggle J, and Yu A (eds) *Handbook of phonological theory*. 2nd edition. Oxford: Blackwell, pp. 348–73.
- Lieberman A, Cooper F, Shankweiler D, and Studdert-Kennedy M (1967) Perception of the speech code. *Psychological Review* 74: 431–61.
- Lisker L and Abramson A (1964) A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 384–422.
- Lleo C and Vogel I (2004) Learning new segments and reducing domains in German L2 phonology: The role of the Prosodic Hierarchy. *International Journal of Bilingualism* 8: 79–104.
- Mielke J (2012) A phonetically-based metric of sound similarity. *Lingua* 122: 145–63.
- Morrison G and Assmann P (2013) *Vowel inherent spectral change*. Berlin: Springer.
- Nearey T and Assmann P (1986) Modeling the role of vowel inherent spectral change in vowel identification. *Journal of the Acoustical Society of America* 80: 1297–1308.
- Ohala J (1981) The listener as a source of sound change. In: Masek CS et al. (eds) *Papers from the parasession on language and behavior*. Chicago, IL: Chicago Linguistic Society, pp. 178–203.
- Palková Z (1997) *Fonetika a fonologie češtiny* [The phonetics and phonology of Czech]. Praha: Karolinum.
- Rubach J (1996) Nonsyllabic analysis of voice assimilation in Polish. *Linguistic Inquiry* 27: 69–110.

- Scheuer S (2002) What makes a foreign accent sound foreign. In: James A and Leather A (eds) *New sounds 2000: Proceedings of the 4th international symposium on the acquisition of second language speech*. Klagenfurt: University of Klagenfurt, pp. 128–58.
- Schwartz G (2010) Phonology in the speech signal: Unifying cue and prosodic licensing. *Poznan Studies in Contemporary Linguistics* 46: 499–518.
- Schwartz G (2013) A representational parameter for onsetless syllables. *Journal of Linguistics* 49: 613–46.
- Schwartz G (2016a) On the evolution of prosodic boundaries: Parameter settings for Polish and English. *Lingua* 171: 37–74.
- Schwartz G (2016b) Word boundaries in L2 speech: Evidence from Polish learners of English. *Second Language Research* 32: 397–426.
- Schwartz G (2017) Formalizing modulation and the emergence of phonological heads. *Glossa: A Journal of General Linguistics* 2(1): 81.
- Schwartz G and Dzierla J (2017) Pre-voicing suppression in the speech of Polish learners of English. Unpublished paper presented at *Approaches to Phonology and Phonetics*, Lublin, Poland.
- Schwartz G, Balas A, and Rojczyk A (2014) External sandhi in L2 segmental phonetics: Final (de)voicing in Polish English. *Concordia Working Papers in Applied Linguistics* 4: 637–49.
- Schwartz G, Aperliński G, Jekiel M, and Malarski K (2016) Spectral dynamics in L1 and L2 vowel perception. *Research in Language* 14: 61–77.
- Schwartz G and Wojtkowiak E (2017) Asymmetries in L2-induced phonetic drift in L1 Polish. Unpublished paper presented at *Accents 2017*, Łódź, Poland.
- Sobkowiak W (2008) *English phonetics for Poles*. 3rd edition. Poznań: Wydawnictwo Poznańskie.
- Steriade D (1993) Closure, release, and nasal contours. In: Huffman and Krakow R (eds) *Nasals, nasalization, and the velum*. San Diego, CA: Academic Press, pp. 401–70.
- Stevens KN and House AS (1963) Perturbation of vowel articulations by consonant context. *Journal of the Acoustical Society of America* 85: 2135–53.
- Strange W (1989) Evolving theories of vowel perception. *Journal of the Acoustical Society of America* 85: 2081–87.
- Strange W, Jenkins J, and Johnson T (1983) Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America* 34: 695–705.
- Traunmüller H (1994) Conventional, biological, and environmental factors in speech communication: A modulation theory. *Phonetica* 51: 170–83.
- Wells J (1982) *Accents of English*. Cambridge: Cambridge University Press.
- Wetzels WL and Mascaró J (2000) The typology of voicing and devoicing. *Language* 77: 207–44.
- Wiese R (1996) *The phonology of German*. Oxford: Clarendon Press.
- Williams D, van Leussen J, and Escudero P (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.
- Wright R (2004) A review of perceptual cues and cue robustness. In: Hayes B, Kirchner R, and Steriade D (eds) *Phonetically based phonology*. Cambridge: Cambridge University Press, pp. 34–57.
- Zajac M (2015) *Phonetic convergence in the speech of Polish learners of English*. Unpublished doctoral dissertation, University of Łódź, Łódź, Poland.
- Zampini M (2008) L2 speech production research: Findings, issues, and advances. In: Hansen Edwards JG and Zampini M (eds) *Phonology and second language acquisition*. Amsterdam: John Benjamins, pp. 219–249.
- Zsiga E (2011) External sandhi in a second language: The phonetics and phonology of nasalization in Korean and Korean-accented English. *Language* 87: 289–345.