


*J. Linguistics* 59 (2023), 623–654. © The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.  
doi:10.1017/S0022226722000275

# All TRs are not created equal: L1 and L2 perception of English cluster affrication<sup>1</sup>

GEOFFREY SCHWARTZ 

*Adam Mickiewicz University*

(Received 9 January 2021; revised 13 May 2022)

This paper describes a perception experiment with L1 Polish and L1 English listeners on the affrication of initial English /tr/ and /dr/ (TR) consonant clusters. The goal was to test phonological predictions formulated within the Onset Prominence (OP) framework. OP offers two distinct structural configurations for representing rising sonority onset clusters. One predicts synchronous cluster articulation in English, giving rise to affrication, while the other predicts asynchronous cluster articulation in Polish. Two groups of listeners performed a two-alternative forced choice identification task on stimuli that included affricated clusters, unaffricated clusters, affricates, and CəC-initial words. For L1 English listeners, the unaffricated cluster-initial items induced the slowest responses. For L1 Polish listeners, the lack of affrication facilitated cluster identification, while the CəC-initial words induced the slowest responses. The results suggest cross-language interaction by which Polish listeners equate L1 unaffricated clusters with L2 CəC-initial words, in accordance with the OP proposal.

KEYWORDS: Phonotactics, Cluster Affrication, Phonetics and Phonology, L2 English, Polish

## 1. INTRODUCTION

Much of the research in the area of consonant cluster phonotactics may be grouped into two basic categories deriving from the theoretical priorities of the researchers. A common goal underlying what might be called a *phonological* approach to phonotactics is to establish and investigate the effects of the ‘goodness’ or ‘markedness’ of clusters, usually characterized in terms of sonority sequencing (Hooper 1976; Selkirk 1984; Clements 1990; Blevins 1995; Parker 2002), or other comparison of features (Dziubalska-Kończak 2014; Orzechowska 2019). This research is quite copious, and aims at uncovering language users’ subconscious knowledge

---

[1] The author was supported by a grant from the Polish National Science Centre (Narodowe Centrum Nauki), project number 2018/29/B/HS2/00088. I am grateful to Adam Olender for his contribution to the early stages of this project. Thanks also go to Jerzy Dzierła, Ewelina Wojtkowiak, and Kamil Kaźmierski. I am also grateful to three anonymous referees, and the associate editor of *Journal of Linguistics*. Any remaining errors are my own responsibility.

about phonotactic constraints or preferences, while spanning a wide range of research paradigms, such as acceptability studies (Albright 2007), surveys of typology and frequency (Blevins 1995), loanword adaptation (Kang 2011), language acquisition (Jarosz 2017), and computer simulation (Hayes & Wilson 2008; Daland et al. 2011). Another approach found in the literature is an *articulatory* perspective on cluster phonotactics (e.g. Browman & Goldstein 1989; Goldstein et al. 2007; Shaw et al. 2009; Tilsen et al. 2012; Hermes et al. 2013; Hermes et al. 2017; Pastätter & Pouplier 2017), which has been aimed at quantifying the extent to which consonants in a cluster are pronounced synchronously, as well as their degree of phonetic coordination with neighboring vowels. The goal of this research is to relate the phonetics of consonant clusters to phonological assumptions about syllable structure. The most notable methodology used in this research is electromagnetic articulography (EMA), which has been employed to test hypotheses about ‘simplex’ vs. ‘complex’ cluster organization that is claimed to govern the syllabic affiliation of consonants found in clusters (see e.g. Shaw et al. 2009 for discussion).

Although there is certainly some overlap between these approaches, a significant divide remains in which many questions go unanswered, or even unasked. Phonological approaches often pay little or no attention to the actual phonetic realization of the clusters they examine. For example, a stop-rhotic cluster such as /tr/, owing to its rising sonority, is typically assumed to be phonologically equivalent regardless of whether the rhotic is a trill or an approximant (e.g. Chabot 2019). Meanwhile, articulatory approaches have been limited by the phonological assumptions underlying their investigation. For example, the assumption that a syllable is organized around its ‘nucleus’ motivates phonetic measures of onset-nucleus coordination, such as center-to-anchor stability between singleton onsets and onset clusters (see Shaw et al. 2009). With this measure, if the time lag between the center of the onset and an anchor point in the vowel is constant across 1 and 2 (or 3) member onsets, it is assumed that the cluster shows complex onset alignment (Shaw et al. 2009: 189). However, center-to-anchor lag describes only onset-nucleus coordination, and says nothing about the relative synchronicity of consonants in a cluster with each other (target-to-target lags, see Hermes et al. 2017). In principle, shorter target-to-target lags should be associated with complex clusters, since the consonants exhibit tight phonetic cohesion, which suggests that they are contained in a single ‘onset’. However, shorter lags do not always mean greater center-to-anchor stability assumed to be associated with complex onsets. In Hermes et al.’s (2017) EMA study of Tashlhiyt Berber and Polish onset clusters, Polish showed greater center-to-anchor stability yet longer target-to-target lags. That is, one measure suggested complex onsets in Polish, while the other suggested simplex onsets, so the assumed link between one of these measures and phonological structure must be reconsidered (see 4.2).

The research described in this paper shares its goals with those of the articulatory approach – to explore the phonetics of cluster phonotactics, albeit in perception rather than production – but at the same time operates on refined phonological assumptions about the nature of the ‘onset’ and the ‘syllable’ as phonological

constituents. Data come from perception of allophonic /tr/ and /dr/ affrication (Cruttenden 2014; Carley & Mees 2020) in English, a process that is clearly indicative of tight phonetic coordination associated with complex onsets. Two groups of listeners, first language (L1) English listeners and L1 Polish listeners, for whom English is a second language (L2), took part in a perception experiment. Participants performed a two-alternative forced choice (2AFC) identification task, which investigated the perceptual effects of affrication. The results to be presented show opposing perceptual forces of this allophonic process as a function of L1 background. L1 English speakers' identification of cluster-initial words is facilitated by affrication, while a lack of affrication is a hindrance. For L1 Polish listeners, lack of affrication facilitated cluster identification. As mentioned above, allophonic affrication is related to tighter phonetic synchronicity of stop-liquid clusters in English. This is in stark contrast to Polish, which shows evidence of asynchronous cluster articulation in rising sonority onsets (Dłuska 1986; Świącieński 2012; Hermes et al. 2017; Schwartz et al. 2021).

Underlying this research is the question of whether and how this kind of phonetic synchronicity, or a lack thereof, may be derived from phonological considerations. The essence of the problem to be investigated is that word-initial stop-liquid and stop-approximant clusters (henceforth TR), which usually are treated as a single 'unmarked' class obeying sonority sequencing, show cross-language differences both in their phonetic synchronicity and phonological behavior, despite the fact that conventional descriptions of syllable structure suggest that the sequence is structurally equivalent across languages. In other words, by most accounts TR clusters are equivalent structures in English and Polish, and many other languages, constituting what is usually thought of as a 'branching onset'. We shall see, however, that the single branching onset category is insufficient to encode the systemic differences to be described in this paper.

In the Onset Prominence representational framework (OP; Schwartz 2010, 2013, 2015, 2016, 2018), there are two separate structural configurations for TR clusters, the choice of which is determined on a language-specific basis (see Section 4). Crucially, the relevant configurations are derived from phonotactic mechanisms that are independently motivated within the OP model. These structures not only govern the relative degree of phonetic synchronicity in TR clusters in English (Section 2.1) and Polish (2.2), which underlie the experimental data described in Section 3, they also capture phonological differences related to the shape of well-formed prosodic words in the two languages (Section 2.3).

## 2. THE PHONETICS AND PHONOLOGY OF TR ONSET CLUSTERS IN ENGLISH AND POLISH

This section will compare the phonetics and phonology of rising sonority onset clusters in English and Polish. The particular focus of the discussion will be on stop-rhotic clusters, with some additional mention of other stop-approximant (/w j l/ as C2) sequences. The basic picture is that these languages differ in the phonetic

synchronicity of their onset clusters, a difference that also appears to be reflected in the phonology.

### 2.1 Cluster-induced allophonic processes in English

The affrication of /tr/ and /dr/ onsets is one of a number of cluster-induced allophonic processes in TR clusters in English, which are indicative of synchronous articulation of the two consonants in the sequence. The phonetic synchronicity of English onset clusters has been documented in articulographic (EMA) studies (Browman & Goldstein 1989; Honorof & Browman 1995; Marin & Pouplier 2010; Tilsen et al. 2012), while the specific allophonic processes that occur are a function of the segmental makeup of the consonant sequences. Both the phonetic and phonological facts of English suggest that onset clusters in the language are ‘complex’ in the sense that the two consonants of the cluster are joined into a single prosodic unit traditionally referred to as an ‘onset’. It may be assumed that the presence of two consonants in a single onset, with no intervening prosodic boundary, is responsible for the phonetic synchronicity that induces the allophonic processes.

The first process we shall examine, the focus of the empirical study described in this paper, is the affrication of /tr/ and /dr/ onsets. From an articulatory perspective, the process may be attributed to the fact that the two consonants in the sequence have the same active articulator, and that the localization of the articulation of the two consonants is in close proximity (alveolar ridge and post-alveolar region). In other words, affrication is a natural by-product of two consonants being produced with single articulatory gesture. Beyond its phonetic properties, TR affrication may be conditioned by phonological factors. Namely, it occurs in onset position. It may therefore be absent when a morpheme boundary interrupts the cluster, as in compounds like *headroom* or *bedrock*, while its presence in words like *bedroom* may indicate resyllabification of the cluster accompanied by the loss of the boundary. The process may also be fed by schwa deletion in words like *lavatory* in British English.

TR affrication is noted in descriptive works on English, as well as in pronunciation textbooks aimed at second language learners. According to Scobbie et al. (2006), the sequences are often pronounced as rhotacized post-alveolar affricates, with rhotacization distinguishing them from the post-alveolar /tʃ/ and /dʒ/. However, these authors do not offer a particular transcription of the clusters. Wells (2011) claims the sequences are the result of a retracted /t/ and fricativized /d/, clearly distinct from the post-alveolar affricates, and he does not advocate transcribing the affrication. Cruttenden (2014) transcribes the sequences as [t̠ʃ] and [d̠ʒ], with the IPA diacritic for raising on the approximant. Carley & Mees (2020: 23) describe the clusters as “phonetic affricates” that are analogous to stop-fricative sequences, but their transcriptions ([t̠ʃ], [d̠ʒ]) do not use the raising diacritic employed by Cruttenden. Note that both Cruttenden and Carley & Mees use the

devoicing diacritic for the /tr/ sequence, and neither of these works uses the symbols for post-alveolar affricates/fricatives. Indeed, none of these authors suggests that the affricated clusters are merged with the post-alveolar affricates /tʃ dʒ/, despite their phonetic similarity.

TR affrication shows parallels with other allophonic process affecting stop-sonorant clusters in English. In one such process, approximants after word-initial voiceless stops become devoiced, as exemplified in words like *cry*, *quite*, *place*, and *cute*. Approximant devoicing has also been noted in English pronunciation textbooks (Cruttenden 2014; Carley and Mees 2020), and studied experimentally by Tsuchida et al. (2000). These authors note that the process occurs primarily after stops, while only partial devoicing may be observed after fricatives. The other process that warrants mention here is found in varieties of English that preserve /j/ before /u:/ after /t/ and /d/. These /tj/ and /dj/ sequences undergo coalescence to [tʃ] (sometimes [tʃ̥]) and /dʒ/, as in *tune* and *dune* (see Cruttenden 2014: 229).

Parallels among affrication, coalescence, and approximant devoicing in English suggest a similar structural interpretation, according to which stop and sonorant clearly share a prosodic constituent. We may observe these parallels between affrication and approximant devoicing, as well as between affrication and coalescence. First, after /t/, affrication may be seen as an instantiation of approximant devoicing, as reflected in the transcriptions mentioned above. Additionally, like affrication, coalescence affects clusters starting with both voiced and voiceless stops. Finally, all of these processes are much more robust in stop-sonorant clusters than fricative-sonorant clusters (Tsuchida et al. 2000). In sum, these three processes observed in English may be unified phonologically into single stop-approximant configuration, which is characterized by a large degree of phonetic synchronicity. In traditional syllabic terms, the English clusters therefore qualify as ‘complex’ onsets in both the phonetic and phonological senses. In what follows, we consider rising sonority clusters in Polish, which show notably different patterns.

## 2.2 Polish TR clusters and their phonetic realization

As is well-known, Polish features a large number of initial consonant clusters that do not show rising sonority. It is this fact that has gained most of the attention of researchers interested in Polish phonotactics – much less attention has been given to the realization of rising sonority clusters. However, available descriptions allow for the generalization that Polish onset clusters are characterized by asynchronous articulation. Classic descriptions of Polish phonetics (Dłuska 1986; Dukiewicz & Sawicka 1995) make no mention of the allophonic processes discussed above for English. Indeed, Dłuska (1986) states that vowel intrusion is quite common in Polish onset clusters, independent of their sonority profile. With regard to rising sonority clusters, intrusive vocoids are most frequent when C2 is /r/, but are also observed when C2 is /l/ or /w/, particularly when C1 is voiced. Vowel intrusion, in

which transitional vocoids between consonants are produced, but generally not perceived (Hall 2006), is a clear sign of asynchronous cluster articulation.

More recent experimental studies have confirmed the asynchronous articulation of Polish rising sonority onset clusters. Świącieński (2012) studied acoustic transitions in Polish Cj sequences. Interestingly, there is some disagreement in the literature about whether such sequences in Polish indeed contain an approximant segment /j/, or whether they are simply singleton consonants with secondary palatalization (C<sup>j</sup>), which is a more traditional transcription (e.g. Gussmann 2007). The acoustic results of Świącieński strongly suggest that these sequences are indeed clusters rather than singletons. The presence of /j/, separate from C1, is evident in the long F2 transitions observed in the acoustic displays. In another study, Cieślak (2015) looked at stop-/l/ and stop-/r/ clusters in Polish and found evidence for asynchronous articulation consisting in both intrusive vocoids, as well as clear separation of the two consonants in the cluster in acoustic displays. Hermes et al. (2017) gathered EMA data on rising sonority clusters (pr, pl, kr, kl) in Polish and found very large target-to-target lags. Finally, in an EMA study, Schwartz et al. (2021) compared target-to-target lags with Polish-English bilinguals speaking in both their languages, and found larger lags in Polish.

The preceding discussion has established that Polish and English differ to some degree in the phonetic realization of TR clusters. We turn now to the question of whether the phonetic differences may reflect distinct phonological structures underlying the clusters in the two languages.

### 2.3 *Polish-English differences in TR onsets: phonetics or phonology?*

At first glance, it may appear as though the phonetic differences described above are simply a matter of language-specific phonetic implementation of a complex or branching onset that is phonologically equivalent in Polish and English. In accordance with this view, two familiar aspects of Polish and English phonetics suggest possible implementational explanations for the allophonic processes observed in English, but not in Polish. These patterns relate to both affrication of /tr/ and /dr/ (2.3.1), as well as approximant devoicing (2.3.2), which will be discussed in turn. In both cases, we shall see that an implementational explanation is problematic, and the observed phonetic patterns more likely reflect distinct phonological structures, which are also manifest in differing requirements for word minimality in the two languages (2.3.3).

#### 2.3.1 *Approximant rhotics in Polish do not induce affrication*

As mentioned before, it may be suggested that the phonetic origins of affrication in English TR clusters are attributable to the approximant realization of the rhotic in English, as opposed to the trill or tap that is common in Polish. In other words, affrication would be thought of by many as an automatic phonetic effect of the

stop-approximant sequence with the same active articulator in which the place of articulation of both consonants is the same or similar. Since stop-tap sequences do not show affrication, but rather are frequently associated with vowel intrusion, the affrication process may be linked to the realization of the rhotic. Many scholars accept that rhotics form a single phonological class independent of their phonetic realization (e.g. Chabot 2019), so allophonic processes induced by approximant (as opposed to tapped or trilled) rhotics would be assumed to be outside the domain of phonology.

In the case of Polish, this argument is untenable. The reason for this is that approximant realizations of Polish /r/ are in fact relatively common (Jaworski & Gillian 2011), but they do not induce affrication in TR clusters, so affrication cannot be considered an automatic phonetic effect of approximant rhotics. One study that documented this was carried out by Cieślak (2015) who describes a phonetic code-switching experiment in which Polish-English bilinguals inserted TR-initial Polish words into sentences produced in an English-language context. The aim of the experiment was to see if the task would induce both approximant rhotics and affricated realizations of the clusters, as would be expected in English. Interestingly, the code-switches did induce approximant realizations of Polish /r/, but they did not induce affrication. Rather, the stop and rhotic in the code-switched productions were interrupted by an intrusive vowel, as is common with L1 tapped realizations.

An illustration of one of the items examined by Cieślak (2015) is given in the spectrogram in Figure 1, which shows the Polish word *dres* 'tracksuit' code-switched in an English-language mode, as pronounced by a female L1 Polish speaker with an advanced level of English. In the figure, effects of the English context are visible, including a lack of prevoicing of /d/ and an approximant realization of /r/, but at the same time affrication is lacking, and an intrusive vocoid is visible (the selected portion). Apparently, the phonetic realization of the rhotic (and the voicing of the /d/) was affected by the code-switching task, but the asynchronous articulation of the Polish cluster was not.

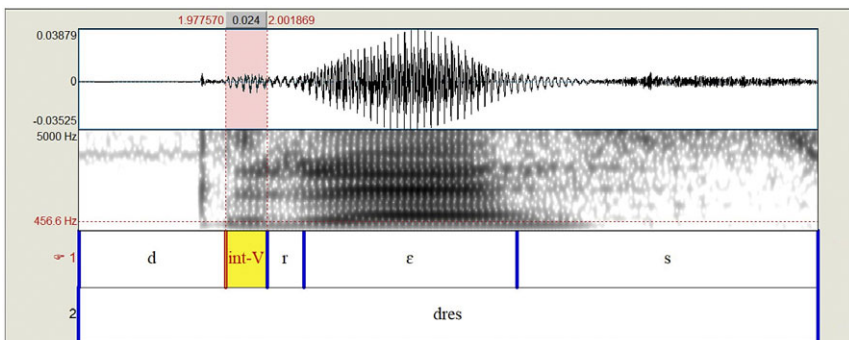


Figure 1

Waveform/spectrogram of Polish *dres* 'tracksuit', with an intrusive vowel and approximant rhotic (after Cieślak 2015).



The presence of approximant rhotics in Polish, accompanied by a lack of affrication of /dr/ and /tr/ clusters, indicates that TR affrication in English cannot be an automatic phonetic effect stemming from the phonetic realization of the rhotic. Rather, it appears that there is a language-specific difference in the phonetic synchronicity of TR clusters. In English, synchronous TR induces allophonic affrication. In Polish, asynchronous TR does not.

### 2.3.2 Polish voicing and the behavior of TR clusters

The next process we will consider is approximant devoicing. As with TR affrication, the difference between the two languages stems from the degree of phonetic synchronicity within the cluster. In English, synchronous clusters induce the allophonic process. In Polish, asynchronous clusters do not.<sup>2</sup> It is suggested here that this is due to distinct phonological configurations for stop-approximant clusters in the two languages. Before doing so, however, it is necessary to address a possible alternative explanation for the lack of approximant devoicing in Polish, which may arise out of Laryngeal Realism (Harris 1994; Iverson & Salmons 1995; Honeybone 2005; Beckman et al. 2013), and its assumptions about laryngeal specifications and what they represent.

The main claim of Laryngeal Realism is that stop consonants with short-lag Voice Onset Time (VOT) are unmarked, both typologically and phonologically. In voicing languages such as Polish, this means that voiceless /p t k/ are assumed to be unspecified for laryngeal features. As such, they should not take part in phonological processes, and the lack of approximant devoicing in Polish may be attributable to the unmarked status of voiceless stops.

The main problem with this explanation is that there is evidence that voicelessness is phonologically active in Polish, so the lack of approximant devoicing cannot be attributed to a lack of laryngeal specification. This evidence consists of a range of phonological and phonetic phenomena, from distributional requirements of laryngeal features in stop-fricative clusters (see e.g. Rubach 1996) to fortis effects in the phonetics of the laryngeal contrast, including f<sub>0</sub> raising (Schwartz et al. 2019; see Kirby & Ladd (2016) for similar effects in French and Italian), and positive VOT category boundaries in the perception of the voice contrast (Keating 1979; Schwartz & Arndt 2018; Schwartz et al. 2019). The perception studies cited here clearly show that a lack of voicing in Polish stops does not induce voiceless percepts, which is incompatible with the claim that voiceless stops are unmarked for laryngeal features.

---

[2] So called trapped sonorants in Polish (see e.g. Strycharczuk 2012), i.e. those that follow a voiceless obstruent and precede another voiceless obstruent or a word boundary (*wiatr* 'wind'; *krtani* 'larynx'), can undergo devoicing. This paper focuses on pre-vocalic TR clusters, so trapped sonorants will not be discussed in detail. For more discussion of trapped sonorants, see Schwartz (2016).



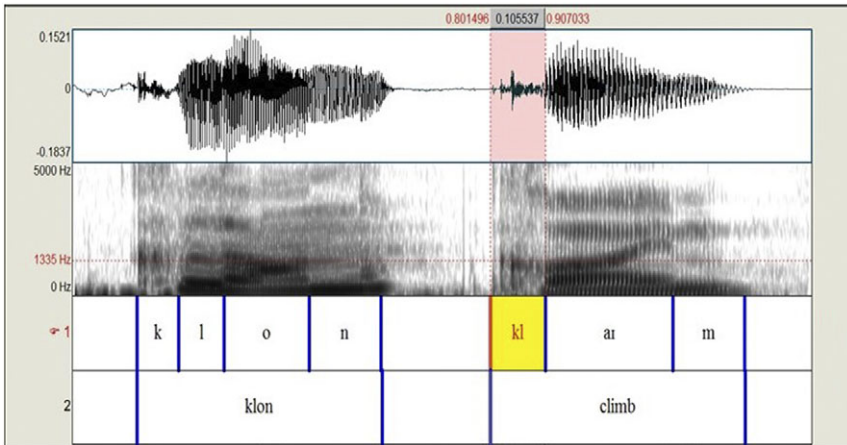


Figure 2

Annotated waveform/spectrogram displays of Polish *klon* ‘maple’ and English *climb*.

Since there is sufficient evidence available to reject the claim that voiceless obstruents in Polish are lacking in laryngeal specification, we may conclude that the lack of approximant devoicing in Polish must be attributable to something other than laryngeal phonology. Rather, it appears to be a function of the asynchronous articulation of the clusters in Polish. By way of illustration, Figure 2 presents an annotated waveform/spectrogram display of Polish *klon* ‘maple’ and English *climb*,<sup>3</sup> highlighting the differences in the phonetic realization of the /kl/ cluster. In the Polish example, the voiced approximant is easily identifiable in the acoustic display – it is annotated as a separate unit from the stop. However, in English, as a result of approximant devoicing, it is difficult, if not impossible, to identify C1 and C2 as separate entities in the acoustic display, and the cluster is annotated as a single unit. In other words, in Polish, the two consonants are realized over two distinct portions of the acoustic signal, one unvoiced and one voiced, while in English the /kl/ sequence resembles a single aspirated stop in the acoustic display. Clearly, the degree of phonetic overlap between the two consonants is much greater in English. Note also that the /k/ in the Polish item shows a rather long VOT measure (over 60 ms), indicative of glottis spreading and a slightly aspirated realization, and presumably attributable to the English proficiency of the speaker, yet this has little or no effect on the /l/. In Section 4, it will be shown how phonetic differences such as

[3] The Polish item was produced by a female L1 speaker (aged 24) from the Wielkopolska province, with advanced proficiency in English. The English item was produced by a male L1 speaker of American English (aged 47) who is also fluent in Polish. The bilingual status of these speakers might have us expect phonetic similarity between the Polish and English items (e.g. the VOT of /k/), thus underlying rendering the cross-language difference in cluster synchronicity visible in the spectrogram.

these may derive from distinct phonological configurations for TR clusters in the two languages.

### 2.3.3 Onset clusters contribute to word minimality in Polish

In addition to the phonetic differences discussed above, there is one important Polish-English difference affecting consonant clusters, the phonological status of which appears to be incontrovertible. Namely, unlike in English, onset clusters in Polish may bear prosodic weight, as evidenced by minimality requirements for morphological inflection (see Comrie 1976; Garrett 1999; Schwartz 2016). In Polish, in order to show inflectional morphology a word must contain either a coda consonant or an onset cluster. That is, for the purposes of Polish morphology, a minimal word is comprised of either (C)VC or CCV. By contrast, CV-shaped words in Polish are sub-minimal for the purposes of morphology, are often pronounced as enclitics, and when produced in isolation frequently show glottal stop insertion after the vowel (CV→CV?). The fact that onset clusters may contribute to prosodic minimality in Polish is compatible with the claim that the two consonants in the cluster are not contained in the same prosodic constituent. The first consonant in the cluster is not linked to the second, and is free to contribute prosodic weight to the word.

The experiment to be described in what follows is intended to shed light on the question of how TR clusters may be encoded in the phonology of English and Polish. The cross-language differences described above suggest distinct structural configurations in the two languages, but are gleaned from data on speech production. No perceptual data on this question has been published.<sup>4</sup>

## 3. L1 VS. L2 PERCEPTION OF ENGLISH TR AFFRICATION

This section will present the results of perception experiment comparing L1 English speakers with L1 Polish learners of English. Participants performed a two-alternative forced choice identification task. The experiment was designed to address the following research questions.

- 1 Do listeners confuse affricated TR clusters with post-alveolar affricates?
- 2 Do listeners confuse unaffricated TR-initial words with CəC-initial words (e.g. *train-terrain*)?
- 3 How are these effects related to L1 background?

The three research questions seek to investigate phonetic issues associated with TR onset clusters, as well as how those issues may relate to language experience and the perception of a second language. RQ1 concerns the possible implications of the

---

[4] Dłuska (1986) notes that intrusive vocoids in Polish are generally not heard by Polish listeners, but does not elaborate with any experimental data.

phonetic similarity between affricated clusters and affricates, as discussed in the previous section, with implications for perception of allophonic processes in L1 and L2. RQ2 concerns the implications of a prosodic difference in word structure between cluster-initial words and words in which the stop and rhotic are separated by a lexical unstressed vowel schwa. RQ2 thus seeks to investigate whether the absence of the expected allophonic cluster affrication may cause confusion between cluster-initial words and words starting with /tər/ or /dər/. Cluster-induced confusion has been studied with speakers from L1 backgrounds with more restrictive phonotactics than English (see Leung et al. 2021; and references therein). Those studies frequently observe perceptual epenthesis (e.g. Durvasula & Kahng 2015), by which listeners hear illusory vowels, apparently to simplify the structure of an L2 consonant sequence that is absent in L1. By contrast, the current study deals with L1 Polish listeners, whose native phonotactic patterns are known to be extremely liberal. Since Polish also lacks a lexical schwa vowel, it is fair to ask whether L1 Polish listeners, upon hearing a /tər/- or /dər/-initial word, might engage in what might be called ‘perceptual schwa deletion’, and hear a cluster when none is present in the input.

Before proceeding, we briefly describe possible interpretations for hypothetical results of our study, which concern both L1 English and L1 Polish participants. With regard to L1 English listeners, confusion between clusters and affricates (Yes for Question 1) would suggest a phonological change in progress, while robust and accurate identification (No for Question 1) would suggest stable categories for the affricates and clusters. Additionally, difficulty in identifying unaffricated clusters would suggest confusion with CəC-initial words (Yes for Question 2), and imply that the synchronous articulation that induces affrication is encoded in the phonology of English. By contrast, if a lack of affrication in clusters does not hinder identification (No for Question 2), the affrication process can be considered to exist outside of the English phonological system.

If TR clusters are structurally equivalent in English and Polish, i.e. they are complex onsets whose phonetic synchronicity is not encoded in the phonology, we should expect Polish listeners to confuse affricates and affricated clusters (Yes for Question 1), since *equivalence classification* (Flege 1987) with unaffricated L1 clusters would render the affricated clusters difficult to acquire. Additionally, we should not expect Polish listeners to confuse cluster-initial and CəC-initial words (No for Question 2), which are clearly structurally distinct under this interpretation. If however, the relative phonetic synchronicity of TR clusters is encoded in the phonologies of Polish and English, we should expect the opposite effects: no confusion between affricates and affricated clusters (No for Question 1), but confusion between unaffricated TR-initial words and CəC-initial words (Yes for Question 2). Under this view, we should expect equivalence classification between asynchronous Polish TR clusters and English sequences containing schwa, hindering acquisition of the latter.

### 3.1 Participants

Two groups of listeners took part in the experiment. The first group of listeners was comprised of 28 (16 male, 12 female) L1 native speakers of North American English. All of the listeners were between the ages of 20 and 25, enrolled as students at a medical university in Poland, where they were attending classes in English. The students had only elementary knowledge of Polish, and much of their social interaction was within their own group. Thus, despite the fact that they lived in Poland, it is not entirely accurate to say that they were immersed in a Polish language setting. Additionally, none of group reported being fluent in any other language except for English. The second group of listeners was comprised of 21 (7 male, 14 female) first year Polish students majoring in English. Their level of L2 proficiency may be classified as B1/B2 according to the Common European Frame of Reference for Languages (CEFR; Council of Europe 2011). All participants signed Informed Consent forms before taking part in the experiment.

### 3.2 Stimuli

Four types of stimuli were created for the present experiment (not including filler trials, see Procedure), based on recordings of six words (*train*, *drive*, *chain*, *jive*, *terrain*, *derive*) made by a male L1 speaker of North American English.<sup>5</sup> These four stimulus types are given below. Relevant acoustic properties of the stimuli are given in Table 1, and spectrograms of the stimuli are given in Appendix A.

- Naturally produced  $\text{tʃ}/\text{dʒ}$ -initial words: *chain/jive*
- Naturally produced  $\text{tər}/\text{dər}$ -initial words: *terrain/derive*
- Naturally produced affricated  $\text{tr}/\text{dr}$ -initial words: *train/drive*
- Hybrid unaffricated *train/drive* items, produced by replacing the affricated cluster with an unaffricated cluster, obtained by deleting the schwa from *terrain/derive* recordings.

The acoustics of the stimuli show effects that are expected based on what we know about articulatory-acoustic relationships (e.g. Stevens 1998). The English rhotic is associated with a low 2<sup>nd</sup> and particularly low 3<sup>rd</sup> formant, and this is evident in the F2 and F3 columns in the table – *train* and *drive* have lower F2 and F3 measures than *chain* and *jive*. Meanwhile, the F3 values for *terrain* and *derive* fall in between those of the affricates and clusters, indicating the start of a transition through the schwa to the rhotic, which had a lower F3 target.<sup>6</sup> English clusters containing rhotics also produced lower-frequency noise spectra than what we see in the

[5] The speaker was a phonetician who was aware of the purpose of the recordings, and made an effort to produce the words in as neutral a manner as possible.

[6] Impressionistically, the *terrain* and *derive* stimuli from this experiment contained a syllabic rhotacized schwa (e.g.  $[\text{t}\partial\text{.e}\text{ɪ}\text{n}]$ ). As suggested by a reviewer, the ‘schwa duration’ column in Table 1 may be thought of as the duration of the alveolar formant transitions preceding the rhotic. The rhotic then induces a drop in amplitude in the vocalic portion of the signal.

item	VOT (ms)	Noise				Skewness	Kurtosis	F1	F2	F3	schwa duration (ms)
		Intensity (dB)	COG (Hz)	SD (Hz)							
<i>train</i>	104	68	4499	2565	0.838	0.939	545	1574	2020	NA	
<i>chain</i>	84	69	5281	1645	0.877	5.26	428	2092	2539	NA	
<i>drive</i>	18	60	1812	2015	2.2	7.05	511	1306	1818	NA	
<i>jive</i>	40	66	5615	1658	-0.22	1.77	415	1890	2660	NA	
<i>terrain</i>	65	65	6289	2606	0.1642	0.55	455	1432	2132	53	
<i>derive</i>	12	61	562	1619	4.34	19.3	377	1707	2364	49	
hybrid <i>train</i>	65	65	6289	2606	0.1642	0.55	545	1574	2020	NA	
hybrid <i>drive</i>	12	61	562	1619	4.34	19.3	511	1306	1818	NA	

Table 1  
Acoustic properties of stimulus recordings<sup>7</sup>

unaffricated (hybrid) clusters, whose noise bursts were taken from *terrain* and *derive*. Additionally, the formant measures and noise spectra contribute to a large acoustic difference between the cluster-initial and affricate-initial items – there is no hint of any *train-chain* or *drive-jive* merger. The center of gravity (COG) of the noise spectrum is, as expected, highest for the alveolar burst in *terrain* (see FN 5 below), while the low spectral center of gravity in *derive* may be a function of the weakness of the burst noise, since its high positive values for skewness and kurtosis indicate that higher frequencies are also well represented in its spectrum.

### 3.3 Procedure

Listeners performed a two-alternative forced choice identification task in E-Prime 2.0. In each trial, an orthographic representation of two choices were displayed on the screen for 500 ms, after which the stimulus, a recording of a single item played.

There were two types of trials presented to listeners, based on the first two research questions provided above. In one of the types, based on RQ#1, listeners heard either an affricated cluster or an affricate and had to choose which of the two they heard. That is, they would see e.g. *train* and *chain* displayed on the screen, and have to select one of these options after having heard a single-word stimulus. In the other type, based on RQ#2, listeners heard either a hybrid (non-affricated) cluster or a CəC-initial word, and had to decide which they heard (*train-terrain*). A third type of trial, forcing listeners to choose between affricate-initial and CəC-initial words (*chain-terrain*) was also included in the experiment to give equal representation among the correct response to each of the six target words, with the goal of minimizing bias toward or against any particular word. These trials, however, were not related to any of the research questions, and were not included in the analysis.

[7] COG is the spectral center of gravity of the frication noise. SD is the standard deviation of the noise spectrum. Skewness refers to the lack of symmetry in the shape of the noise spectrum. Kurtosis is a measure of the sharpness of the spectral peak of the noise.

A summary of the trial types analyzed in this study, along with the research question they relate to, is given below.

- Naturally produced cluster stimulus (*train, drive*) with affricate distractor (RQ1)
- Affricate stimulus (*chain, jive*) with cluster distractor (RQ1)
- Unaffricated cluster stimulus (*hybrid train, hybrid drive*) with CəC-word distractor (RQ2)
- CəC-word stimulus (*terrain, derive*) with cluster distractor (RQ2)

Trials were organized into two blocks, while the left-right arrangement of the correct choices was reversed in the second block, to control for possible effects of handedness on the part of the participants. Each block contained 16 trials related to this experiment, along with 60 fillers that were part of unrelated experiments dealing with vowel quality and consonant voicing. Thus in total, each participant heard 32 items with affricate-affrication trials. Participants listened to the stimuli using high quality headphones in a quiet environment. Listeners entered their responses using the <1> and <0> keys, and were instructed to keep a finger on each of these keys, and to respond as quickly as possible. E-Prime recorded accuracy, as well as response time, which was measured from the onset of the voiced portion of the audio signal. Trials with response times of less than 150 ms and greater than 5000 ms were eliminated, taken to be false alarms and hesitations, respectively.

### 3.4 Analyses

Statistical analyses were carried out with the help of the SPSS statistical software (IBM corporation, 2017). Accuracy was analyzed with a mixed-effects binary logistic regression model, with a Group\*Target interaction term as the main fixed predictor of interest, and by-participant random slopes and intercepts. The acoustic properties of the stimuli, as well as the token frequency of the target word in the iWeb corpus of English (<https://www.english-corpora.org/iweb/>), were included as control variables (for frequency scores from the iWeb corpus, see Appendix B). Response Time (RT) observations were converted to their  $\text{Log}_{10}$  values to eliminate the skewness in their distribution.  $\text{Log}_{10}\text{RT}$  of correct responses was analyzed as a dependent variable in a linear mixed-effects model with a similar structure to the logistic model.

### 3.5 Results

Mean identification accuracy for the two groups is given in Figure 3. As is clear in the figure, identification accuracy was near ceiling level for both groups of listeners. The only noticeable deviation from this pattern is *derive*, which was identified correctly only 89% of the time by the Polish listeners. This effect approached but did not quite reach significance ( $\beta = -1.42$ ,  $\exp\beta = 0.243$ ,  $\text{std.e.} = 0.741$ ,  $t = -1.91$ ,  $p = .056$ ; reference level = *drive*).

ALL TRS ARE NOT CREATED EQUAL

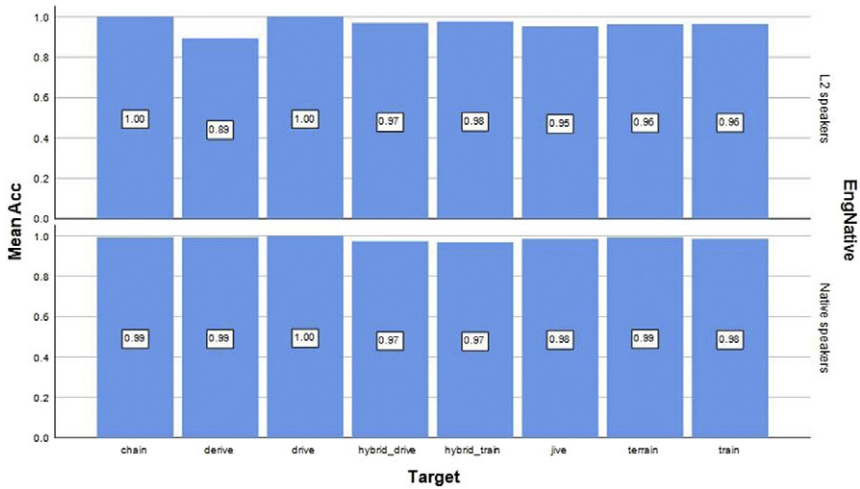


Figure 3

Proportion of correct responses as a function of stimulus item. Top panel: L2 speakers. Bottom panel: native speakers.

Mean raw RT results as a function of Target item are provided in Figure 4. As can be observed in the figure, L2 speakers were slower overall in their responses (this was also observed in the experiments making use of the filler items), and showed different effects of target item on RT than native speakers. The estimates of the linear model looking at the Group\*Target interaction are given in Table 2. The estimates represent the differences in LogRT between each Target, which was sum-coded, and the overall mean within each group. A full coefficient table is given in Appendix C.

In Table 2, the between-group difference in the effects of the hybrid items with unaffricated clusters stands out. Both the voiced and voiceless hybrid items (unaffricated *train* and unaffricated *drive*) induced slower responses (positive estimates) for the L1 English speakers, and faster responses (negative estimates) for the L2 speakers. The affricate-initial items (*chain* and *jive*) induced faster response times across both groups. We can also see that the C<sub>0</sub>C-initial items (*terrain* and *derive*) induced slower RTs for the L2 speakers, but had no significant effects on RT in the native speaker group. Finally, the naturally produced affricated clusters (*train* and *drive*) induced quicker responses for native listeners, while for L2 listeners, only the voiceless cluster had an effect, inducing slower responses.

### 3.6 Discussion

As pointed out above, the most important result of the perception study concerns the effects of affrication, or lack thereof, on the identification of cluster-initial English



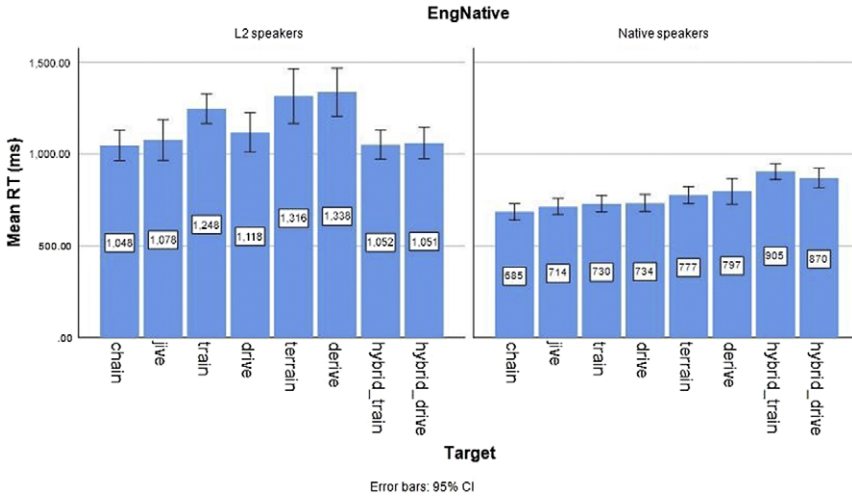
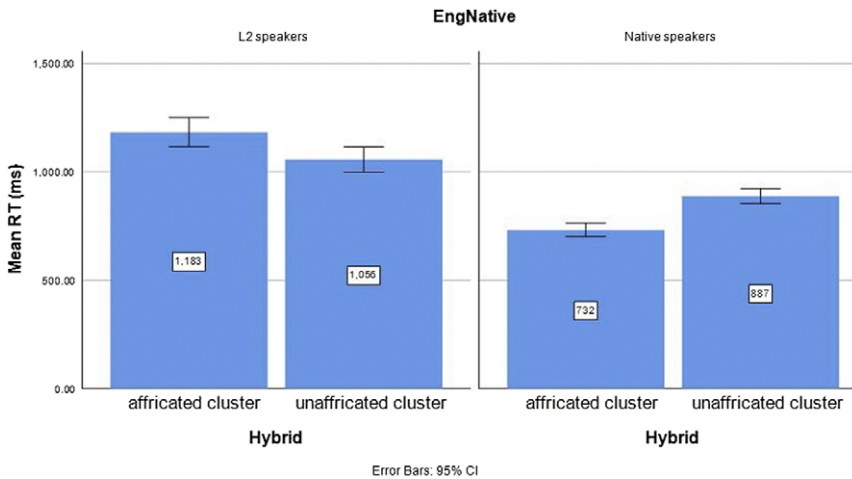


Figure 4 Mean response times for each stimulus item. Left panel: L2 speakers. Right panel: Native speakers. Error bars represent 95% confidence intervals.

Group	Target	Estimate	Std. Error	<i>t</i>	<i>p</i>
Native speakers	<i>train</i>	-0.02	0.009	-2.183	.029
	<i>terrain</i>	0.007	0.009	0.792	.428
	<i>jive</i>	-0.031	0.009	-3.437	.001
	<i>hybrid train</i>	0.068	0.007	9.876	<.001
	<i>hybrid drive</i>	0.042	0.007	6.179	<.001
	<i>drive</i>	-0.02	0.009	-2.17	.03
	<i>derive</i>	0.003	0.009	0.377	.706
	<i>chain</i>	-0.05	0.009	-5.433	<.001
L2 speakers	<i>train</i>	0.05	0.011	4.493	<.001
	<i>terrain</i>	0.06	0.011	5.306	<.001
	<i>jive</i>	-0.037	0.011	-3.312	.001
	<i>hybrid train</i>	-0.044	0.008	-5.326	<.001
	<i>hybrid drive</i>	-0.05	0.008	-5.94	<.001
	<i>drive</i>	-0.012	0.011	-1.112	.266
	<i>derive</i>	0.066	0.011	5.931	<.001
	<i>chain</i>	-0.033	0.011	-2.934	.003

Table 2 Contrast estimates of linear model for LogRT. Estimates represent the difference *t* between the LogRT for each individual target item and the within-group grand mean.

words by native listeners and by L1 Polish learners of English. Non-affricated clusters facilitated cluster identification for the Polish listeners, while hindering it for the native listeners. To offer a more direct insight on this finding, Figure 5 presents mean raw RT results for cluster-initial items as a function of affrication,



*Figure 5*

Mean RT as a function of affrication in cluster-initial targets. Left panel: L2 speakers. Right panel: Native speakers. Error bars represent 95% confidence intervals.

sorted by group. In what follows, we consider possible interpretation of these results.

At first glance, it appears that differing effects of affrication could be explained in terms of cross-language comparison at the allophonic phonetic level, without any phonological implications or consequences. The absence of the expected allophonic process associated with English TR clusters made the hybrid items more difficult for native listeners to identify, hence the slower response times. By contrast for Polish listeners, the lack of affrication may be claimed to have rendered the items more phonetically similar to unaffricated /tr dr/ in L1, thereby inducing more rapid identification.

The problem with this interpretation is that the unaffricated items are acoustically quite different from what we would expect for initial /tr/ and /dr/ in Polish. In particular, Polish /r/ is associated with intrusive vocoids that interrupt the cluster (Dłuska 1986, Cieślak 2015). Thus, Polish-accented productions of *train* and *drive* would yield [t<sup>ɹ</sup>rejɲ] and [d<sup>ɹ</sup>rajf], respectively. These pronunciations are phonetically more similar to English *terrain* and *derive* than to affricated *train* and *drive*, especially considering the fact that Polish /r/ is frequently realized as an approximant (Jaworski & Gillian 2011). As a consequence, Polish listeners' lower accuracy score for *derive* (Figure 3, top panel) and slower response times for *derive* and *terrain* (Figure 4, left panel), may be assumed to reflect confusion between these words and *drive* and *train*, respectively. Thus, Polish listeners apparently had trouble perceiving the schwa in the CəC-initial words, just as they frequently do not hear intrusive vocoids in their own L1 Polish clusters containing /r/.

While the above discussion points to phonetic and phonological factors responsible for the experimental results, an alternative explanation may be suggested. It is possible that Polish listeners' slower responses for *terrain* and *derive* was due to the relatively low frequency of the CəC-initial words relative to the cluster-initial words (see Appendix B), biasing listeners against the former. However, such an interpretation would be complicated by the fact that the least frequent word by far, *jive*, did not induce slower response times (see Table 2).

In a further attempt to tease apart the effects of the phonological shape of the target word from possible effects of frequency, an additional linear mixed effects model was fitted, in which a new factor, Phonological Shape (affricate-initial, CəC-initial, cluster-initial; the factor was sum-coded), and Word Frequency were compared as predictors in their interaction with Group. The results are shown in Table 3, which provides the coefficients of the model. Frequency was not a significant predictor of response time for either group (although there is a trend in this direction for the native speakers), while the phonological effects (slower responses for CəC-initial items by the Polish listeners; faster responses for affricate-initial items for native listeners) held. Thus, it appears that phonological shape of the experimental items had the most robust effects on response times in this study.

The fundamental picture that emerges from these results is that cross-language interaction governs the group-based differences. Native speakers were confused by the unaffricated cluster-initial items. Polish listeners also were confused by the stimuli. However, for the most part their confusion was not induced by the L2 allophonic process of affrication. Rather, Polish listeners had difficulty identifying CəC-initial items, presumably because they resembled between unaffricated cluster-initial items that appear in their L1. That is, asynchronous L1 clusters hampered identification of L2 items containing schwa. Since cross-language interference is attributable to *equivalence classification* (Flege 1987, 1995), by which bilinguals and L2 learners perceptually merge categories across languages, the results provide evidence that TR clusters are not phonologically equivalent in Polish and English, but that there is a structural connection between Polish cluster-initial and English

Model Term	Coefficient	Std. Error	<i>t</i>	<i>p</i>
Intercept: Polish*Affricate	3.027	0.0516	58.653	<.001
L1Eng*CəC	-0.093	0.066	-1.404	.16
LEng*Cluster	-0.054	0.0704	-0.768	.443
L1Eng*Affricate	-0.14	0.0656	-2.134	.033
Polish*CəC	0.098	0.0124	7.897	<.001
Polish*Cluster	0.017	0.0147	1.144	.253
Freq*L1Eng	-0.012	0.007	-1.75	.08
Freq*Polish	-0.006	0.0085	-0.672	.502

Table 3

Results of additional linear model comparing interaction effects of phonological shape and word frequency with Group.

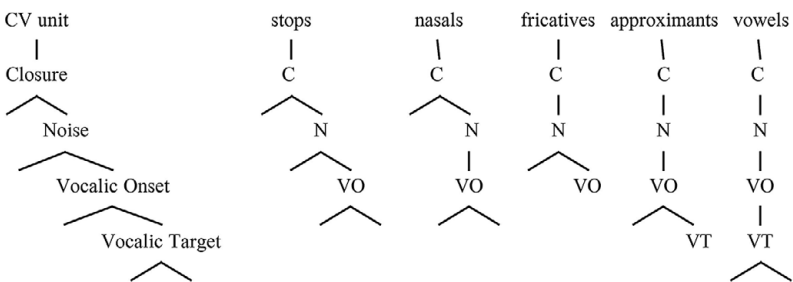
CəC-initial words. In what follows, we shall show the origins of this connection within the Onset Prominence representational environment.

#### 4. TWO TYPES OF TR CLUSTERS IN THE ONSET PROMINENCE FRAMEWORK

This section will provide a brief presentation of the Onset Prominence framework (Schwartz 2010, 2015, 2016, 2018), with particular attention to the mechanisms that produce structural configurations by which consonant clusters may be represented.

Onset Prominence representations are derived from a hierarchy of phonetic events associated with a stop-vowel sequence, typologically the most common ‘syllable’ type across languages. The fundamental building block in OP is therefore a prosodic unit, a stop-vowel CV, that provides the materials from which ‘segmental’ representations may be extracted. The stop-vowel hierarchy is shown in the tree structure on the left in (1). Each layer of the tree on the left in (1) is labeled for the phonetic event in the stop-vowel sequence from which it is derived. At the top of the hierarchy is Closure (Closure; C), the defining property of stops. The next level down is Noise (N), which is derived from aperiodic noise associated with stop release bursts, affrication, aspiration, and frication. Below Noise is the Vocalic Onset (VO) level, derived from the CV transition in the stop-vowel sequence. VO is derived from periodicity and formant movement associated with the CV transition. At the bottom of the hierarchy is the Vocalic Target (VT) node, which encodes the more or less steady formant targets associated with vowels.

#### (1) The Onset Prominence hierarchy (left) and OP manner categories



The structures in (1) reveal OP’s perspective on the relationship between prosodic and segmental units in phonology. From the tree on the left, categories of manner of articulation are derived, as shown in the remaining trees. These categories are defined in terms of binary nodes, which encode the phonetic events from the stop-vowel hierarchy that are present in the articulation of a given segment type. For example, stops constitute the top three layers, which are binary in the stops’ structure in (1), nasals lack noise bursts, so their Noise node is unary, fricatives lack Closure, as shown by the unary C node, and so on.

The representations in (1) also illustrate two inherent structural ambiguities built into the OP model. The first is the status of the VO node, which derives from the initial vocalic portion of a CV sequence directly following stop release. Acoustically, this part of the signal is vocalic, yet at the same time it bears important perceptual cues to the identity of the preceding consonant. In (1), VO is the only active node in approximants, but is also shown in the trees for obstruents and nasals. Alternatively, however, VO may be absent from the representations of nasals and obstruents, and/or present in the representation of vowels. In the OP model, VO is parametrized, with wide-ranging empirical consequences (for details, see Schwartz 2013, 2016). The other ambiguity stems from the unary nodes in the segmental trees, which represent phonetic properties that are absent in the production of a given manner category (e.g. fricatives lack Closure, nasals lack Noise bursts, etc.). Since the building block of OP representation is a stop-vowel sequence, each individual segment type is somehow ‘marked’ with respect to the CV unit tree in (1). Thus, a unary node reflects a form of ‘markedness’, a mismatch between the ‘unmarked’ CV unit and individual segmental trees. Such mismatches may motivate adjustments to the representations, with implications for phonotactics, to be discussed briefly in what follows (for more thorough discussion, see Schwartz 2016).

Since the fundamental building block in the OP framework is a CV unit, a consonant is by definition an ‘onset’ (cf. Scheer 2004), while consonant clusters (and ‘codas’) are by definition derivative entities. The default ‘onset’ status of consonants is reflected in the fact that the higher nodes of the OP hierarchy are derived from phonetic properties associated with obstruents (Closure and Noise) and approximants (VO), while vowels are at the bottom of the tree. In what follows, we shall provide a brief description of the OP phonotactic mechanisms that underlie the formation of consonant clusters. We shall see that different mechanisms are responsible for TR clusters in English as opposed to Polish, leading to different structural configurations that are predictive of the cross-language differences in cluster synchronicity investigated in this paper.

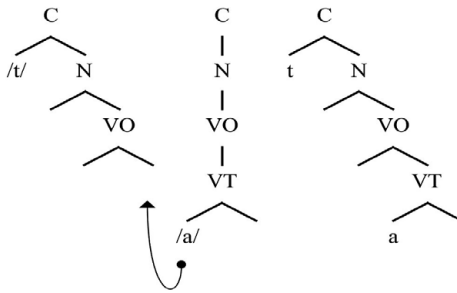
#### 4.1 *Absorption in rising sonority clusters in English*

The first and most basic phonotactic mechanism in the OP model is *absorption* (Schwartz 2016: 37). Absorption merges two trees when the tree to the left is at a higher hierarchical level. By absorption, a stop-vowel sequence may be expected to merge into a single CV unit, as in (2).<sup>8</sup>

---

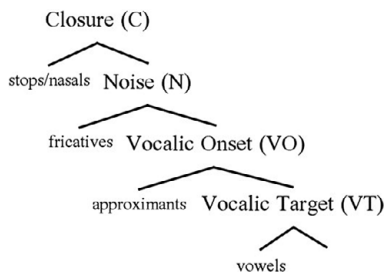
[8] The segmental symbols in (2) are used as shorthand for place and laryngeal specifications, the details of which are not directly relevant for the discussion here. Also, the convention in OP (Schwartz 2016) is to use brackets around segmental symbols when they appear on a single segmental tree, while structures that contain more than one segment are shown without brackets.

## (2) Absorption in a /ta/ sequence (after Schwartz 2016: 37)



Absorption may also be responsible for TR-type rising sonority clusters. According to the sonority sequencing principle, from the onset to the peak of a syllable, sonority should increase. The OP hierarchy directly encodes a version of this principle, defined not in terms of sonority, but rather its mirror image: consonantal strength. Stated briefly, structures with higher level binary nodes are stronger, more prominent as ‘onsets’, while more sonorous segments are lower in the hierarchy. An OP version of the sonority/strength hierarchy, in which each level is defined by the highest binary node in the structure, may be read off the trees in (1). The OP version of the hierarchy, from least to most sonorous, runs as follows: stops/nasals → fricatives → approximants → vowels.<sup>9</sup> The structure from which this hierarchy is described is given in (3), which is simply a presentation of the OP CV unit, with terminal labels corresponding to the manner categories from (1).

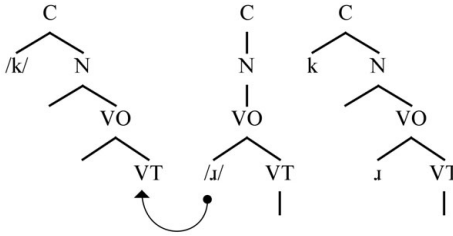
## (3) OP ‘sonority’ hierarchy derived from manner categories in (1)



[9] The place of nasals in sonority sequencing is debatable. Most languages disallow stop-nasal onsets, suggesting that nasals are relatively low in sonority. However, nasals are among the most commonly found ‘syllabic’ consonants, so theories equating sonority with syllabic peaks would describe them as quite sonorous. Krämer & Zec (2020) propose that there are two types of nasals – high sonority and low sonority. Their ‘high sonority’ nasals would be represented differently in OP, but that question is beyond the scope of this paper.

An illustration of the absorption mechanism resulting in a /kɪ/ cluster, as in English *cry*, is given in (4). In the leftmost tree, we see the structure for the stop. In the center we have the structure for the rhotic. The arrow indicates the absorption mechanism, the result of which is shown in the tree on the right, in which the stop and rhotic are contained in a single iteration of the OP hierarchy. Note that the /kɪ/ cluster is structurally equivalent to the singleton stop<sup>10</sup> – it contains active C, N, and VO nodes.

(4) Absorption in a /kɪ/ onset cluster in English *cry*



The absorbed configuration is associated with tight phonetic cohesion between the consonants in the cluster, resulting in the allophonic processes described for English in Section 2, approximant devoicing in the case of /kɪ/. When a language obeys sonority sequencing in its onset clusters, we may assume that only absorbed onsets are allowed.<sup>11</sup>

#### 4.2 Adjoined clusters in Polish

While the OP hierarchy in (2) can encode rising sonority in onsets, including clusters produced by means of absorption, the model allows for a additional mechanisms that may open the door to sonority-violating cluster types. In addition to allowing sonority violations, these mechanisms also produce TR clusters with a distinct configuration from the one we see in (4).

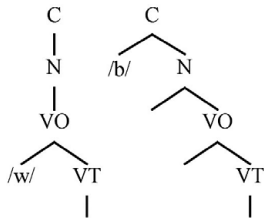
In (5), we see an approximant-stop sequence /wb/. These consonants cannot be joined via absorption, since the tree to the right is a higher-level structure.

[10] A reviewer makes an interesting point that at some level clusters like this in English are made of two components, even if they look here to be structurally equivalent to singleton stops. From the perspective of OP, English speakers would be assumed to have internalized the absorption process that is responsible for merging /t/ and /ɪ/ trees. What this means is that the linguistic inventory of English speakers includes both an absorbed TR tree (giving rise to affrication) as well as separate segmental trees for both the /t/ and the /ɪ/.

[11] English, and many familiar European languages, obey sonority sequencing, but allow one commonly encountered exception in the form of s-initial onset clusters. S-clusters may be derived by a third mechanism, which will be mentioned here, but not discussed in detail. For discussion of s-stop clusters in the OP environment, see Schwartz (2021).

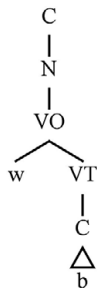


(5) OP representations of /w/ and /b/



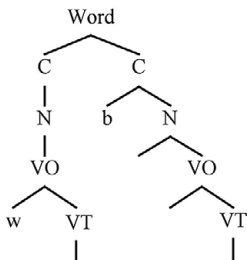
In Polish, /wb/ is an attested onset cluster, as in the word *tba* /wba/ ‘head (pejorative; gen sg)’. In the OP model, there are two possibilities for combining these two structures. In one configuration, the second consonant is *submerged*, resulting in a structure where the /b/ lies underneath the structure of the /w/. This is shown in (6).

(6) *Submerged* /wb/ cluster



The other possibility is that the consonants may be *adjoined* at a higher level of structure, as shown in (7). The next higher level is labeled ‘Word’, reflecting assumptions about the shape of minimal prosodic words in Polish (Schwartz 2016: 58).

(7) *Adjoined* /wb/ cluster

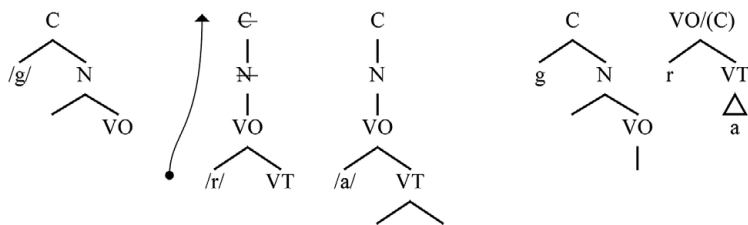


In the OP environment, the submersion and adjunction mechanisms, in addition to their role in cluster phonotactics shown in (6) and (7), also govern a range of other aspects of phonological systems. On the basis of evidence discussed in Schwartz (2016), Polish is posited as an adjunction system. This evidence includes

sub-segmental phonetic details such as coda stop release, prosodic organization characterized by fixed and relatively weak stress, phonetic as opposed to phonological vowel reduction, and the behavior of segmental phonetics as a function of prosodic position (Wojtkowiak & Schwartz 2019; Wojtkowiak 2020). Thus, we assume that if a TR cluster in Polish is not absorbed, it must be adjoined (rather than submerged), and what requires explanation is why Polish TR clusters are not absorbed.

This origins of an explanation may be found in the OP structural ambiguities discussed at the beginning of this section: the status of the VO node, and the presence of unary nodes. In Polish, VO is posited as a crucial element in the representation of vowels, with a range of empirical implications outlined in Schwartz (2016). A potential consequence of vocalic VO affiliation in Polish is that vowels and sonorant consonants are both housed at the VO level, creating ambiguity between the two categories. This ambiguity, in turn, motivates a process of *promotion* (Schwartz 2016: 55), which eliminates the sonorants' unary Closure and Noise nodes, which themselves are 'marked' in relation to binary nodes in the OP system, and raises the VO node that remains to the level of Closure. Sonorant promotion in the Polish word *gra* 'game' is illustrated in (8).

(8) TR cluster in Polish *gra* 'game' with promoted /r/



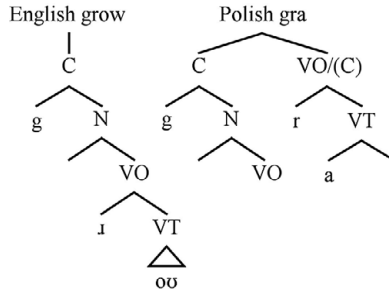
Notice that the /r/ and /a/, represented in the second and third trees from the left, are both at the VO level, so the /a/ cannot be absorbed into the /r/. In other words, it is the need to absorb the vowel that motivates promotion, which in turn allows the sonorant to absorb the following vowel as shown in the rightmost tree. At the same time, however, the promoted sonorant cannot be absorbed into the stop, so the ensuing cluster must be adjoined.<sup>12</sup>

In (9), the English word *grow*, with an absorbed stop-liquid /gr/ cluster is presented alongside the Polish word *gra*, with an adjoined stop-liquid cluster. The two distinct structural configurations, for what textbook descriptions would consider the "same" cluster, encode the different prosodic status of onset clusters in the two languages discussed in 2.3.3. In Polish, the minimal structure required for inflectional morphology contains two trees built down from the Closure level – either (C)VC or CCV

[12] The representations in (8) also illustrate an assumption that phonotactic mechanisms in OP proceed from right to left, at least in the languages whose phonotactics OP has given a detailed treatment: English, Polish, Tashlhiyt Berber (see Schwartz 2015 for an account of Tashlhiyt syllabification).

(Schwartz 2016: 58). In (9) – both consonants in the Polish cluster appear as singleton onsets, each in a separate tree. In English, onset clusters, and onsets in general, play no role in defining minimality for prosodic words.

(9) English *grow* vs. Polish *gra* ‘game’: absorbed vs. adjoined TR clusters



With regard to phonetics, the structures in (9) encode the language-specific differences in cluster synchronicity described in Section 2, and investigated in the experiment described in Section 3. In the English tree, the two consonants are contained in a single iteration of the OP hierarchy, and should exhibit greater phonetic synchronicity than in Polish, where the consonants are housed in separate trees.

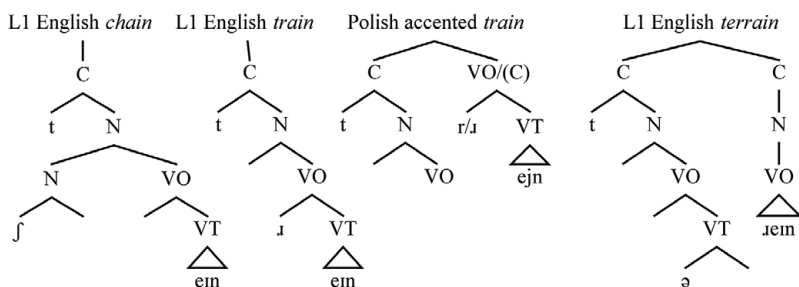
Note also in (9) that the ‘syllable’ and the ‘nucleus’ have no formal status in the OP model, but are emergent entities that evolve differently in different languages. Both of these words are one-syllable words within the confines of language-specific phonological intuition, yet in OP they are structurally distinct. Likewise, OP structures are not projected from a syllabic ‘nucleus’, but are built down from the Closure level of the hierarchy.<sup>13</sup> For this reason, the phonetic predictions of the structures in (9), when considered against the EMA studies discussed in Section 1 (Shaw et al. 2009; Hermes et al. 2017), encompass only target-to-target lag measures, which describe the cohesion between consonants in a cluster, and not center-to-anchor measures, which describe onset-nucleus coordination. From the OP perspective, Hermes et al.’s (2017) conflicting EMA findings for Polish (center-to-anchor stability suggesting complex onsets vs. long target-to-target lags suggesting simplex onsets) should be resolved in favor of the target-to-target lags, since center-to-anchor stability makes reference to nuclei, which have no formal status in the model.

[13] This is the fundamental difference between OP and a CVCV approach (e.g. Scheer 2004), in which all consonant clusters contain an empty nucleus, although both approaches share the crucial assumption that the CV unit is fundamental building block of phonology. The English structure in (9) is analogous to structures resulting from CVCV’s ‘infrasegmental government’, by which empty nuclei are rendered invisible. The Polish structure in (9) would resemble a CVCV sequence without infrasegmental government. Therefore, the CVCV approach is also capable of capturing the cross-language difference, in what some would consider phonetic detail, that is the focus of this paper.

4.3 *TR affrication and the lack thereof*

Returning to the data examined in this paper, relevant OP structures for the experimental items are given in (10). In the two trees on the left we see OP representations for affricate-initial *chain*<sup>14</sup> and cluster-initial *train*, respectively, in L1 English. On the right we see L1 English *terrain*, and in the second structure from the right we see Polish-accented *train*. Recall from the experiment that Polish listeners had the most difficulty identifying CəC-initial words. This is presumably because they would attribute the English schwa to the type of intrusive vowel that is common in Polish /tr/ and /dr/ clusters. In this connection, note the structural parallel in (10) between the Polish-style adjoined cluster (Polish-accented *train*) and the CəC-initial words – both structures contain two separate trees built down from Closure.

## (10) OP structures for experimental items



At the same time, we may also assume that the English affricates have distinct representations, since they did not cause confusion for either listener group. The proposed representation in (10), with an additional left-branching node off of the Noise level, is intended to capture both the sibilant noise associated with post-alveolar frication, as well as the inability of affricates to form onset clusters in English. Nevertheless, working out the specific details about how affricates should be represented in OP remains a task for future research.

## 5. CONCLUSION

This paper has presented the results of a perception experiment examining the effects of TR affrication, or the lack thereof, on the identification of cluster-initial words in English by both L1 and L2 listeners whose L1 is Polish. The most interesting results of the study are those bearing on RQ2 and RQ3, in which L1 Polish and L1 English listeners showed opposite effects of items with unaffricated TR clusters, and their potential for confusion with /tər/- and /dər/-initial words. These findings are compatible with the claim that TR clusters in English and Polish

[14] The structure for *chain* is only a tentative proposal, since at this point in its development, phonologists working in OP have not yet had a chance to do a thorough investigation of the possible consequences of various possibilities available for the representation of affricates.

are characterized by distinct structural configurations, which govern their relative phonetic synchronicity and the probability of allophonic processes such as TR affrication, as well as the prosodic behavior of the cluster in terms of word minimality. It has been shown how the distinct structural configurations may be derived in the Onset Prominence representational framework.

APPENDIX A – SPECTROGRAMS OF STIMULUS ITEMS

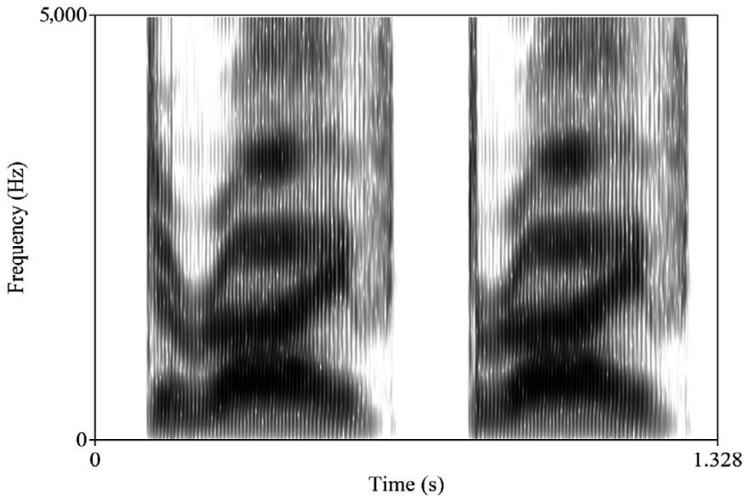


Figure 6  
*Derive* (left) and *hybrid drive*

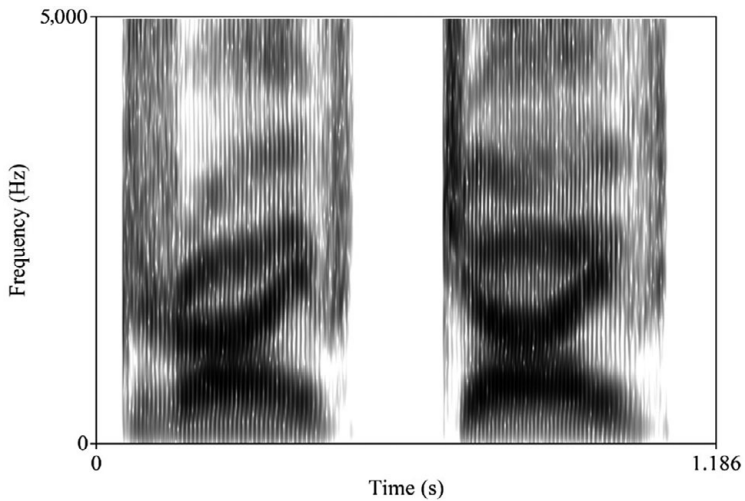
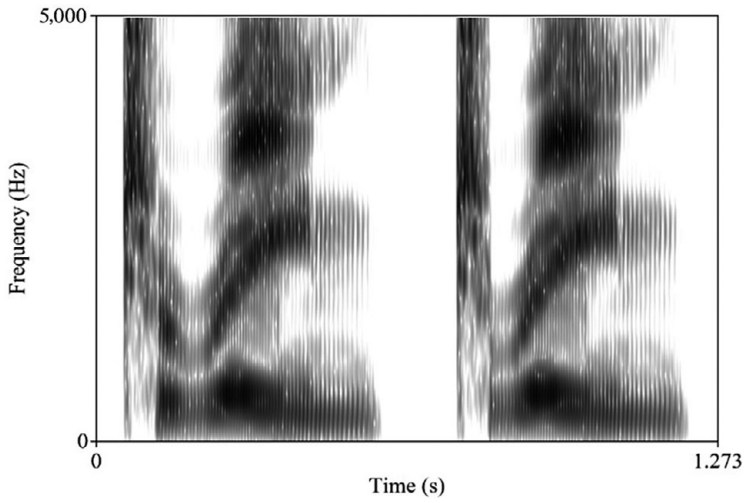
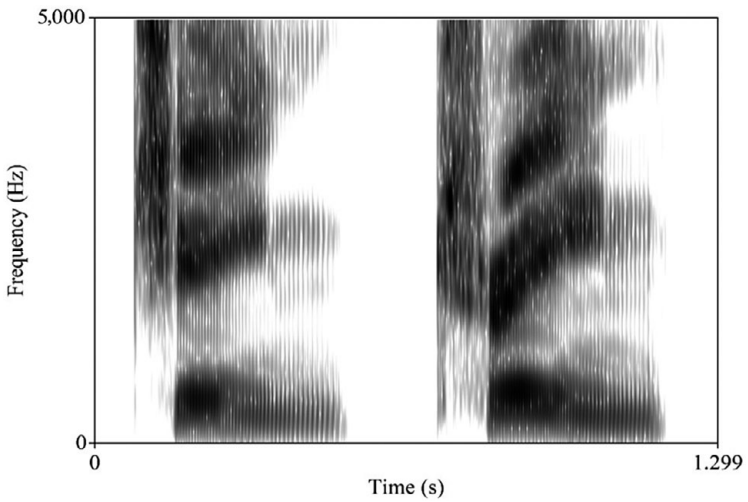


Figure 7  
*Drive* (left) and *jive*



*Figure 8*  
*Terrain (left) and hybrid train*



*Figure 9*  
*Chain (left) and train*

ALL TRS ARE NOT CREATED EQUAL

APPENDIX B – TOTAL FREQUENCY OF OCCURRENCE OF EXPERIMENTAL ITEMS IN IWEB CORPUS ([HTTPS://WWW.ENGLISH-CORPORA.ORG/IWEB/](https://www.english-corpora.org/iweb/))

*jive* 10029  
*derive* 57837  
*terrain* 209361  
*chain* 716050  
*train* 955542  
*drive* 2367157

APPENDIX C – ADDITIONAL STATISTICAL INFORMATION

Model Term	Coefficient	Std. Error	<i>t</i>	<i>p</i>	95% C.I. - lower	95% C.I. - upper
Intercept	2.847	0.022	127.82	<.001	2.803	2.891
[target=train]* [EngNative=No]	0.154	0.038	4.01	<.001	0.079	0.229
[Target=chain]* [EngNative=No]	-0.03	0.013	-2.13	.033	-0.057	-0.002
[Target=chain]* [EngNative=Yes]	0.252	0.038	6.58	<.001	0.177	0.327
[Target=derive]* [EngNative=No]	0.023	0.013	1.67	.093	-0.004	0.051
[Target=derive]* [EngNative=Yes]	0.174	0.038	4.54	<.001	0.099	0.249
[Target=drive]* [EngNative=No]	0	0.013	0.01	.993	-0.027	0.027
[Target=drive]* [EngNative=Yes]	0.137	0.037	3.66	<.001	0.063	0.21
[Target=hybrid_drive] *[EngNative=No]	0.062	0.012	5.16	<.001	0.039	0.086
[Target=hybrid_drive] *[EngNative=Yes]	0.142	0.037	3.79	<.001	0.069	0.215
[Target=hybrid_train] *[EngNative=No]	0.088	0.012	7.26	<.001	0.064	0.111
[Target=hybrid_train] *[EngNative=Yes]	0.15	0.038	3.90	<.001	0.074	0.225
[Target=jive]* [EngNative=No]	-0.011	0.013	-0.82	.411	-0.039	0.016
[Target=jive]* [EngNative=Yes]	0.246	0.038	6.41	<.001	0.171	0.321
[Target=terrain]* [EngNative=No]	0.027	0.013	1.95	.051	0	0.055
[Target=terrain]* [EngNative=Yes]	0.236	0.038	6.16	<.001	0.161	0.311
[Target=train]* [EngNative=No]						

Table 4

Full coefficient table of linear model described in 3.5 (cf. estimates in Table 2).



## REFERENCES

- Albright, Alan. 2007. Natural classes are not enough: Biased generalizations in novel onset clusters. Cambridge, MA: MIT ms.
- Beckman, Jill, Michael Jessen & Catherine Ringen. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics* 49, 259–284.
- Blevins, Juliette. 1995. The syllable in phonological theory. In John A. Goldsmith, (ed.), *The handbook of phonological theory*. Blackwell Handbooks in Linguistics no. 1, 206–244. Oxford: Blackwell Publishers.
- Browman, Catherine & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6, 201–251.
- Carley, Paul & Inger Mees. 2020. *American English Phonetics and Pronunciation Practice*. London: Routledge.
- Chabot, Alex 2019. What’s wrong with being a rhotic. *Glossa – a journal of general linguistics* 4(1), 38.
- Cieślak, Joanna 2015. English-Polish differences in the organization of consonant clusters: a study in phonetic code-switching. MA Thesis, Adam Mickiewicz University in Poznań.
- Clements, G.N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary E. Beckman (eds.), *Papers in Laboratory Phonology I. Between the Grammar and Physics of Speech*, 283–333. Cambridge: Cambridge University Press.
- Comrie, Bernard. 1976. Irregular stress in Polish and Macedonian. *International Review of Slavic Linguistics* 1. 227–240.
- Cruttenden, Alan 2014. *Gimson’s Pronunciation of English*. London: Routledge.
- Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis & Ingrid Norman. 2011. Explaining sonority projection effects. *Phonology* 28, 197–234.
- Dluska, Maria, 1986. *Fonetyka Polska* [Polish phonetics]. Warsaw: Państwowe Wydawnictwo Naukowe.
- Dukiewicz, Leokadia & Irena Sawicka. 1995. *Gramatyka współczesnego języka Polskiego – fonetyka i fonologia* [A grammar of modern Polish – phonetics and phonology]. Krakow: Wydawnictwo Instytutu Języka Polskiego PAN.
- Durvasula, Karthik & Jimin Kahng. 2015. Illusory vowels in perceptual epenthesis: the role of phonological alternations. *Phonology* 32(3), 385–416.
- Dziubalska-Kończak, Katarzyna. 2014. Explaining phonotactics using NAD. *Language Sciences* 46, 6–17.
- Flege, James E. 1987. The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 162–177. [https://doi.org/10.1016/S0095-4470\(19\)30537-6](https://doi.org/10.1016/S0095-4470(19)30537-6)
- Flege, James E. 1995. Second language speech learning – theory, findings, and problems. In Winnifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language research*, 233–277. Timonium, MD: York Press.
- Garrett, Edward. 1999. Minimal words aren’t minimal feet. *UCLA Working Papers in Linguistics* 111, 92–122.
- Goldstein, Louis, Ioana Chitoran & Elisabeth Selkirk. 2007. Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. In Jurgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 241–244. Dudweiler, Saarbrücken: Pirrot.
- Gussmann, Edmund. 2007. *The phonology of Polish*. Oxford: OUP.
- Hall, Nancy. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23, 387–429.
- Harris, John. 1994. *English Sound Structure*. Oxford: Blackwell.
- Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39, 379–440.
- Hermes, Anne, Doris Mücke & Martine Grice. 2013. Gestural coordination of Italian word-initial clusters: The case of “impure s”. *Phonology* 30 (1), 1–25.
- Hermes, Anne, Doris Mücke & Bastian Auris. 2017. The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics* 64, 127–144.

- Honeybone, Patrick. 2005. Diachronic evidence in segmental phonology: The case of obstruent laryngeal specifications. In Marc van Oostendorp & Jeroen van de Weijer (eds.), *The internal organization of phonological segments*, 319–354. Berlin: Mouton de Gruyter.
- Honorof, Douglas & Catherine Browman. 1995. The center or edge: How are consonant clusters organized with respect to the vowel? In Kjell Elenius & Peter Branderud (eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 552–555.
- Hooper, Joan B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.
- Iverson, Gregory & Joseph Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12, 369–396.
- Jarosz, Gaja. 2017. Defying the stimulus: Acquisition of complex onsets in Polish. *Phonology* 34 (2), 269–298.
- Jaworski, Sylwester & Edward Gillian. 2011. On the phonetic instability of Polish /r/. *Poznan Studies in Contemporary Linguistics* 47(2), 380–398.
- Kang, Yoonjung. 2011. Loanword phonology. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume & Keren Rice (eds.), *Companion to Phonology*, 2258–2282. Malden, MA: Wiley Blackwell. <https://onlinelibrary.wiley.com/doi/10.1002/9781444335262.wbctcp0095>
- Keating, Patricia. 1979. A phonetic study of the voicing contrast in Polish. PhD dissertation, Brown University.
- 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, 2400–2411.
- Krämer, Martin & Draga Zec. 2020. Nasal consonants, sonority and syllable phonotactics: the dual nasal hypothesis. *Phonology* 37, 27–63.
- Leung, Alex Ho-Cheong, Martha Young-Scholten, Wael Almurashi, Saleh Ghadanfari, Chloe Nash & Olivia Outhwaite. 2021. (Mis)perception of consonant clusters and short vowels in English as a foreign language. *International Review of Applied Linguistics in Language Teaching*. Published online by Mouton de Gruyter, August 9, 2021. <https://doi.org/10.1515/iral-2021-0030>
- Marin, Stefania & Marianne Pouplier. 2010. Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control* 14, 380–407.
- Orzechowska, Paula. 2019. *Complexity in Polish phonotactics: On features, weights, rankings and preferences*. Singapore: Springer Nature.
- Pastätter, Manfred & Marianne Pouplier. 2017. Articulatory mechanisms underlying onset-vowel organization. *Journal of Phonetics* 65, 1–14.
- Parker, Steven. 2002. Quantifying the sonority hierarchy. Ph.D. dissertation, University of Massachusetts at Amherst.
- Rubach, Jerzy. 1996. Non-syllabic analysis of voice assimilation in Polish. *Linguistic Inquiry* 27, 69–110.
- Scheer, Tobias. 2004. *A lateral theory of phonology. Vol 1: What is CVCV, and why should it be?* Berlin: Mouton de Gruyter.
- Schwartz, Geoffrey. 2010. Phonology in the speech signal – unifying cue and prosodic licensing. *Poznan Studies in Contemporary Linguistics* 46(4), 499–518.
- Schwartz, Geoffrey. 2013. A representational parameter for onsetless syllables. *Journal of Linguistics* 49 (3), 613–646.
- Schwartz, Geoffrey. 2015. Who needs a nucleus? Tashlhiyt Berber syllabification in the Onset Prominence representational environment. *Poznan Studies in Contemporary Linguistics* 51(2), 247–290.
- Schwartz, Geoffrey. 2016. On the evolution of prosodic boundaries – parameter settings for Polish and English. *Lingua* 171, 37–74.
- Schwartz, Geoffrey. 2018. Towards a typology of consonant synchronicity. Paper presented at LabPhon 16 session on phonotactics. <http://labphon16.labphon.org/files/workshop/se-3-abstracts/002.pdf>.
- Schwartz, Geoffrey. 2021. Spit or sbit? ST-type clusters in the speech of Polish learners of English. *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 138 (4): 187–199.
- Schwartz, Geoffrey & Daria Arndt. 2018. Laryngeal Realism vs. Modulation Theory – evidence from VOT discrimination in Polish. *Language Sciences*, 69, 98–112. <https://doi.org/10.1016/j.langsci.2018.07.001>
- Schwartz, Geoffrey, Ewelina Wojtkowiak & Bartosz Brzoza. 2019. Beyond VOT in the Polish laryngeal contrast. *Proceedings of the 19th International Congress of Phonetic Sciences*. Melbourne, Australia. [https://icphs2019.org/icphs2019-fullpapers/pdf/full-paper\\_256.pdf](https://icphs2019.org/icphs2019-fullpapers/pdf/full-paper_256.pdf)

- Schwartz, Geoffrey, Anne Hermes & Radosław Świąciński. 2021. The phonology of initial cluster coordination – a cross-language articulographic study. Talk given at the 18<sup>th</sup> Old World Conference in Phonology.
- Scobbie, James, Olga Gordeeva & Benjamin Matthews. 2006. Acquisition of Scottish English phonology: an overview. *QMU Speech Science Research Centre Working Papers*, WP-7.
- Selkirk, Elisabeth. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
- Shaw, James, Adamantios Gafos, Phillip Hoole & Chakir Zeroual. 2009. Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability in articulation. *Phonology* 26, 187–215.
- Stevens, Kenneth. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.
- Strycharczuk, Patrycja. 2012. Sonorant transparency and the complexity of voicing in Polish. *Journal of Phonetics* 40(5), 655–671.
- Świąciński, Radosław. 2012. Acoustic aspects of palatalization in English and Polish – a study in Laboratory Phonology. Ph.D. dissertation. UMCS Lublin.
- Tilsen, Sam, Draga Zec, Christina Bjorndahl, Beckey Butler, Marie-Josée L’Esperance, Alison Fisher, Linda Heimisdottir, Margaret Renwick & Chelsea Sanker. 2012. A cross-linguistic investigation of articulatory coordination in word-initial consonant clusters. *Cornell Working Papers in Phonetics and Phonology* 3. 51–82.
- Tsuchida, Ayoko, Abigail C. Cohn & Masanobu Kumada. 2000. Sonorant devoicing and the phonetic realization of [spread glottis] in English. *Working Papers of the Cornell Phonetics Laboratory* 13, 167–181.
- Wells, John. 2011. how do we pronounce train? (Blog post, 22 Mar. 2011). John Wells’ Phonetic Blog (<http://phonetic-blog.blogspot.com/2011/03/how-do-we-pronounce-train.html>).
- Wojtkowiak, Ewelina. 2020. Prosody-segment interactions in the acoustics of Polish front vowels. *Studies in Polish Linguistics* 15(3), 151–175.
- Wojtkowiak, Ewelina & Geoffrey Schwartz. 2019. Prosody-segment interactions in the acoustics of Polish plosives. In Calhoun, Sasha [et al.] (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia 2019, 3285–3289. Australasian Speech Science and Technology Association Inc.

*Author’s address:* Faculty of English, Adam Mickiewicz University  
ul. Grunwaldzka 6, 60–780 Poznań POLAND  
[geoff@amu.edu.pl](mailto:geoff@amu.edu.pl)