

## PHONOLOGY IN THE SPEECH SIGNAL – UNIFYING CUE AND PROSODIC LICENSING

GEOFFREY SCHWARTZ  
*Adam Mickiewicz University, Poznań*  
*geoff@ifa.amu.edu.pl*

### ABSTRACT

This paper is offered in commemoration of Prof. Edmund Gussmann, who passed away sadly and unexpectedly just a few short weeks before the 41st Poznań Linguistic Meeting, where the paper was presented. The PLM session, *Competing Explanations in Phonology*, was the type of gathering at which Prof. Gussmann would thrive, advancing his strong theoretical position that phonetics is irrelevant for phonological theory (Gussmann 2004). Prof. Gussmann argued for this view in an animated and sometimes provocative manner, but he always did so with charm and good nature. My own views on the role of speech in phonology differ sharply from Prof. Gussmann's. I am nevertheless quite grateful for his perspective, which has indeed changed the way I think of speech. Under the influence of Government Phonology, I have adopted a phonological view of the acoustic signal, which seeks to challenge phoneticians with new hypotheses about the way speech interacts with grammar. This paper explicates this perspective, and applies it to a recent case, cue vs. prosodic licensing, in which “phonetic” and “phonological” explanations seemed to be at an impasse. Thanks in part to Prof. Gussmann's strong theoretical position, I have developed a new theory of constituency that offers a vehicle with which we may reconcile competing views on the underpinnings of phonological licensing.

KEYWORDS: Phonetics; phonology; explanation; prosodic licensing; licensing by cue.

### 1. Phonetic vs. Phonological explanations

Experimental research in the speech sciences is a much more recent endeavor than the impressionistic study of phonetics and phonology, blossoming only in the second half of the twentieth century. As one would expect, early research into speech perception tended to test hypotheses that originated in established linguistic theory. Two of these hypotheses (see e.g. Goldinger et al. 1994), a linearity condition by which a phonological segment must correspond with a portion of the signal, and an invariance condition that a given phonological feature must have a constant signature in the acoustic signal, were quickly falsified. Difficulties in the segmentation of speech, especially utterances containing glides, proved the linearity condition untenable – boundaries in the signal rarely

correspond with individual segments. Additionally, coarticulation was found to cause a large amount of acoustic variability for a single phoneme, a finding that invalidated the invariance hypothesis. The failure of speech to meet the linearity and invariance conditions posed a serious challenge to phonological theory, which had always relied on the supposed discreteness of phonological units. Consequently, many researchers, both on the phonetic and phonological side, came to the conclusion that it was undesirable and even impossible to reconcile the discrete symbolic aspects of phonological grammars with the dynamic and continuous phenomena observed in speech.

Over the last quarter of the 20th century, these issues seemed to carry with them a general communication failure between phoneticians and phonologists. Phoneticians became more and more interested in gradient aspects of speech, occasionally denying the validity of phenomena that had traditionally been described categorically. Phonologists, on the other hand, tended to dismiss experimental studies, suggesting that either they were methodologically flawed, or that they touched upon non-grammatical aspects of speech, such as orthographic influence or hyper-correctness. The debate (Manaster-Ramer 1996; Port 1996) over “incomplete neutralization” of final voice contrasts is a case in point. A series of experimental phonetic studies (e.g. Slowiaczek and Dinnsen 1985; Dinnsen and Charles-Luce 1984) found small but systematic acoustic differences in apparently neutralized word final voicing contrasts in languages such as German, Catalan, and Polish. On the basis of such findings the suggestion was advanced that contrast neutralization as understood in the phonological sense is impossible. Instead of arguing that phonological theories must be revised to account for such findings, phonetics researchers apparently sought to challenge the very foundations of phonological study. Thus, in Dinnsen and Charles-Luce (1984), we read that “*contrary to all phonological accounts*, the underlying voice distinction was not neutralized (Dinnsen and Charles-Luce 1984: 49; emphasis mine)”. Never did they consider the possibility of a phonological description in which incomplete neutralization was possible. Meanwhile, scholars on the other side of the divide adopted a defensive position. For example, Manaster-Ramer (1996) protests that if incomplete neutralization indeed exists “we must rethink the whole process of collecting and evaluating claims about the phonetics and phonology of the world’s languages (Manaster-Ramer 1996:480)”. Clearly, many phonologists were slow to acknowledge the possibility of small but systematic phonetic differences or of asymmetries between speech production and speech perception. This failure is echoed in Labov’s account (1997) of resistance to the concept of near-merger, which he cites as a “dramatic example of how a theoretical framework can prevent the recognition of facts” (Labov 1997: 368). The advent of Optimality Theory in the 1990s spawned new interest in the role of phonetics in phonology, offering a convenient vehicle to incorporate perceptual and articulatory constraints directly into grammar. However, as we shall see in the next section, this research tradition seemed to be more concerned with replacing traditionally assumed phonological explanations with phonetically oriented accounts, claiming that phonological analyses are inferior because they lack phonetic grounding. Thus, the divide between “phonological” phonologists and

phonetically-based phonologists has remained a deep one in which scholars from the two camps have been in competition. Rather than seeking out areas of compatibility, the primary concern became to invalidate the claims of the opposing camp.

In the heat of the debate on the role of speech in phonology, linguists seem to have lost sight of some basic scholarly tenets, impairing their ability to reconcile theoretical positions that are only superficially incompatible. Although many phonologists claim that phonology is an autonomous component of grammar that is free from the effects of speech performance, their representational devices typically make reference to the physical properties of speech. In other words, phonologists rely on established phonetic knowledge, and indeed take that knowledge for granted. This is a problematic basis for scholarly endeavor. It implicitly assumes that all phonetics has been “done”, and can be excised from phonological theory. Phonetics, however, is an active scientific field in which new discoveries are regularly made. Often these discoveries demand that phonology must revise some of its theoretical assumptions. Near-merger and incomplete neutralization, for example, suggest that a successful phonological framework must have some way of representing asymmetrical aspects of grammar in which subtle contrasts may exist but are not actively perceived. Phonology has been slow to incorporate this notion (see, however, van Oostendorp 2006 on incomplete neutralization).

Assuming that phonology is indeed an active component of grammar, phonological explanations should in principle be preferable to phonetic ones. This idea finds motivation in well-known principles of speech perception. Briefly stated, the auditory system is known to act in a “phonological” manner. It is an information reduction device that works in both the spectral and temporal domains. The entire frequency range of human audition (up to about 20 kHz) may be reduced to a relatively small number of critical bands. In the temporal domain, onset boosts serve to focus perceptual attention on smaller portions of the signal, again reducing the amount of acoustic input to be processed. Unfortunately, the phonological aspects of audition seem to have been overlooked in current phonetically-based approaches. Consider Flemming’s (2002) use of auditory representations in the formulation of OT constraints on contrast. In the representation of vowels, Flemming uses numerical scales intended to reflect the acoustic values of vowel formants. For example, in his representation of the vowel /i/, Flemming offers the following specifications: F1=1, F2=6, F3=3 (Flemming 2002: 21). While it faithfully represents the acoustic properties of the vowel, this representation does not seem to capture the linguistic ramifications of its auditory specifications. For example, a change of the F1 value from 1 to 2 would entail a perceptible change in vowel height. At the same time, if we lowered the F3 value to 2 while keeping the other values constant we would still be dealing with a /i/-like vowel. The functional primacy of F1 over F3 is not captured – Flemming’s representations are at heart phonetic and not phonological<sup>1</sup>.

<sup>1</sup> A *PSiCL* reviewer who has confessed to seeing “no merit in GP-style representations” offers the work of Paul Boersma and Andrew Wedel, in addition to Flemming, as “simple mechanisms for deriving categorical phonological patterns from quantifiable phonetic signals”. This criticism is entirely beside the point. The fo-

In sum, recent history has shown a real impasse among scholars concerning the role of phonetics in phonology. Phoneticians and phonetically-oriented phonologists have seemingly adopted a strategy of using phonetics to invalidate rather than improve phonological analyses. At the same time, phonological phonologists have seemed reluctant to consider new phonetic knowledge, adopting instead a defensive position that excludes the possibility of incorporating phonetically grounded innovations. The debate about the underpinnings of phonological licensing, which we shall turn to in the following section, represent an instructive example of this failure to communicate.

## 2. Cue vs. prosodic licensing – background

Empirical observation of phonological patterns revealed that certain segments, features, contrasts, or other phonological entities may be subject to restrictions on where they can occur. If a given feature appears in a certain position, it is said to be licensed. The criteria that determine patterns of licensing represent another area in which traditional and phonetically-oriented phonologists have experienced serious difficulties in communication.

Traditionally, phonologists made generalizations about licensing with respect to syllable positions. The view that licensing is dependent on syllable position has come to be known as Licensing by Prosody (e.g. Ito 1986). A common empirical pattern from this perspective is that licensing of consonant features including place and laryngeal specifications is observed in onset position but not in coda position. For example, a number of well-studied languages have processes of word-final devoicing, with a similar pattern in word-internal codas. These processes neutralize the [voice] contrast in favor of the voiceless obstruent. Thus, when addressing questions of licensing, scholars have posited that the feature [voice] is licensed in onset position (e.g. Rubach 1990) or before a tautosyllabic sonorant (accounting for voice contrasts in onset clusters, e.g. Lombardi 1995). As we can see, syllable-based positions found in traditional structures offer a number of useful generalizations from which phonological explanations may be formed.

As it happens, however, phonetics researchers have not been able to provide a precise physical definition of what a syllable is (see discussion in e.g. Krakow 1999). Once again we are faced with a situation in which a hypothesis with origins in traditional phonological study has not found phonetic support. Syllabic structures thus represent a parallel with the linearity and invariance conditions discussed in the previous section. A falsified hypothesis about the phonetic underpinnings of phonological structure left an

---

cus here is on the nature of the representations themselves. That is, I am not interested in deriving phonological patterns so much as characterizing them. The representations offered in this paper are constructed on the basis of acoustic signatures (in the tradition of Harris 1994), but I make no claims whatsoever about their quantification.

apparent compatibility gap between traditional phonology and phonetically based phonology. Instead of trying to bridge that gap, scholars concentrated on invalidating the claims of the rival theoretical camp. As a response to the failure to identify physical correlates of syllabic structures, phonetically-oriented phonologists established an alternative view of licensing, Licensing by Cue (Steriade 1997 et seq.), that makes no reference to syllable positions. The primary tenet of the theory is that segments, features, and contrasts are licensed in positions in which their perceptual cues are salient. In a large majority of cases, these positions correspond with those offered by the proponents of Licensing by Prosody. However, since the cue-based positions found independent phonetic motivation in speech perception research, the cue approach is argued to be superior to a syllable-based approach.

Additionally, phonetically-oriented researchers have identified important counter-examples to the prosodic generalizations. One example involves retroflexion on stops: in a number of languages, including Gujarati (Dave 1977, cited in Steriade 1997) and Norwegian (Hamann 2003), retroflex stops are licensed in post-vocalic position, while the retroflexion may not appear in onsets. From the phonetic point of view, this is indeed expected, since the articulatory approach to retroflex articulations generally requires significant duration, while their release may be instantaneous. As a result, the acoustic patterns associated with retroflexion (a lowering of F3 and F4; Ladefoged and Maddieson 1996: 26) are much more robustly housed on preceding vowels than on following ones (Ladefoged and Maddieson 1996: 27). In other words, retroflexion is more perceptually salient in post-vocalic coda position than in pre-vocalic onset position. Steriade (1997) cites Lithuanian as an additional counter-example to the generalizations of Licensing by Prosody. This language has word-final devoicing and neutralization of voice contrasts before obstruents. At the same time, voice contrasts may occur in coda position in hetero-syllabic obstruent-liquid clusters (*dump.les* vs *dumb.las*; Steriade 1997: 19). Steriade concludes that the VOT cue to voice contrasts present in pre-sonorant position, independent of syllabic affiliation, determines the licensing of the contrast.

On the whole, it appears as if the Licensing by Cue approach is superior in that it can account for cases like retroflexes that are problematic for Licensing by Prosody. What is missing from Licensing by Cue, however, is any mechanism that would predict the rarity of cases, like retroflexes, in which VC transitions are more salient than CV transitions. This rarity may of course be explained on the basis of phonetic studies – the boost in auditory response at stimulus onset predicts Onset licensing in the vast majority of cases (Wright 2004). This prediction, however, is not incorporated in the phonology of Licensing by Cue. Ideally, we should have a more restrictive system that expresses coda licensing as the unusual case that it is. Licensing by Prosody cannot in its present form explain the licensing of retroflex-based contrasts in stops, but it at least predicts that they should be rare.

The position of phonetically-based OT in the licensing debate, perhaps surprisingly, represents a problematic interpretation of experimental phonetic results. As a result of

the failure to identify reliable phonetic correlates of syllabic structures, proponents of cue licensing seem to have come to the conclusion that such correlates do not exist. Thus, in Hayes et al (2004: 23) we read that “nothing about perception, articulation, or processing leads us to expect any licensing asymmetries on the basis of syllable positions”. What they failed to consider was the possibility that experimental phonetics was looking for the correlates of the wrong syllabic structures. Indeed for a true phonetically-based phonologist, speech must contain cues to prosodic structure. If those cues are not expressible with traditional representations of structure, it is the representations that must be refined.

### 3. Onset Prominence – the structure of segmental representations

Although the cue and prosodic accounts of licensing may be at first glance incompatible, they share one very important aspect. They both explain the problem in terms of positional behavior of phonological entities. The difference is that in the cue account, the position is defined in specifications for segmental properties – pre- or post- sonorant position – while prosodic licensing defines the position from the top down. If we are to reconcile these two descriptions, we must create an environment that may unify these two perspectives. That is, prosodic information must be built into the representation of segments.

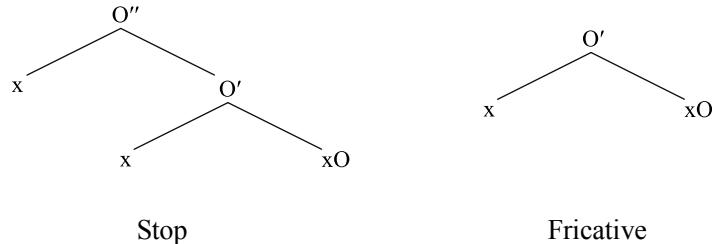
This is the strategy adopted by the theory of Onset Prominence (OP; Schwartz 2009), which builds on insights into the structural nature of segmental phonology. These insights are developed in two recent works. In the first, Golston and Hulst (1999), a model is posited in which stricture (manner of articulation) is structure, eliminating the need for a syllabification algorithm. Linear order and syllabic structure may be read directly from segmental representation. Unfortunately, the proposal offers only a rough sketch of what these segmental representations are meant to look like. Golston and Hulst’s insight about manner is flushed out in a much more explicit fashion in a second work that has been influential for the OP theory, Pöchtrager (2006), which offers an innovative structural definition of manner classes in which stops and fricatives are distinguished by the number of structural layers they contain. These structures are given in (1). A contrast traditionally classified being based on a feature [continuant] (or an element {?}) has been recast as a structural difference.

The trees in (1), in which stops are more structurally complex than fricatives, find an interesting parallel in the acoustic signal.<sup>2</sup> Stops are constructed of two auditory events, silent closure and noise, while fricatives are made of only noise. Updating the structures in (1) to capture their auditory specification gives us new trees as in (2), in

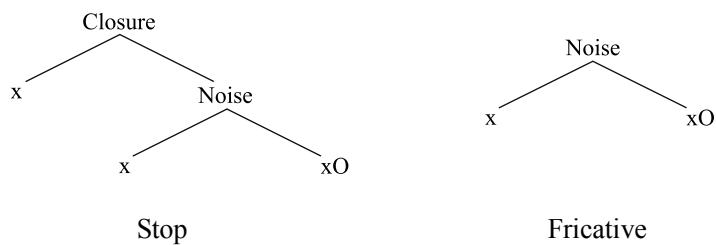
---

<sup>2</sup> Pöchtrager is a traditional Government Phonologist in that he considers the acoustic signal to be irrelevant for explaining phonological behavior. Thus, the link between his structures and the signal is unintended, perhaps even coincidental.

(1) Structural representation of stop–fricative distinction (after Pöchtrager 2006)



(2) Stop–fricative distinction amended for auditory specification



which the two layers of prosodic structure represent two distinct phonetic entities, stop closure (characterized by silence) and obstruent noise (aperiodic turbulence in the signal).

The trees in (2) illustrate the important assumption of the Onset Prominence model that constituent structure is itself a phonetic object, constructed from explicitly defined phonetic events. This idea presents a stark departure to the traditional idea of a “bare skeleton, lacking in any phonetic property and responsible for temporal sequencing” (Scheer and Szigetvári 2005: 51). In my view, refining views of structure offers a fruitful environment to model the relationship between speech and grammar.

Before we present further development of OP structures, we need to address the question of where such structures come from. The position adopted here is attributable to an evolutionary view in which phonemic representations emerge on the basis of auditory input (Bybee 2001; Blevins 2004). We shall qualify this view somewhat with a stipulation that the primary or basic input from which representations are constructed is found in initial positions. Since initial positions are frequently prominent, we are therefore assuming that segments evolve on the basis of the most prominent auditory inputs. An important corollary to our position on the emergence of phonemes is that phonetic context must be built into segmental representations. In phonetic terms there is no such

thing as a segment in isolation;<sup>3</sup> a segment's prosodic position is present in all auditory exemplars.

The prosodic frame from which segmental representations may be derived is given in (3). This tree may be thought of as the equivalent of a universal CV structure constructed of four layers. In addition to Closure and Noise seen in (2), there is a layer labeled Vocalic Onset, corresponding to the initial portion of a vowel that provides important perceptual information for the identification of the onset consonant. This portion houses acoustic cues for consonant place (formant transitions, e.g. Wright 2004), laryngeal specification (e.g. VOT, f0, spectral tilt; e.g. Maddieson 1997), and manner (amplitude rise time for stop–glide distinctions; Shinn and Blumstein 1984). Below the Vocalic Onset layer, we posit the Rhyme, comprised of vocalic formant targets and an offset position corresponding to a period of perceptual saturation (Wright 2004: 44, Figure 2.4). The tree structure in (3) is built down from stimulus onset in an attempt to capture facts related to auditory response, in particular the boost at onset (Wright 2004). Thus, higher levels of structure reflect greater perceptual sensitivity, while rhymes are associated with decreased auditory response. Placing the rhyme at the bottom of the tree allows us to capture cross-linguistic preferences for sonorous nuclei – sonorous segments are the only type of sounds that are consistently robust enough acoustically to be heard in perceptually dull portions of the speech stream.

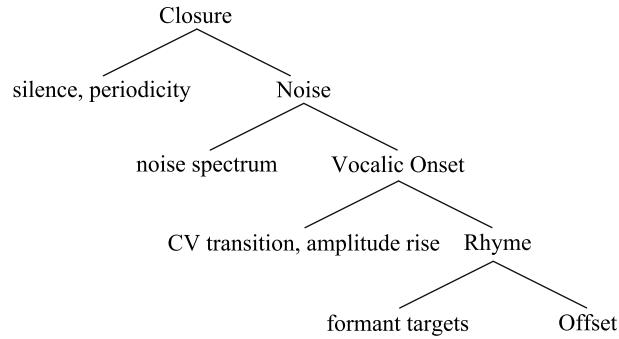
The final important property of the tree in (3) is that individual layers of structure are inherently ordered in the acoustic signal. Thus, in initial positions, Closure always precedes Noise, and Noise always precedes Vocalic Onset.<sup>4</sup> This sequencing allows us to capture the notion that the linear order of segments, and by extension phonotactic facts associated with syllable positions, may be read directly off prosodic structure (Golston and Hulst 1999).

Manner distinctions and may be easily derived from (3). Manner is structural and does not need to be specified melodically. Further, we have a principled structural definition of the consonant–vowel distinction: vowels contain melodic specification under the Rhymal structure, while consonants are specified at higher levels. A summary of the manner distinctions, which resemble scales of phonological strength, are given in (4). Stops contain all three layers of onset structure, fricatives and nasals contain two, while approximants contain only one.

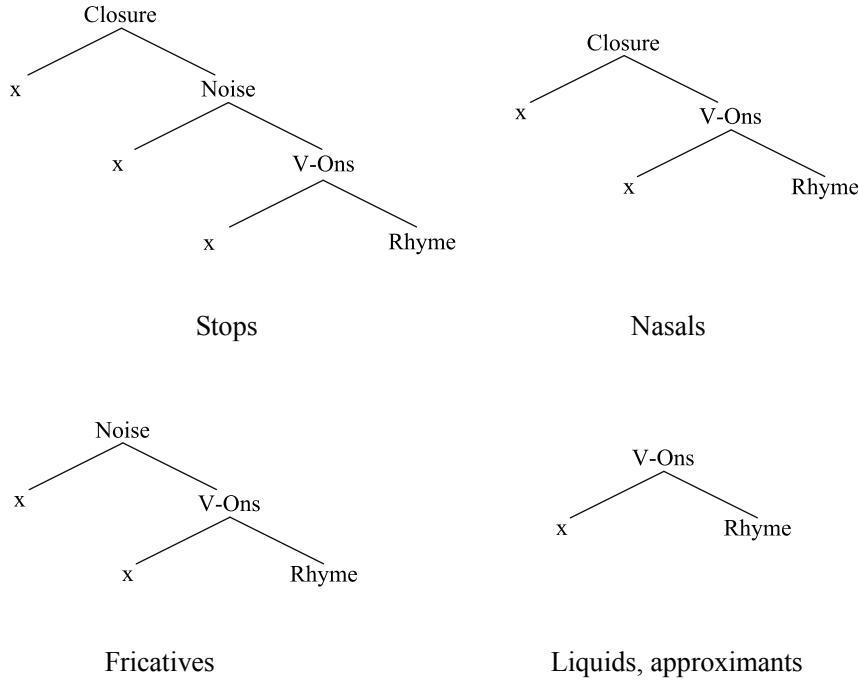
<sup>3</sup> Phonetic events such as stop bursts or fricative noise may, of course, be produced without a neighboring sound. However, such cases also represent a sort of phonetic context, namely silence. Obstruent noise preceded and followed by silence will always have measurable durational properties that are frequently different from those found in vocalic contexts.

<sup>4</sup> Two types of sounds require some clarification of this claim. The first is ejectives, in which a silent period of glottal closure occurs after noise produced at the oral constriction. We must therefore specify *oral* closure as the top layer of structure. Glottal closure associated with ejectives (especially stops) often produces a short silent period before Vocalic Onset, which may be seen as a perceptual foregrounding device to the laryngeal cues at vocalic onset (Ladefoged and Maddieson 1996). Pre-aspirates also show different sequencing of the auditory properties specified in (3), but rarely occur in initial position (Ladefoged and Maddieson 1996), and derive historically from hetero-syllabic clusters or geminates (Blevins 2004).

## (3) The context of OP representations, an Onset-Rhyme constituent



## (4) Onset Prominence, manner and strength



The trees in (4) constitute an upside-down version of the sonority hierarchy, and as such offer some perspective on phonotactic patterns (see Schwartz 2009). Simply stated, the less structure segment has, the more sonorous it must be. The levels OP structure may be seen to represent auditory sensitivity. Closure, since it allows for auditory recovery,

facilitates the perceptual boost at stimulus onset. The magnitude of this boost is largely dependent on manner of articulation – stops produce greater boosts than fricatives, which in turn produce larger boosts than approximants (Wright 2004). Thanks to this boost, stop consonants are salient despite their transient acoustic properties. This salience is represented as structure. Vowels have less structure, and thus owe their salience to their sonorous nature. Returning to phonotactics, OP structures (just like sonority) predict two types of cluster. Consonant clusters that decrease in Onset Prominence (such as stop–liquid) may (but do not have to) be joined in a single constituent, while those with equal or increasing OP (such as sibilant-stop) must span two constituents. The two types of cluster show distinct behavior (e.g. in loanword adaptation; Fleischhacker 2005), so they must be represented differently.<sup>5</sup>

At this point, we must turn to melodic and laryngeal specifications, which we assume are housed on terminal nodes of structure (marked above with x). Since the V-Onset layer is shared by all consonants, we posit this layer as the default location for consonantal melodic specification.<sup>6</sup> A proposed representation for a voiced stop is offered in (5). We place the voiced specification on the V-Ons node, labeled here as the element {L},<sup>7</sup> to capture the transitional nature of the VOT cue to voicing. The acoustic effects of melodic specifications may be assumed to spread to the higher levels of onset structure. Thus, in languages with L-voicing (see e.g. Harris 1994), we observe low frequency periodicity, the acoustic signature associated with the L-element, during the Noise and Closure phases of an initial stop.

To sum up this section, Onset Prominence builds prosodic positions into segmental representation. The perspective has been shown elsewhere (see Schwartz 2009) to be particularly insightful for phonotactics and phonological strength. As we shall see in the following section, we are also afforded a useful perspective on issues of licensing. OP allows us to capture the syllabic generalizations of prosodic licensing, while at the same time incorporating the phonetic grounding of cue licensing.

#### 4. Licensing, neutralization, and Onset Prominence

Our structures as presented so far derive segmental representations from auditory input tokens in initial positions. To gain some perspective on the licensing issue, however, we

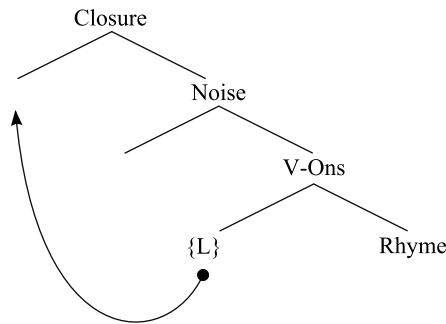
---

<sup>5</sup> The discussion here has been added for the benefit of an anonymous *PSiCL* reviewer who questions the claim that OP structures can account for phonotactic patterns. The key benefit of the OP view of phonotactics is that it captures sonority-based generalizations without invoking sonority. For arguments against sonority as a phonological entity, see Ohala (1992) and Harris (2006).

<sup>6</sup> Vowels, of course, are specified below the Rhymal layer.

<sup>7</sup> For the purposes of this paper, it is irrelevant whether we posit a monovalent laryngeal element or a binary [voice] feature. Considering the works that have been influential in developing the OP theory, the privative strategy seems like the more natural one. However, I can see no reason why OP representations cannot be implemented with binary features.

(5) A fully voiced stop, with low frequency periodicity spreading onto closure



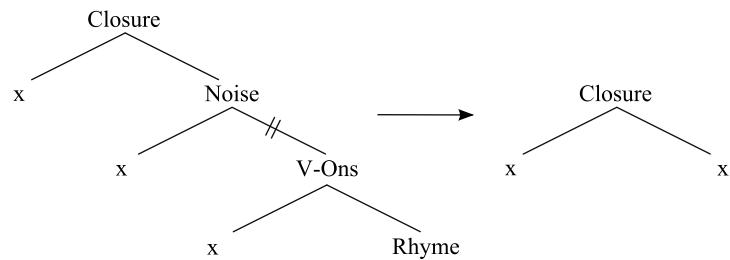
must examine what happens to these structures in other positions. As we shall see, the OP structures allow us to unify prosodic positions for licensing and neutralization, while at the same time capturing the phonetic environments of the cue-based approach.

#### 4.1. Vocalic Onset specification in licensing and neutralization

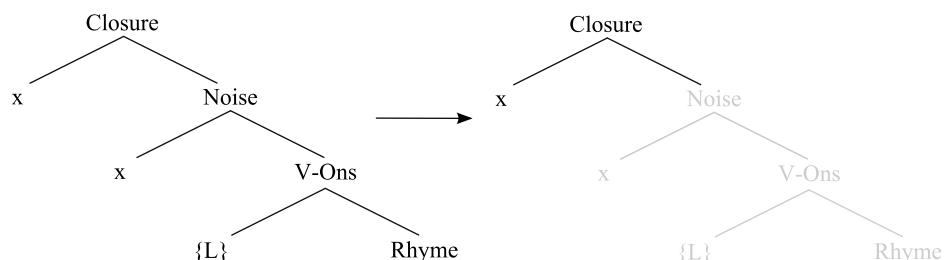
The key feature that separates OP consonant structures from traditional ones is the presence of the Vocalic Onset layer, which represents something of a transitional stage between the traditional constituents of onset and rhyme. This layer may be thought of as the phonetic glue that holds universal Onset-Rhyme sequences together. It houses CV formant transitions, frequently the most reliable cue for listeners in identifying place of articulation (Wright 2004), as well as manner and laryngeal cues. In non-prevocalic position, this layer of structure is necessarily missing from the signal. What remains may be thought of an ambiguous phonological or perceptual object: incomplete onset structure combined with weakened structural housing for melodic specification.

To visualize what this object might look like, we turn once again to the structures in Pöchtrager (2006), who employs a principle of Structure Minimality (Pöchtrager 2006: 65), by which a unary node is reduced to its daughter. The representation of a non-prevocalic stop is given in (6). On the left, we see the original tree, with a double bar through the node linking the V-Onset level. On the right, we see what is left of the tree when the remaining unary nodes are trimmed. As noted by Steriade (1997), the most common environments for neutralization of laryngeal contrasts are word-finally and before an obstruent. The structure in (6) unifies these two generalizations into a single structural representation of a consonant in neutralizing positions. In both cases, we are dealing with trimmed onset structure. In (7), this is illustrated for a voiced stop that is subject to devoicing. We can see that the  $\{L\}$  specification loses its structural housing in the absence of the V-Onset node. The neutralization of the voice contrast falls out naturally from this structure.

## (6) Pruning of a non-prevocalic stop



(7) Pruned onset structure for a voiced stop



Two fundamental issues arising from the representation in (7) must be addressed at this point. First, the structure in (7) predicts that final devoicing and neutralization should always occur, since in final position the structural housing for the laryngeal specification is subject to pruning. There are of course many languages that preserve voice contrasts in this position. Second, Final Devoicing also occurs in languages like German, in which “voiced” obstruents are only passively voiced. In such cases we must assume that there is no  $\{L\}$  specification (or equivalent binary feature) to be lost.

The preservation of final laryngeal contrasts may be executed in one of two ways. In a language such as French, which has L-voicing, we hypothesize that the Onset structure undergoes a process of listener reconstruction (Ohala 1981), restoring the prosodic housing for the laryngeal specification on the V-Ons node. On the articulatory side, this reconstruction is typically implemented as a simple release of the final consonant into a short vocoid, often perceived as a schwa (Walker 2001). Consonant release is frequent historical source of the schwa vowel (Hall 2006; Silverman 2009). These representations capture the process as the restoration of Onset structure, as well as the fact that inserted schwa may be systematically different from lexical schwa in languages as English (Davidson 2007). The release stage may be assumed to be one at which Vocalic Onset, essentially an element of consonantal representation, is restored by means of a

short vocoid, but there is no categorical insertion of a lexical vowel segment with rhymal melodic specification. In other words, schwa epenthesis first and foremost restores consonantal structure rather than inserting a vocalic segment. Diachronically, of course, a segmental interpretation of the inserted vocoid may be adopted, producing alternations. This, however, must be seen as a later development.

We now turn to cases such as German. Assuming (after e.g. Harris 2009) that the /b, d, g/ series in German are plain obstruents lacking specification for {L}, then devoicing in final position cannot be the result of a rule that removes a laryngeal element. Rather, it is the lack of passive voicing that occurs in pre-sonorant position. In other words, German final devoicing is merely a phonetic effect that does not influence phonological representations. Like many other genetically related languages, German is often described as distinguishing plain vs. fortis (or aspirated) obstruents. The plain series is unspecified for laryngeal properties, while the fortis series is marked with a feature such as [spread glottis], [-voice] or an element {H}.<sup>8</sup> The question that remains does not concern the effects of devoicing on the plain series so much as what becomes of fortisness in final position. There is no consensus on this issue (see discussion in Harris 2009). Some scholars claim that the fortis series loses its fortisness, while others claim it does not. What is clear, however, is that the fortis obstruents in final position are systematically different from those in initial position. This systematicity may be easily represented in OP structures – in initial position, onset trees remain intact, while in final position they are pruned. What has been reported as incomplete neutralization may also be represented as some kind of subtle phonetic trace of the original laryngeal specification. This possibility is expressed by the faded grey portions of the onset tree in (7) – perhaps pruning only weakens, but does not completely wipe out the structural specification.<sup>9</sup>

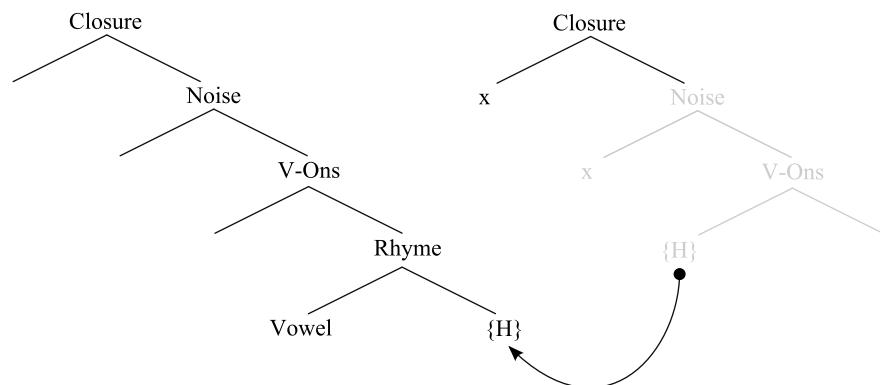
In a language such as English, the situation is somewhat more complicated. Since final consonants are frequently unreleased, we cannot talk about the restoration of Onset Structure (except perhaps in hyper-articulated speech) as in French. Nevertheless, the voice contrast is maintained, most reliably implemented in the durational properties of the preceding vowel as well as the consonant (Port and Dalby 1982). The vowel is shorter before voiceless consonants as in *beat* [bit], and longer both before voiced ones [bi:d] as in *bead*, and when no consonant follows [bi:] as in *bee*. This pattern is frequently described as pre-fortis clipping (e.g. Wells 1990). Since the final consonant has its Onset structure pruned, its fortisness (or voicelessness) can only be implemented on the preceding rhyme. The spreading of fortisness serves to shorten the preceding

<sup>8</sup> Again, this paper will not offer any opinions in support of one laryngeal specification over any other. They will be used here interchangeably.

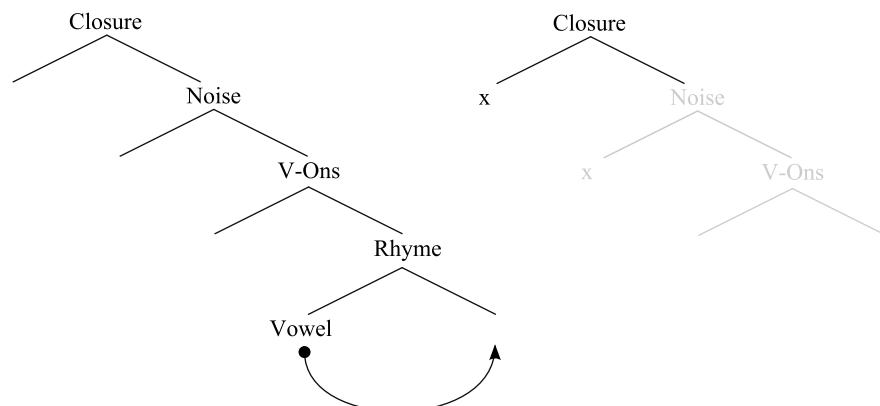
<sup>9</sup> The idea of a perceptual trace of pruned OP, which bears some semblance to traces in syntax, is a difficult thing to test experimentally. However, the possibility of such a phenomenon is not so revolutionary as it may appear. Consider studies of near-merger (Labov 1997), in which subjects claim that two sounds are “the same”, but nonetheless produce small phonetic differences. These distinctions must be perceived subconsciously, since nearly merged contrasts may persist over generations. This “trace” perhaps represents such a state of subconscious perception.

vocoid. In (8) we offer an OP perspective on this process. The implementation of the laryngeal contrast is housed on the preceding rhyme.<sup>10</sup> Fortis consonants (8a) spread their laryngeal specification to the preceding vowel – the result is pre-fortis clipping. When the final consonant is lenis (8b), there is no laryngeal specification to spread, and the offset rhymal position is occupied by the vowel, reflecting its durational properties.

(8a) Pre-fortis clipping



(8b) Longer vowel before lenis C



<sup>10</sup> Pöchtrager (2006) eliminates the element H, replacing it with a licensing relation that he calls m-command. As a result of m-command, the fortis stops occupy an additional position in the constituent trees, reflecting their longer phonetic duration. A similar principle may be assumed to be at work in the OP representations. However, a formal account of fortisness is beyond the scope of this paper. Nevertheless, the representational strategy for pre-fortis clipping remains constant – it is clearly the effect of the final consonant on the preceding rhyme. For lenis consonants, when this effect is missing, the extra rhymal position is occupied by vocalic melody.

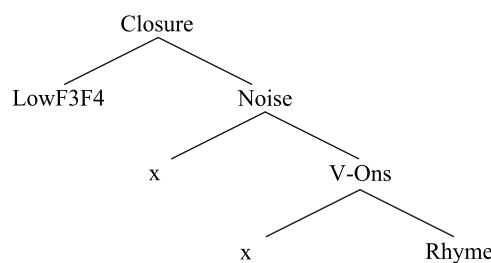
#### 4.2. Onset Prominence and VC (coda) licensing

In the structure in (7), we have taken a first step in unifying cue and prosodic accounts of licensing and neutralization. We have built a prosodic entity in which we can express the distribution of perceptual cues associated with melodic features or elements. At the same time, we have established a structural mechanism, positional pruning of Onset Prominence, that captures the insights of syllable-based approaches to licensing. It is time now to turn to Steriade's counter-examples to the predictions of Licensing by Prosody. I shall argue that representational flexibility in OP structures provides a faithful representation of phonetic facts and a prosodically insightful perspective on cue licensing.

Steriade's case for licensing by cue is heavily based on licensing of contrasts between retroflex and non-retroflex obstruents. Perceptually, retroflex obstruents differ from most other types in that their transitional cues (Low F3 and F4) are present to a much greater degree on a preceding vowel than on a following vowel. At the same time the release of retroflex stops shows formant transitions that are largely indistinguishable from other coronals (Steriade 2001). Thus, the positional salience of this particular class of segments is quite unusual.

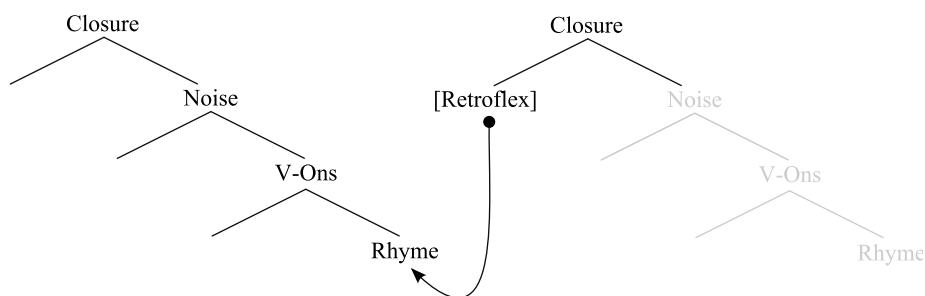
This difference is easily represented in OP representations that feature multiple layers of consonantal structure. Remember that most melodic specification for consonants is assumed to be housed on the V-Onset node, reflecting the robust acoustic cues found in pre-vocalic position. Retroflex stops, on the other hand, may be assumed to be annotated with melody on the Closure node. The unusual behavior of retroflexes thus falls out from their unusual representational properties. The representation for the retroflex plosive /t/ is given in (9). The melody associated with retroflexion is posited as the formant pattern found in VC transitions. At the same time, the V-Ons node lacks retroflex specification. This reflects the articulatory facts mentioned earlier – due to the dynamic nature of their articulations, retroflex plosives show release cues that are only minimally different from those of other coronals. Since the retroflex melody is housed on closure but not release, the formant pattern only becomes perceptually robust when it may spread to a preceding vowel.

(9) A retroflex stop /t/



In (10) we see a CVC sequence ending in a retroflex /t/. The stop, of course, is pruned in non-prevocalic position. The pruning however, does not affect the [Retroflex] specification. Retroflexion, because of the articulatory difficulties it presents, may be assumed to spread onto the preceding vowel, which offers robust formant structure for the realization of the lowered F3 and F4.

(10) CVC sequence ending in a /t/

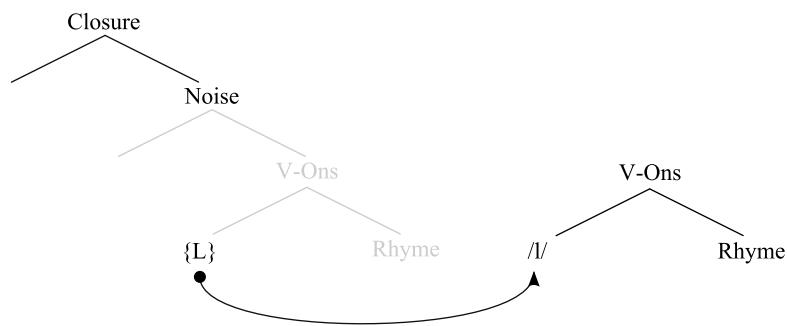


The final case we shall discuss are hetero-syllabic consonant clusters of the type stop + liquid. Steriade cites Lithuanian as a language in which the voice contrast is maintained in the stop members (the coda) of these clusters, since their laryngeal cues may be realized on the following sonorant containing the V-Ons node. From this perspective, the syllable position is irrelevant with regard to licensing. Since Lithuanian has final and pre-obstruent devoicing, these cases (contrasts such as *dump.les* vs. *dumb.las*) represent a counter-example to the predictions of traditional prosodic licensing. The OP structure for the voiced coda (in the form *dumb.las*) is given in (11). Note that the following V-Ons node built into the representation of the sonorant offers a structural home for the realization of the preceding obstruent's laryngeal specification. At first glance, the structure in (11) predicts a similar progressive spreading of laryngeal specifications in pre-obstruent prediction, since obstruents too contain a V-Ons node for the realization of laryngeal specifications. However, following obstruents have their own laryngeal melodic specifications that may spread leftwards.

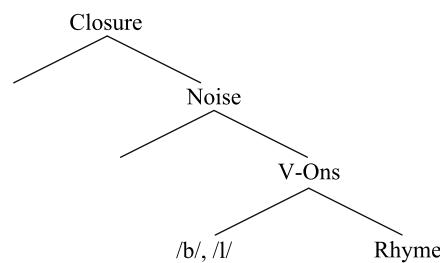
Cases such as Lithuanian provide an interesting perspective on an important aspect of the OP framework concerning the status of the syllable. In the theory, the syllable must be seen as a language-specific emergent unit that may be derived from the universal Onset–Rhyme constituent in (3). The hetero-syllabic status of the stop–liquid clusters is a case in point. In a large number of languages, such clusters are exclusively tautosyllabic, a fact often attributed to a syllable contact constraint (e.g. Vennemann 1988) demanding that sonority must fall in coda–onset clusters. The tautosyllabicity of stop–liquid clusters falls out naturally from our structural specifications for manner of

articulation. Since Onset–Rhyme constituents are built down from stop closure, the V-Ons node of may be shared by the stop and the liquid to form a single constituent with two onset segments (traditionally thought of as a “branching” onset). This is represented in (12). In languages where stop–liquid clusters are hetero-syllabic, we may posit a simple parametric setting limiting the amount of melody a single layer of structure can house. Comparing the structures in (11) and (12) we are afforded a perspective on the mechanisms from which language-specific syllable structures may develop. In some cases segmental trees are permitted to merge, while in others they are not. The Onset–Rhyme constituent depicted in (3) is a universal structure. The segmental content of such structures is dependent on processes affecting the onset trees.

(11) A heterosyllabic /b.l/ cluster in Lithuanian



(12) Tautosyllabic stop–liquid cluster



In sum, the representational perspective presented here allows us to unify cue and prosodic licensing. With OP structures we may faithfully represent the phonetic properties of the acoustic signal, while at the same time form the type of predictive generalizations that are crucial for phonology.

### 5. Conclusion – phonologically-based phonetics and Prof. Gussmann’s legacy

The theory of Onset Prominence has been developed with the aim of reconciling phonetic and phonological explanations of phonological phenomena. To achieve this goal, we must re-evaluate the speech signal to establish which speech events are categorical and therefore phonological in nature. By refining the traditional syllable-based representations of phonological constituents, we may form a new set of hypotheses to serve as the subjects of experimental study. In my view the Onset structures presented in this paper offer a useful starting point. The structures here allow for faithful representation of the acoustic signal that is nevertheless phonological in nature. Closure, Noise, and Vocalic Onset, the phonetically specified layers of Onset Prominence share an important physical property. They produce landmarks in the speech signal (Shattuck-Hufnagel and Veilleux 2007) that are more or less discrete in nature. Since phonology is by definition discrete, any attempt to reconcile phonology with speech should be constructed on the basis of (quasi-)discreteness in speech.

Rather than the recent focus on “phonetically-based phonology”, the Onset Prominence theory suggests a research program of “phonologically-based phonetics”, by which phoneticians are encouraged to refocus their experimental priorities on the categorical, rather than gradient, aspects of speech. In particular, I would call for more study of cues to manner of articulation. Since there is good evidence that manner may be equated with structure, precise phonetic description of manner categories may provide new insights on the phonetics of structure. Some research of this type has been done (e.g. Shinn and Blumstein 1984). However, phonetically-based phonology in most of its incarnations seems to have focused on spectral properties such as formant frequencies that represent place of articulation.

The perspective that gave birth to the idea of phonologically-based phonetics as opposed to phonetically-based phonology is directly attributable to Prof. Gussmann’s strong view that speech is irrelevant for phonology. Though I never had a chance to discuss these ideas thoroughly with him, I think that Prof. Gussmann would have endorsed the research program. I wish that he could have taken comfort in the fact that unknowingly and despite a famed lack of interest in phonetics, Edmund Gussmann has had a significant influence on at least one phonologically-based phonetician.

### REFERENCES

Blevins, J. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Bybee, J. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Dave, R. 1977. “Retroflex and dental consonants in Gujarati: A palatographic and acoustic study”. *Annual Report of the Institute of Phonetics, University of Copenhagen (ARIPUC)* 11. 27–155.

Davidson, L. 2007. "The relationship between the perception of non-native phonotactics and loanword adaptation". *Phonology* 24. 261–286.

Dinnsen, D. and J. Charles-Luce. 1984. "Phonological neutralization, phonetic implementation and individual differences". *Journal of Phonetics* 12. 49–60.

Fleischhacker, H. 2001. "Cluster-dependent epenthesis asymmetries". *UCLA Working Papers in Linguistics* 7 (*Papers in Phonology* 5.) 71–116.

Flemming, E. 2002. *Auditory representations in phonology*. New York: Routledge.

Golston, C. and H. van der Hulst. 1999. "Stricture is structure". In: Hermans, B. and M. van Oostendorp (eds.), *The derivational residue in phonological Optimality Theory*. Amsterdam: Benjamins. 153–173.

Goldinger, S., D. Pisoni and P. Luce. 1996. "Speech perception and spoken word recognition: Research and theory". In: Lass, N. (ed.), *Principles of experimental phonetics*. St. Louis: Moseby-Year. 277–327.

Greenberg, S. 1996. "Auditory processing of speech". In: Lass, N. (ed.), *Principles of experimental phonetics*. St. Louis: Moseby-Year. 362–408.

Gussmann, E. 2004. "The irrelevance of phonetics". *Corpus* 3. 125–152.  
<<http://corpus.revues.org/index203.html>>

Hall, N. 2006. "Cross-linguistic patterns of vowel intrusion". *Phonology* 23. 387–429.

Hamann, S. 2003. "Norwegian retroflexion: Licensing by cue or prosody?" *Nordlyd* 31(1). 63–77.

Harris, J. 1994. *English sound structure*. Oxford: Blackwell.

Harris, J. 2006. "The phonology of being understood – further arguments against sonority". *Lingua* 116. 1483–1494.

Harris, J. 2009. "Why final obstruent devoicing is weakening". In: Nasukawa, N. and P. Backley (eds.), *Strength relations in phonology*. Berlin: Mouton de Gruyter. 9–46.

Hayes, B. and D. Steriade. 2004. "The phonetic basis of phonological markedness". In: Hayes, B., R. Kirchner and D. Steriade (eds), *Phonetically based phonology*. Cambridge: Cambridge University Press. 1–33.

Ito, J. 1988. *Syllable theory in prosodic phonology*. New York: Garland.

Krakow, R.A. 1999. "Physiological organization of syllables : A review". *Journal of Phonetics* 27. 23–54.

Labov, W. 1997. *Principles of linguistic change: Internal factors*. Oxford: Blackwell.

Ladefoged, P. and I. Maddieson. 1996. *The sounds of the world's languages*. Cambridge, MA: Blackwell.

Lombardi, L. 1995. "Laryngeal neutralization and syllable well-formedness". *Natural Language and Linguistic Theory* 13. 39–74.

Maddieson, I. 1997. "Phonetic universals". In: Hardcastle, W. and J. Laver (eds), *The handbook of phonetic sciences*. Cambridge: Blackwell. 619–640.

Manaster-Ramer, A. 1996. "A letter from an incompletely neutral phonologist". *Journal of Phonetics* 24. 477–489.

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

Ohala, J. 1992. "Alternatives to the sonority hierarchy for explaining segmental sequential constraints". *Papers from the parasession on the syllable*. Chicago: Chicago Linguistic Society. 319–338.

Oostendorp, M. van. 2006. "Incomplete devoicing in formal phonology".  
<<http://vanoostendorp.nl/>> (Last accessed 10 Oct 2010.)

Pöchtrager, M. 2006. The structure of length. (Unpublished PhD dissertation, University of Vienna.)

Port, R. 1996. "The discreteness of phonetic elements and formal linguistics: A response to A. Manaster-Ramer". *Journal of Phonetics* 24. 491–511.

Port, R. and J. Dalby 1982. "Consonant/vowel ratio as a cue for voicing in English". *Perception and Psychophysics* 32. 141–152.

Rubach, J. 1990. "Final devoicing and cyclic syllabification in German". *Linguistic Inquiry* 21. 79–94.

Scheer, T. and P. Szigetvári. 2005. "Unified representations for the syllable and stress". *Phonology* 22. 37–75.

Schwartz, G. 2009. *Phonology for the listener and language learner*. Poznań: Wydawnictwo Naukowe UAM.

Shattuck-Hufnagel, S. and N. Veilleux. 2007. "Robustness of acoustic landmarks in spontaneously-spoken American English". *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken.

Shinn, P. and S.E. Blumstein. 1984. "On the role of the amplitude envelope for the perception of [b] and [w]". *Journal of the Acoustical Society of America* 75. 1243–1252.

Silverman, D. 2009. "Release and reduction: Two origins of schwa". Paper presented at the 17th Manchester Phonology Meeting.

Slowiaczek, L. and D. Dinnsen. 1985. "On the neutralizing status of Polish word-final devoicing". *Journal of Phonetics* 13. 325–341.

Steriade, D. 1997. Phonetics in phonology: The case of laryngeal neutralization. (Ms., University of California, Los Angeles.)

Steriade, D. 2001. "Directional asymmetries in place assimilation – A perceptual account". In: Hume, E. and K. Johnson (eds.), *The role of speech perception in phonology*. New York: Academic Press. 219–250.

Vennemann, T. 1988. *Preference laws for syllable structure*. Berlin: Mouton.

Walker, D. 2001. *French sound structure*. Calgary: University of Calgary Press.

Wells, J. 1990. "Syllabification and allophony". In: Ramsaran, S. (ed.), *Studies in the pronunciation of English*. London: Routledge. 76–86.

Wright, R. 2004. "Perceptual cue robustness and phonotactic constraints". In: Hayes, B., R. Kirchner and D. Steriade (eds.), *Phonetically based phonology*. Cambridge: Cambridge University Press. 34–57.

Wright, R., S. Frisch, and D. Pisoni. 1997. Speech perception. Research on spoken language processing. (Progress Report No. 21.) Indiana University.

**Address correspondence to:**

Geoffrey Schwartz  
 School of English  
 Adam Mickiewicz University  
 A1. Niepodległości 4  
 61-874 Poznań  
 Poland  
 geoff@ifa.amu.edu.pl