

*Acta Linguistica Hungarica*, Vol. 57 (4), pp. 381–397 (2010)  
DOI: 10.1556/ALing.57.2010.4.2

## AUDITORY REPRESENTATIONS AND THE STRUCTURES OF GP 2.0

GEOFF SCHWARTZ

School of English  
Uniwersytet im. Adama Mickiewicza w Poznaniu  
ul. Wieniawskiego 1  
61-712 Poznań  
Poland  
geoff@ifa.amu.edu.pl

**Abstract:** This paper presents a proposal for the implementation of auditory specification within structures inspired by GP 2.0 (Kaye–Pöchtrager 2009). In the theory, constituent structure itself is a phonetic object built down from Onset structure. This strategy allows for a faithful representation of the acoustic signal and an insightful model of consonant strength and coda formation. Perceptual considerations suggest that a privative approach to auditory representation can account for phonological behavior, as well as enable us to form testable hypotheses for experimental phonetic study.

**Keywords:** phonology, Government Phonology, auditory representations, constituent structure, phonetics

### 1. Introduction

The relationship between Government Phonology (GP) and phonetics may at first glance appear somewhat paradoxical in nature. On the one hand, most scholars working in GP would argue that phonology represents an autonomous system independent from speech. In this tradition we may encounter some very strong statements. Scheer (2001) has written that “phonology is everything but the phon”. Gussman (p.c.) has stated that “ease of articulation is a moronic term”. Others have held that a ‘phonetics–phonology interface’, or at least one involving a phonetic level of representation, does not exist. On the other hand, the elemental primes of GP must be phonetically interpretable (Harris–Lindsey 1995), produc-

ing a physical signature in the speech signal. Thus, while a single SPE feature may not stand alone as a perceptible segment, a GP element can and must do so. The paradox is that in GP speech falls outside of phonology, but speech provides the building blocks of its representations.

This, I would argue, is not a paradox at all, but rather a fruitful perspective for the cognitive representation of the speech signal. Taking this perspective to its logical conclusion, we arrive at the position that phonology must subsume much of what is traditionally attributed to phonetics. Though this view has not achieved widespread acceptance, it has been expressed by a number of phonetically inclined scholars (Ohala 1990; Donegan 2002). In GP, this perspective is developed by Harris and Lindsey (1995) as the autonomous interpretation hypothesis, a “minimalist take on the phonetics–phonology interface” (Harris 2004), in which phonological representations are mapped directly onto the speech signal without an intervening level of categorical phonetics.

When GP has concerned itself with speech, the focus has been largely auditory rather than articulatory. Notably, a Jakobsonian perspective has been adopted in which the acoustic signal represents the only shared experience between speaker and hearer. At first glance, the mapping of acoustic features to their phonological specifications seems to be a simple matter. The situation becomes more complex, however, when we consider the fact that speech perception is in large measure a phonology-dependent language-specific phenomenon. It is well known that speakers of different languages often hear the same acoustic input differently—listeners adjust the ambiguous acoustic signal and “reconstruct” phonological representations (Ohala 1981). Consequently, if we accept the tenet that phonological objects are housed in the acoustic signal, our representation of the signal must include devices that allow for language-specific listener-induced adjustments.

This paper will offer a brief presentation of a GP-inspired theory of representation, emphasizing the perceptual considerations behind a proposal in which constituent structure is constructed from concrete auditory properties. Section **2** will discuss structure as a phonetic phenomenon, proposing auditorily-specified modifications to the structures of GP 2.0 (Kaye–Pöchtrager 2009). Section **3** discusses the primary role of onsets in constituent formation. Section **4** illustrates how the emergence of marked structures may be seen as adjustments to onset specification. Section **5** discusses further implications of privativeness for auditory representations.

## 2. Structure in the signal and GP 2.0

Phonetically-based phonology has largely eschewed structure. Citing the failure to identify reliable phonetic correlates of syllabic constituents, this tradition claims that “[n]othing about perception, articulation, or processing leads us to expect any licensing asymmetries among syllable positions” (Hayes–Steriade 2004, 23). These scholars do not seem to consider the possibility that the failure to find perceptual correlates of prosodic categories may be due to flawed experimental hypotheses borne out of traditional representations. Speech contains prosodic cues. If these cues do not correspond with traditional representations of structure, it is the representations that must be refined.

Close observation of the speech signal suggests clear-cut acoustic definitions of both melody and structure. Melody, by definition, specifies the timbre or quality of a sound, characterized by spectral properties such as formant frequencies, bandwidths, and the shape of the obstruent noise spectrum. Viewing these acoustic features requires a spectral display such as a spectrogram. More general auditory properties, such as silence associated with stop closure, amplitude rise time, the presence or absence of aperiodic noise, and periodicity, give structure to the signal. These types of cues, which are visible on a simple acoustic waveform, allow us to delineate speech into identifiable chunks.

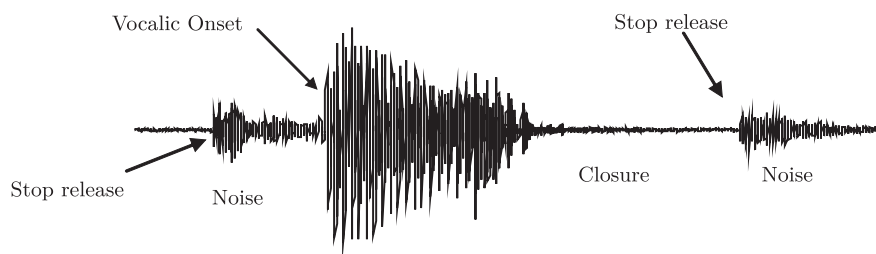


Fig. 1

Waveform of English *taught* as spoken by a native speaker of British English

Figure 1 shows a waveform of the word *taught* as spoken by a native speaker of British English. A waveform display does not represent spectral information, so looking at this picture we have little idea of the melodic specifications. Nevertheless, it is not difficult to glean from this display that we are dealing with an Unvoiced Stop–Vowel–Unvoiced Stop sequence. In other words, the manner of articulation, the “structure” of the utterance, is clearly visible. Turning our attention to the

articulatory–acoustic relationship, we may note that spectral cues are largely dependent on place of articulation, while manner is reflected in the more general acoustic properties observable in a waveform. We therefore arrive at the following insight which will guide us as we present our theory: place is spectral, manner is structural.

The auditory phenomena visible in Figure 1 share one very important property: they produce boundaries in the signal that are discrete or almost discrete in nature. Phonology deals with discrete entities, yet the continuous nature of the speech signal has led some to believe that phonetics and phonology must remain separate domains. If a plausible integration of speech into phonology is to be implemented, it is reasonable to assume that it should be concentrated around those parts of the signal that actually contain discrete or nearly discrete entities.

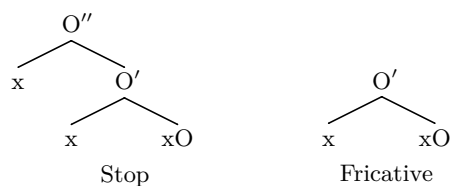
Building on Jensen (1994), who eliminates the stop element **?** in favour of an additional structural position, Pöchtrager (2006) establishes separate structural configurations for stops and fricatives, which are given in (1). Fricatives are single layered structures with two skeletal points. Stops are two-layered structures containing three. The trees in (1) give formal expression to the notion that manner of articulation is structural rather than melodic, providing quite a novel perspective on the nature of segmental representation. The difference between, for example, /k/ and /x/, is no longer expressed in terms of a melodic feature ([cont]) or element (**?**). Rather, the difference lies in the structure of the segment itself.

A crucial feature of Pöchtrager's trees is that they are not linked with phonetic properties by means of traditional association lines. Rather, the trees are **annotated** with melody that may or may not be projected to other structural positions. This view represents a starting point for the theory to be presented here. By eliminating the association lines that link constituent structure with phonetic properties, we allow for the possibility that structure **itself** is a phonetic object, constructed from the types of acoustic building blocks we saw in Figure 1: delineative properties associated with manner of articulation. In other words, we may build constituent structure out of speech-based landmarks: stop closures, aperiodic noise, and vocalic onsets.

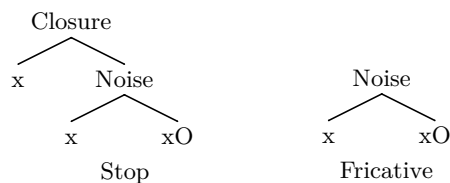
The acoustic signal suggests a similar perspective on the stop–fricative distinction. Acoustically, stops are indeed more structurally complex than fricatives. A released stop contains a silent closure period followed by a portion of aperiodic noise, with the stop release produc-

ing a quasi-discrete boundary between them. This boundary is visible on a waveform as an abrupt rise in amplitude. The fricative, on the other hand, produces only aperiodic noise with a slow amplitude rise. The stop contains two auditory events, the fricative only one. Under the assumption that the nodes of the trees presented in (1) must represent a concrete phonetic property we may redraw these trees as in (2).

- (1) Binary branching structures for stops and fricatives (after Pöchtrager 2006)

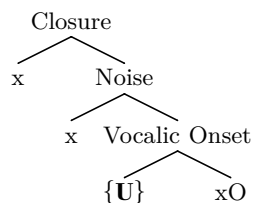


- (2) Stop and fricative structures from (1) modified for auditory specification of structure



From a perceptual perspective, the trees in (2) are incomplete. Acoustic information housed on the formants of a following vowel is often the most reliable cue that listeners have in identifying place of articulation of a consonant (e.g., Wright 2004). As a result, we must add a layer at the bottom of our tree, defined phonetically as Vocalic Onset, as in (3), which posits a hypothetical representation for a /p/. The formant transitions are specified in terms of spectral (melodic) annotations on the V-Onset terminal.

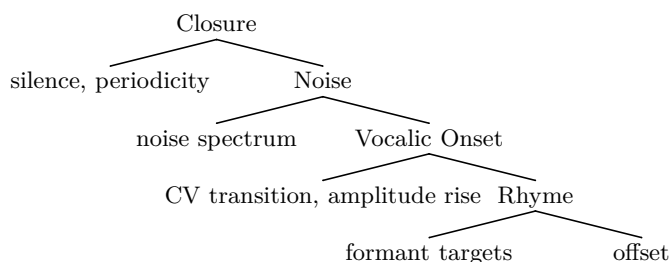
- (3) Representation of a /p/



The three layers of structure in (3) are defined as concrete phonetic properties associated with the articulation of onsets. Rhymes will be attached under the xO (Pöchtrager's 'onset head'), an assumption that will be motivated in due course. The hierarchical arrangement of the tree reflects inherent temporal sequencing of onset cues in initial positions—stop closure precedes aperiodic noise,<sup>1</sup> which precedes vowel onset. The terminal nodes may be annotated for spectral or melodic properties, as in (4), which gives a general outline of our basic Onset–Rhyme constituent.

The top three layers of the structure in (4) represent the inherent sequencing of auditory properties in onset position. Assuming that manner of articulation is a structural specification, we may read manner classes directly off the tree in (4), producing a non-arbitrary scale of consonant strength, or **Onset Prominence** (Schwartz 2009), that can be read directly off our representations. Stated briefly, strength is structure—stops contain the most, then fricatives, then sonorants, then vowels.<sup>2</sup>

(4) Basic Onset–Rhyme structure



### 3. The primacy of onsets in constituent formation

Content et al. (2001), in a study investigating the possible psycholinguistic correlations of ambisyllabicity, offer a novel view of syllabic parsing. Their experiments cast doubt on the assumption that listeners syllabify

<sup>1</sup> Preaspirated segments are not problematic for this claim. Preaspirates produce an acoustic sequence of noise-closure-noise, in which the pre-aspiration is assumed to be part of another structural constituent.

<sup>2</sup> Due to space restrictions we will not discuss laryngeal categories in this paper. I would suggest that longer closures/noise associated with fortis/unvoiced obstruents may be represented with some sort of structural reinforcement that guards against lenition.

an utterance by locating boundaries between syllables, suggesting an **onset hypothesis** for syllabic segmentation. The essence of the hypothesis is that the parsing of onsets and offsets are two distinct tasks, of which onset identification is dominant. The two cognitive operations, it is argued, need not be coupled temporally—the beginning of one constituent does not necessarily imply the end of another. If we assume that the representation of, for example, a stop, contains both onset cues and place cues, we may represent this asymmetry. Place cues may spread leftwards to the preceding rhyme, while the onset cues remain stationary.

This primacy of onsets is reflected in the constituent structure given in (4). As we proceed down the tree from Closure to Rhyme, the functional role of the attached phonetic material decreases. Wright (2004) discusses perceptual issues relating to phonotactics. He presents a schematic diagram (*op.cit.*, 44; Figure 2.4) of auditory response to a CVC sequence. In the diagram, heightened auditory response aligns with the release burst of the initial stop and the formant transitions on the early portion of the vowel. Following this boost, auditory nerve fibres endure a period of saturation in which sensitivity is reduced, aligning with the rhyme of the stimulus syllable. The sequence of acoustic properties associated with onsets may be projected on an auditory response schematic to produce a perceptual theory of constituent structure. In such a theory, stop closure, which is followed by aperiodic noise and then formant transitions, represents the facilitating force behind the onset boost (*ibid.*, 45). Rhymes, on the other hand, correspond with auditory saturation, and are placed at the bottom of the tree.

This view of structure provides a provocative listener-oriented view of the preference for sonorous syllable peaks. Vowels are preferred as nuclei not because of any built-in phonological property called “sonority”. Rather, they occupy a “nuclear” position since they are the only sound class that is sufficiently robust perceptually to overcome the saturation that occurs after an onset boost. Additionally, we may note the informational burden of consonants (onsets) and vowels in speech perception (see e.g., Wright et al. 1997). Vowels generally bear a lesser functional load in the formation of lexical contrasts, and serve as acoustic background for the perception of consonants. Further, performance phenomena show an interesting tendency: in citation forms undershoot in the production of vowels may be observed (Lindblom 1990, 429), but onsets are always pronounced faithfully. These familiar principles from speech perception suggest that a perspective on phonological structure in which constituents

are built down onsets may offer testable hypotheses about the relationship between phonology and speech.

For an illustration of the phonological insight offered by the placement of the Rhyme at the bottom of the tree in (4), consider the common sound change of velar coronalization (often called palatalization). Before we look at the structures, we must establish two important phonetic premises regarding the motivation behind velar–coronal shifts, and how such shifts are to be represented. The first is the notion that velar coronalization is a sound change with perceptual, rather than articulatory, motivation.<sup>3</sup> Traditionally, a shift of /ki/ to /tʃi/ involves the spreading of a palatal feature or element from the vowel to the consonant. The velar is assumed not to contain the palatal element, so the shift to /tʃ/ is assumed to be an articulatory assimilation. Flemming (2002, 104) offers an alternative explanation. Dorsal stops before /i/ are generally produced with a great deal of affrication, a feature which may be reinterpreted as sibilant noise. Since all sibilants are coronal, the shift in place of articulation is merely a by-product of the auditory enhancement.

The second premise concerns how to represent a /k/. Velar consonants are well known to produce a wedge-like formant pattern on neighbouring vowels (the velar “pinch”), characterized by a convergence of the second and third formants. This convergence is a feature we may also observe in the vowel /i/.<sup>4</sup> Since the F3–F2 convergence is a key ingredient in the element {**I**} (Harris–Lindsey 1995), an auditory representation of /k/ may be assumed to contain {**I**} on the V-Onset node,<sup>5</sup> a claim supported by processes in which /j/ strengthens to /k/ in languages such as Cypriot Greek (Kaisse 1992) and Swiss-Romansch (Kamprath 1987).

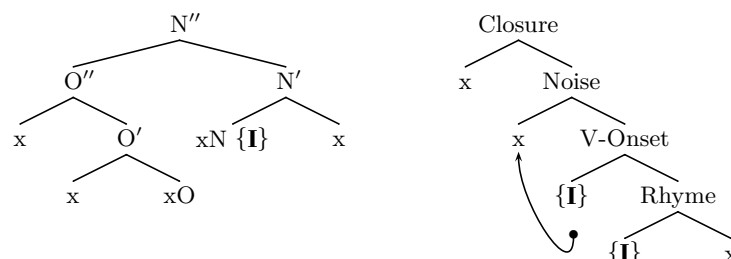
<sup>3</sup> Velar and coronal gestures are performed by different tongue muscles powered by different cranial nerves (for speech physiology see e.g., Palmer 1993), calling into doubt the notion that a /k/ > /tʃ/ shift has any articulatory motivation.

<sup>4</sup> The closeness of the convergence is greater for the velar transitions than for the vowel /i/. However, in both cases F2 and F3 are within 3 bark of one another and may be assumed to be perceptually converged (Chistovich et al. 1979).

<sup>5</sup> Standard versions of GP generally treat velars as lacking melody. From a phonetic perspective this may be reflected in the fact that the noise spectrum of velar obstruents is largely context-dependent (Stevens 1998). The perspective described here would posit that velars may be annotated for their formant transitions that resemble /i/. In other words, the velar pinch is treated here as melodic, even if velar noise may lack melody.



- (5) /ki/ in GP 2.0 (left) and in an Onset-based constituent (right)



Keeping these considerations in mind, compare the two representations of the sequence /ki/ given in (5). On the left, we see /ki/ as it would appear in GP 2.0. The onset and nucleus are members of separate constituents that merge only at a higher level. On the right, we see an onset-based representation, in which there are two consecutive **{I}** annotations. The first is from the /k/'s formant transitions, the second is from the vowel. This representation captures the motivation for the common sound changes involving /k/ sequences: two consecutive **{I}** specifications represent a perceptually unstable violation of the Obligatory Contour Principle (OCP). The most common change is to shift the /k/ to a coronal /tʃ/, which entails moving the **{I}** specification off the V-Onset node onto the Noise node. Other shifts are also attested. Ukrainian backed the vowel, simply eliminating the second of the two consecutive **{I}** specifications and yielding /ki/. A shift of /ki/ to /tʃi/ (also observed in the Slavic languages) is achieved by removing the V-Onset **{I}**, but without subsequent linking to the Noise node.

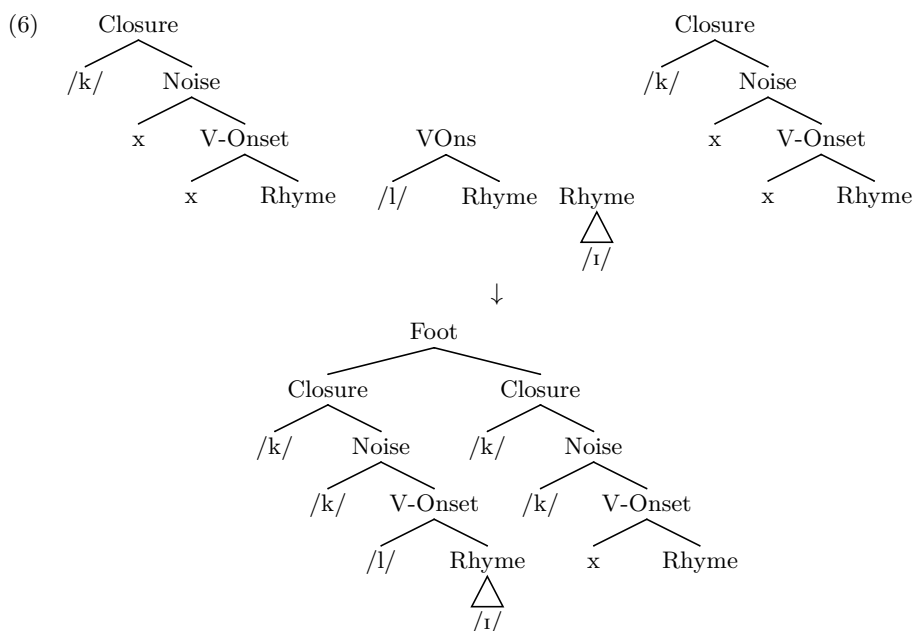
The Onset-based constituent structures proposed here offer a useful frame in which layers of structure (and melodic annotations) receive concrete acoustic specifications. Sound change may be visualized here as listener-induced adjustments to these structures, e.g., the /k/ > /tʃ/ shift represented in (5) moves a melodic specification. In the following section we shall look at structural adjustments.

#### 4. Constituent formation and adjustments to Onset specification

The basic principle behind constituent formation is that Onset Prominence must decrease within a constituent. Since each of the levels in the tree represents concrete auditory information, the presence of a cue at a higher level implies that a new constituent must begin. When this

condition is met only open syllables may occur, since every VC sequence represents a rise in Onset Prominence. In syllabic typologies, open syllables (usually CV) are the most common.<sup>6</sup>

In (6) we see this principle at work in an onset-based representation of the English word *click*. Spectral annotation is specified with segmental symbols intended as shorthand for illustrative purposes. The /kɪ/ in *click* represents a single constituent over which Onset Prominence decreases. The final /k/ comprises a new constituent. This delineation clearly mirrors the acoustic signal, in which stop closure provides a discrete break between /kɪ/ and the final /k/.



A parallel may be drawn here with GP tradition, in which a final stop is typically analysed as the onset to an “empty nucleus”.<sup>7</sup> Although I will

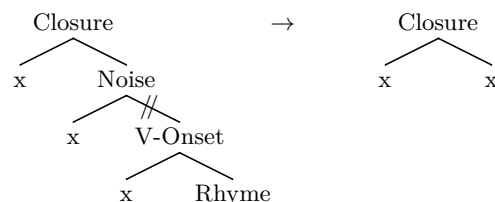
<sup>6</sup> Since there is no single “Onset” position in our structures, our theory is unable to make a typological distinction between CV syllables and V syllables, and indeed from this perspective it is not at all clear that we should. Assuming that constituent structure is a phonetic entity, “onsetless” syllables are impossible. Vowel-initial syllables must be assumed to be vowels with the V-Onset layer reconstructed.

<sup>7</sup> The model here is entirely compatible with GP when it comes to the empty nucleus. The perspective, however, is somewhat different. In our model, every

have very little to say about nuclei *per se*, I would like to claim that the “coda as onset” assumption is not so abstract as sometimes assumed, and has phonetic grounding, at least in the case of final stops. Acoustically, since it is preceded by silence, a final released stop is an onset, and is heard as one by speakers of many languages. When a final stop is unreleased, its melody is housed on the preceding rhyme, and its Onset structure is absent from the signal.

The structures in Pöchtrager (2006), which serve as a starting point for the onset-based perspective proposed here, are remarkable for their resemblance to the trees employed in Minimalist Syntax. He indeed borrows a key syntactic principle, Structure Minimality (*op.cit.*, 65), by which a unary node is reduced to its daughter. This results in the pruning of trees to eliminate unary nodes. In our Onset-based model of phonotactics we may explore this principle as a way of modelling positional effects. In particular, the absence of the V-Onset layer in non-prevocalic positions creates unary nodes in our trees, resulting in “lenited” onsethood, reduced Onset Prominence. For an illustration, consider (7), which illustrates this process for a stop in non-prevocalic position. The absence of the V-Onset layer (crossed out with two lines) and everything it dominates results in a single layered structure for a stop.

(7) Pruning of a non-prevocalic stop



The tree in (7) captures the perceptual ambiguity of a non-prevocalic stop as an onset or a coda—it is heard as a coda by speakers of some languages and as onset by others. Faced with such an acoustic structure, listeners are presented with the difficult choice of whether to apply reconstructive rules to restore the Onset Prominence of the final /k/. Listener reconstruction (Ohala 1981)<sup>8</sup> results in vowel epenthesis; failure to reconstruct

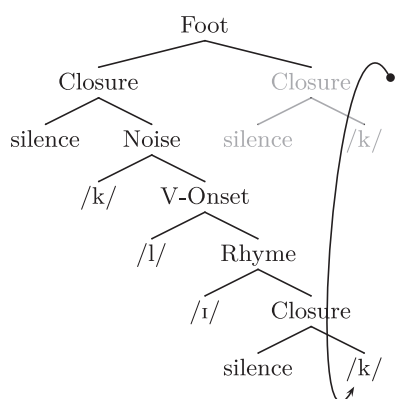
---

consonantal representation projects a following vowel on the V-Onset node. An “empty nucleus” in this view is a consonant with its V-Onset layer removed.

<sup>8</sup> The principle of listener reconstruction is a potentially powerful tool that at some level may be invoked to account for just about any phonological puzzle. However,

the Onset leaves the spectral properties of the /k/ with minimal structural housing. Such a /k/ is perceptually weak, and may be subsumed under the preceding rhyme as in (8). Remember that the lower levels under the Rhyme correspond with the period of saturated auditory response according to Wright's (2004) schematic diagram. Subsumption of pruned onset structures thus represents the inherent perceptual weakness of coda consonants.

(8) Subsumption of final stop in *click*

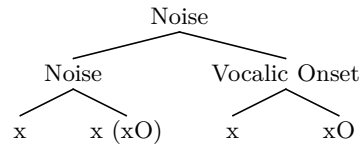


We may use our structures to represent an additional case of acoustic ambiguity, notably sibilants, which have long been known to show unexpected patterns of constituent membership. For instance, when they occur in onset clusters, sibilants typically behave in a coda-like fashion. The high intensity aperiodic noise associated with sibilants has been found to be difficult for listeners to locate temporally (Bregman 1990). Blevins and Garrett (2004) refer to sibilant noise as an acoustic distraction, causing “auditory stream decoupling” that may cause metathesis. We may represent this “distraction” in terms of an adjunct to the Noise layer of structure as in (9).

---

invoking the principle makes claims that are inherently testable and falsifiable by means of experimental study. The notion of perceptual ambiguity in the signal has been observed many times, but it is still not well understood. Some particularly interesting cases involving near-merger (e.g., Labov 1994) suggest that speakers may produce subtle phonetic distinctions, and even learn them in the acquisition process, without being able to explicitly discriminate between the categories (cf. *op.cit.*, 366 on *toe-too* in Norwich English).

(9) Proposed representation of a sibilant fricative



The three layers of Onset structure in our theory represent auditory properties with inherent temporal sequencing in the acoustic signal. Closure necessarily precedes aperiodic Noise, which in turn precedes V-Ons. The Noise adjunct disturbs the sequential descent from left to right down the tree and should be expected to confuse listeners. The potential confusion is reflected in the configuration at which Noise occurs at the same level as Vocalic Onset.

### 5. Towards a privative theory of auditory representation

The representational theory presented here, like much of the work that inspired it, seems inherently suited to privative specifications. Privativeness has largely been seen as a property of phonological, rather than phonetic, representations. It is inherently categorical—a given prime is either present or it is not. Phonetic studies, on the other hand, have devoted significant attention to scalar or gradient properties associated with the acoustic signal. With such priorities it seems that there is no room for privativeness in phonetics; one rarely if ever encounters descriptions of “privative” phonetic properties. However, many perceptual principles suggest that auditory cues might best be represented in a privative fashion. Categorical perception might be taken as an example. If listeners divide a VOT continuum into discrete categories, we may conclude that each category must represent an auditory property that is linguistically meaningful. In addition, the presence of a perceptual boost at stimulus onset suggests that the auditory system acts in a phonological manner, reducing the amount of acoustic information that needs to be processed in speech comprehension. In the frequency domain, the presence of critical bands indicates that a range of acoustic frequencies may produce a unified perceptual correlate. With these considerations in mind, we may assume that a privative characterization of auditory representations might provide us with a fruitful tool for modelling the phonetics–phonology interface.

Flemming is one of the most influential of a relatively small number of scholars to define phonological representations in terms of their auditory properties. Flemming (2002) uses scalar auditory dimensions that function within Optimality Theoretic constraints on contrasts. While the numerical nature of acoustic measurements suggests that auditory representations are inherently scalar, Flemming's model provides no way of attaching linguistic significance to an auditory property. For example, for the vowel /i/, the following representation is posited:  $F1 = 1$ ,  $F2 = 6$ ,  $F3 = 3$  (*op.cit.*, 21). This representation, while it has clear motivation as a way of describing the acoustic properties of /i/, gives something of a hollow interpretation of the auditory dimensions it employs. It does not, for instance, make predictions as to what would happen if we were to make adjustments to a single auditory dimension in the representation. For example, changing the F1 value to 3 would create a noticeable change in vowel height. At the same time however, if we changed the F3 value to 1 and kept the other values constant, the resulting vowel would still have largely an /i/-like quality. Flemming's representations do not allow us to make this prediction. The functional primacy of F1 over F3 is not incorporated into the model, which may be argued to be phonetic rather than phonological.

Our approach to auditory representations might establish perceptual cues as privative building blocks based not on continuous scales of single formant values, but on formant values in relation to other spectral properties. This perspective builds on two streams of research in speech perception that have received little attention in phonetically-based phonology. The first is Modulation Theory (Trautmüller 1994), according to which linguistic content is housed within modulations to a carrier signal (see also Harris 2009). Other research has investigated the role of spectral convergences rather than individual formants. If the carrier signal is a schwa-like periodic signal with spectral peaks that are evenly spaced, formant convergences represent significant modulations. Listeners hear convergence of spectral peaks when the frequency difference between the peaks is less than 3 Bark (Chistovich et al. 1979).<sup>9</sup> Later studies (Hoe-meke–Diehl 1994; Fahey et al. 1996) have shown that listener perception of vowel height is categorical, based on a convergence of the first formant with the fundamental. Similarly, the perceptual role of single formants

<sup>9</sup> Because of differences in pitch sensitivity, the acoustic distance (in Hertz) between a “converged” F2 and F3 may be much greater than in the case of the F1– $f_0$  convergence. The auditory distance (in Bark), however, will be similar.

may also be seen from a privative perspective. A measurement of 300 Hz for F1 is not capable of carrying linguistic meaning—what is meaningful is that the F1 is low with respect to a schwa-based midpoint.

One area where a privative perspective on perceptual cues may reap benefits is rounding harmony, a process that serves to increase the perceptual salience of the feature that triggers it (Kaun 2004), and finding motivation in the thesis that Rhymes correspond with perceptually dull portions of the signal. Kaun observes that mid vowels are better triggers for harmony than high vowels. Since rounding on mid vowels is less robust it may require additional duration to be reliably perceived. Let us assume the presence of an /u/-like prime, such as an element {**U**}, that may be broken down into two (or more) privative cues: a lowered F2 that is present in both mid and high rounded vowels, and an F1– $f_0$  convergence found only in the high vowels. An /u/, which contains two privative cues, has an inherently more robust {**U**} realization than /o/, so its rounding is less likely to spread by means of a perceptually motivated harmony process.

## 6. Concluding remarks

The onset-based theory of representation briefly outlined here offers a new perspective on the phonetics–phonology relationship. Inspired by GP’s minimalist view of the mapping between phonology and speech, we have implemented concrete acoustic specifications for constituent structure. These structures offer a universal acoustic frame from which the marked patterns of individual languages may be derived, and a predictive environment for new experimental studies to enhance our understanding of the relationship between speech and phonology.

## References

- Blevins, Juliette–Andrew Garrett 2004. The evolution of metathesis. In: Hayes et al. (2004, 117–56).
- Bregman, Albert S. 1990. Auditory scene analysis. The perceptual organization of sound. MIT Press, Cambridge MA.
- Chistovich, Ludmilla A.–Ruslan L. Sheikin–Valentina V. Lublinskaja 1979. ‘Centres of gravity’ and spectral peaks as the determinants of vowel quality. In: Björn Lindblom–Sven Ohman (eds): *Frontiers of speech communication research*, 143–57. Academic Press, London.

- Content, Alain–Ruth Kearns–Uli Frauenfelder 2001. Boundaries versus onsets in syllabic segmentation. In: *Journal of Memory and Language* 45: 177–99.
- Donegan, Patricia J. 2002. Phonological processes and phonetic rules. In: Katarzyna Dziubalska-Kořaczyk–Jarosław Weckwerth (eds): *Future challenges for natural linguistics*, 57–81. Lincom, Munich.
- Fahey, Richard P.–Randy L. Diehl–Hartmut Traunmüller 1996. Perception of back vowels: Effects of varying F1–F0 Bark distance. In: *Journal of the Acoustical Society of America* 99: 2350–7.
- Flemming, Edward 2002. *Auditory representations in phonology*. Routledge, New York.
- Harris, John 2004. Release the captive coda: The foot as a domain of phonetic interpretation. In: John Local–Richard Ogden–Rosalind Temple (eds): *Phonetic interpretation (Papers in Laboratory Phonology 6)*, 103–29. Cambridge University Press.
- Harris, John 2009. Why final devoicing is weakening. In: Kuniya Nasukawa–Philip Backley (eds): *Strength relations in phonology*, 9–46. Mouton de Gruyter, Berlin & New York.
- Harris, John–Geoff Lindsey 1995. The elements of phonological representation. In: Jacques Durand–Francis Katamba (eds): *Frontiers of phonology: Atoms, structures, derivations*, 34–79. Longman, Harlow.
- Hayes, Bruce–Robert Kirchner–Donca Steriade (eds) 2004. *Phonetically based phonology*. Cambridge University Press, Cambridge.
- Hayes, Bruce–Donca Steriade 2004. The phonetic bases of phonological markedness. In: Hayes et al. (2004, 1–33).
- Hoemeke, Kathryn A.–Randy L. Diehl 1994. Perception of vowel height: The role of F1–f0 distance. In: *Journal of the Acoustical Society of America* 96: 661–74.
- Jensen, Sean 1994. Is ? an element? Towards a non-segmental phonology. In: *SOAS Working Papers in Linguistics and Phonetics* 4: 71–8.
- Kaisse, Ellen 1992. Can [consonantal] spread? In: *Language* 68: 313–32.
- Kamprath, Christine K. 1987. Swiss-Romansch glide, extrasyllabicity, and the sonority scale. In: *Texas Linguistic Forum* 28: 28–40.
- Kaun, Abigail 2004. The typology of rounding harmony. In: Hayes et al. (2004, 87–116).
- Kaye, Jonathan D.–Markus Alexander Pöchtrager 2009. GP 2.0. Paper presented at the GPRT6, Piliscsaba, Hungary, April 25, 2009.
- Labov, William 1994. *Principles of linguistic change: Internal factors*. Wiley-Blackwell, Oxford.
- Lindblom, Björn 1990. Explaining phonetic variation: A sketch of the H&H theory. In: William J. Hardcastle–Alain Marchal (eds): *Speech production and speech modeling*, 403–40. Kluwer, Dordrecht.
- Ohala, John J. 1981. The listener as a source of sound change. In: Carrie S. Masek–Roberta A. Hendrik–Mary Frances Miller (eds): *Papers from the parasession on language and behaviour (CLS 17)*, 178–203. Chicago Linguistics Society, Chicago.
- Ohala, John J. 1990. There is no interface between phonology and phonetics: A personal view. In: *Journal of Phonetics* 18: 153–71.



- Palmer, John M. 1993. *Anatomy for speech and hearing*. Lippincott, Williams & Wilkins, Baltimore.
- Pöchtrager, Markus Alexander 2006. *The structure of length*. Doctoral dissertation, Universität Wien.
- Scheer, Tobias 2001. *On locality and phonology in phonology*. Ms. (<http://www.unice.fr/dsl/tobias.htm>; date of access: Jan 13, 2008)
- Schwartz, Geoff 2009. *Onset prominence and phonotactics*. Poster presented at the 17th Manchester Phonology Meeting.
- Stevens, Kenneth N. 1998. *Acoustic phonetics*. MIT Press, Cambridge MA.
- Traunmüller, Hartmut 1994. *Conventional, biological, and environmental factors in speech communication: A modulation theory*. In: *Phonetica* 51:170–83.
- Wright, Richard 2004. *A review of perceptual cues and cue robustness*. In: Hayes et al. (2004, 34–57).
- Wright, Richard – Stefan A. Frisch – David B. Pisoni 1997. *Speech perception. Research on spoken language processing (Progress Report No. 21)*. Indiana University.