*Poznań Studies in Contemporary Linguistics 45(1), 2009, pp. 149–168*
*© School of English, Adam Mickiewicz University, Poznań, Poland*
*doi:10.2478/v10010-009-0008-4*

# LISTENER ORIENTED REPRESENTATIONS IN NATURAL PHONOLOGY

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

## ABSTRACT

While Natural Phonology has long contended that phonemes are specified for their phonetic properties, followers of the theory have concentrated primarily on phonological processes, instead of delving into the details of pronounceable representations. In the area of representation, NP has thus failed to pursue its claim that systematic articulatory and perceptual phenomena below the level of segmental contrast must be treated phonologically. By building an explicit model of representation in NP, we may help the theory to meet one of its primary challenges: "to confirm the hypothesis that speech processing is categorical, or phonological, down to the level of the actual phonetic (pronounceable) representation" (Donegan 2002: 79). Prominence Phonology (Schwartz, in press) is an NP-inspired model that seeks to take Donegan's call to action to heart, introducing new and phonetically explicit representations based upon scalar yet monovalent elemental primes. This paper introduces these representations with the goal of refining our view of the signal so as to develop a phonological view of speech.

KEYWORDS: Natural Phonology; auditory representations; element theory.

## 1. Introduction

One of the major challenges for any phonetically inclined theory of phonology is to establish a language of phonological entities that may be defined in terms of their physical and perceptual properties. This challenge emerges out of the longstanding difficulties encountered by speech perception research in the search for the "units of perception" (e.g. Wright et al. 1997). If we assume that universal phonological primes indeed exist, we should expect them to make themselves evident in the acoustic signal. Presumably under the influence of the alphabetic writing systems and transcription conventions such as the IPA, segments have long been considered as basic units. However, the linearity problem and the lack of acoustic-phonetic invariance (see e.g. Lass 1994) represent a dark cloud over segments' primitive status in determining the relationship between signal and symbol. The status of the syllable may also be questioned. In some languages they seem to qualify as units of processing, but this is far from being a clear universal. Even "phonetically defined" features of the type introduced in *SPE* (*Sound Pattern of English*; Chomsky and Halle 1968) cannot be taken for granted. Thus, for example, a physical characterization of a traditional feature such as [voice] has proven elusive, as evidenced by the large degree of cross-language variation found in the implementation of voice contrasts.

In the search for common ground between the cognitive units of language and their physical correlates, phonemic contrast has traditionally played a significant role. Phonetic features that are predictable by rule have been seen as non-contrastive, and not afforded significant status in phonological representations. Such phenomena as aspiration were thus relegated to the "phonetic implementation" component of the grammar, and a number of levels of representation were assumed. In accordance with this view, aspiration and the long-lag VOT that accompanies it, which constitute the primary perceptual ingredient in initial English voicing contrasts (Lisker and Abramson 1964), are considered to be the product of "phonetic implementation" rules. In other words, [voice] is seen as a universal feature with different physical implementations.

This, ultimately, cannot be a satisfactory explanation, since in a phonetically grounded phonological model, systematically different physical properties must be attributable to systemic differences in phonological representation. According to Pierrehumbert (2000: 116), "every thorough study that has looked for a difference between two languages in details of phonetic implementation has found one". In other words, the "phonetic implementation" view is untenable, often leaving important physical properties unconsidered. As a result, phonetic implementation rules must be excised from the grammar and replaced by phonological representations that incorporate non-distinctive properties.

In order for a phonetically grounded phonological model to be workable, it is necessary to refine our view of the signal. Instead of relegating phonetic details to an implementation component of the grammar, the goal is to search the speech signal for entities that are discrete and therefore by nature phonological. This view of the signal has been espoused in Natural Phonology (henceforth NP; Stampe 1973; Donegan and Stampe 1979). In NP, all representations are seen as "pronounceable". If this is indeed the case, then the signal should contain cues to all the phonologically relevant information in an utterance. Thus, the challenge for NP "is to confirm the hypothesis that speech processing is categorical, or phonological, down to the level of actual phonetic (pronounceable) representation" (Donegan 2002: 79). Implicit in this hypothesis is a notion that is easily overlooked considering the traditional divisions of labor between phonetics and phonology. In short, the acoustic signal, despite its continuous nature, contains categorical entities that are phonological in nature. The task of the researcher is to identify these entities.

This paper will outline a set of representations developed in Prominence Phonology (Schwartz, in press), a listener oriented model inspired by NP. Prominence Phonology operates on the basis of three major assumptions that underlie the workings of the model. The first is that all phonological representations are specified only for their auditory properties.[1] Secondly, we posit monovalent elements (Harris and Lindsey 1995) instead of binary features as the substance from which phonological representations are constructed. Finally, the model interprets the universality of a CV sequence as the primary building block of phonological structure. Due to space limitations, these assumptions will not be defended in detail here (for thorough argumentation, see Chapter 2 of Schwartz, in press). Rather, this paper will focus on the auditory anatomy of Prominence Phonology's elemental primes. As we shall see, by considering certain auditory properties that are privative in nature, we gain a perspective on speech that is truly phonological. Since the view of phonological representations presented here represents a departure from traditional views, much space in this paper is devoted to explication, at the expense of empirical support. Nevertheless, these representations are constructed in a phonetically explicit manner that is both experimentally and empirically testable.

The rest of this paper will proceed as follows. Section Two will argue for a privative, categorical view of listener orientation in phonology. Section Three will discuss the melody-structure opposition from an auditory perspective. Section Four will introduce a selection of the cues employed in the Prominence model as the building blocks of phonological elements. Section Five will illustrate some interactions between elemental realization and process application in the Prominence model.

## 2. A privative take on listener orientation

Flemming (2002) is one of a relatively small number of major works to define phonological representations in terms of their auditory properties. Flemming uses scalar auditory dimensions that function within Optimality Theoretic constraints on contrasts. While the nature of auditory phenomena may be scalar, Flemming's model provides no obvious 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 (Flemming 2004: 238). This representation, while it has clear phonetic motivation as a way of describing the auditory 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

[1] This is a strong hypothesis that, as one *PSiCL* reviewer has noted, represents something of a break from Classical NP. In my view, however, auditory properties of speech (such as critical bands and onset boosts) are quantal and lend themselves to categorical study in a way that articulation does not. Computers may learn to produce perceptible spoken language with a completely different articulatory apparatus. Articulation, of course, motivates a large number of phonological processes. However, all processes must be perceptually licensed in order to take hold in the grammar of a language. Nevertheless, I do not believe that this view is incompatible with NP. It merely offers a new strategy for the representation of processes.

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. As a result, they are better seen in phonetic rather than phonological terms.

A more promising "phonological" approach to auditory representations might establish 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 the OT-inclined "phonetically-based phonology" literature. The first is Modulation Theory (Traunmüller 1994), according to which linguistic content is housed within modulations to a carrier signal. In this view, a particular acoustic property may bear linguistic currency only if it represents a significant modulation of the carrier (see also Harris 2006). The second area of research that is relevant for this discussion has investigated the role of spectral convergences rather than individual formants. If the carrier signal is assumed to be a schwa-like periodic signal with spectral peaks that are evenly spaced, formant convergences represent significant modulations that may bear linguistic content. Perceptual studies have found that listeners hear convergence of spectral peaks when the frequency difference between the peaks is less than 3 Bark (Chistovich et al. 1979).[2] Later studies (Homeke and 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 frequency. In a phonological model that is based on these assumptions, perceptual cues must be privative rather than scalar in nature.

For an illustration of how privative cues might operate, consider Flemming's representation of /i/ discussed above. When we raised F1 from 1 to 3, there was a noticeable difference in vowel height, but the lowering of F3 from 3 to 1 did not entail a significant difference in vowel quality. To represent /i/ in terms of privative cues, we will posit the following building blocks in terms of the convergences of two spectral peaks: F1f0 and F3F2. Representation in terms of privative formant convergences allows us to predict the effects of making the above-mentioned changes in F1 or F3 in the case of /i/ discussed above. Raising F1 would result in a loss of the F1f0 convergence, reflecting the noticeable change in vowel quality. If F3 were lowered, the F3F2 convergence would still be present, and the vowel would retain its /i/-like quality. If we take Flemming's formant parameters at face value, we have no way of predicting that the functional role of the F1 dimension is significantly greater than that of the F3 dimension – we should expect a difference of 1 on the two dimensions to have comparable auditory effects. By specifying our auditory cues in terms of formant convergences we may avoid this problem.

[2] 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 F1f0 convergence. The auditory distance (in Bark), however, will be similar.

Nevertheless, privative formant convergences are not sufficient for describing a number of vowel contrasts. For example, our model needs to be able to represent the difference between an F1 of 1 and an F1 of 2 according to Flemming's representations. Such a difference may be used to characterize a contrast between /i/ and /ɪ/, where both vowels are high and presumably marked by a convergence of F1 and f0. As a consequence, in addition to formant convergence cues, we shall propose additional cues based on the frequencies of single formants. Returning to the F1 difference between /i/ and /ɪ/ mentioned earlier, on top of the F1f0 spectral convergence may observe another cue, LowF1, that is present when the F1 value is more than 1–1.5 Bark below the central value for F1.[3] The LowF1 cue is present in /i/ but absent in /ɪ/, whose F1 is only slightly lower than the central value. Thus, the tense vowel is signaled by the presence of both LowF1 and the F1f0 spectral convergence, while the lax vowel contains only the convergence. In the case of /i/, LowF1 may be seen to increase the salience of the convergence.

When we look at the relationship between LowF1 and F1f0, we are presented with an implicational interaction – if F1 is low, there must be a convergence of F1 and f0. In other words, LowF1 implies F1f0,[4] so in some sense it may be seen as "redundant" from the perspective of the convergence. However, it is indeed the redundancy that contributes to the realization of a subtle phonetic difference. Building redundancies into phonological representations has a number of benefits. Most importantly, it provides phonology with a useful parallel to the role of redundancy in speech perception. Wright et al (1997: 24) characterize this role as follows:

> Since the main goal of speech is the communication of ideas from the talker to the hearer (normally under less than optimal conditions), it is not surprising that spoken language is a highly redundant system of information transmission. While redundancy of information in a transmission system implies inefficient encoding, it facilitates error correction and recovery of the intended signal in a noisy environment and insures that the listener recovers the talker's intended message.

If one considers the fact that speech perception occurs in real time, it is easy to see how linguistically relevant acoustic information can be missed when listening conditions are not ideal. Thanks to redundancy the listener is able to overcome difficult listening conditions. Indeed, "inefficient encoding" must be an essential property of phonological representations, for without redundancy oral communication would be impossible.

[3] This central value may be thought of in terms of uniform tube models (e.g. Chapter 4 of Johnson 1997). According to this model, 500 Hz for F1 shall be thought of as a reference for adult male speakers.

[4] The reverse statement here is not necessarily true. Remember that a convergence is perceived when two peaks fall within 3 Bark of each other. However, the LowF1 is defined as a single spectral peak that is low in terms of a single reference point, the central value of the formant.

## 3. Melody and structure in auditory terms

In autosegmental approaches to phonology, melodic primitives are linked with structural positions. Implicit in this strategy is the assumption that it is the melodic primitives that specify phonetic properties such as voicing or place of articulation, whereas the structural positions themselves are devoid of phonetic content. This distinction between melody and structure is widely assumed, even among "phonetically based" approaches to phonology. For example, Steriade (1997) presents a cue-based account of laryngeal neutralizations in various languages, which is presented largely as a refutation of a "licensing by prosody" (e.g. Ito 1986) approach that relates the presence or absence of laryngeal contrasts to questions of syllable structure. A Naturalist approach would question the assumed separation of melody and structure, since by removing cue-based approaches from prosodic licensing (or vice versa), we deprive ourselves of the impetus to thoroughly investigate the relationship between the two areas.

When we consider the auditory properties of speech, clear definitions of melody and structure emerge. Cues associated with specific formant frequencies and bandwidths are melodic in nature, describing the specific timbre or quality of sounds. At the same time, more general auditory properties, such as duration, amplitude, periodicity, and the formant robustness, give structure to the signal. At the same time, spectral properties may have perceptual consequences for structure. For example, a raised spectral tilt contributes to subjective loudness and has been found to be a cue to stress in many languages (Sluijter and van Heuven 1996, 1997; Crosswhite 2003). Thus, both melody and structure may be specified in terms of psychophysical properties found in the speech signal. They represent two separate types of auditory information that are nevertheless subject to certain interactions. This has been reported in speech perception research, but has received relatively little attention in phonological theory, where bare skeletal slots have been the norm.

Natural Phonology, on the other hand, has long assumed that melody and structure interact "in the act of speaking" (Donegan 2002: 57). Implicit in this assumption is the idea that structural positions, in addition to melodic properties, must find some expression in the speech signal. Unfortunately, this idea has yet to be flushed out explicitly in NP. An explicit phonological account can provide a new perspective for phonetic studies, which have largely failed to identify the physical correlates of phonological structure. A primary goal of Prominence Phonology is therefore to provide explicit representations whose validity may be tested by means of experimental phonetic study. While the difficulties of phonetic research in identifying phonological entities in speech has led many to assume a strict separation of phonetics and phonology, Prominence Phonology seeks to alter the priorities of phonetic study in the hope that a new perspective will finally enable us to produce a plausible account of the relationship between signal and symbol.

The primitive elements of Prominence Phonology have been proposed with an eye toward predicting the behavior and interactions of melody and structure as they are reflected in speech. Thus, we may posit a hierarchy of elements that is based on the nature of the auditory cues from which they are constructed. At the bottom of the hierarchy lie elements that are purely melodic, whose cues are defined by spectral properties. An example is the element **I**, which is constructed of cues that purely spectral in nature. The element **A** is something of a hybrid element in this model. It contains formant cues that are spectral, but it is also specified for general auditory properties such as loudness and duration. Such properties are inherently structural. At the top of the hierarchy we find the onset element **ʔ** with cues that denote solely structural properties.

Some works in Element Theory, noting the special behavior of certain elements, propose reductions in the elemental inventory, replacing these elements with structural positions. For instance, Jensen (1994) eliminates the element **ʔ** (denoting occlusion or "stopness" in traditional accounts), while Pöchtrager (2006) replaces **H** with additional structural positions. Such a strategy is a necessity under the traditional assumption that all elements are melodic, and that structural positions lack phonetic specification. The account to be developed here explains the unusual behavior of certain elements in terms of their auditory anatomy, so the "structural" nature of **ʔ** (Jensen 1994) falls out directly from its cues that denote structural properties.

## 4. A selection of auditory cues

This section will take a look at some of the perceptual cues that have been posited for elemental representations in Prominence Phonology. Due to space restrictions, we will limit ourselves to the cues associated with four basic elements: **I**, **U**, **A**, and **ʔ**. It is important to note that in the theory presented here, elemental realization is scalar, specified with the privative presence or absence of perceptual cues.

### 4.1. Spectral cues

In keeping with the discussion in the previous section, cues are seen here as being either melodic or structural in nature. We shall then observe for instance that the elements **I** and **U**, characterized in Prominence Phonology as Spectral Elements, are indeed defined by purely spectral properties. While textbooks in acoustic phonetics generally concern themselves with numerical measures of formant frequencies, Prominence Phonology takes a privative perspective on spectral properties in the signal. Privativity of spectral properties is possible under the assumptions of Modulation Theory (Traunmüller 1994), whereby linguistic content is contained in modulations to a schwa-like carrier signal. In this view, a given modulation is either present or absent in the realization of an element. As a result, instead of continuous physical scales for auditory properties, we may assume discrete auditory categories that are inherently phonological.

### 4.1.1. Formant convergence cues

The significance of formant convergences, as opposed to the frequency values of single formants, has a long history in auditory perception research, going back as far as the 1950's (Delattre et al 1952). Syrdal and Gopal (1986) proposed a model of vowel perception based on auditory representations in Bark (a psychophysical measure of critical bands in hearing). Their model proposed formant convergences, based on Chistovich et al.'s (1979) critical distance experiments, when two spectral peaks were 3 Bark or less apart.

The first convergence, which has been found to be a perceptual indicator of tongue height (Homeke and Diehl 1994; Fahey et al. 1996), is between the first formant (F1) and the fundamental frequency or pitch (f0). High vowels contain this convergence, which will be labeled in the present model as F1f0. This cue is present in both **I** and **U**, allowing us to provide a unified account of processes affecting chromatic (Donegan 1978/1985) vowels.

The other important formant convergence is F3F2, denoting a distance of less than three Bark between the second and third formants. This convergence is present in front vowels, unified with the element **I**. This convergence can also be found in the formant transitions of velar consonants, as well is in the formant structure of some rhotics.

A third convergence, F2F1, is somewhat more difficult of characterize. This convergence does not seem to universally represent any articulatory parameter (Syrdal and Gopal 1986). It may be found in back vowels, but only when they are low and mid. As a result, it may be seen as an ingredient in the element **U**, but since it requires a high F1, it eliminates the LowF1 and F1f0 cues from **U** realization. Alternatively, it may posited for low vowels, but only when they are not front.

### 4.1.2. Single Formant Cues

The frequencies of single formants are the most commonly employed acoustic phenomena in phonetics, appearing in every textbook. Schwa is often assumed to present a set of defaults for F1, F2, and F3 at 500 Hz, 1500 Hz, and 2500 Hz, respectively for male speakers with an average size vocal tract (e.g. Johnson 1997). Low vowels raise F1. Front vowels raise F2. Lip rounding lowers all formants, including F3.

In accordance with Modulation theory (Traunmüller 1994), single formant cues are specified in Prominence Phonology only when they differ by more than one Bark (auditory critical band) from the default values produced by schwa. Thus, these formant frequencies are only relevant when they are either low or high in relation to a reference point.

LowF1 is present in vowels that are high and tense (or +ATR), and serves to enhance the F1f0 cue. HighF1 is present in low vowels and contributes to the realization of the element **A**. LowF2 is present in labial articulations, as such it is an ingredient in the element **U**. HighF2 enhances the F3F2 cue that is present in front vowels. It is also present in coronal and velar formant transitions.

The perceptual significance of F3 when it not part of a spectral convergence may be determined language-specifically. It seems that while the F3 cues are present universally, they are not always perceptually salient enough to be exploited by a language. This is probably due to the fact at the frequency range generally associated with the third formant, pitch sensitivity in the auditory system is relatively low. As a result, it is more difficult for listeners to distinguish differences in F3 than it is for differences in F1 or F2. These facts presumably contribute to the difficulty that speakers of many languages have in hearing the difference between /l/ and /r/, segments that are generally distinguished by the frequencies of F3. As a result of the relative subtlety of single F3 cues, they play a somewhat limited role in elemental realization. The assumption is that they are universally present, but not necessarily universally adopted by languages. When they are adopted, they often serve to reinforce elemental realizations based upon other cues.

LowF3 is a cue associated with lip rounding as well as retroflexion. HighF3 is present in front vowels so it will be posited as an ingredient in **I**. The HighF3 cue is also present in coronal formant transitions (Harris et al. 1957). However, because of the fact that coronal gestures are frequently quite rapid (Browman and Goldstein 1990), the transitions they produce may be less salient than those of other places of articulation (cf. Hume et al 1997).

An additional cue for **I** an **U** is associated with very tense articulations of extreme peripheral vowels. The cue will be referred to in Prominence Phonology as Weak Harmonics (WH). Acoustically, WH is characterized by a drop off in amplitude of all harmonics above the fundamental frequency. This type of harmonic structure may be seen to enhance the F1f0 cue by increasing the relative amplitude of the fundamental. Thus it will be posited as a cue to the color elements **I** and **U**. The WH cue also is present in nasals, whose harmonic structure weakens at higher frequencies.

To illustrate the WH cue, Figure 1 (overleaf) shows FFT spectral displays of the vowels in the English words *cooed* (solid peaks) and *could* (dashed peaks). Each of the peaks in the spectra corresponds to a harmonic of periodic vocal fold vibration. Notice how the harmonic structure of the vowel in *could* is much stronger than that of *cooed.* A similar pattern can be found in pairs of the *sheep–ship* type. Though both vowels are high and front, in the tense vowel significantly more energy is present at frequencies between F1 and F2. These harmonic properties mark a slight different voice quality associated with tense and lax vowels in American English. Weak harmonics signal a slightly breathy voice quality and lesser overall amplitude, while strong harmonics entail greater sonority. Voice quality has been found to contribute to the tense-lax distinction in many dialects of American English (DiPaolo and Faber 1990).

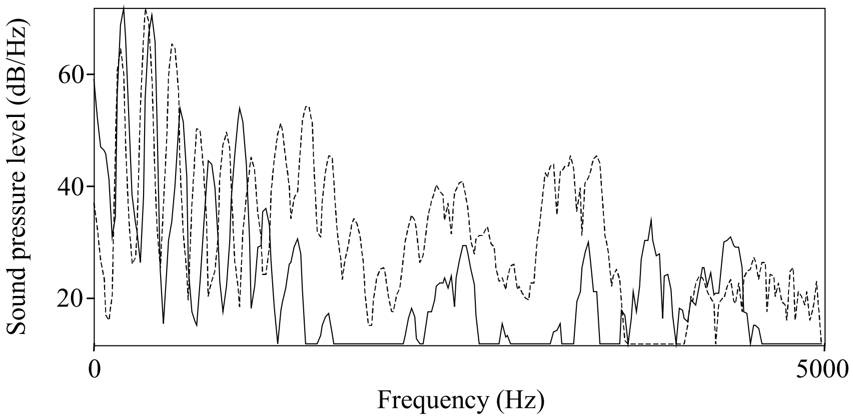

Figure 1. FFT spectra of *could* (dashed peaks) and *cooed* (solid peaks) produced by a speaker of American English. Notice the weaker harmonic structure in the tense vowel.

### 4.2. Structural cues

Structural cues describe more general auditory properties than melodic cues. These may include both static aspects of the signal such as frication, loudness, and duration, as well as dynamic properties such as amplitude rise time and formant transitions. In Prominence Phonology, which posits that onsets and nuclei are the only structural positions (cf. Scheer 2004), the element **A** specifies the inherent nuclear properties of a segment, while the element **ʔ** specifies how well a segment fits into the onset position. These two elements are referred to as *Beat Structure Elements*.

#### 4.2.1. Onset cues

If it is indeed possible to relate the speech signal to phonological structure, the amplitude envelope must play a significant role. When the amplitude of stimulus rises quickly at onset, a perceptual boost occurs (Wright 2004), enhancing the perceptibility of contrasts in onset position. Amplitude rise time has been found to help listeners distinguish phonetic categories (e.g. stops from glides; Shinn and Blumstein 1984). Prominence Phonology takes this finding one step further, assuming that rapid amplitude rises help delineate structural positions. They are assumed here to be important ingredients in the element **ʔ**, which in the present model denotes the structural position of onset.

Rapid Rise 1 (RR1) is the first of two amplitude rise cues to **ʔ**. This cue corresponds to the release of stop consonants, activating auditory nerve fibers with characteristic frequencies above around 1500 Hz.[5] In order for the RR1 cue to be present, the amplitude rise must be rapid. Fricatives and sonorant consonants lack this cue, since they are characterized by a more gradual rise in amplitude. Fricatives and sonorants are thus less prominent as onsets than plosives, a notion that is reflected in the fact that they are less frequently encountered in languages' segmental inventories (Maddieson 1984). The second rapid rise cue (RR2) corresponds to vocalic onsets. It is posited as a separate from RR1 because of the tonotopic organization (Greenberg 1996) of auditory nerve fibers (ANFs). Different ANFs are activated by different frequencies. Since noise bursts tend to be higher in frequency than vocalic formant onsets, two onset boosts are possible, one in the formant frequency range and the other in the burst frequency range. RR2 denotes a rapid rise of amplitude in the frequency range below about 1500 Hz

Figure 2 shows a waveform of the English word *cot* with an intensity contour superimposed on the waveform. The two separate RR cues associated with the initial [kʰ] are labeled on the display. Notice the steep slope of the intensity rises both for stop release and for vocalic onset. In the case of aspirated stops in English, it is worth noticing that the long VOT provides temporal separation of the two RR cues, increasing their perceptual salience. When stops are unaspirated, the temporal delay between RR1 and RR2 is minimal.

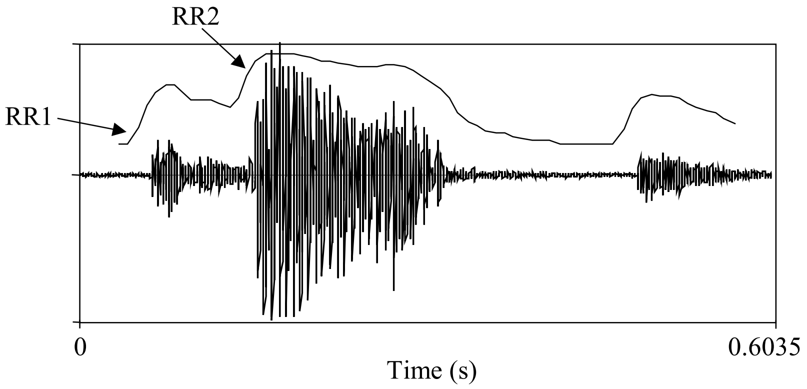

Figure 2. Waveform and intensity contour of English *cot*.

Figure 3 shows a waveform and intensity contour of the English word *done*. In this case the amplitude rise on the vocalic onset is gradual (compare with Figure 2), so we shall

[5] Labial stops may lack this cue, since the spectrum of their burst noise is generally lower (Stevens and Blumstein 1981), and frequently lower in amplitude.

assume that RR2 is absent.[6] Comparing the two figures provides an interesting perspective on the lenis-fortis distinction in languages like English – the fortis stop contains an additional structural cue to the onset position. Thus, we find a perceptual parallel with Jensen's (1994) and Pöchtrager's (2006) proposals, which associate **ʔ** and **H** with additional structural positions in the phonological skeleton.

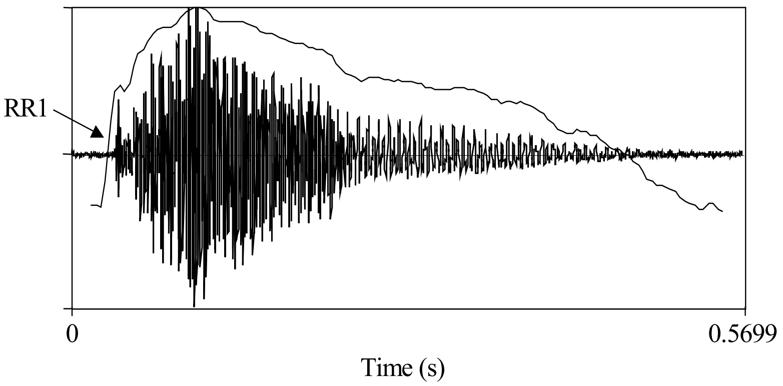

Figure 3. Waveform and intensity contour of English *done*.

An additional cue to **ʔ** is Silence, which allows full auditory recovery, thus facilitating an onset boost (Wright 2004). Silence is of course associated with the closure period of stop consonants. Silence is present in voiceless stops. Voiced stops are marked by periodicity during closure, as well as shorter closure, and thus do not contain the Silence cue. When there is some periodicity during the closure period, recovery is less complete, and a smaller onset boost is observed (Wright 2004: 45). A fourth cue that is posited for the realization of **ʔ** is Aperiodic Noise. While noise *per se* is not directly involved in producing onset boosts, it is a by-product of many onset articulations and is often present during onset boosts.

The final cue that we shall posit for **ʔ** is the CV formant transition. Formant transitions have long been examined for their melodic properties – frequently they are listeners' best tools for identifying the place of articulation of consonants. Additionally, however, formant movement may help provide structural information to listeners. Consider

[6] The contrast posited here between the presence and absence of RR2 is admittedly somewhat speculative. I am unaware of any published experimental studies that document it. Informally, I have observed it in numerous tokens in English obstruents (vocalic rise time is rapid for voiceless obstruents), and in Polish fricatives (vocalic rise time is rapid for voiceless but not voiced fricatives). The phenomenon seems aerodynamically plausible, and calls out for experimental study.

here the example of glides. In autosegmental approaches to phonology, vowels and glides are seen as identical expressions, the difference lying in alignment with structural positions. This alignment is indeed reflected in the signal by the presence of formant movement, which defines glides as consonantal. The present model takes this notion a step further, and assumes the simple presence of CV transitions provides structural information for listeners – formant movement denotes an onset, while spectral steady states are associated with nuclei.

### 4.2.2. Nuclear cues

In Prominence Phonology, the element **A** is assumed to represent the structural position of nucleus. Five cues are proposed in the realization of **A**, which all are present in low vowels, universally preferred as nuclei. Four out of these five cues are structural in nature, while the final cue (HighF1) is melodic. In addition, three of the structural cues (loudness, robust formant structure, duration) may interact with melodic properties. Thus, while **ʔ** is entirely structural and **I** and **U** are entirely melodic, the element **A** represents a true hybrid of melody and structure. This postulation of **A** finds parallels in Pöchtrager (2006:8), who suggests that in GP the element **A** would be better described by means of a structural interpretation.

Loudness is probably the most problematic of the cues presented here in that it is difficult to state in privative terms. Nevertheless, for the purposes of representation we shall make that stipulation that it is indeed privative. Loudness is assumed to be present in non-high vowels, which are generally quite high in amplitude (Lehiste and Peterson 1959), as well as in sibilant noise, which occurs at frequencies to which the auditory system is quite sensitive. Periodicity is another structural cue, present of course in all voiced sounds. Robust Formant Structure, based on formants with relatively narrow bandwidths (Johnson 1997), helps listeners to distinguish oral vowels from nasals and other sonorant consonants such as laterals. Segments that are entirely vocalic generally have a greater robustness to their formant structure, a notion that can also provide cues to listeners as to whether or not they are hearing a nucleus.

Duration is posited as an additional structural cue to **A**. This is something of a problematic stipulation because of the large amount of durational variability due to factors such as rate of speech. In Prominence Phonology, duration as a nuclear cue is defined as a spectral steady state that is long enough to be perceived as such. Thus, the cue will be referred to as Steady State Duration (SSD), which is in essence the inverse of the CV formant movement cue for **ʔ**. The idea is that listeners perceive a correlation between nuclei and spectral steady states. The SSD cue implies a necessary interaction of spectral and melodic properties.

The SSD cue, while still in need of further experimental confirmation, is suggested by an interesting asymmetry in Polish speakers' perception of American English (AE) diphthongs (Bogacka 2005). Polish listeners perceive AE /aɪ/ as a native vowel-glide sequence /aj/, where as /oʊ/ is not heard as the corresponding native vowel-glide sequence /ow/, but rather as a single vowel /o/. While both AE diphthongs are long in duration, only the one with lesser formant movement /oʊ/ is heard as single vowel, suggesting that in Polish nuclei may be specified as having a spectral steady state. Schwartz (2007b) provides further evidence of a requirement for SSD in Polish. In a cross language study of vowel quality, the experiment found that Polish realizations of /i/, /a/, and /u/ were marked by a significantly greater spectral steady state as a percentage of overall vowel duration than the corresponding English vowels. A perception test employing the Silent Center paradigm (Jenkins and Strange 1999) suggested that the dynamic specification that English speakers rely on in vowel identification is less important in Polish, reflecting that language's tendency for pure vowels and spectrally steady nuclei.

When the first formant of vowel is high, it raises the amplitude of harmonics in a frequency range that is clearly distinct from the fundamental. This indicates an absence of the F1f0 convergence. As a result the first formant and the fundamental may contribute independently to the subjective loudness of the vowel. A high first formant is thus said to raise the spectral tilt of the sound source. This spectral property has been found to contribute to subjective loudness (Sluijter and van Heuven 1996, 1997; Crosswhite 2003) and can be a cue to stress in many languages. The HighF1 cue is present in low vowels, the most sonorant of all speech sounds.

### 4.3. Summary of cues in elemental realization

Table 1 provides a summary of the cues presented thus far as privative building blocks of four elements, **I**, **U**, **A**, and **ʔ**. The presence of a cue contributes 1 to the realization of an element.

Table 1. Auditory anatomy of four elements in Prominence Phonology.

| Element | Cue |
|---|---|
| **ʔ** – Onset | RR1, RR2, Silence, Noise, CV transition |
| **A** – Nucleus | Loudness, Periodicity, Robust Formant Structure, SSD, HighF1 |
| **I** – Palatal | F1f0, F3F2, LowF1, HighF2, HighF3, WH |
| **U** – Labial | F1f0, LowF1, LowF2, LowF3, WH |

Promenince representations are constructed as follows. Each cue contributes 1 to the realization of an element. For example a plosive stop contains all 5 cues to **ʔ** and is thus specified **ʔ**=5. A phonological process in Prominence Phonology is an adjustment to these representations, so spirantization of the aforementioned plosive entails an adjustment of **ʔ** realization from 5 to 3, corresponding to the removal of the Silence and RR1 cues.

Table 2 presents a selection of segments and their projected elemental settings (for additional explanation, see Schwartz, in press). Notice that latent realizations are possible of elements for which a given segment is not traditionally associated. These latent realizations are the result of cues that may be shared by more than one element.

Table 2. Projected realizations of selected segments for the four elements described above.

| Segment | **ʔ** | **A** | **I** | **U** |
|---|---|---|---|---|
| /i/ | 0–1 | 3 | 5–6 | 2 |
| /ɛ/ | 0–1 | 4 | 2–3 | 0 |
| /æ/ | 0–1 | 5 | 2 | 0 |
| /a/ | 0–1 | 5 | 1 | 1 |
| /u/ | 0–1 | 3 | 2 | 4–5 |
| /w/ | 1 | 2 | 2 | 5 |
| /ɹ/ | 1 | 2 | 1–2 | 1–2 |
| /n/ | 2 | 2 | 2 | 1 |
| /m/ | 2 | 2 | 1 | 2 |
| /h/ | 2 | 1 | 0 | 0 |
| /f/ | 3 | 1 | 0 | 2 |
| /s/ | 3 | 2 | 1 | 0 |
| /ʃ/ | 3 | 2 | 2 | 0 |
| /b/ | 3–4 | 0 | 0 | 1 |
| /d/ | 4 | 0 | 0 | 0 |
| /g/ | 4 | 0 | 1 | 0 |
| /p/ | 4–5 | 0 | 0 | 2 |
| /t/ | 5 | 0 | 0 | 0 |
| /k/ | 5 | 0 | 1 | 0 |

## 5. Elemental realization and phonological processes.

This section will provide a brief outline of how representations in Prominence Phonology reflect the realization of phonological elements in terms of the auditory cues discussed in the previous section. Because of space restrictions we shall limit ourselves to specifications and processes for only two of the elements discussed so far: **ʔ** and **A**. For a more complete account, see Schwartz (in press).

Segments in Prominence Phonology are seen as emergent propreties (cf. Bybee 2001) that are extracted from speech on the basis of auditory input during the language acquisition process. Although speech input produces a great deal of variation, "canonical" segmental representations are assumed to contain those properties that are common to input tokens in prosodically prominent positions. Because they emerge from speech input, segments are necessarily language-specific entities. A vowel such as /i/ must be assumed to have been extracted from auditory input in each of the languages in which it appears. In this view, language-specific differences in the realization of the "same" segment do not result from differences in "phonetic implementation". Rather, such differences imply the evolution of slightly different elemental representations.

The assumption that segments emerge from auditory input has an additional important consequence for representation in the Prominence framework. It follows that under this assumption segments must emerge with built-in specifications for both the onset element **ʔ** and the nuclear element **A**, in addition to any melodic properties. Thus, all segments are specified for how well they fit the structural positions of onset and nucleus. Those segments that produce all of the cues to **ʔ** (unvoiced stops) are the most prominent onsets, while those producing all the cues to **A** (low vowels) represent the most prominent nuclei.

Elemental specifications for segments are constructed according to the number of cues to a given element that are present when given segment is articulated in a prominent position. The presence of a cue contributes 1 to the realization of an element. Canonical segmental representations are specified for the cues present in prominent positions. In other positions where these cues may be physically absent, processes adjust elemental representations accordingly, unless listeners either reconstruct them in perception (Ohala 1981).

### 5.1. Processes affecting **A**

The role of listener reconstruction comes to the forefront in processes affecting the nuclear element **A**. Donegan (2001) considers the perception of phonological processes, where listeners "filter" out non-distinctive properties to reconstruct. One case she discusses are English words such as *bend*. In such words, despite the fact that the vowel is nasalized, listeners hear the vowel in *bend* as being the same as the vowel in *bed*. Nasalization entails an inherent lenition of a vowel's nuclear properties, obscuring the robustness of formants and reducing overall amplitude (Johnson 1997). By filtering out nasalization, listeners restore the vowel's inherent settings for **A.** Vowel lowering before nasal consonants, a well documented historical process (e.g. Italian /vino/ vs. French /vɛ̃/ 'wine'), may be analyzed in a similar way. The lowering restores the prominence of **A** that was obscured by the nasal.

## 5.2. Processes affecting **ʔ**

Prominence Phonology offers a somewhat new perspective on phonological structure, modeling a principle from speech perception research known as the primacy of onsets (Greenberg 1996; Content et al. 2001). Two important assumptions fall out from this principle. First, phonological structure is seen as a series of binary oppositions between onsets and nuclei, by which onsets represent portions of the signal that receive an auditory boost in perception (e.g. Wright 2004). Segments are specified for their inherent onset prominence in terms of the cues to **ʔ** discussed in Section 4. These specifications may be thought of as a kind of "anti-sonority" that has been invoked in an account of phonotactic patterns (Schwartz 2007a, 2008). However, instead of being an external look-up table whose justification is inherently circular (Harris 2006), the realization of **ʔ** is characterized by explicit and testable auditory specifications.

In addition to phonotactics, onset prominence provides an interesting perspective on the application of a number of phonological processes. Phonological processes in Prominence Phonology, in accordance with NP, make no reference to segments (Donegan 2002: 61), but rather take place within the domain of beat, a universal constituent structure comprised of an onset and a nucleus that are phonetically specified as the elements **ʔ** and **A**, respectively. A process is seen here as an adjustment to elemental representation, rather than a derivational substitution. Since representations in Prominence Phonology are specified for structural as well as melodic element, they offer insightful descriptions of what takes place during process application. Consider for example the English word *meant*, which may be pronounced as [mɛ̃t]. To account for this pronunciation segmentally requires an opaque interaction. The vowel must nasalize before the nasal consonant is deleted, otherwise the environment for nasalization is wiped out. In Prominence Phonology, an /n/ is specified for its inherent properties as an onset, with a realization of 2 for the onset element **ʔ**, corresponding to the RR2 and CV transition cues that /n/ produces in initial position. When /n/ is projected in non-initial position in a word such as *meant*, these cues are absent and the /n/ loses its inherent onset properties. In other words, the /n/ loses is structural specification, so its nasal specification must align with the preceding nucleus, producing a nasalized vowel. In this view, instead of two processes, one that nasalizes the vowel and another that deletes the nasal consonant, we have single process affecting the element **ʔ**.

The structural domain of adjustments to elemental realization allows us to make predictions about process application based on the presence or absence of cues to **ʔ**. Compare the words *man* and *manner* in American English. In many dialects, the former is almost always produced with a raised and nasalized vowel, while in the latter this process may be blocked.[7] In a Prominence account, the /n/ in manner may maintain its **ʔ**

[7] The raising of /æ/ before nasals may be an auditory enhancement of nasalization processes. Nasalization is cued by weakened energy above a fairly low frequency nasal formant (cf. the WH cue to **I**). Raising will lower F1, leaving a greater portion of the spectrum attenuated, ensuring the presence of the WH cue.

realization because its onset cues are preserved in pre-vocalic position, aligning the nasality with the second syllable and blocking effects on the preceding vowel. In dialects where *man* and *manner* have a similar raised vowel, the /n/ can be seen to have completely lost its onset cues, producing syllabification of the latter as [mæn.əɹ].

The element **ʔ** also offers an insightful account of vowel epenthesis. In the Prominence view, epenthesis is not seen as the insertion of a segment, but as the restoration of CV formant transition and RR2 cues to the onset element **ʔ**. Indeed, a segmental approach to epenthesis may be promlematic. Davidson (2007) found that vowels inserted by English speakers to break up foreign consonant clusters differ systematically from lexical schwas. Such a finding supports the assumption that the real motivation behind epenethesis is not to avoid articulatory difficulty, but rather to make the initial segment in a cluster more perceptible.

The primacy of onsets may have even further reaching implication for phonological process application (see Chapter 6 of Schwartz, in press). It allows us to make an interesting prediction with regard to which processes may occur under various external conditions affecting speech. In careful speech we may observe processes such as vowel reduction that lenite nuclei, while onsets are never reduced. Only in casual speech do we see onset lenition.

## 6. Summary

This paper has outlined the auditory representations posited in Prominence Phonology, a listener oriented model of Natural Phonology. These representations offer a truly categorical and phonological view of the speech signal, and seek to offer an explicit and testable account of what is commonly referred to as the relationship between phonetics and phonology.

## REFERENCES

Bogacka, A. 2005. "Why Poles can hear Sprite but not Coca-Cola". Talk given at the annual meeting of the German Linguistic Society, Cologne.

Browman, C.P. and L. Goldstein. 1990. "Tiers in articulatory phonology, with some implications for casual speech". In: Kingston, J. and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press. 341–376.

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

Chistovich, L.A., R.L. Sheikin and V.V. Lublinskaja. 1979. "'Centres of gravity' and spectral peaks as the determinants of vowel quality". In: Lindblom, B. and S. Ohman (eds.), *Frontiers of speech communication research*. London: Academic. 143–157.

Content, A., R. Kearns and U. Frauenfelder. 2001. "Bondaries versus onsets in syllabic segmentation". *Journal of Memory and Language* 45. 177–199.

Crosswhite, K. 2003. “Spectral tilt as a cue to word stress in Macedonian, Polish, and Bulgarian”. Paper presented at the 15th International Congress of Phonetic Sciences, Barcelona.

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

Delattre, P., A. Liberman and F. Cooper. 1955. “Acoustic loci and transitional cues for consonants”. *Journal of the Acoustical Society of America* 27(4). 769–773.

Delattre, P.C., A.M. Liberman, F.S. Cooper, and L.J. Gerstman. 1952. “An experimental study of the acoustic determinants of vowel colour: Observations on one- and two-formant vowels synthesized from spectrographic patterns”. *Word* 8. 195–210.

Delgutte, B. 1997. “Auditory neural processing of speech”. In: W.J. Hardcastle and J. Laver (eds.), *The handbook of phonetic sciences*. Oxford: Blackwell Publishers. 507–538.

Di Paolo, M. and A. Faber. 1990. “Phonation differences and the phonetic content of the tense-lax contrast in Utah English”. *Language Variation and Change* 2. 155–204.

Donegan, P.J. 1978. “On the natural phonology of vowels”. (Unpublished PhD dissertation, University of Ohio. Also *Ohio State University Working Papers in Linguistics* 23. Republished 1985: New York: Garland.)

Donegen, P.J. 2001. “Constraints and processes in phonological perception”. In: Dziubalska-Kołaczk, K. (ed.), *Constraints and preferences*. Berlin: Mouton de Gruyter. 43–68.

Donegan, P.J. 2002. “Phonological processes and phonetic rules”. In: Dziubalska-Kołaczyk, K. and J. Weckwerth (eds.), *Future challenges for Natural Linguistics*. Munich: Lincom. 57–81.

Donegan, P.J. and D. Stampe. 1979. “The study of Natural Phonology”. In: Dinnsen, D. (ed), *Current approaches to phonological theory*. Bloomington: Indiana University Press. 126–173.

Fahey, R.P., R.L. Diehl and H. Traunmüller. 1996. “Perception of back vowels: Effects of varying F1–F0 Bark distance”. *Journal of the Acoustical Society of America* 99. 2350–2357.

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

Flemming, E. 2004. “Contrast and perceptual distinctiveness”. In: Hayes, B., R. Kirchner and D. Steriade (eds.), *Phonetically based phonology*. Cambridge: Cambridge University Press. 232–276.

Greenberg, S. 1996. “Auditory processing of speech”. In: Lass, N. (ed.), *Principles of experimental phonetics*. St. Louis: Mosby. 362–408.

Harris, J. 2006. “On the phonology of being understood: Further arguments against sonority”. *Lingua* 116. 1483–1494.

Harris, J. and G. Lindsey. 1995. “The elements of phonological representation”. In: Durand, J. and F. Katamba (eds.), *Frontiers of phonology: Atoms, structures, derivations*. Harlow: Longman. 34–79.

Harris, K., H. Hoffman, A. Liberman, P. Delattre and F. Cooper. 1958. “Effect of third-formant transitions on the perception of voiced stop consonants”. Journal of the Acoustical Society of America 30(2). 122–126.

Hoemeke, K.A. and R.L. Diehl. 1994. ‘‘Perception of vowel height: The role of F1-f0 distance”. *Journal of the Acoustical Society of America* 96. 661–674.

Hume, E., K. Johnson, M. Seo, G. Tserdanelis and S. Winters. 1999. “A cross-linguistic study of stop place perception”. Paper presented at the 14th International Congress of Phonetic Sciences, San Francisco.

Jensen, S. 1994. “Is ʔ an element? Towards a non-segmental phonology”. *SOAS Working Papers in Linguistics and Phonetics* 4. 71–78.

Johnson, K. 1997. *Acoustic and auditory phonetics*. Cambridge, MA: Blackwell Publishers.

Lehiste, I. and G.E. Peterson. 1959. "Vowel amplitude and phonemic stress in American English". *Journal of the Acoustical Society of America* 31. 428–435.

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.

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

Pierrehumbert, J. 2000. "What people know about sounds of language". *Studies in the Linguistic Sciences* 29(2). 111–120.

Scheer, T. 2004. *A lateral theory of phonology: What is CVCV and why should it be?* Berlin: Mouton.

Schwartz, G. 2007a. "Prominence in beat structure". *Poznań Studies in Contemporary Linguistics* 43(1). 129–152.

Schwartz, G. 2007b. Vowel quality and its holistic implications for phonology. (Ms., Adam Mickiewicz University.)

Schwartz, G. 2008. "Anti-sonority: Phonotactics in terms of an onset element". Paper presented at the 5th Old World Conference in Phonology, Toulouse.

Schwartz, G. In press. *Phonology for the listener and language learner*. Poznań: Wydawnictwo Uniwersytetu im. Adama Mickiewicza.

Stampe, D. 1973. A dissertation on Natural Phonology. (Unpublished PhD dissertation, University of Chicago. Republished with annotations 1979: New York: Garland.)

Stevens, K.N. and S.E. Blumstein. 1981. "The search for invariant acoustic correlates of phonetic features". In: Eimas, P. and J. Miller (eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Lawrence Erlbaum. 1–38.

Syrdal, A. and H. Gopal. 1986. "A perceptual model of vowel recognition based on the auditory representation of American English vowels". Journal of the Acoustical Society of America 79(4). 1086–1100.

Traunmüller, H. 1994. "Conventional, biological, and environmental factors in speech communication: A modulation theory". *Phonetica* 51. 170–183.

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
al. Niepodległości 4
61-874 Poznań
Poland
geoff@ifa.amu.edu.pl