The Phonetic Grounding of Phonology

JANET B. PIERREHUMBERT

Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60208-4090, USA

ABSTRACT

Phonological theory makes claims about the entities of phonological representation (such as features, phonemes, or syllables); about the constraints or rules which describe the well-formed combinations of these entities; and about the way in which constraints or rules interact to determine the space of well-formed abstract representations. Work on the phonetic grounding of phonology has eroded the distinction between entities and constraints. In several current frameworks, entities are viewed as emergent from the kind of articulatory and perceptual factors which also provide grounding for rules and constraints. This paper argues for a deeper penetration of phonetic substance into formal theory by suggesting that phonetic substance shapes constraint interaction. This means that the general form of the phonological grammar reflects the actual nature of speech. The paper first reviews the phonetic properties of vowels and consonants. The ability of vowels to carry tones and prosodic prominence, their organization into a metric space, and their behavior with respect to spreading and assimilation, all reflect their phonetic nature. Such phenomena lie at the heart of framework revolutions over the last decades. The paper then takes up the issue of cumulative versus noncumulative constraint interactions. In a framework with noncumulative interactions, such as Chomsky and Halle (1968) and Optimality Theory, some constraints logically override others. Cumulative effects appear in aspects of phonology involving overall similarity and sequential probability. Noncumulative effects frequently appear in areas of phonology involving serial encoding, where alternative analyses presumably compete in the brain through mutual inhibition. The fact that both cumulative and noncumulative interactions occur suggests an "object-oriented" approach to phonology, in which the way a constraint interacts depends on what kind of a constraint it is.

---

1. Introduction

A traditional theory of phonology has three components: an inventory of phonological entities, a set of rules or constraints for describing the relations of these entities, and a framework for describing how the rules or constraints interact with each other to determine the well-formedness of particular words or phrases. Chomsky and Halle (1968) provides a particularly clear example of these three components. The primitive entities are distinctive features laid out in a matrix, with phonemes described as columns of distinctive features in the matrix. Marking conventions and transformational rules describe the way that that features in a sequence relate to each other. The phonological cycle, with extrinsic ordering of the rules within the cycle, organizes the interaction of transformational rules with each other.

By the early 1960's, Jakobson, Fant, and Halle (1951), Fant (1960), and other works had already made a persuasive case that feature and phoneme inventories are grounded in human capabilities for speech production and speech perception. Chomsky and Halle (1968) build on this understanding by attempting a universal inventory of phonetically grounded distinctive features. However, the framework of rule interaction is not in any way grounded in facts about the physical world. In the strongly modular technical treatment developed in this work, the phonological grammar carries out formal manipulations on the uninterpreted symbols of the theory. It computes a surface phonological representation which provides an interface to a separate, and non-linguistic, phonetic implementation module. This concept of the relationship between the phonetic substance and the phonological grammar is reiterated in Chomsky and Lasnik (1995) as a defining feature of the Minimalist Program.

In work subsequent to Chomsky and Halle (1968), the distinction between entities and rules or constraints began to erode. Natural Phonology (Stampe 1973, Hooper 1976) proposed that phonological processes are grounded in natural phonetic processes, such as coarticulatory effects. The use of intrinsically ordered hierarchies also provides many additional examples of phonetic substance being imported into rules and constraints. Chomsky and Halle (1968) already proposed a distinct rule set, the stress rules, which had the power to manipulate integer-valued features incrementally throughout the phonological cycle, and the use of sonority scales in treatments of syllable structure provides another important case. Scales take the form of intrinsically ranked constraint hierarchies in Optimality Theory, and Kirchner (1997) proposes a particularly radical use of scales, with continuously valued phonetic functions (such as degree of articulatory laziness) figuring directly in the OT grammar. In fact , in West Coast Optimality Theory (Kirchner 1997; Flemming, 1995), phonemes are not viewed as entities in the theory, but instead emerge epiphenomenally from the interaction of phonetically grounded marking constraints. However, in work in the generative tradition a uniform framework is proposed for handling the interaction of rules or constraints with each other. For example, in Declarative Phonology (see Scobbie 1993 and Bird 1995), the constraints are all held to be surface true and are combined using first-order logic. Any two constraints will interact in this way, no matter what aspects of the sound pattern they may describe. In classic Optimality Theory (as first laid out in Prince and Smolensky, 1993), the constraints are ordered in a prioritized list and a higher-ranked constraint overrules a lower-ranked constraint absolutely. Again, this is the case no matter what the phonetic substance of the constraints. As Kirchner (1997) explains, though scalar constraints are readily accommodated in the OT architecture, it does not support the gradient or cumulative interactions of scalar constraints with each other. Thus, the formal apparatus provided by the framework for combining constraints is viewed as a claim about the specific nature of the human language faculty. Possible phonetic grounding of the way that constraints interact has not received the kind of attention that phonetic grounding of features, segments, and constraints has received.

In this paper, I will suggest that the distinction between constraints and frameworks is no more clear-cut than the distinction between constraints and phonemes or features. Widely discussed theories of the "phonetic" grounding of phonemes and features now involve assumptions about perception, memory, and motor control as well as about articulation and acoustics. Such factors are as likely to shape the phonological framework as to shape the entities of phonology; there is no substance-free part of phonology. Thinking about the phonetic grounding of features, phonemes, and constraints has already proved productive; careful analysis of the consequences of substance for constraint interaction is also likely to prove productive, eventually leading us to an explicit and predictive understanding of phonological systems which a prioristic assumptions about grammar and competence cannot provide. In particular, it appears likely that interactions between constraints depend on their actual substance, with some types of factors showing cumulative interactions and others showing overriding (or "winner-takes- all" interactions). The hybrid architecture towards which these observations point is reminiscent of object-oriented programming languages which have been developed in the world of graphics and web programming.

In advancing this viewpoint, I do not mean to deny that implicit knowledge of sound structure involves multiple levels of representation. As further discussed in Pierrehumbert (1999), different levels of implicit knowledge involve generalizations over different types of events and entities; this is the case both in the synchronic description of the adult grammar and in language acquisition. The kind of detailed phonetic knowledge which includes probability distributions of phonetic parameters represents generalizations over speech events. Infants appear to have made substantial progress on this level of encoding even before they have a lexicon, and they indeed depend on this level of encoding to develop a lexicon. The implicit knowledge of phonotactic constraints which allows adult speakers to make well-formedness judgments of novel words or to adapt loan words to native patterns of stress and syllabification arises by generalizing over phonetically encoded words in the lexicon. Principles of morphophonological alternation are even more abstract, arising as generalizations over relations amongst words. Such knowledge can only emerge when a speaker has a good-sized lexicon and has sufficiently structured lexical entries to reveal the relationships amongst entries. Thus, we observe a ladder of abstraction from phonetics to morphophonology. An important contribution of generative theory has been to draw attention to the interest and importance of characterizing the most abstract of these levels.

Urging the importance of substance at all levels of phonology does not entail a rejection of formal phonology. Phonological theory has been handicapped, I would suggest, by the assumption that formalism and substance are in some way inconsistent or opposed. This assumption appears to arise from a narrow view of formalism as equatable with logic or formal grammar, an assumption addressed further in Pierrehumbert et al. (2000). However, the more formalized fragments of phonetic theory, such as the acoustic theory of speech production developed in Fant (1960), provide a standard of formal explicitness and coherence which generative phonology has only sporadically achieved. Phonetic theory contrasts with phonological theory not in the extent to which it is formal, but rather in the extent to which it uses the formalism of continuous mathematics as opposed to relying on logic exclusively. Thus, I will not draw any distinction between formal and substantive theories of sound structure; the crucial distinction instead is the one between syntactic theories (theories which use logic or grammar to manipulate uninterpreted categories) and theories in which logic or grammar is supplemented with continuous mathematics, or in which the behavior of categories depends on their interpretation.

Continuous mathematics can be built outward from foundations of logic, and it is also possible to view logic as pared back from continuous mathematics. In particular, the theory of integers can be constructed on a foundation of set theory, and provides in turn a foundation for the theory of real and complex numbers. Logic in the form of Boolean algebra provides a formal foundation for the theory of statistics. Both intuitive observation, and the successes to date of generative phonology, provide arguments that the more abstract areas of phonology are more categorical than phonetics is. To a phonetician whose work requires Laplace transforms or signal detection theory, treatments of phonology using the mathematically restricted resources of logic alone surely have a surprising degree of success, in terms of the coverage of the phenomena they address. This suggests that the cognitive system makes a simplified or impoverished use of the wealth and complexity of the phonetic domain – a very interesting fact about language which we have by no means gotten to the bottom of. The large scale analyses made possible by recent technology, however, show that even core areas of phonology as not as categorical as more intuitive methods of data collection once suggested. Stochastic variation – and implicit knowledge of stochastic variation – has been demonstrated both for phonotactics (see e.g. Hay et al. in press, Treiman et al., 2000) and for some morphophonological alternations. Word-specific allophonic details have also been demonstrated, most relating to word frequency, degree of morphological decomposability, and paradigm uniformity. (See e.g. Phillips 1984, Bybee 2000, Hay 2000, Steriade 2000, and papers reviewed in Pierrehumbert, in press) Thus, although the most abstract areas of phonology appear to be more categorical than phonetics is, that does not mean that they are absolutely categorical. We need to formal tools for treating the phonological grammar as approximately categorical. This paper is intended as a contribution in this direction.

2. Vowels: A Case Study

As a case study in phonetic and cognitive grounding, let us take a fresh look at vowels. Vowels are at once one of the best understood and most problematic areas of phonetic and phonological theory. Aspects of their phonetic character to which I will allude are covered in textbooks such as Catford (1977), Clark and Yallop (1990), and Johnson (1997). What is less well appreciated is the extent to which the problems they present for phonological theory relate directly to their phonetic character. The more closely one considers vowels, the more difficult it is to separate entities, relations, and interactions; this provides support for our suggestion that these are not in fact separable.

2.1 Phonetic grounding of the vowel space

Vowels, like all other phonological entities, can function as elements of lexical contrast only insofar as they are both possible to produce and possible to perceive. That is, they have a dual articulatory-acoustic character. The vowel space whose various regions provide the vocalic phonemes of different languages arises from many particular facts about articulation, acoustics, and cognitive representation. The phonological potentialities of vowels would be quite different if any of these facts were otherwise.

Articulatorily, vowels are distinguished from other speech sounds by a relatively open configuration of the vocal tract, and by the typical presence of vocal fold oscillation. These characteristics are related to each other, because vocal fold oscillation is best sustained when the airflow from the lungs is not significantly impeded by obstructions in the vocal tract. The acoustic outcome is a class of sounds with a clear fundamental frequency (F0) and a clear resonant structure. The existence of a recoverable F0 during vocalic regions of speech arises both from the vibration of the vocal folds and from the fact that this vibration is the only significant sound source – turbulent noise from other sources does not disguise the periodicity of the signal. Titze (1994) reviews the relationship between laryngeal control parameters and voice source outcomes. Catford (1977) is a useful reference for the role of aerodynamics, specifically the conditions for laminar and turbulent flow, in linguistic phonetics.

The F0 of vowels and the vocal tract resonances are to a large extent independently controllable. This is a result of the relative lack of coupling between the vocal tract and the voice source (e.g. the pressure variation at the glottis which excites the vocal tract resonances). See Fant (1960) for the original exposition of the separation between source and filter for vowels, and Fant et al. (1985) and Lin (1987) for exploration of the moderate amount of coupling which occurs. Just as important, the F0 of vowels is a perceptually robust property which can be recovered even if many frequency components (including even the fundamental frequency component itself) are missing. This robustness results from the way that F0 information is transmitted by the auditory nerve and decoded by the auditory cortex. The consequence is that F0 is perceptible under varying vocal tract shapes and circumstances of speech transmission. Because F0 is both controllable and recoverable independently from the vocalic resonances, it functions phonologically as a separate channel of information. The F0 channel can be used for tone and intonation systems. The vocalic channel is available for lexical distinctions even if F0 is being used independently for phrasal intonation.

In this regard, the human voice contrasts with many other sound sources, including musical instruments whose sounds are produced and perceived by humans. For string instruments, the characteristics of the resonator are fixed and only the F0 is controllable. For flutes, woodwinds, and brass instruments, the sound source and the resonances are tightly coupled and separate control of F0 and resonances is not as a result possible. If the human voice resembled musical instruments, we could not have the simultaneous phonological distinctions in vowel quality and tone or intonation which pervade human languages.

The resonances of vowels – that is, the formants – organize the vowels into the familiar vowel space. This space is two-and-a-half dimensional, in the sense that two dimensions (corresponding to the first and second formants) does most of the work of capturing distinctions amongst vowels and showing their relatedness; however, a third dimension, related to the existence of the third formant, is relevant to complex vowel systems and a more complete perceptual characterization, although it is not fully independent of the other two dimensions. How does this two-and-a- half dimensional space arise? We must consider the interaction of many different factors.

First, consider the overall length and configuration of the vocal tract. For an adult male, the length is approximately 16.5 cm from the larynx to the lips. For a neutral vowel, or schwa, the vocal tract may be approximated as a uniform half-open tube. The result of these two factors is that resonances corresponding to lengthwise excitation of the tube arise every 1000 Hz beginning at 500 Hz. (or at 500 Hz, 1500 Hz, 2500 Hz, etc.). For vowels produced with one or more constrictions, the formants are displaced; however, since the vocal tract is still in "one-tract mode" (cf. Mrayati, Carré, and Guérin 1988), the overall density of resonances with respect to frequency is still approximately one per 1000 Hz.

The fact that these resonances are perceptible to the listener and establish the form of the vowel space is due to the fact that they are higher than the typical rate of vibration of the vocal folds. The spectral peak which reflects each resonance is indicated by the relative amplitude of harmonics in the region of the spectral peak. For example, if the vocal folds vibrate at 100 Hz, then a resonance at 500 Hz is reflected in the relative amplitude of the 10 harmonics in the range 100 Hz to 1000 Hz. The spectral envelope may be visualized as a comb whose teeth have been trimmed into the outline of the spectrum – with the outline itself resembling a series of mountain peaks. The location of the mountain peaks is recoverable if each peak is delineated by the lengths of a sufficient number of teeth. This is typically the case in speech. In contrast, if you attempt to trim a comb to a very complex outline (in the sense of a outline which has more peaks than the number of teeth in the comb can individuate), then the outline you intended will not be evident to anyone else who may view the comb. An unusual situation in which something like this occurs is soprano singing. An operatic soprano may easily produce notes whose frequency is higher than that of the first formant for a vowel in the lyrics. In this case, the vowel quality generally becomes unintelligible. The unintelligibility of soprano vowels is not a problem in opera because the main purpose is aesthetic and members of the audience can in any case familiarize themselves with the lyrics through the libretto. If such a situation were the norm – either because the human vocal tract was much shorter, or because the rate of vibration of the vocal folds was much higher, than is usually the case – the resonances of the vocal tract would not provide a communicatively functional channel of information. Only the information of the F0 itself would still be available for the phonology.

There are actually an infinite number of lengthwise modes of the vocal tract, not just the two-and-a-half that structure the vowel space. Why can the upper resonances be ignored in a linguistic description? One factor is the spectrum of the voice source. The effective spectrum of the voice source, after radiation from the mouth is taken into account, rolls off at approximately 6 dB per octave (more or less, depending on voice quality and individual laryngeal physiology). This means that the spectral energy of the source at 4000 Hz is 18 dB down from the energy at 500 Hz. Now, human speech is normally heard against the background noise of the ambient environment, and a signal-to-noise ratio of 18 dB represents highly favorable listening conditions. The upper regions of the vowel spectrum normally fall through the noise floor, and even the perceptibility of the third formant is compromised by typical background noise such as rushing water, wind, and rustling leaves. Thus the relationship of the voice source spectrum to typical background noise explains the unimportance of the higher vowel formants in the linguistic description. Through evolution, the relative importance of the first two or three formants has even been exaggerated in perception. Humans are most sensitive to the frequencies in exactly the range in which these formants fall (see Johnson, 1997 for graphs and references).

The situation would be different if the human voice source had different spectral properties. Imagine that the voice source resembled a pulse train, like the sound of a woodpecker hammering. Such a source could arise in space aliens who had hard structures in the oscillating mechanism which provided the sound source. In this case, the voice spectrum would be essentially flat and higher formants would be far more pronounced in the signal.

So far, I have discussed the resonances which arise from lengthwise excitation of the vocal tract. What about cross-wise excitation? The cross-modes of the vocal tract are not important in the acoustics of vowels because the human vocal tract is much longer than it is across. The cross modes thus have high relatively frequencies and can be disregarded for the same reason as the upper formants: they are not effectively excited by the human voice source; they are easily masked by background noise; and they fall in a region in which the human auditory system is not especially sensitive. In this respect, human vocal tract acoustics contrasts with the acoustics of bells. Because the lengthwise and cross-wise dimensions of bells are more similar to each other – all relate to resonances which may be both effectively excited and perceived – it is not possible to develop the kind of simple acoustic model we have for vowels, and much more complicated models are necessary. A space alien with a controllable bell-shaped resonator would have a rather different phonological potential.

It is because the cross modes may be neglected that the vocal tract may be approximated as one-dimensional, a point developed fully in Fant (1960). That is, to first approximation, the acoustics may be modeled in terms of plane wave propagation along a tube. In short, the problem of determining the resonances for vowels is approximately equivalent to the problem of finding the eigenvalues of a variably loaded string.

An extremely important point – which often proves challenging for the beginning student of phonetics – is that the entire set of vocalic resonances is determined by the entire configuration of the tube. For a variably loaded string, the load at all points affects any given resonance; changing the load at any one point potentially affects all the resonances. This fact takes somewhat different forms in different models. In the highly schematic models of Stevens (1972, 1989), vowel articulations are approximated by a small number of tubes of differing cross-section. The resonances associated with each cavity are computed separately, and the response of the whole system is determined by correcting for the coupling between the cavities as well as for the radiation at the mouth. The region of validity of this approach is centered on extreme vowels (e.g. /i/, /u/, /a/), for which the minimal cross-sectional area is a great deal smaller than the cross-section elsewhere. Even for these extreme vowels, however, the (adjusted) resonances associated with the back of the vocal tract end up being radiated and the entirety of the vocal tract is relevant to the outcome; this is in contrast to obstruent consonants, in which the constriction is so extreme that the resonances of the cavity in back of it have very little impact on the acoustic outcome. The DRM model developed in Mrayati et al. (1988) has a region of validity centered on the so-called neutral vowel, representing a vowel of uniform cross-sectional area in the general region of schwa. Vowels which differ from the neutral vowel by a single moderate constriction may have their resonances estimated as perturbations from the resonances of a half-open tube . The perturbation diagrams in this paper bring out plainly the fact that resonances arise as properties of the whole system, and the paper contains a useful discussion of "one tract mode" which delineates the physical regime in which the spectral peaks in the signal relate to resonances determined by the entire length of the vocal tract. A regime distinction and the basic physical properties of the vocalic regime also embedded in the more comprehensive and advanced articulatory models, such as the model whose behavior is analyzed in Maeda (1979), Maeda (1991) and Boë et al. (1992).

With this background, we now come back to the fact that only a few of the actual resonances of the vocal tract are perceptually available and linguistically relevant. This means that many different tubes in principle yield equivalent acoustic outcomes; a landmark paper by Atal et al. (1978) uses the mathematical concept of a fiber to characterize these equivalences. The extent and implications of these equivalences remain an area of active research, since the invertability or noninvertability of the articulatory to acoustic mapping has profound consequences for the cognitive representations. Insofar as the vocal tract is appropriately modeled as an arbitrary tube of variable cross-section, extensive fibers will result, meaning that the tube shape is highly underdetermined by the formant patterns. If the tube shape is highly indeterminate, then it may be argued that the fundamental categories are acoustic goals, and that the articulatory characterization is variable and unstable in comparison to the acoustic one.

Boë et al. (1992) argue that the inversion problem is considerably less severe than arbitrary n-tube models predict. The actual properties of the articulatory system substantially reduce the set of configurations which are available to generate any given formant pattern. They present calculations over a large database of French vowels, using the model of Maeda (1979). They conclude that a characterization of vowels in terms of constrictions (e.g. locations of minimal cross-sectional area) shows a fairly good relationship to the vowel categories as defined by the formant structure.

However, their calculations still do not provide a one-to-one, or fully invertible, mapping between formant patterns and articulatory gestures. Indeterminacy arises at two levels. The first is the mapping from the formant patterns to the location of the constriction. As noted in Maeda (1991), the vowel /o/ has two different constriction locations in X-ray studies of production. Boë et al extend this observation, with all three of the vowels (/u/, /o/, and /œ/) showing a bimodal distribution of estimated locations for the tongue body constriction. Other vowels show a range of estimated constriction locations of up to 2 cm, corresponding to about one eighth of the length of the vocal tract. Some of this variation is systematic. Most of the vowels show a significant tradeoff between the degree of the primary constriction and the lip opening. A related study of English vowels by Gay et al. (1992) reports a tradeoff between location and degree of the constriction for all six vowels analyzed.

A second level of indeterminacy arises in the mapping from the computed constrictions to the underlying articulatory gestures. The Maeda (1979) model used by Boë et al. has five articulatory control parameters, representing a careful and thorough distillation of the effective dimensions of articulatory control. This is still two or three more parameters than the two-to-three dimensional formant space, an obvious hint of the potential for compensatory control strategies. As discussed by Boë et al, there are well known tradeoffs between the jaw and the tongue in forming constrictions. This was already established by the bite-block experiments of Lindblom et al. (1979), and the reader can verify this fact by producing the vowel /a/ with a relaxed open jaw and with a clenched jaw. With sufficient tongue root retraction, /a/ can be produced despite the closed jaw position. This paper in fact has been widely interpreted as providing an argument for viewing vowels as acoustic goals, a line of interpretation which has been picked up by Lindblom and colleagues in a number of papers on vowel inventories as sets of acoustic contrasts. The calculations summarized by Boë et al. also show a range of three centimeters in overall vocal tract length for different vowels; this length is treated as a unitary variable without discussion of how lip protrusion and larynx positioning interact to determine it.

In short, vowels are produced by complex articulatory strategies which recruit the muscles of the jaw, tongue, lips and larynx in order to determine the length and shape the vocal tract as a whole. A description at the level of constriction degree and location shows less pervasive compensatory effects than the articulatory description; nonetheless, vowels differ in the extent to which they may be associated with a well-defined constriction. For some vowels, more than one place of articulation is possible while for others, trading effects can be identified.

I am not aware of an inventory of articulatory and acoustic variability for obstruent consonants which has the range and systematicity of Boë et al. (1992). Nonetheless, some observations appear to be possible. The characterization of stops and fricatives is dominated by the place of the constriction which is responsible for generating noise. This constriction is so tight that the resonances of the cavity behind it may be disregarded in computing the acoustic outcome (see Fant, 1960); this dramatically reduces the potential for compensatory interactions between different regions of the vocal tract. Tight constraints determined by the aerodynamics in many cases leave little latitude for variability in the constriction associated with the consonant. A complete closure is required for the pressure build-up and release of plosives. Fricatives require narrow closures to produce turbulent noise, and in strident fricatives, this closure is rather accurately positioned to direct a jet of air against the teeth. The suggestion that the place of articulation is more tightly defined for obstruent consonants than for vowels also receives support from the number of different locations which appear to be specifically targeted by different languages, or to contrast within languages. For example, for coronal stops, dental, alveolar, and retroflex points of articulation must be distinguished. As discussed in Vilain et al. (1998), the place of articulation of a consonant is maintained in articulation even when coarticulation with different vowels reorganizes the rest of the vocal tract shape. This suggests that the exact place of articulation of the consonant is functionally more important than that of the vowel.

The aspects of vowel phonetics just presented have obvious implications for the possible vowel inventories of natural languages (e.g. for the sets of "things" which may be manipulated in the phonology). It structures the vowel space via the first three formants, and predicts that articulatory configurations which do not produce distinct acoustic outcomes in this space cannot be linguistically distinct. Because the vocal tract is in one-tract mode for vowels, with the resonances predicted as a group from the overall configuration of the vocal tract, it predicts that these dimensions are not fully orthogonal, e.g. not independently controllable. It explains why certain vowels (such as /i/ and /I/) are closely related phonologically, while other pairs of vowels are not. Equally, the aspects of vowel phonetics just discussed shape the phonological rules and constraints. I have already mentioned the ability of vowels to support phonologically contrastive tones, which derives from their affinity for vocal fold vibration, from the relative lack of coupling of the voice source to the vocal tract resonances, and from the robust perceptibility of F0. The separability of F0 from vocalic quality also provides a foundation for rules or constraints which align map tonal melodies across varying material, or which move tones from one location to another. Such effects provided the seed of autosegmental phonology, and thus lead us to the past and outstanding problems in the role of vocalic behavior in general in the phonological grammar. This is the topic of the next subsection.

2.2 Vowels and phonological frameworks

The intrinsic properties of vowels shape vowel inventories and shape constraints involving vowels. In addition, over the history of generative phonology typical phonological behaviors of vowels which can be traced to their intrinsic properties have repeatedly motivated framework shifts. But vowels continue to present problems for phonological frameworks, and many of these problems relate directly to their phonetic characteristics.

Consider the rise of nonlinear (autosegmental and metrical) phonology during the late 1970s and early 80's. (See Goldsmith 1990 for a review) This framework originated in response to limitations of the Chomsky and Halle (1968) framework in the area of tone and stress. Tone presented problems for the Chomsky-Halle framework because it acts more like an independent channel of information than like a distinctive feature of phonemes. On the one hand, it can only cooccur with vowels (or occasionally, with vowel like consonants); on the hand, the span of a tone can be either more or less than a single vowel. In addition, constraints or rules targeting tones tend to be quite insensitive to the particular vowels on which the tones are expressed, and in intonation languages such as English the tonal selection is entirely separate from the lexicon. All of these properties of tone systems directly reflect that fact that vowels have clear periodicity and that this periodicity is independently controllable from the shape of the vocal tract, as discussed above. The study of stress systems in metrical phonology also brings together several properties of vowels. One is their ability to bear distinctive tones, an important correlate of phrasal prosody in intonation languages. Equally important is their gradable character. Degrees of stress are related to the degree to which vowel targets are expressed, including their duration, their voice quality, and their formant extrema. Although metrical effects on consonants have now been documented experimentally, (e.g. Pierrehumbert and Talkin, 1991), it is difficult to imagine that these are sufficiently large or regular to provide the sufficient evidence in and of themselves for the language learner to acquire the complex metrical structures which exist in languages. In short, it appears unlikely that metrical structures would have evolved in languages in their present form without the gradable properties of vowels.

Tonal features are those which behave most independently from other features. However other features exhibit many of the same behaviors to some extent. The theory of feature geometry undertakes to extend the approach of autosegmental phonology by providing a theory of how segmental features are organized into tiers which support constraints on association and spreading analogous to those first established for tone features. A hierarchical structuring of the features, with constraints on its use, is proposed to capture the fact that the segmental features exhibit systematic groupings and dependencies in their autosegmental behavior. The tree is organized according to articulatory properties.

Since the theory of feature geometry was first proposed, it has undergone several different and in some ways inconsistent lines of development. One line of development, motivated mainly by work on the phonological patterning of consonants, treats the organization of the nodes in terms of active articulators; see Sagey (1986) and Halle (1988). The most obvious success of this line of development is the treatment of phonologized coarticulation of consonants, for example the homorganic constraint for nasal-obstruent clusters which is found in many languages. Description of this widespread and natural process required a complex and unconstrained use of alpha-notation in the SPE framework. In feature geometry, the process is described through spreading of the Place node, which may dominate any of the various place features. Spreading of the Place node permits us to capture the fact that the place of articulation of the obstruent, whatever it may be, is shared by the nasal; the spreading arises via the timing and duration of the articulatory gesture, along the lines of Articulatory Phonology (Browman and Goldstein, 1989). An assumption which figures critically in the success of the analysis is the assumption that minor features which subclassify the major place features are located under the major features in the featural tree. For example, the features [anterior] and [distributed], which subdivide the coronal articulations, are located under the feature [coronal]. Similarly, the distinction between labio-dentals and bilabials is located under [labial] because this distinction is only relevant when the lips are the active articulator. With this organization of minor features under major ones, the minor features are dragged along when either the specific major place feature, or the place node in general, is spread. In general, this treatment of subclassification connection advances a connection between markedness and assimilation/dissimilation. More complex or marked variants of typologically unmarked sounds are described through sproating of tree structure. Neutralization of a minor featural contrast to the unmarked member of a class is treated by pruning just those features which would be dragged along in an assimilation.

Coarticulatory effects between vowels and consonants give rise to the hope of extending this framework to cover vowels. Indeed, the whole enterprise of phonology, which is to describe the entire system of sound contrasts, plainly requires a comprehensive system of representation. In Clements and Hume (1995), feature geometry is extended to cover vowels as well as consonants. Since this proposal is the most successful of its kind, I will consider it here in some detail. I will make note both of some successes of the model, exemplifying the suggestion in the Introduction that phonology is approximately categorical, and some limitations which suggest that the treatment is overly syntactic compared to the actual state of affairs. These limitations relate specifically to the fact that vowels have a less well defined place of articulation than consonants.

In the Clements and Hume feature geometry, vowels as well as consonants have a place and an aperture degree. In a departure from the work just mentioned, place is interpreted as the location of the minimal cross-sectional area in the vocal tract, rather than in terms of a primary active articulator. Degree of opening is described using an intrinsic hierarchy of degree features which is formally equivalent to an integer scale. Empirical findings limit this scale to four values, representing high, upper-mid, lower-mid, and low vowels respectively. All sounds, both consonants and vowels, are held to have a C-node and a V-node. The V-node is under the C-node in the tree, and in the case of consonants with a secondary articulation such as labialization or pharyngealization, it encodes the (basically vocalic) character of this secondary articulation. A rather complicated phonological argument is presented that ordinary vowels have a C-node dominating the V-node, and I refer the reader to Clements and Hume rather than summarize this argument here.

This treatment responds to some of the aspects of vowel phonetics discussed above. The use of an height scale permits a treatment of proximity and gradience in one dimension, roughly corresponding to F1. This scale can be exploited to describe a collapse of height distinctions in a prosodically weak position, a capability exploited in Wetzels (1995). In addition, the interpretation of place in terms of the location of the constriction (or minimum cross-sectional area) acknowledges the fact that the concept of a dominant active articulator is not well motivated for vowels. Insofar as the location and size of a cross-sectional minimum dominates the acoustic outcome, it provides a more well-behaved parameter along which to organize the cognitive/linguistic model. This choice of parameter, which shifts us from articulatory control per se to functional goals of articulation, already represents a move towards a acoustic characterization. Thus Clements and Hume may be understood to claim that vocal tract area functions provide a middle ground between motor control and acoustics which is well-behaved for both consonants and vowels.

There are a few phonemes whose nature and behavior appears to be extremely problematic for the Clements/Hume proposal. One is the low front vowel /æ/, which differs from its lower-mid counterpart in having a higher F1 and a lower F2, thereby displaying a well-defined location in the acoustic vowel space. The Clements/Hume proposal treats front vowels as coronal (e.g. as involving a constriction in the region of the tongue blade and palate). This is critical to their treatment of interactions of vowels with consonants through familiar spreading and cooccurrence rules. However, /æ/ is too open to meet the definition of coronal in terms of the location of a minimum in the cross-sectional area. The Mrayati et al. model suggests that the minimal cross-sectional area for this vowel should be immediately above the larynx. This does not represent an active gesture of the articulators, or course, but rather the fact that the cross-section of the vocal tract at this point is intrinsically smaller than that arising from the tongue and jaw positioning for the /æ/. The results of Gay et al. confirm that the vowel can be produced without any active constriction whatsoever. Thus, the vowel obviously does not meet the definition of [coronal], and out of eight phenomena Clements and Hume discuss in support of the suggestion that front vowels are [coronal], all in fact involve only vowels which are mid to high. /æ/ nonetheless can function with the front vowels in phonological processes involving vowels. For example, it acts as a trigger of front harmony in Finnish and alternates with higher front vowels in English. The Clements/Hume proposal appears to offer no way to handle such behavior, since the location of the cross-sectional minimum for /æ/ fails to group it other phonemes. Insofar as any grouping is suggested by the cross-sectional minimum, it would be with the pharyngeals.

The English rhotic, which can function either as a consonant or as a vowel, raises similar problems. Its characteristically low third formant can be produced by any of various combinations of coronal constriction, pharyngeal constriction, and lip rounding. In general, as discussed in McCarthy (1988) and Clements and Hume (1995), OCP-Place effects (e.g. phonological constraints disfavoring segments at the same point of articulation in close proximity) are taken to be diagnostic of nodes in the feature geometry. For example, the fact that English syllables cannot begin in /pw/, /bw/, or /fw/ provides evidence that a /w/ shares the node [labial] with the labial obstruents. Similarly, a shared [coronal] property blocks syllables beginning in /tl/ or /dl/. /r/, however occurs both with labials (as in "brick") and coronals (as in "tree"); in fact to my knowledge it does not participate in any OCP effects at all. The situation appears to be that due to its variable and multiple constrictions, /r/ does not present a clear case of any single place of articulation.

The phonetic situation presented by the English rhotic is also exemplified by the three French vowels displaying a bimodal distribution of the location for the minimal cross-sectional area, namely /u/, /o/, and /œ/, as shown in Boë et al. 1992. Recall also that most vowels (unlike consonants) exhibit tradeoffs between the location and the degree of the cross-sectional area.

A more general limitation of the feature geometric treatment of vowels relates to the presupposition that the articulatory attributes which appear as nodes have equal phonological importance for all degrees of constriction. However, as one moves along the sonority scale from obstruents to vowels, the balance of articulatory and acoustic factors in determining phonological contrasts and behaviors changes. Articulatory attributes are particularly pertinent to the phonological behavior of obstruents. As argued by Lindblom (1992) among others, acoustic targets are clearly pertinent to vowel phonology. Along the way from obstruents to vowels, many problematic cases of articulatory-acoustic interaction arise. For example, for nasal consonants, place of articulation is critical in shaping syntagmatic constraints. Nasal consonants tend to nasalize preceding vowels because the lowering of the velum is anticipated. But for nasalized vowels themselves, acoustic considerations are critical in deriving the fact that no language has more nasalized vowels than oral vowels, as shown in Maeda (1993). Arguments presented by Maeda also show that acoustic factors lead to reorganization of the tongue gestures for French vowels which are nasalized, in order to maintain constrastiveness. In short it is not the case the phonemes in general are mainly articulatory, or mainly acoustic in nature. Phonemes are located in well-behaved regions of the articulatory-to-acoustics mapping, and the role of articulatory and acoustic factors in shaping these regions is somewhat different for different classes of phonemes

Discussion in Clements and Humes indirectly tacitly acknowledges some of these difficulties by abandoning the goal of handling subclassification through subbranching under the major articulator nodes, as exemplified above for [anterior] and [distributed]. As they correctly observe, many featural dependencies involved in characterizing subclassification and markedness do not lend themselves to being localized in the tree structure in this way, as they involve acoustics and aerodynamics. These dependencies must be part of the phonological grammar – since different languages permit different elaborations of the featural contrasts, both globally and as a function of prosodic position. Thus the issue remains open of what formal representations are needed to capture these dependencies as they are encoded in the cognitive system, and it appears doubtful that it will be possible to solve this problem with as syntactic an approach as feature geometry.

A further connection between the phonetic substance of vowels and consonants and the very core of the phonological framework can be adduced by considering a curious assymmetry in featural spreading rules. Vowel-vowel assimilation rules, such as vowel harmony rules very readily cross consonants, whereas consonant-consonant place assimilation rules show much less ability to cross vowels. This assymmetry is the primary motivator of the Clements/Hume suggestion that all phonemes have a V-node positioned hierarchically under the C-node. It is also extensively discussed in Gafos (1998, 1999). (I will adopt his suggestion that apparent counterexamples involving total identify of the two consonants are better understood as cases of copying not spreading, and I also assume this interpretation of the reduplicative utterances produced by babies.) The phonetic grounding of this effect is surely the difference I have already made much of, in the extent to which vowels and consonants have a specific place of articulation. As the studies of coarticulation in Vilain et al. (1998) show, different vocalic contexts can reorganize the articulations by shaping the vocal tract away from the constriction, provided only that the obstruent's place of articulation is achieved. Although Vilain et al (1998) do not discuss vowel harmony or V-V assimilation, the extension to this situation is clear. The connection I would like to draw here is that a syntactic treatment of this effect, as proposed by Clements and Hume, depends on a very specific and controversial understanding of the No-Crossing constraint, a core principle of autosegmental phonology which controls the operation of all rules and the well-formedness of all representations. See their paper for discussion of this controversy. This provides an example of how phonetic substance impacts the very nature of the phonological framework.

3. Phonetic Grounding versus Cognitive Grounding

The phonetic grounding of phonology is often conceived of terms of the speech acoustics and the physical characteristics of the articulatory system. Further consideration of the points I have raised about vowels, however, shows that it is difficult to draw a line between physical and cognitive factors. For example, the discussion of why vowels are organized in a two and a half dimension formant space relies both on general facts about physics (such as the resonances of acoustic tubes), specific facts about speech production (such as the nature of vocal fold vibration), low level facts about perception (such as the response characteristics of the basilar membrane), and facts about cognition (such as the fact that perceptual spaces are encoded in general encoded in the brain in terms of maps of physical spaces) (See Barinaga, 1999) The fact that physics grades smoothly into cognition should hardly be a surprise, if one abandons Cartesian dualism. The brain is part of the physical world. Any theory of cognition is describing synoptically the behavior of a complex physical system. Cognition and perception are not really separable, and the perceptual system has evolved to provide an organism with ecologically relevant information. As an example, it is useful for the ear to encode resonances, not only in speech processing but also elsewhere, because resonances provide information about the type of source responsible for a sound. Even for mice and guinea pigs, it is useful for auditory signals to be perceived in terms of physical objects that are responsible for sounds. This usefulness was no doubt a fact in the natural selection that gave rise to the mammalian auditory system in its present form.

Does phonetics describe speech as a physical phenomenon, with phonology describing it as a cognitive phenomenon? This traditional division of labor proves to be an encumbrance in discussing the physical grounding of phonology. There is no particular point on the continuum from the external world to cognitive representations at which it is sensible to say that phonetics stops and phonology begins. In particular, all proposals to date about the phonetic grounding of phonology explicitly or implicitly rely on some cognitive factors.

Steven's quantal theory (Stevens 1972, 1988) provides a case in point. Quantal theory advanced the claim that the point vowels /a/, /i/, /u/ are typologically preferred because they represent regions of the articulatory space in which variability in production has a minimal acoustic impact. The facade of this theory is a set of calculations about the relationship between variability in vocal tract shape and variability in formant structure, calculations which plainly fall in the purview of "phonetics". Critical to the argument, however, is the assumption that human category formation is dominated by considerations of perceptual invariance. As pointed out by Carré et al (1994), the assumption that languages need efficient and well-behaved control strategies leads to different predictions about the preferred regions of the acoustic/articulatory space.

Another example of a phonetic explanations which involves cognitive factors is provided by Steriade's work on positional neutralization for stop consonants (Steriade 1993). Over many languages, stop consonants display more phonological contrasts in positions in which they are released into a vowel than in positions in which they are unreleased. The spectral and temporal characteristics of the stop release contains a great deal of information about the place and voicing of the stop. When it is missing, phonological contrasts in place and voicing tend to be neutralized. Languages such as Urdu which have a full set of stops in all positions often require stop releases whether or not a vowel follows.

This example of phonetic explanation involves many factors. The stop release is acoustically salient in part because the aerodynamic aspects of stop production; air pressure builds up during the stop closure and it is the difference between the oral and atmospheric air pressure which causes the turbulent flow of the release. The perceptibility of the release is enhanced by the quiet stop gap which precedes it; the automatic gain control system of the ear has the result that a sound of a given amplitude which follows silence will appear to be much louder than if it followed a loud signal. Thus, if a recording of speech were simply time-reversed, the stop burst would appear less salient because the original stop gap would now follow instead of preceding it. But it is also necessary to consider cognitive factors. One issue is how large or reliable an objective difference is needed to maintain a category distinction in a language. Unreleased stops are still objectively different from each other; the formant transitions into the stop differ by place of articulation, and the length of a preceding vowel as well as the amount of voicing during the closure differs according to the stop voicing. The mere fact that unreleased stops are less distinct than released ones does not explain why a language would collapse them. All languages have some sounds which are more similar to each other than other sounds, but that does not mean that languages persistently collapse their weakest distinctions until no more distinctions are left. For example, Hindi and Urdu maintain distinctions between retroflexed and nonretroflexed coronal stops, despite their relatively great acoustic similarity.

A major theme of research in nonlinear phonology has been the locality of constraints or rules. Both autosegmental and metrical phonology undertake to define phonological structures in terms of which natural constraints can be naturally stated on contiguous sequences of elements. For example, metrical theory states stress regularities in terms of constraints on the stress of adjacent syllables; this provides one of many arguments for the syllable as a unit of linguistic description. Long distance constraints over noncontiguous or unrelated aspects of a phonological description emerge as unnatural. It is very typical for a language to have a constraint on coda-onset sequences in the middle of disyllabic or multisyllabic words. (Medial clusters in words such as "pantry" and "pamper" are highly restricted.) It is very atypical for a language to have a constraint relating the word initial onset to the word final coda. (The onset and coda clusters of words such as "strict" are not highly correlated, and the correlation is still less for the beginnings and ends of longer words such as "brilliant" and "contradict").

Extremely local constraints such as the nasal homorganic rule are often viewed as grounded in facts about coarticulation. But coarticulation cannot fully explain the pervasiveness of locality conditions within phonology. For example, Pierrehumbert and Beckman (1988) discuss at length the locality conditions on tonal implementation; some of these regularities affect tones which are as much as a second apart in time, but the conditions are still local because the tones are sufficiently proximate in the abstract phonological description. The grounding of locality is thus likely to relate also to higher level cognitive factors. Possible factors include the duration of working iconic memory during speech perception; working buffers for speech motor control; and general principles of object formation in memory. The course of language acquisition may also play a rule, since children begin speaking with small prosodic units (especially, the core CV syllable) which are later differentiated into a more complex repertoire. This emergence of complex units from simple ones may also be a source of locality constraints in the adult system.

Cognitive factors are widely admitted to play a role in the phonetic grounding of phonology. However, such factors are often latent or implicit in the analysis. If included overtly, they are often described heuristically rather than formally, making it difficult to determine their predictions exactly. One practical result of this state of affairs has been more conspicuous progress on the phonetic grounding of phoneme inventories and sequential constraints than on the possible grounding for phonological grammars viewed as abstract systems. Work by Lindblom and colleagues on vowel dispersion theory, and work by Ohala on positional contrastiveness, provide conspicuous examples. A second practical result has been an association of substance-based approaches to phonology with scientific reductionism. Ohala in particular is associated with the view that phonological taxonomy can never be explanatory in itself, and that an explanatory theory of phonology necessarily involves a reduction of phonology to independently motivated theories of biology and physics. The viewpoint I am urging here, in contrast, is not a reductionist one; there is no reason that substance based theories must be reductionist. Instead, I would view languages as abstract systems with multiple levels of representation. These systems represent nature's solution to multiple constraints, including physical, biological, and cognitive constraints. By constructing an inventory of the force of these constraints for each level, we have the hope of a more incisive and predictive theory of what a language can be.

4. Cumulative versus Noncumulative Interactions

One of the most important lines of division amongst phonological frameworks is that between frameworks with cumulative (or "monotonic") interactions and frameworks with noncumulative (or "non-monotonic") interactions. Let me emphasize that I am concerned here with cumulative interactions of different constraints or dimensions of description with each other. Cumulative effects on single scales, such as stress or sonority, are widely precedented and appear to be consistent with any of the major frameworks. Cumulative interactions both impact the framework directly – since the primary job of the framework is to specify the interactions – and they are more difficult to handle with intrinsic scales or hierarchies.

In a cumulative framework, every single rule or constraint is true everywhere, that is, every form respects all constraints. In a noncumulative framework, information does not add together. Some information can override other information, or the combination of multiple factors can result in qualitatively different outcomes. In Chomsky and Halle (1968), the phonological rules are extrinsically ordered in a list, with each having the capability to overwrite the outcome of the rule before. This amounts to a claim that the generic mode of interaction is noncumulative, with cumulative effects only occurring when rules happen not to overlap in their effects. This assumption is inherited in Optimality Theory (Prince and Smolensky, 1993), though it takes a different form. Constraints are rank-ordered on a single list in OT, with higher ranking constraints represented high priority goals whose satisfaction may be achieved through violation of lower ranked constraints. In contrast, Natural Phonology (Stampe, 1973 and Hooper, 1976) and Declarative Phonology (Scobbie, 1993; Coleman, 1998) both advance the claim that phonological knowledge is cumulative.

To achieve a more intuitive understanding of this difference, consider some everyday information systems. An example of a cumulative system would be the Department Chair's handbook – in an ideal university. Such a handbook may have a section specifying procedures to be followed to appoint an outside faculty member with to a permanent position, and another section specifying procedures to be followed in appointing citizens of foreign countries. If the manual is fully coherent, the set of procedures to be followed in appointing a foreign citizen to a permanent post is simply the union of these two sets; no contradictions arise in combining the procedures, and so there is no need for a mechanism to adjudicate the contradictions. An example of a noncumulative system would be the correspondence between the Department Chair and the Dean. Many conflicting assertions may arise in this correspondence. Overarching principles adjudicate these contradictions; a memo from the Dean in many cases takes priority over a memo from the Chair, and the Dean's most recent memos take priority over previous ones. The difference between a cumulative and a noncumulative system is not the same as the difference between a continuous and a discrete system. Among continuous systems, streams flowing into a river provide an intuitive (if imperfect) image of a cumulative system, whereas phenomena such as earthquakes and sexual reproduction provide images of noncumulative interactions – combining factors has a result other than the mere combination.

Although the contrast between cumulative and noncumulative interaction is often presented as a basic mathematical parameter of a framework, it is important to understand that it is more a matter of broad scientific conception. The reason is that even cumulative categorical frameworks rely heavily on disjunction (or its logical equivalent, the material conditional) in order to formulate constraints that are universally true. For example, consider the nasal homorganic rule, as formulated in a hypothetical language in which the regularity is exceptionless. We can't say that any two sounds in this language agree in place of articulation – if we formulated such a constraint, we would be saying that languages have only a single place of articulation, which is patently false. Instead we say that IF a nasal precedes an obstruent THEN it agrees in place of articulation; or equivalently, for any two sounds, either they agree in place, or they do not represent a nasal-obstruent sequence. (See Bird, 1995, for development and application of this point.) Exactly the same logic would permit us to handle any kind of exceptionality. Obviously, any noncumulative set of grammatical rules or constraints can be compiled out into a large set of constraints whose domain of application is restricted by a suitable antecedent in a conditional expression. This larger and more atomic set of constraints is then exceptionless, and so information from multiple constraints can be combined monotonically. The issue, then, is not whether a cumulative framework is possible, but to what extent is it insightful and predictive. In short, are the grammatical constraints in such a framework so minute that they fail to capture scientific generalizations and prove unable to predict productive behavior?

The same question can also be turned in the other direction, with respect to the ability of noncumulative frameworks to handle gradient cumulative interactions. A particularly important class of cases arises when considering gradient cumulative interactions. When two scales interact cumulatively, their interaction creates a mathematical cross-product, leading to a combinatoric explosion of cases which can cause assumptions of categoriality to break down. It is of course always possible for a noncumulative grammatical framework to simulate gradient cumulative interactions by positing an extremely large number of categories and by specifying related outcomes for situations which are de facto similar to each other. However, the discipline provided by some kind of continuous metric space may begin to look like the path of simplicity. The dimensions of such a space may show far more universality and typological interest than the many, many categories of the categorical approximation. The issue is thus whether the numerous categories are independently justified and whether the grammar which results extends properly to novel cases.

Consideration of the phonetic and cognitive grounding of phonology leads one to expect that some kinds of information would combine cumulatively and other information would combine noncumulatively. That is, the actual substance of a rule or constraint should affect how it interacts with other rules or constraints. This is not a property of any framework in generative phonology proposed to date. Because the mode of interaction depends on the actual phonetic substance, it is also inconsistent with any framework in which formal and substantive properties of phonology are segregated in separate modules.

In general, we expect cumulative effects related to probability and to similarity. Let us consider probability first. It is a general law that the probability of complex events can be computed as the joint probability of their simpler components. Imagine that the probability of a pipe freezing on a day in January is 1 in 1000. Imagine also that the probability of a squirrel falling on your head while you are walking home in January is 1 in 1,000,000. If these events are independent – that is, they happen at random without any relation to each other – then the probability that you will experience both events on the same day is the product of the two probabilities, of 1 in 1,000,000,000. Of course, these types of events may not actually be independent. It is possible that they are negatively correlated, because pipes are more apt to freeze on unusually cold days, and on such days, squirrels tend to stay asleep in their nests rather than being active. The actual joint probability would depend on the strength of the negative correlation, but the general concept of a joint probability still pertains.

The area of phonotactics presents an analogous situation. Complex lexical forms are made up of subparts, each of which has its own probability of occurrence in the lexicon. Several studies now demonstrate that the likelihood that a particular combination of phonological components is attested depends on the likelihood of the components themselves, and the existence of (positive or negative) correlations for combining them. For example, Pierrehumbert (1994) showed that the single best predictor of whether a medial cluster (such as /lfr/ in "palfrey") is attested in English is the likelihood of the coda and the likelihood of the onset as independent components. Frisch (1996) reports similar results for onsets and codas as beginnings and ends of complex monosyllabic words. Such probabilities are also reflected in well-formedness judgments of nonsense words (Coleman and Pierrehumbert, 1997; Pierrehumbert et al, 1998; Treiman et al, 2000) and in other experimental paradigms tapping implicit phonological awareness.

Another area in which cumulative effects are to be expected involve the perception of similarity. An extensive literature on categorization shows that the categorization of a percept depends on its overall similarity to experienced examples of cognitive categories. Overall similarity in turn is the cumulative reflection of similarity in relevant attributes or features. Two things are extremely similar if they are well-matched in many attributes and less similar if less well matched or matched in fewer of the relevant attributes; see Tversky (1977) for a more exact mathematical development.

OCP-Place provides a case in point. OCP-Place, or the Obligatory Contour Principle for Place of Articulation, refers to the tendency in many languages to avoid homorganic consonants in close proximity. As observed already in Lightner (1973) and demonstrated more fully in Pierrehumbert (1992) and Frisch (1996), the strength of this tendency is sensitive to the overall similarity of the consonants in question. These papers report statistical modeling of OCP-Place in Arabic, a language with a rich consonant systems and a strong OCP effect in its triconsonantal verbal roots. The effect is stronger for homorganic consonants which are highly similar, such as /s/ and /z/; it is weaker for dissimilar homorganic consonants. Pierrehumbert (1992) proposes an explicit connection between the lexical statistics and the perception of the consonants. It can be shown that OCP-Place in Arabic is strongest between consonants which are adjacent in the root. It is weaker, though still statistically demonstrable, between the first and third consonants of the triconsonantal root. This weakening of the effect with intervening material cannot be handled in a grammatical formalism, such as the autosegmental formalism treatment outlined in McCarthy (1988). It does follow from the assumption that intervening material acts as a masker, interfering with the perception of similarity of the first and third consonants to each other. Even if the consonants are highly similar, this similarity is less salient than if the consonants were adjacent. Therefore violations of the constraint involving nonadjacent consonants are more tolerated.

In contrast to phonotactics, there are other areas of phonology in which multiple effects do not cumulate but instead appear to compete, with one competitor winning absolutely. The area which I would like to consider here is prosodic parsing. Prosodic parsing is the process by which prosodic structures (such as syllables and metrical feet) are assigned to words or phrases. Following Ito (1986), Archangeli (1991) and subsequent work, I view the prosodic structures of individual words as the outcome of a successful parse of those words. If a string is not parsable in a language, then it cannot serve as a word. Prosodic parsing is also involved during speech perception in segmenting the incoming speech stream and supporting lexical access.

In both domains, prosodic parsing exhibits noncumulative interactions of the form "winner takes all". Study of the typology of stress systems (see Hayes 1995) reveals that many stress systems have defaults which do not relate smoothly to their core constraints. For example, a language may have a stress system by which the last heavy syllable of a word (if any) is stressed, otherwise the initial syllable. It is extremely difficult to view initial stress as any sort of extrapolation over the location of heavy syllables near the end. We cannot define a similarity space in which the different provisions of such a stress system add up to give the result. Optimality Theory readily handles such a situation by specifying two unrelated goals (stress the initial syllable, stress a heavy syllable) and allowing one to take absolute priority over the other in all situations in which it is invoked at all. In Declarative Phonology, treatment of the same regularity involves a conspicuous inelegance, because the rule for initial stress must refer to heavy syllables in the antecedent of the conditional.

Similar effects are also found in speech perception. Hay, Pierrehumbert and Beckman (to appear) report experimental results on the perceived well-formedness of nonsense words. For nonsense words which could in principle be parsed in more than one way – either as monomorphemic, affixed, or compound – the perceived well-formedness proved to be a function of the single best parse. It is noteworthy that this effect is demonstrated in the exact same experiment in which the well-formedness of nasal-obstruent clusters is shown to stochastically gradient. This outcome has an analogue in visual perception, where ambiguous stimuli such as a Necker cube are seen in only one way at a time. Similarly, for syntactically ambiguous sentences, listeners typically perceive only a single reading in a given context and may have great difficulty becoming aware of structural alternatives.

A cognitive basis for this class of results may be identified through a comparison to results in cognitive neuroscience on serial encoding. Experimental studies using monkeys have made it possible to develop schematic neural circuits for the perception, memory, and production of ordered events. In an experimental paradigm described in Houk and Wise (1995), and Beiser and Houk (1998), monkeys perceive a sequence of flashing lights. They are taught to remember the sequence and to reproduce it by pushing buttons when they receive a prompt after a short delay. This task requires the monkey to perceive the stimulus, encode it into memory, and reproduce it from memory. The neural circuitry required to accomplish this task has several important properties. First, each serial order in the monkey's repertoire requires the establishment of a separate neural circuit which spans levels of the brain from the cortex through the cerebellum. Even encoding the light pattern before executing it involves the cerebellum as well as the cortex. Second, these circuits interact with each other through mutual inhibition. That is, alternative serial orders compete and the winning competitor is the one which is encoded and reproduced.

Prosodic parsing can be viewed abstractly as a serial encoding problem. Parsing a single stimulus as a trochee (a stressed syllable followed by an unstressed syllable) is analogous to parsing a visual signal as a red light followed by a green light. A linguistic constraint requiring that all words begin with a stressed syllable (as in Finnish) is analogous to a situation in which neural circuitry has been established only for light sequences beginning with a red light. Prosodic systems differ from the stimulus space in the monkey experiments in falling in the auditory rather than the visual domain, and in having more different entities occurring in more different orders. They resemble the stimulus space in the monkey experiments in having a permanent repertoire of temporally organized patterns which can be individually activated in perception and production. Therefore, generic results about serial encoding should pertain to prosodic parsing. The "winner takes all" constraint interactions displayed in prosodic parsing follows from this fact.

5. Conclusions

The goal of generative phonology is to develop an formally explicit answer to the question "what is a possible language". This goal has had the result that generative frameworks reflect the computational zeitgeist of the time in which they were formulated. Chomsky and Halle (1968) reflects the serial architecture of the computers of that time. The constraint oriented frameworks of Declarative Phonology and Optimality Theory reflect the rise of constraint based programming languages such as Prolog. Both serial and constraint based computer languages are limited in the naturalness with which they express the relationship of human cognition to physical reality. This relationship is more directly reflected in the object-oriented programming languages which have recently been developed for user interfaces which are intuitive to naive users, for web-based and other applications. In these programming languages, both low-level and high-level objects are defined and these objects have characteristic behaviors. Different types of objects have different behaviors, related to how they are conceptualized and how they function.

I think that future phonology will resemble object oriented programming. The objects in the theory will be at variable levels of physical concreteness or cognitive abstraction. The way that each object interacts with others will depend on what it is, and will reflect the physical interactions that contribute to its nature as well as the neurophysiology involved in representing that object in the brain.

References

Archangeli, D. (1991). Syllabification and prosodic templates in Yawelmani. *Natural Language and Linguistic Theory, 9*, 231-283.

Atal, B. S., Chang, J. J., Mathews, M. V. & Tukey, J. W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer sorting technique. *Journal of the Acoustical Society of America*, 63, 1535-1555.

Barinaga, M. (1999). The Mapmaking Mind. *Science, 285*, 189-192.

Beiser, D. & Houk, J. C. (1998). Model of cortical-basal ganglionic processing: encoding the serial order of sensory events. *Journal of Neurophysiology, 79,* 3168-3188.

Bird, S. (1995) *Computational phonology: a constraint-based approach*. Cambridge, Cambridge University Press.

Boë, L.J., Perrier, P. & Bailly, G. (1992). The geometric vocal tract variables controlled for vowel production: proposals for constraining acoustic-to-articulatory conversion. *Journal of Phonetics*, *20*, 27-39.

Broe, M. & Pierrehumbert, J., eds. (2000) *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge, UK, Cambridge University Press.

Browman, C. & Goldstein. L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*, 201-251.

Bybee, J. (2000). The phonology of the lexicon: evidence from lexical diffusion. In Barlow, M. and Kemmer, S. (eds.) Usage-Based Models of Language. Stanford, CSLI Publications.

Carré, R. & Mrayati, M. (1994). Vowel transitions, vowel systems, and the Distinctive Region Model. *In Levels in Speech Communication: Relations and Interactions*. Elsevier.

Catford, J. C. (1977) *Fundamental Problems in Phonetics*. Indiana University Press, Bloomington.

Chomsky, N. & Halle, M. (1968) *The Sound Pattern of English*. New York, Harper Row.

Chomsky , N. & Lasnik, H. (1995). The Theory of Principles and Parameters. In Chomsky, N. The Minimalist Program. Cambridge MA, MIT Press, 13-128.

Clark, J.E. & Yallop, C. (1990) *An Introduction to Phonetics and Phonology*. Oxford: Basil Blackwell.

Clements, G.N. (1985). *The geometry of phonological features*. *Phonology Yearbook, 2,* 225-252.

Clements, G.N. & Hume, E. (1995). The Internal Organization of Segments. In *J. Goldsmith (ed.) The Handbook of Phonological Theory. Blackwell Handbooks in Linguistics*. Blackwell Publishers, 245-306.

Coleman, J.S. (1998) *Phonological representations – their names, forms, and powers*. Cambridge, Cambridge University Press.

Coleman, J. & Pierrehumbert, J. (1997). Stochastic Phonological Grammars and Acceptability. 3rd Meeting of the ACL Special Interest Group in Computational Phonology: Proceedings of the Workshop, 12 July 1997. Association for Computational Linguistics, Somerset NJ. 49-56.

Fant, G. (1960) *Acoustic Theory of Speech Production*. Mouton, The Hague.

Fant, G., Lin, Q. & Gobl, C. (1985). Notes on glottal flow interaction. STL-QPSR 2-3/1985. 21-45.

Flemming, E. S. (1995) *Auditory representation in phonetics*. PhD dissertation, UCLA.

Frisch, S. (1996) *Similarity and Frequency in Phonology*. PhD dissertation, Northwestern University.

Gafos, D. (1998). Eliminating Long-Distance Consonantal Spreading. Natural Language and Linguistic Theory, 16, 223-278.

Gafos, D. (1999) The Articulatory Basis of Locality in Phonology. Outstanding Dissertations in Linguistics. New York, NY: Garland Publishers.

Gay, T., Boë, L.-J., & Perrier, P. (1992). Acoustic and Perceptual Effects of Changes in Vocal Tract Constrictions for Vowels. J. Acoust. Soc. Am. 92, 1301-1309.

Goldsmith, J. (1990). Autosegmental and Metrical Phonology. Basil Blackwell, Cambridge MA.

Halle, M. (1988). The Immanent Form of Phonemes. In Hurst, W. (ed.) Giving Birth to Cognitive Science: a Festschrift for George. A. Miller. Cambridge University Press, Cambridge UK. 167-83.

Hay, J. (2000) Causes and Consequences of Word Structure. PhD dissertation, Northwestern University.

Hay, J., Pierrehumbert, J. & Beckman, M. (in press). Speech Perception, Well-Formedness, and Lexical Frequency. Papers in Laboratory Phonology VI. Cambridge, UK. Cambridge University Press.

Hayes, B. (1995) *Metrical Stress Theory: Principles and Case Studies*. University of Chicago Press, Chicago.

Hooper, J. (1976) *Introduction to Natural Generative Phonology*. Academic Press, New York.

Houk, J. C. & Wise, S. P. (1995). Distributed Modular Architectures linking Basal Ganglia, Cerebellum and Cerebral Cortex: their role in planning and controlling action. *Cerebral Cortex, 5:2,* 95-110.

Ito, J. (1986) Syllable theory in prosodic phonology. Garland Publishing, NY.

Jakobson, R., Fant, G. & Halle, M. (1951) *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates*. MIT Press, Cambridge.

Johnson, K. (1997) *Acoustic and Auditory Phonetics*. Blackwell Publishers.

Kirchner, R. (1997). Contrastiveness and Faithfulness. *Phonology, 14:1,* 83-111.

Lin, Q. (1987). Nonlinear interaction in voice production. *STL-QPSR 1/87*, Royal Institute of Technology, Stockholm. 1-12

Lightner, T. (1973). Against Morpheme Structure Rules and Other Things. In Issues in Phonological Theory. Mouton, The Hague, 53-60.

Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In C. A. Ferguson, L. Menn, and C. Stoel-Gammon, (eds.) Phonological Development: Models, Research, Implications. Timonium, MD: York Press, 131-163.

Lindblom, B., Lubker, J. & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics, 7,* 147-161.

Maeda, S. (1979). An articulatory model of the tongue based on a statistical analysis. *Journal of the Acoustical Society of America, 65,* S22.

Maeda, S. (1991). On articulatory and acoustic variabilities. *Journal of Phonetics, 19,* 321-331.

Maeda, S. (1993). Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In M. Huffman and Krakow, R. (eds.) Phonetics and Phonology Vol. 5: Nasals, Nasalization, and the Velum. Academic Press, New York.

McCarthy, J.J. (1988). Feature geometry and dependency: a review. *Phonetica, 43,* 84-108.

Mrayati, M., Carré, R. & Guérin, B., (1988). Distinctive Regions and Modes: a New Theory of Speech Production. *Speech Communication, 7,* 257-286.

Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language, 60,* 320-42.

Pierrehumbert, J. (1992). Dissimilarity in the Arabic Verbal Roots,. Proceedings of the 23rd Meeting of the Northeastern Linguistic Society Graduate Student Association, U. Mass. Amherst.

Pierrehumbert, J. (1994). Syllable Structure and Word Structure. In Keating, P. (ed.) Papers in Laboratory Phonology III, Cambridge Univ. Press, Cambridge, UK. 168-188.

Pierrehumbert, J. (1999). What people know about sounds of language. *Studies in the Linguistic Sciences, 29(2),* 111-120.

Pierrehumbert, J. (in press). Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, J. and Hopper, P. (eds.) Frequency effects and the emergence of linguistic structure, 137-157. Amsterdam: John Benjamins.

Pierrehumbert, J. & Beckman, M. (1988) Japanese Tone Structure. Linguistic Inquiry Monograph Series 15, MIT Press, Cambridge, MA.

Pierrehumbert, J., Beckman, M., & Ladd, D. R. (2000). Conceptual Foundations of Phonology as a Laboratory Science. In Carr, P., Burton-Roberts, N., and Docherty, G. (eds.) *Phonological Knowledge*, 273-303. Oxford University Press, Oxford.

Pierrehumbert, J. & Talkin, D., (1991). Lenition of /h/ and glottal stop. in G. Doherty and D. R. Ladd, (eds.) *Papers in Laboratory Phonology II,* 90-117. Cambridge Univ. Press, Cambridge UK.

Prince, A. & Smolensky, P. (1993). *Optimality Theory*. MS Rutgers and University of Colorado.

Sagey, E. (1986) *The Representation of Features and Relations in Nonlinear Phonology*. MIT PhD dissertation. Published 1991 by Garland Publications, New York.

Scobbie, J. (1993). Constraint violation and conflict from the perspective of Declarative Phonology,. *The Canadian Journal of Linguistics, 38:2*, 155-168.

Schane, S. (1984). The Fundamentals of Particle Phonology. Phonology Yearbook 1, 129-155.

Stampe, D. (1973). *A Dissertation on Natural Phonology*. PhD dissertation, University of Colorado.

Steriade, D. (1993). Closure, Release, and Nasal Contours. In Huffman, M. and Krakow, R. (eds) Phonetics and Phonology 5: Nasals, Nasalization, and the Velum. Academic Press, New York, 401-470.

Steriade, D. (2000). Paradigm Uniformity and the phonetics-phonology interface. In Broe, M. and Pierrehumbert, J. (eds) Papers in Laboratory Phonology V: Acquisition and the Lexicon. Cambridge UK, Cambridge University Press, 313-335.

Stevens, K.N. (1972). The quantal nature of speech: Evidence from articulatory- acoustic data. In E. David and P. Denes (eds.) Human Communication: a unified view. McGraw-Hill, New York. 51-66.

Stevens, K.N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17,* 3- 45.

Titze, I. (1994) *Principles of Voice Production*. Englewood Cliffs NJ., Prentice Hall.

Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (2000). English speaker's sensitivity to phonotactic patterns. In M. Broe and J. Pierrehumbert (eds) Papers in Laboratory Phonology V Cambridge, UK: Cambridge University Press, 259-273.

Tversky, A. (1977). Features of Similarity. *Psychological Review, 84:4*, 327-352.

Vilain, A., C. Abry, and P. Badin (1998). Coarticulation and degrees of freedom in the elaboration of a new articulatory plant. In R. H Mannel and J. Robert-Ribes (eds.) Proceedings of the Fifth International Congress on Spoken Language Processing Vol. 7, Sydney Australia Dec. 1998. Australian Speech Science and Technology Association Inc., 3147-3150.

Wetzels, W. L. (1995). Mid-vowel alternations in the Brazilian Portuguese verb. *Phonology, 12,* 281-304.