Functionalism

Matthew Gordon
University of California, Santa Barbara

1. Introduction

Investigation into the role of functional factors in shaping sound systems has developed into an important area of research in phonological theory. Growing out of earlier research exploring the articulatory, perceptual, and processing underpinnings of phonology, the last decade has witnessed the birth of a research program attempting to integrate functional, especially phonetic, explanations into formal analyses of phonological phenomena. This program of phonetically-driven phonology has been spurred by the advent of Optimality Theory, which allows for the direct grammatical encoding of phonetic naturalness in the form of constraints banning phonetically unnatural sounds and sequences of sounds.

While an overarching appreciation for the role of phonetic and other functional factors unites all work within phonetically-driven Optimality Theory (henceforth OT), there are disparate areas of research and viewpoints represented in the framework. Some work focuses principally on the role of articulation, other research attaches primary importance to perception, while other work appeals to processing factors. Some research focuses on the role of phonetic factors in predicting cross-linguistic markedness patterns, whereas other research explores correlations between phonetics and phonology on a language specific basis. Approaches also differ in terms of the predicates manipulated by the constraints; some favor analyses in which constraints are expressed using discrete phonological constructs while others assume that continuous phonetic variables are directly encoded in the constraints. Some researchers assume that phonetic considerations alone are sufficient to predict phonological patterns, while others assume that raw phonetic factors are mediated by measures of phonological simplicity.

2. The groundwork for phonetically-driven phonology

Interest in the role of phonetic factors in shaping phonological systems long antedates the rise of phonetically-driven OT. A substantial body of research conducted during the last three decades of the 20th century by phoneticians explores phonetic motivations for recurring patterns in sound inventories. In one of the earlier works in this research program, Liljencrants and Lindblom (1972) advance the hypothesis that vowel inventories are guided by a preference for vowels to be maximally distinct from each other in the perceptual domain. In order to quantity perceptual distinctness, they convert formant values expressed in Hertz in the acoustic dimension to a perceptual measure of frequency calculated in mels. As hypothesized, Liljencrants and Lindblom find a fairly close match between frequently occurring vowel systems and perceptual distinctness.

Later work by Lindblom and Maddieson (1986) builds on Liljencrants and Lindblom's perceptually based approach by attributing some role to articulatory factors in forging sound systems. Focusing on consonants, they propose a model in which languages prefer perceptually divergent sounds within regions of similar articulatory difficulty. Languages first exploit the subspace consisting of articulatorily simpler sounds, choosing sounds within the simple articulatory space that are maximally distinct from a perceptual standpoint. Once the articulatorily basic subspace is perceptually saturated, inventories are expanded through introduction of more complex articulations. Space within this second tier of articulatory difficult is then carved up according to perceptual distinctness until no more sounds may be added without jeopardizing other distinctions. At this point, any inventory expansion necessitates exploitation of the most difficult articulatory subspace. In this way, perceptual and articulatory factors conflict: maximizing perceptual distinctness comes at the price of greater articulatory difficulty, while minimizing articulatory effort reduces perceptual distinctness. This conflict between maximization of perceptual differentiation and minimization of articulatory complexity is a recurring theme of much work in phonetically-driven phonology.

Other work by various phoneticians tackles the phonetic motivations, both articulatory and perceptual, behind various phonological phenomena (see Ohala 1997 for an overview). Much of this work is appealed to by later researchers working within the framework of formal phonetically-driven phonology.

3. Optimality Theory and phonetic motivations in phonology

The advent of Optimality Theory (Prince and Smolensky 1993) in the 1990s sparked a large body of research attempting to integrate phonetic explanations directly into the OT formalism as constraints on naturalness. Early work in phonetically-driven OT attempted to ground implicational statements of markedness in acoustic and articulatory factors, arguing that a phonetically-informed model of phonology is both more explanatory and offers better empirical coverage than alternative approaches not appealing to phonetics.

3.1. Universal markedness scales in phonetically-driven OT: the case of laryngeal neutralization

Steriade's (1997) account of laryngeal neutralization provides a cogent example of the formal implementation of phonetically-driven phonology using OT. Steriade's work explores the hypothesis that observed implicational hierarchies in laryngeal neutralization sites correspond closely to hierarchies of perceptual salience: laryngeal contrasts are maintained in positions where they are less perceptible only if the same contrasts also exist in contexts of greater salience. To illustrate the basic patterns in need of explanation, many languages, such as Greek and Lithuanian, only have voicing contrasts in obstruents occurring before a sonorant, including vowels and sonorant consonants, e.g. áu**kl**e 'governess' vs. au**gl**ingas 'fruitful', a**km**uõ 'stone' vs. au**gm**uõ 'growth'. In other positions, including word-finally and before an obstruent, the voicing contrast is neutralized: to voiceless word-finally and to the voicing specification of a following obstruent word-medially, e.g. daũ**g** [**k**] 'much', a**tg**al [**dg**] 'back' vs. dè**gt**i [**kt**] 'burn-inf.' Other languages are less stringent in their minimal requirements of salience for voicing contrasts to be preserved. Thus, in Hungarian, a voicing contrast in obstruents is found not only in presonorant position but also word-finally. Yet another neutralization pattern, less stringent than the other two, is found in many varieties of Arabic and allows for voicing contrasts not only in presonorant position and word-finally, but also after a sonorant. Finally, the possibility of voicing contrasts occurring in all contexts, including when not adjacent to either a preceding or following sonorant is attested in Khasi. We thus have a hierarchy of voicing neutralization sites, as in (1), where languages differ in their cut-off points between permissible and impermissible locations for voicing contrasts in obstruents. Note that the division between languages lacking voicing contrasts and those only allowing contrasts in presonorant position is included for the sake of completeness

(1) Hierarchy of environments for laryngeal neutralization

Voicing contrasts allowed ◄————————► Voicing contrasts disallowed
nowhere presonorant word-finally postsonorant preobstruent everywhere

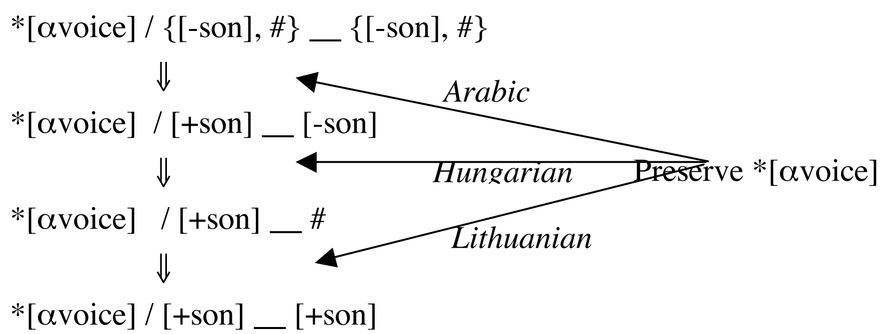
Steriade makes two important observations about this implicational hierarchy of voicing neutralization sites. First, she notes that neutralization accounts appealing to the syllable (e.g. Lombardi 1995) fail to explain many of the patterns in (1). In syllable-based accounts, laryngeal neutralization affects consonants in coda position of a syllable, since coda position is unable to license independently linked laryngeal features. The Hungarian type pattern, whereby voicing contrasts are limited to final position and to presonorant position cannot be explained with reference to the coda, since word-final obstruents are codas and are thus erroneously expected to undergo neutralization parallel to word-medial obstruents. Nor is it descriptively adequate to say that word-final consonants are extraprosodic and thus not codas, since word-final consonants are prosodically active in the calculation of the minimal word requirement, which is CVC in Hungarian. Similarly, in Lithuanian, only a subset of coda consonants, those occurring before obstruents and word-finally, undergo neutralization. It is thus insufficient to state that codas undergo neutralization in Lithuanian.

Rather, Steriade observes that perceptual considerations predict the Lithuanian type neutralization pattern and, more generally, the hierarchy of neutralization sites. Neutralization is more likely in contexts where laryngeal features are difficult to implement in a perceptually salient manner. Drawing on the results of studies on the perception of voicing (e.g. Raphael 1981, Slis 1986), Steriade suggests that the perceptual salience of laryngeal features in different environments depends on the acoustic properties associated with those environments. The accurate perception of obstruents, in particular stops, relies heavily on cues realized on transitions from the obstruents to adjacent vowels. Focusing on voicing, these contextual cues include the following: burst, which is less intense for voiced obstruents than for voiceless ones, voice-onset-time, which is negative for voiced stops and either zero or positive for voiceless stops, as well as fundamental frequency and first formant values during adjacent vowels, both of which are lower in proximity to voiced obstruents. Internal cues to obstruents, i.e. properties temporally aligned with the consonant constriction itself, are less numerous and generally less salient perceptually; these internal cues to laryngeal features include voicing, present for voiced obstruents but for voiceless ones, and closure duration, typically shorter for voiced obstruents than for voiceless ones.

Presonorant position is superior to preobstruent or final position for realizing a laryngeal contrast saliently, since several transitional cues are present: voice-onset-time, the burst, and fundamental frequency and first formant values at the offset of the consonant. Final position is better than preobstruent position since obstruents are more likely in this context to have an audible release and thus a burst and a discernible closure. Preobstruent position is worst from a perceptual standpoint, since the only cues to laryngeal features in this position are the internal cues of voicing, and if audibly released, closure duration.

Steriade posits a series of constraints whose ranking is fixed based on scales of perceptibility: constraints banning a laryngeal feature in a less salient context are ranked above constraints banning that laryngeal features in a more salient context. A faithfulness constraint requiring preservation of underlying laryngeal features is interleaved on a language specific basis to predict the laryngeal neutralization pattern characteristic of a given language. This schema (slightly modified from Steriade's analysis) is depicted in (2) for the feature [voice].

(2) Ranking of constraints governing voicing contrasts



Steriade characterizes her constraints in terms of [αF], where [F stands for the relevant laryngeal feature, in this case [voice]. She adopts this notation rather than one referring to either a positive or negatively stated feature, arguing that the perceptibility of the laryngeal contrast is at stake rather than only a positively or only a negatively specified feature values. This analysis is also consistent with the fact that laryngeal neutralization characteristically produces laryngeally unspecified consonants whose surface properties are those that are easiest to implement in a particular environment, voiced when preceding a voiced sounds and voiceless when preceding a voiceless sound or in final position, where aerodynamic considerations militate against voicing.

The fact that the output of laryngeal neutralization is context dependent indicates that Steriade's constraints are not wholly reliant on perceptual factors. Rather, Steriade suggests that the constraints refer to the ratio of effort required to implement a contrast in a perceptually salient manner: contrasts are more likely to be banned in contexts in which great effort must be expended for minimal perceptual rewards.

3.2. Language specificity in phonetic conditioning factors: the case of syllable weight
While Steriade's work focuses on the explication of universal markedness scales, other work within phonetically-driven OT tackles the issue of whether cross-linguistic variation in phonological patterns are also predictable on phonetic grounds. In his study of weight-sensitive stress, Gordon (2002) tests the hypothesis that closed syllables have different phonetic properties in languages in which they are phonologically heavy, e.g. Finnish, from languages, in which they are light, e.g. Khalkha Mongolian. As a starting point in the study, Gordon suggests that languages tend to adopt weight distinctions that are phonetically sensible, where a distinction's phonetic effectiveness is a function of the degree to which it offers maximal separation of heavy and light syllables.

Gordon tests various potential parameters along which phonetic effectiveness can be quantified, ultimately finding that a measure of perceptual energy of the rime, loudness integrated over time, matches up well with weight distinctions in a number of languages. Crucially, languages that treat CVC as light differ from those that treat CVC as heavy in the relative phonetic effectiveness of different distinctions. In languages with heavy CVC, the inclusion of CVC in the set of heavy syllables improves the degree of phonetic separation of heavy and light syllables relative to other candidate distinctions, in particular, the distinction that treats only CVV and not CVC as heavy. In languages with light CVC, on the other hand, treating CVC as a heavy syllable type reduces the phonetic effectiveness relative to other weight distinctions. Gordon's work builds on earlier work by Broselow et al. (1997) exploring language specific correlations between syllable weight and phonetic properties. Broselow et al., however, find a close correlation between coda weight and a simple measure of phonetic duration in languages with light CVC (Malayalam in their study) and heavy CVC (Hindi and Arabic). Broselow et al. find that vowels in closed syllables are substantially shorter than their counterparts in open syllables in Malayalam unlike in the examined languages with heavy CVC. They suggest that the shortening of vowels in closed syllables in Malayalam is attributed to mora sharing between the coda consonant and the nucleus. In languages with heavy CVC, on the other hand, there is no mora sharing between a nucleus and a coda, in keeping with the absence of a phonetic distinction in vowel length between open and closed syllables.

3.3. Directionality of the phonetics-phonology relationship

Besides the difference between Gordon and Broselow et al.'s studies in the phonetic parameters found to correlate with weight, the two works differ in the nature of the relationship assumed to obtain between phonetics and phonology. Broselow et al. take the position that languages tailor their phonetic systems in order to enhance the realization of phonological weight. Gordon, on the other hand, pursues the hypothesis that weight systems are constructed on the basis of a language's phonetic properties.

These two models of the phonetics-phonology interface are difficult to tease apart since they both predict a correlation between phonetics and phonology. One way to tease apart the two hypotheses is to look for an independent and language specific property of languages that could explain the observed phonetic patterns independent of weight. Gordon claims that the match between the phonetics and phonology of weight is attributed to a more basic property of languages: syllable structure. He finds that the languages that treat CVC as heavy have a higher proportion (in type frequency) of relatively intense codas, including sonorants and voiced consonants, than languages that treat CVC as light. Gordon suggests that the large number of high intensity codas in languages with heavy CVC increases the aggregate energy profile of CVC, thereby increasing the likelihood that it will be phonologically heavy. The observed indirect link between syllable structure and phonological weight criterion would be accidental in a model that assumes that phonology only influences but is not influenced by phonetics.

3.4. Phonological simplicity in phonetically-driven phonology

3.4.1. Simplicity in syllable weight

Gordon's work on weight explores another factor that emerges as relevant in quantitative studies of phonetically-driven phonology. He finds that certain hypothetical weight distinctions in fact provide a closer match to the phonetic map than some of the actual attested distinctions. For example, in Khalkha Mongolian, a distinction treating only long vowels and syllables containing /a/ followed by a coda nasal (CVV, CaN heavy) is phonetically superior to all other weight distinctions including the exploited distinction between heavy CVV and all other syllables. Gordon suggests that there is a bias against the {CVV, CaN} heavy distinction and others like it even if they are phonetically effective, since such distinctions manipulate highly asymmetrical weight categories. In the case of the {CVV, CaN} heavy distinction, reference must be made to multiple phonological dimensions: vowel length, vowel quality, and type of coda. Attested phonological dimensions are simpler in terms of the dimensions they manipulate, either number of timing positions in the case of the distinction that treats both CVV and CVC heavy, vowel length, or vowel quality. Gordon thus proposes that languages employ a criterion of phonological simplicity in addition to the criterion of phonetic effectiveness when evaluating potential weight distinctions: in his model, languages adopt the phonetically most sensible among the distinctions that do not exceed a complexity threshold, which Gordon tentatively formulates with reference to the number of associations between timing slots and features.

3.4.2. Simplicity in obstruent voicing patterns:

Gordon's appeal to a notion of phonological simplicity is shared with work by Hayes (1999) on the phonetic naturalness of obstruent voicing. Based on results from an aerodynamic modelling experiment, Hayes finds that the relative naturalness of stop voicing is contingent upon a number of factors, two of which I focus on here: place of articulation and the context in which the stop occurs. First, ease of voicing is correlated with frontness of the constriction. Bilabials facilitate voicing because they are associated with a relatively large oral cavity, which delays the equalization of oral and subglottal pressure that triggers cessation of vocal fold vibration. Velars, on the other hand, inhibit voicing since pressure builds up rapidly behind the closure thereby eliminating the necessary aerodynamic conditions for voicing. The second factor that predicts ease of voicing is the context in which the stop occurs. Voicing is facilitated in a postnasal context where the raising of the velum for the nasal increase the size of the cavity behind the oral closure and the potential for some air leaking through the nasal cavity delays the cessation of voicing. Voicing is slightly more difficult following a non-nasal sonorant, still more difficult in initial position, where low subglottal pressure inhibits voicing, and most difficult after an obstruent, where intraoral pressure is already high. Combining the two dimensions of frontness and environment yields a matrix of stop voicing naturalness (expressed in arbitrary units based on aerodynamic modelling), as in (3), where larger numbers indicate increased difficulty of voicing.

(3) Phonetic map of obstruent voicing

Environment	b	d	g
[-son]__ (after obst)	43	50	52
#__ (initial)	23	27	35
[+son, -nas] __ (after non-nasal sonorant)	10	20	30
[+nas]__ (after nasal)	0	0	0

While Hayes finds that cross-linguistic patterns of stop voicing line up well with the aerodynamic modelling results, phonologies of individual languages for the most part are sensitive to only one of the dimensions relative for predicting voicing ease: either context or place of articulation. For example, Latin bans voiced obstruents after another obstruent while Dakota's only voiced stop is the bilabial [b]. Strikingly absent are systems that are simultaneously sensitive to environment and place of articulation in predicting stop voicing patterns, even if these patterns are phonetically well-grounded. For example, we do not find languages that ban all voiced stops after an obstruent, both /b/ and /d/ but not /g/ in initial position, and /g/ but not /b/ and /d/ after a non-nasal sonorant.

Hayes suggests that the explanation for this gap in attested patterns lies in their complexity relative to other slightly less phonetically natural but nevertheless more symmetrical patterns. Hayes' procedure for integrating complexity and naturalness differs from Gordon's in assuming that phonetic naturalness is compared across constraints that are formally similar in terms of the features they manipulate (differing only in the substitution of a single predicate, e.g. switching feature values, addition or loss of feature, etc.). The phonetically most natural of the constraints within each family of closely related constraints are those that are exploited by actual languages. A crucial difference between Hayes' metric of simplicity and the one adopted by Gordon is that phonetic effectiveness in Hayes' approach is only evaluated across formally similar constraints, unlike in Gordon's work, which assumes that phonetic effectiveness is compared across all potential constraints regardless of their formal similarity.

3.5. Continuous phonetic variables and constraint formulation: the case of contour tones

The works discussed up to this point have in common that their constraint formulation relies on discrete phonological predicates, i.e. features, timing positions, syllables. Some work in phonetically-driven OT, however, has posited constraints referring directly to continuous phonetic variables, such as duration, frequency, and distance. The incorporation of gradience into the formal analysis has proved beneficial in at least two areas. First, certain phenomena appear to be sensitive to finer grained distinctions than traditional discrete representations are able to differentiate. Second, the application of many processes is dependent on speech rate, a factor that is not easily modelled using conventional phonological categories.

Zhang's (2002, 2004) analysis of contour tone distributions implements a set of constraints referring to continuous phonetic dimensions. Drawing on a cross-linguistic survey of contour tones in 187 languages, Zhang finds that certain syllables are more conducive to supporting contour tones than others. The first relevant dimension concerns the rime. Contour tones most prefer to dock on syllables containing a long vowel (CVV) than on any other syllable type, followed by short voweled syllables ending in a sonorant coda (CVR). Contour tones on short voweled open syllables and on short voweled syllables ending in a coda obstruent are comparatively rare. Zhang also finds that many languages preferentially allow contour tones on stressed syllables but not on unstressed syllables. Another predictor is syllable position: final syllables are more likely to tolerate tonal contours than non-final syllables. Finally, some languages are sensitive to number of syllables in a word, such that shorter words are more receptive to carrying contour tones than longer words.

Zhang proposes that all of these distributional skewings are sensible if one considers the phonetic requirements of tone. Tonal information is recoverable from not only the fundamental frequency, but also from the lower harmonics, which occur at frequency multiples of the fundamental. Sonorants are far better suited to carrying tone than obstruents due to their more energetic harmonic structure. Vowels are ideal carriers of tone since they have the greatest intensity in their harmonic structure. Because contour tones require a greater duration than simple tones to be executed in a perceptually recoverable manner, Zhang argues that it is not surprising that many languages restrict contour tones to CVV and others limit contours to CVV and CVR. It also follows that stressed vowels, final vowels, and vowels in shorter words should be better equipped to support contour tones. Stressed vowels are characteristically longer than unstressed vowels, final vowels are longer than non-final vowels, and vowels are longer in shorter words than in longer words.

Zhang posits a formula for predicting the ability of a syllable to carry a contour tone: $C_{CONTOUR} = a \cdot \text{Dur(V)} + \text{Dur(R)}$. According to this formula, the contour tone carrying ability is a function of the duration of a sonorant coda plus the duration of the vowel multiplied by some value (*a*) greater than one, which reflects the greater ability of a vowel to support a contour relative to a sonorant consonant. The actual value of *a* is not crucial for present purposes (see Zhang 2002 for discussion).

Whether a given syllable can support a contour tone or not is a function of the $C_{CONTOUR}$ value for the rime and the type of contour involved. Thus, rising tones require larger $C_{CONTOUR}$ values than falling tones since they take longer to execute, and complex tones require larger $C_{CONTOUR}$ values than contour tones since there are more tonal targets to reach. Formally, the tone bearing ability of different syllables is captured through a family of constraints of the form *CONTOUR(*T*)- $C_{CONTOUR}$(*R*), where a tone*T* is banned for a rime possessing an insufficiently large $C_{CONTOUR}$ value to support the tone. These constraints interact with faithfulness constraints requiring that underlying tones surface, PRES (T), and constraints banning excess length in the rime beyond that minimally required in a given prosodic context, *DUR. Parallel to the CONTOUR(*T*)- $C_{CONTOUR}$(*R*) constraints, both PRES (T) and *DUR refer to continuous values reflecting; in the case of PRES (T), the degree to which a surface tone is perceptually divergent from its corresponding input and, in the case of *DUR, the amount of the durational difference between the surface rime and the duration characteristic of a given prosodic position when not supporting a tonal contour.

Depending on the ranking of these three constraints relative to each other different output patterns emerge. If all members of both the PRES (T) and *DUR families outrank the relevant *CONTOUR(*T*)- $C_{CONTOUR}(R)$ constraint for a given contour tone, then that underlying tonal contour will surface without any lengthening of the rime to accommodate the tone. If all the *DUR constraints are undominated and *CONTOUR(*T*)- $C_{CONTOUR}(R)$ has priority over some but not all PRES (T) constraints, the tonal contour will be reduced in order to allow for its effective realization. If *CONTOUR(*T*)- $C_{CONTOUR}(R)$ and *DUR outrank all PRES (T), the contour will be completely eliminated. Yet another possibility is for at least some of the *DUR constraints to be outranked by *CONTOUR(*T*)- $C_{CONTOUR}(R)$ and PRES (T). This produces different patterns of lengthening to accommodate the contour tone, where the degree of lengthening depends on which of the *DUR constraints are outranked. A final possibility is a compromise between preserving vestiges of the underlying tonal contour and minimizing lengthening to accommodate the contour; this pattern reflects the ranking of *CONTOUR(*T*)- $C_{CONTOUR}(R)$ above some but not all PRES (T) and *DUR constraints.

An advantage of an analysis employing constraints referring to continuous variables is its ability to more closely capture surface forms than traditional analyses using less finely grained discrete predicates. For example, non-neutralizing lengthening of a short vowel in order to accommodate a contour tone can be represented in Zhang's approach. In a moraic analysis, lengthening in a language with contrastive vowel length can be captured in terms of mora count only if it neutralizes the underlying length distinction. Furthermore, the number of distinctions relevant to the phonology often exceeds the number that can be represented in traditional discrete phonological models. Thus, differences in the ability of various syllable types to carry contour tones are typically captured using moras, such that contour tones can be decomposed into level tones, each of which must be associated with its own mora. However, because the number of moras is limited by phonemic contrasts in length and segment count, certain tonal distribution facts cannot easily be accommodated by moraic models. For example, Zhang cites Mende as a language in which long vowels can carry the complex tone LHL in monosyllabic words but not in longer words. Similarly short vowels can carry both LH and HL contours in monosyllables, only HL contours in the final position of longer words and no contours in other environments. This type of pattern which makes reference to both type of contour and syllable count cannot be captured by a moraic model, in which mora count is consistent across different syllable positions and different word lengths. In Zhang's direct phonetics approach, the Mende patterns emerge naturally, since both number of syllables and syllable position influence the same phonetic variable, duration, which is referred to by a single constraint family. The difference between rising LH and falling HL tones also is predicted given that rising tones take longer to execute than falling tones.

Despite the descriptive richness permitted by a formal approach appealing directly to phonetics, there are some assumptions that such an analysis must make. First, because constraints refer to continuous phonetic properties rather than discrete phonological entities in Zhang's approach, it must be assumed that speakers normalize across different speech rates and styles. If this were not the case, then a constraint such as *CONTOUR(*T*)-C$_{\text{CONTOUR}}$(*R*) could potentially be violated by a form at fast speech rates but honored by the same form in slower speech, thereby yielding different phonologies at different speech rates. Zhang thus assumes that the values manipulated by constraints are determined on the basis of some canonical speech rate and style.

On the other hand, despite the apparent consistency found across speech rates, certain phenomena are rate dependent and suggest the need for constraints referring to absolute durations imposed by physiological limitations. For example, Kirchner (2004) discusses the OT modelling of lenition processes dependent on speech rate in Florentine Italian.

Another issue that a direct phonetics approach must address is the fact that the set of attested contrasts in any language is a small subset of those logically predicted to occur given a set of constraints manipulating continuous variables. To account for this fact, Flemming (1995, 2004) and Kirchner (1997) suggest that the set of contrasts is limited by considerations of perceptual distinctness, such that phonetic differences must be sufficiently salient if they are to be exploited as a phonological contrast. For example, Flemming assumes a family of constraints governing the perceptual distance between different formants. These MINDIST constraints compete with constraints requiring that articulatory effort be minimalized in keeping with Lindblom's (1986, 1990) Theory of Adaptive Dispersion, which assumes that phonological systems are the result of compromise between the conflicting goals of increasing the number of phonological contrasts while simultaneously minimizing articulatory effort and maximizing perceptual distinctiveness.

4. Other functional factors in phonology

4.1. Speech processing and phonology

In addition to purely phonetic factors, there are other functional considerations that appear to play a role in shaping phonological systems. One such factor is the mechanism of speech processing. In work investigating consonant cooccurrence restrictions in Arabic roots, Frisch et al. (2004) and Frisch (2004) suggest that similar consonants are avoided because they are more easily confused in both perception and production than dissimilar consonants. In order to make explicit this confusion, Frisch assumes Dell's (1986) connectionist model of phonological encoding in which different levels of phonological structure, e.g. features, segments, syllable position, word, are represented as distinct but interlinked tiers each consisting of activation nodes. A node associated with a given property is activated upon hearing or planning utterances containing that property or, in gradient fashion, other similar properties. For example, the activation node corresponding to the segment /k/ is strongly activated by any word containing the sound /k/ and less strongly activated by the occurrence of a word containing a different voiceless stop. Because featurally similar segments overlap in their activation patterns, there is potential for them to be mistaken for each other. Frisch et al. (2004) quantify similarity in terms of number of natural classes shared by the segments in question. Segments that share a greater number of natural classes are more similar to each other and thus less likely to cooccur in the same root.

4.2. Frequency in phonology

One of the factors relevant in many connectionist models of speech processing is word frequency; nodes associated with more frequent properties have lower thresholds of activation required for firing. As a result, frequent items are more likely to be produced or perceived when activated by items sharing similar properties. The relevance of frequency effects in speech production and perception finds independent support from psycholinguistic studies and plays an important role in Bybee's (2001) model of phonology. Bybee assumes that words may have different phonological representations in the mental lexicon according to their frequency. More frequent words are pronounced differently from less frequently words, tending in particular to undergo phonological reduction. For example, a very common word like 'every' is more likely to lack a vowel in the second syllable than a less frequently occurring word such as 'cursory'. In Bybee's model, these reduction phenomena gradually become incorporated into the lexicon leading to different distributions in surface pronunciation.

5. The synchronic vs. diachronic role of phonetics in phonology

One of the major outstanding issues in work on the phonetics-phonology interface concerns the question of whether phonetic considerations play an active role in synchronic phonologies or whether phonetic factors merely are at work on the diachronic level gradually causing languages to drift in the direction of greater phonetic naturalness. This evolutionary perspective on the role of phonetics in phonology has been espoused by a number of researchers (e.g. Ohala 1981, Hyman 2001, Blevins and Garrett 2004). Given that phonological unnatural phenomena exist in various languages, the position that phonetic factors govern all synchronic properties would appear to be untenable. Rather, the existence of seemingly phonetically unmotivated phenomena suggests that speakers have the ability to acquire patterns that could not be learned through phonetic experience. Thus, a phonetically-informed synchronic model of phonology must assume that the acquisition process entails both inductive learning through exposure to the ambient language as well as phonetic experimentation to determine which patterns are articulatorily easy to implement and perceptually recover.

In practice, it is difficult to find evidence that teases apart the synchronic vs. evolutionary view of phonetically-driven phonology. One promising avenue of investigation employs psycholinguistic experiments to determine whether speakers actively employ phonetic criteria in grammaticality judgments. This line of research, which is in its infancy, involves presenting listeners with phonological patterns differing in their phonetic naturalness and then observing how well the listeners acquire the presented patterns. If speakers were sensitive to phonetic considerations in constructing a grammar, they would be predicted to master phonetically natural patterns more easily than phonetically unmotivated patterns. If, on the other hand, phonetically natural patterns were not more easily acquired, the evolutionary view of phonetically-driven phonology would find support.

Recent research using psycholinguistic experiments has already shed some suggestive results. Psycha et al. (2003) presented listeners who are native speakers of English one of three artificially constructed patterns of vowel harmony. Crucially, because English does not have vowel harmony, any results could not be attributed to interference from preexisting knowledge of a harmony system. In one condition, the presented forms illustrated a phonetically natural rule of palatal harmony of the type found in many natural languages (e.g. Finnish) in which suffixes have two allomorphs varying in backness depending on the backness of the root vowel. In another condition, listeners were given forms instantiating a phonetically unnatural process of palatal disharmony in which the suffixal vowel has the opposite backness values of the root vowel. Finally, the third pattern involved an arbitrary type of palatal harmony in which certain front vowels (i, æ, ʊ) trigger a front vowel suffix, while others (ɪ, u, a) trigger a back vowel suffix. Both the phonetically natural harmony process and the phonetically unnatural disharmony process are formally simple in terms of manipulating a single phonological predicate, the backness value. The arbitrary rule of harmony, on the other hand, is formally complex since it requires reference simultaneously to height and backness of the vowels conditioning harmony in the suffix.

After a training session in which examples of harmony were presented aurally, listeners were asked for their grammaticality judgments on a series of novel forms differing in their wellformedness according to the learned harmony rule. Results indicated difficulty in acquiring the formally complex and arbitrary rule of vowel harmony as the correctness of listeners' grammaticality judgments hovered at chance levels for this type of harmony, significantly worse than performance for the other two types of harmony systems. Psycha et al. also found that the percentage of correct responses for listeners exposed to the phonetically natural harmony system was greater than for speakers presented with the phonetically unnatural but formally simple alternation. This difference, however, did not reach statistical significance, though the authors suggest that significance could be reached given a larger subject pool.

Using a somewhat different experimental paradigm, Wilson (2003) also attempted to address the role of naturalness in the acquisition process. Listeners were presented with one of two different nasal harmony processes. In one condition, listeners heard tokens containing a suffix with two allomorphs, [-na] and [-la], where the occurrence of each was conditioned by the nasality of the final consonant of the stem following a well attested and natural type of nasal harmony system found in natural languages: a nasal consonant triggered the [-na] variant whereas an oral consonant triggered the [-la] variant. The other group of listeners were given forms in which the [-na] allomorph was triggered by a final dorsal consonant and the [-la] allomorph was conditioned by a non-dorsal consonant, a less natural and unattested type of harmony system. After a training session in which the relevant grammar was illustrated, listeners were presented novel forms either conforming or failing to conform to the patterns from the training session, and asked whether they had heard these forms previously or not. Wilson found that listeners were far more accurate in recognizing forms conforming to the phonetically more natural rule of nasal harmony conditioned by the nasality of the final root consonant than the rule conditioned by the dorsality of the final consonant. In a follow-up experiment, listeners were presented with forms illustrating a process of nasal disharmony in which a nasal consonant in the root triggered the non-nasal [-la] allomorph. Nasal disharmony is attested in several languages (Alderete 1997, Suzuki 1998). In keeping with the results of Psycha et al. (2003), listeners were better able to recognize grammatical forms illustrating disharmony than another group of listeners exposed to an arbitrary rule of in which the [-la] allomorph was conditioned by a dorsal consonant in the root. Wilson does not make a direct comparison of results for the nasal harmony and nasal disharmony conditions.

Zhang and Lai (2005) also delved into the relative productivity of phonetically motivated and phonetically unmotivated processes in their study of Mandarin tone sandhi. Mandarin possesses two types of tone sandhi, one with a much clearer phonetic motivation than the other. The phonetically natural sandhi involves simplification of the complex dipping (213) tone to a simple falling (21) tone in a phrasal context preceding another word with either a high level (55) tone, a rising (35) tone, or a falling (51) tone. This type of sandhi is presumably the natural result of truncating the tone in a phrase-medial context in which there is less time to execute all three tonal targets required for the canonical realization of the dipping tone. The phonetically less natural tone sandhi changes the dipping tone to a rising (35) tone before an immediately following dipping tone. Zhang and Lai presented subjects pairs of words, in which the first contained the dipping tone and the second contained one of four tones, three of which trigger the phonetically natural sandhi and one, the dipping tone, which triggers the less natural sandhi. The pairs of words differed in that, for some, both words were real, for others both were artificial, and, for still others, only one of the two words was real. Subjects were asked to apply tone sandhi immediately upon presentation of the word pairs. Results indicate that the temporal lag between the presentation of the words and the speakers' application of sandhi was greater in the case of the less natural sandhi for both real and nonse words. Furthermore, among the nonse words, the more natural sandhi was produced with greater phonetic accuracy in terms of contour shape than the less natural sandhi process. Zhang and Lai's results are thus consistent with the view that phonetically natural phenomena have a privileged status in terms of ease of acquisition.

In summary, rigorous research into the synchronic productivity of phonetic conditioning factors is still in its early stages. Results are not completely conclusive but thus far offer some support for the view that speakers have access to phonetic knowledge in constructing phonologies.

6. Conclusions

Exploration of the functional bases for phonological patterns is a productive area of research, since many cross-linguistic distributional facts about phonology appear to be explainable in terms of independent biases in speech articulation, perception, and processing. Many of these functional factors have been incorporated into formal phonological analyses using the constraint based framework of Optimality Theory. Despite important advances in our understanding of the role of functional factors in shaping phonological systems, there are still critical questions remaining to be answered about how and whether phonetic and processing explanations should be implemented in formal models of phonology reflecting synchronic linguistic knowledge.

References

Alderete, John (1997). Dissimilation as local conjunction. *Proceedings of the North East Linguistic Society* 37, 17-31.

Blevins, Juliette and Andrew Garrett (2004). The evolution of metathesis. In Bruce Hayes, Donca Steriade, and Robert Kirchner (eds.), *Phonetically based phonology*. New York: Cambridge University Press, 117-156.

Broselow, Ellen, Susan Chen, and Marie Huffman (1997). Syllable weight: convergence of phonology and phonetics' *Phonology* 14, 47-82.

Bybee, Joan. (2001). *Phonology and language use*. New York: Cambridge University Press.

Dell, Gary (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review* 93, 283-321.

Flemming, Edward (1995). *Auditory representations in phonology*. UCLA PhD dissertation.

Flemming, Edward (2004). Contrast and perceptual distinctiveness. In Bruce Hayes, Donca Steriade, and Robert Kirchner (eds.), *Phonetically based phonology*. New York: Cambridge University Press, 232-276.

Frisch, Stefan. (2004). Language processing and segmental OCP effects. In Bruce Hayes, Donca Steriade, and Robert Kirchner (eds.), *Phonetically based phonology*. New York: Cambridge University Press, 346-371.

Frisch, Stefan, Janet Pierrehumbert, and Michael Broe (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22, 179-228.

Gordon, Matthew (2002). A phonetically-driven account of syllable weight' *Language* 78, 51-80.

Hayes, Bruce (1999). Phonetically-driven phonology: the role of optimality theory and inductive grounding. In M. Darnell, E. Moravscik, M. Noonan, F. Newmeyer, and K. Wheatley (eds.), *Proceedings of the 1996 Milwaukee Conference on Formalism and Functionalism in Linguistics*, *volume 1*. Amsterdam: Benjamins. 243-285.

Hyman, Larry (2001). The limits of phonetic determinism in phonology: *NC revisited. In Elizabeth Hume and Keith Johnson (eds.) *The role of speech perception in phonology*. San Diego: Academic Press. 141-85.

Kirchner, Robert (1997). Contrastiveness and faithfulness. *Phonology* 14, 83-111.

Kirchner, Robert (2004). Consonant lenition. In Bruce Hayes, Donca Steriade, and Robert Kirchner (eds.), *Phonetically based phonology*. New York: Cambridge University Press, 313-345.

Liljencrants, J. and B. Lindblom, B. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48, 839-62.

Lindblom, Björn. (1986). Phonetic universals in vowel systems. In J.J. Ohala and J.J. Jaeger (eds.), *Experimental phonology*. Orlando, FL: Academic Press.

Lindblom, Björn. (1990). Phonetic content in phonology. *PERILUS* 11: 101-18.

Lombardi, Linda (1995). 'Laryngeal neutralization and syllable well-formedness' *Natural Language and Linguistic Theory 13*, 39-74.

Ohala, John (1981). The listener as a source of sound change. In Carrie Masek, Roberta Hendrick, and Mary Frances Miller (eds.), *Papers from the Parasession on Language and Behavior, Chicago Linguistics Society*. Chicago: Chicago Linguistics Society, 178-203.

Ohala, John (1997). The relation between phonetics and phonology. In William Hardcastle and John Laver (eds.), *The Handbook of phonetic sciences*. Oxford: Blackwell Publishers, 674-694.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms. Rutgers University and University of Colorado at Boulder.

Psycha, Anne, Pawel Nowak, Eurie Shin, and Ryan Shosted. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In G. Garding and M. Tsujimura (eds.), *West Coast Conference on Formal Linguistics* 22. Cambridge, MA: Cascadilla Press, 423-435.

Raphael, L. (1981). Durations and contexts as cues to word-final cognate opposition in English, *Phonetica* 38, 126-147.

Slis, Iman (1986). Assimilation of voice in Dutch as a function of stress, word boundaries and sex of speaker and listener. *Journal of Phonetics* 14, 311-326.

Steriade, Donca (1999). Licensing laryngeal features.*UCLA Working Papers in Phonology* 3, 25-146.

Suzuki, Keiichiro (1998). *A typological investigation of dissimilation*. U of Arizona PhD dissertation.

Wilson, Colin. (2003). Experimental investigation of phonological naturalness. In G. Garding and M. Tsujimura (eds.), *West Coast Conference on Formal Linguistics* 22. Cambridge, MA: Cascadilla Press, 533-546.

Zhang, J. (2002). *The effects of duration and sonority on contour tone distribution: typological survey and formal analysis*. New York: Routledge.

Zhang, Jie. (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution. In Bruce Hayes, Donca Steriade, and Robert Kirchner (eds.), *Phonetically based phonology*. New York: Cambridge University Press, 157-190.

Zhang, Jie and Yuwen Lai (2005). The phsychological reality of Mandarin tone sandhi. Poster presented at the annual LSA meeting, January 5-9, 2005. Oakland, CA