9082123468

I let stand the 17 prefix on the figures, as I've seen it in other CUP publications. This one is different, go ahead and take it out.

17

The phonetics-phonology interface

John Kingston

17.1 Introduction

Phonetics interfaces with phonology in three ways. First, phonetics defines distinctive features. Second, phonetics explains many phonological patterns. These two interfaces constitute what has come to be called the 'substantive grounding' of phonology (Archangeli & Pulleyblank 1994). Finally, phonetics implements phonological representations.

The number and depth of these interfaces is so great that one is naturally moved to ask how autonomous phonetics and phonology are from one another and whether one can be largely reduced to the other. The answers to these questions in the current literature could not differ more. At one extreme, Ohala (1990b) argues that there is in fact no interface between phonetics and phonology because the latter can largely if not completely be reduced to the former. At the opposite extreme, Hale & Reiss (2000b) argue for excluding phonetics entirely from phonology because the latter is about computation, while the former is about something else. Between these extremes are a large variety of other answers to these questions, ranging from Blevins's (2004) claim that phonetics motivates sound change but does not otherwise regulate synchronic sound patterns to Browman & Goldstein's (1995) assertion that phonological representations are merely an assemblage of phonetic units of a grain coarse enough to be reliably categorized (see also Hayes 1999).

These examples aren't a comprehensive list of current points of view, nor do they represent the principal alternatives from which one might choose. They instead merely show that the field has reached no consensus about what the interface is, nor has it even agreed that one exists at all. The field therefore cannot agree about how distinctive features are defined, phonological patterns explained, or phonological representations implemented by the phonetics. The confident assertions about the three interfaces with which I began this paper are not self-evident truths to everyone, much less any particular phonetic definitions, explanations, and implementations.

These disagreements could be nothing more than the consequences of lacking the evidence needed to choose between competing hypotheses. But I think the disagreements reflect something more than the commonplace struggle between hypotheses. There is a broader and deeper dispute here about how one can reliably separate the phonetic from the phonological.

In this chapter, I lay out some of the difficulties one encounters in trying to answer this question for each of the three ways that phonetics interfaces with phonology: definition, explanation, and implementation. I do so by displaying some of the enormous richness of the interchange between phonetics and phonology. This richness is what forestalls any simple solution to the division of labor between these two components of the grammar. Nonetheless, where it is possible to do so, I show solutions – even incomplete ones – to the problem of dividing labor between phonetics and phonology. I have omitted many technical details in discussing the various cases presented in this chapter when doing so would not impair understanding or when they may be easily found in the sources that I cite.

The problem of defining distinctive features apparently can be solved by starting with the phonetics and working up to the phonology. Solving the problem of explaining phonological patterns is not so simple, because more than one force is at work even in apparently simple cases, these forces are phonological as well as phonetic, and finally they compete with one another. The phonological pattern that results represents the often delicate resolution of this competition. Finally, the problem of implementing phonological representations cannot to be solved by simply reversing the solution to defining their constituents and working down from the phonology to a pronunciation or percept. Instead, the phonetic implementation also determines what kind of phonological representation is possible in the first place.

17.2 Definition

17.2.1 Resolving the variability problem

Phoneticians and phonologists have worked hard to define distinctive features phonetically. Landmarks in this effort are the acoustic-auditory definitions in *Preliminaries to Speech Analysis* (Jakobson, Fant, & Halle 1952), the articulatory alternatives in Chapter 7 of the *Sound Pattern of English* (Chomsky & Halle 1968), and most recently, the combined acoustic and articulatory definitions in *The Sounds of the World's Languages* (Ladefoged & Maddieson 1996) and *Acoustic Phonetics* (Stevens 1998). Distinctive features would be easy to define phonetically if some articulatory or acoustic property or properties could be observed every time a distinctive feature has a particular value in an utterance's phonological representation. Unfortunately, this is not the case. Rather than invariant phonetic realizations, distinctive feature values are realized differently in different languages,

contexts, speaking styles, and even speakers (Kingston & Diehl 1994; cf. Stevens & Blumstein 1978; Sussman, McCaffrey, & Matthews 1991). How then can distinctive features be defined phonetically?

Some research suggests that distinctive feature values are in fact polymorphous, in that their phonetic realizations bear at best a family resemblance to one another (Kluender 1994; Kingston 2003). However, two recent approaches find invariance by stepping away from the detail of a particular utterance's phonetic realization. The approaches differ in the direction they recommend one should step away to find this invariance: Articulatory Phonology recommends one step back from the utterance's articulatory detail to the speaker's plan for the utterance (Browman & Goldstein 1995), while Auditorism instead recommends one step forward from the utterance's acoustic detail to the acoustic properties' auditory effects (Kingston & Diehl 1994, 1995; Kingston, Diehl, Kirk, & Castleman, in preparation). The two approaches resemble one another in finding invariance by moving to a description of the utterance with many fewer dimensions than are necessary to describe its articulatory or acoustic realization.

17.2.1.1 Articulatory Phonology

In Articulatory Phonology, an utterance is represented as a collection of *gestures*.[1] Gestures specify that the vocal tract be constricted at a particular location to a particular degree for particular interval of time. As such, a gesture specifies the speaker's goal in the time interval during which the gesture is active, and this specification evokes the coordinated action of the various articulators whose movements achieve that goal. Because the gesture specifies the goal rather than the movements of individual articulators, an articulator can contribute different amounts to achieving that goal in different contexts. For example, the goal in [b], [p], or [m] is to close the lips, and the upper and lower lips and the jaw all move to accomplish this goal, but each moves differently depending on neighboring vowel because the vowel's gestures are active at the same time as the lip closing gesture and they compete for control over these articulators (Sussman, MacNeilage, & Hanson 1973; Macchi 1988). This competition is resolved by the *task dynamics*, which calculates articulator movements by resolving the demands imposed by all the gestures that are active at any moment in time. These different combinations of upper and low lip and jaw movement that contribute to closing the lips next to different vowels are *motor equivalent*, because they all succeed in achieving that goal. The gesture that specifies that goal in the first place is then the desired step back from the variable realization of that goal by the different combinations of individual articulator movements to the invariant specification of the goal itself.[2]

17.2.1.2 Auditorism

Auditorism finds invariance in the listener's percepts rather than the speaker's goals. Just as the individual articulators' movements vary, so too

do their acoustic consequences. Listeners could therefore perceive each token as a different value, yet they don't do so. They don't because different arrays of acoustic properties are perceptually equivalent to one another. For example, a stop with a relatively short delay in voice onset (VOT) following its release and a relatively high onset frequency for the first formant (F1) of the following vowel is equally likely to be perceived by an English speaker as [+voice] as one with a relatively longer VOT and a relatively lower F1 onset frequency (Lisker 1975; Summerfield & Haggard 1977; Kluender 1991; Benkí 2001). Kingston, et al. (in preparation) argue that acoustic properties can be perceptually equivalent like this when their auditory effects are similar enough that they integrate perceptually with one another (see also Kingston & Diehl 1994, 1995; Kingston & Macmillan 1995; Kingston, Macmillan, Walsh Dickey, Thorburn, & Bartels 1997; Macmillan, Kingston, Thorburn, Walsh Dickey, & Bartels 1999). In the example, a shorter delay in voice onset and a lower F1 onset frequency both create the percept that low frequency energy occurs near the stop release.

Perceptual equivalence could arise from a source other than the auditory similarity of acoustic properties. The properties could be perceptually equivalent simply because listeners have experienced them covarying reliably (see Holt, Lotto, & Kluender 2001): stops with shorter voice onset delays usually have lower F1 onset frequencies, too. However, Kingston, et al. (in preparation, also Kingston & Diehl 1995) obtained the same responses from listeners to non-speech analogues in which acoustic properties are manipulated in the same way as in the speech signals. Because these stimuli aren't recognized as speech, they should not evoke the listeners' experience with the reliable covariation of acoustic properties in the speech signals they mimic. Listeners should only respond in the same way to the non-speech analogues if their acoustic properties are auditorily similar enough to integrate perceptually. Moreover, if speech sounds are to contrast reliably with one another, speakers may be enjoined to produce articulations whose acoustic correlates integrate perceptually with one another.

Perceptual integration thus achieves the same result as the motor equivalence embodied in gestures: a variable, high-dimensional description is reduced to an invariant, low-dimensional one whose units correspond to the contrastive units of which phonological representations are composed. Distinctive features may therefore emerge out of human's speaking or listening behavior, i.e. either out of the motor equivalence of different combinations of articulations or out of the perceptual integration of different combinations of acoustic properties. If this is correct, then distinctive features can be obtained without the phonological component of the grammar having to impose formal constraints requiring structural symmetry such as those argued for by Hayes (1999). At a gestural or auditory level of description, much of the phonetic particularity that phonological constraints typically ignore has already been lost.

17.2.2 Articulatory or auditory targets?

We are burdened here with an embarrassment of riches, two ways of getting distinctive features to emerge out of the phonetics. Is there any reason to decide that speakers' targets are articulatory or auditory? Speakers' compensation for artificial perturbations of articulations and natural covariation between articulations suggests that their targets are auditory rather than articulatory.

As long as the perturbations of articulations are not too extreme, speakers immediately and successfully compensate for them. For example, when the jaw is prevented from moving by a bite block between the molars, speakers still constrict the vocal tract in the same locations and to the same degree in producing vowels, and the vowels differ very little acoustically from those produced without the bite block (Lindblom, Lubker, & Gay 1979; Fowler & Turvey 1980; Kelso & Tuller 1983).[3] Similarly, if a light load is randomly and infrequently applied to the lower lip at the moment when a bilabial closure is initiated, the speaker exerts more force to lower the upper lip more as well as to overcome the load on the lower lip, and the closure is achieved (Abbs, Gracco, & Cole 1984). Both results demonstrate that different combinations of articulations are motor equivalent, and they suggest that speakers' targets are local constrictions of the vocal tract, as are Articulatory Phonology's gestures.

Other results, however, suggest that speakers' targets are global configurations of the vocal tract. For example, when the upper lip is prevented from protruding in a rounded vowel, speakers compensate by lowering their larynges more (Riordan 1977). As speakers lower their larynges anyway in pronouncing rounded vowels (Lindblom & Sundberg 1971), this additional lowering simply exaggerates an articulatory movement they already make. This finding suggests that the speakers' target is a long resonating cavity rather than a local constriction at the lips. A similar effort to keep resonator length constant can be observed in unperturbed pronunciations of American English [u] and [ʃ], where lip protrusion trades off with tongue backing from token to token (Perkell, Matthies, Svirsky, & Jordan 1993; Perkell, Matthies, & Zandipour 1998).

All these results can also be interpreted as evidence that the speaker is trying to produce a particular acoustic or auditory effect. Variation in lingual articulations of American English [ɹ] provides further evidence that the speaker's target is auditory: speakers use the more efficient bunched articulation after lingual consonants and a retroflexed articulation elsewhere (Guenther, Espy-Wilson, Boyce, Matthies, Zandipour, & Perkell 1999; cf. earlier studies suggesting this is variation between speakers: Delattre & Freeman 1968; Westbury, Hashi, & Lindstrom 1998; Alwan, Narayanan, & Haker 1997). [ɹ] can be pronounced in both ways because they both lower F3 extremely. Bunching produces a long constriction on the palate where F3 has a velocity maximum, and retroflexing creates a large sublingual cavity from which a low F3 arises. Neither the local constriction nor the global configuration of the vocal tract is the same in these two articulations.

The most compelling evidence that speakers' targets are auditory rather than articulatory comes from studies in which auditory feedback about the sound is perturbed rather than its articulation (Houde & Jordan 1998, 2002; Jones & Munhall 2002, 2003). Houde & Jordan (1998, 2002) altered auditory feedback to listeners gradually such that the vowel [ɛ] in *pep* came to sound increasingly like the higher vowel [i]. In response, speakers shifted their articulations of [ɛ] toward the lower vowels [æ] or [a], undoing the alteration. Speakers also compensated by shifting their articulations on trials where the feedback about *pep* was replaced by noise. They did so, too, in pronouncing [ɛ] in words other than *pep*, and for other vowels, even though feedback wasn't altered for [ɛ] other than in *pep* or for other vowels. These other shifts show that speakers have auditory rather than articulatory targets, and that these targets are determined in relation to the auditory targets of other sounds in the same class.

move word

All these results are compatible with the hypothesis that speakers' targets are auditory rather than articulatory, while only some of them are compatible with the opposing hypothesis. They thus suggest that the invariants from which distinctive features emerge are the auditorily similar effects of covarying acoustic properties and not the motor equivalences of different combinations of articulations.

17.3 Explanation

17.3.1 Introduction

Phonetic explanations of phonological patterns are built from physical, physiological, and/or psychological properties of speaking and listening. For example, /g/ is missing in Dutch or Thai but not /b/ and /d/ because it is much harder it is to keep air flowing up through the glottis when the stop closure is velar rather than bilabial or alveolar (Ohala 1976; Javkin 1977). Stops intrude between nasals or laterals and following fricatives in many American English speakers' pronunciations of words such as *warm[p]th*, *prin[t]ce*, *leng[k]th*, and *el[t]se* because voicing ceases and in the case of the nasal-fricative sequences the soft palate rises before the oral articulators move to the fricative configuration (Ohala 1971, 1974, 1981). The velar stop [k] palatalizes to [k^{j}] before [i] because the consonant coarticulates with the vowel, and it eventually affricates to [tʃ] because [k^{j}] is auditorily similar to [tʃ] (Plauché, Delogu, & Ohala 1997; Guion 1998; Chang, Plauché, & Ohala 2001).

no line break

Although all of these phonological patterns are peculiar to particular speech communities or even individuals (many languages have /g/ as well as /b, d/; stops don't intrude between nasals or laterals and fricatives in South African English (Fourakis & Port 1986), and [k] often remains unpalatalized and unaffricated despite coarticulating with [i]), they recur in unrelated speech communities, and they are phonetically possible in all speech

remove comma after "English" and add parens around citation

communities. They recur and are always phonetically possible because all humans who aren't suffering from some speech or hearing pathology possess essentially the same apparatus for speaking and listening. Indeed, as Ohala has repeatedly shown, these and many other phonological patterns can be reproduced in the laboratory with speakers and listeners whose languages don't (yet) exhibit them. Explanations of this kind are highly valued because they are built on generalizations of properties that can be observed any time the affected sound or sounds are uttered or heard, and they are in many instances built on generalizations of properties that can be observed in other domains than speaking and listening. This section illustrates how such explanations are constructed for languages' synchronic sound inventories and for the diachronic changes they undergo.

17.3.2 Explaining inventory content

17.3.2.1 Introduction

There is considerable evidence that the contents of segment inventories can be explained phonetically. Languages have the oral, nasal, and reduced vowels they do because vowels must be dispersed perceptually in the vowel space, certain vowel qualities are more salient than others, and a long vowel duration makes it possible for a listener to hear nasalization while a short duration prevents the speaker from reaching a low target. These factors don't completely explain the contents of these inventories, but they will form a part of any eventual complete explanations. The lesson is that the contents of segment inventories, even apparently compact subsets of inventories such as these, are determined by many more than just one factor. These factors may conflict with one another, and a balance must be struck between them when they do.

17.3.2.2 Oral vowels: the facts to be explained

Languages' vowel inventories resemble one another closely. This section begins by describing these resemblances among the oral vowel inventories in the areally and genetically balanced database of 451 languages in the UCLA Phonological Segment Inventory Database (UPSID, Maddieson & Precoda 1992). Liljencrantz & Lindblom (1972) and subsequent work by Lindblom (1986) established that two factors contribute to these resemblances: the vowel space is limited[4] and vowels mutually repel one another within that space. Following up proposals in Stevens (1989), Schwartz, Boë, Vallée, & Abry (1997a,1997b) added a third factor: languages prefer vowels that are made salient by the close proximity of two of their formants, an effect called 'focalization'. These forces successfully predict that inventories of certain sizes are preferred over both smaller and larger ones but fall short in predicting which vowels are most likely to occur in an inventory of a given size.[5]

The short and long oral vowels were extracted from each of the 451 languages in UPSID. All secondary articulations were stripped off, and the

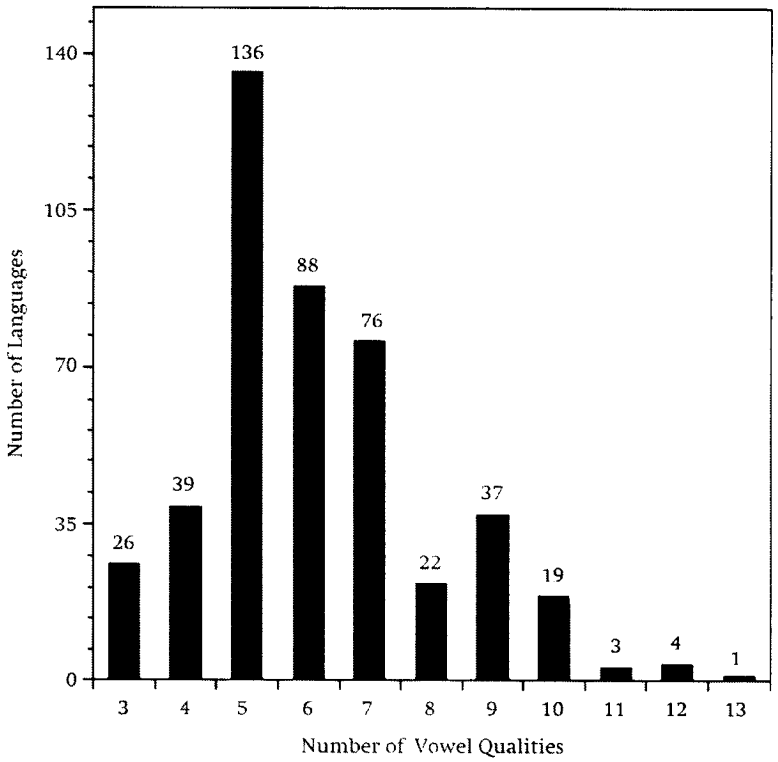


Figure 1: The number of languages having between three and thirteen oral vowel qualities, based on the short and long oral vowel inventories in UPSID.

distinct short and long vowel qualities were counted. If a language's long vowels distinguished more qualities, their number was used to represent how many vowels that language had; otherwise, the number of distinct short vowel qualities was used.[6]

The smallest number of distinct vowel qualities was three and the largest number thirteen. The histogram in Figure 17.1 shows how many languages have each number of vowels within these two extremes.

There is a very clear mode at five vowels, which are found in 136 languages. Even though many languages have either six or seven vowels, only about 65% as many have six vowels as have five vowels and only about 56% as many have seven.

Fully two-thirds of the languages in the sample (300/451) have the three most common numbers of vowels: five, six, and seven. The strength of this preference is emphasized by how few languages have four or eight vowels: in each instance just under 29% as many languages have five or seven vowels,

respectively. A surprisingly large number of languages have nine, substantially more than have eight.

17.3.2.3 Peripheral vs. central vowels

Aside from these preferences for a certain number of vowels, the remaining facts to be explained are the preferred arrangements of vowels in the vowel space for inventories of different sizes. The 45 oral vowel qualities distinguished in UPSID were divided into 15 *peripheral* vowels and 29 *central* qualities: [i u ɪ ʊ e̞ o̞ e o ɛ ɔ æ a̟ a ɑ ɒ] vs. [y ɯ ʏ ɯ̟ ø̞ ɤ̞ ø̞> ɤ̞< e̞> o̞< o ɤ e> o< œ ʌ ɞ ɨ ʉ ɪ ʊ ɘ̞ ɵ̞ ə ɵ ɜ ɞ ɐ ɐ>]. The symbols < and > indicate advanced and retracted pronunciations, respectively.

At the top of the vowel space, only [i] and [u] are peripheral. For non-low peripheral vowels, the tongue body is either as far forward or backward as possible, and the lips are unrounded if the tongue is the front, but rounded if the tongue is back. In low peripheral vowels, the tongue is as low as it can get. This cluster of properties shows that peripheral vowels are actually defined acoustically rather than articulatorily: if non-low, they have the highest or lowest F2 and F3 frequency values for their tongue height, or if low, they have the highest F1 frequency values. Central vowels include those articulated with the tongue body in a central position as well as vowels in which the tongue body is fully front or back, but the lips are rounded when the tongue is front or unrounded when it's back. This definition is also acoustic rather than articulatory in that F2 and F3 frequencies are neither lowered nor raised particularly when the tongue body is central or when the lips are rounded in front vowels or unrounded in back ones. Thus, peripheral vowels are acoustically farther apart from one another in the vowel space than central vowels.[7]

Following Schwartz, et al. (1997a), the vowel inventories for each of the languages with from three to ten vowels were then classified into patterns by how many peripheral and central vowels they have – the 8 languages with more than 10 vowels are ignored in the rest of this discussion. A pattern is identified by a 'P(eripheral)+C(entral)' formula. The results are shown in Figure 17.2, where each panel corresponds to a number of vowels, and the bar heights indicate the proportion of languages with that many vowels which have a particular pattern.[8]

In the vast majority of five vowel languages, all five vowels are peripheral (Figure 17.2c). In all but three of the 127 languages with this 5+0 pattern, front unrounded and back rounded high and mid vowels contrast and there is one low vowel. The remaining 9 languages have four peripheral vowels and one central vowel, the 4+1 pattern.

Figure 17.2d shows that the most common pattern among six vowel languages is 5+1, which occurs in 64 languages. It is distantly followed by 6+0, at just a third 5+1's frequency pattern, in 22 languages.

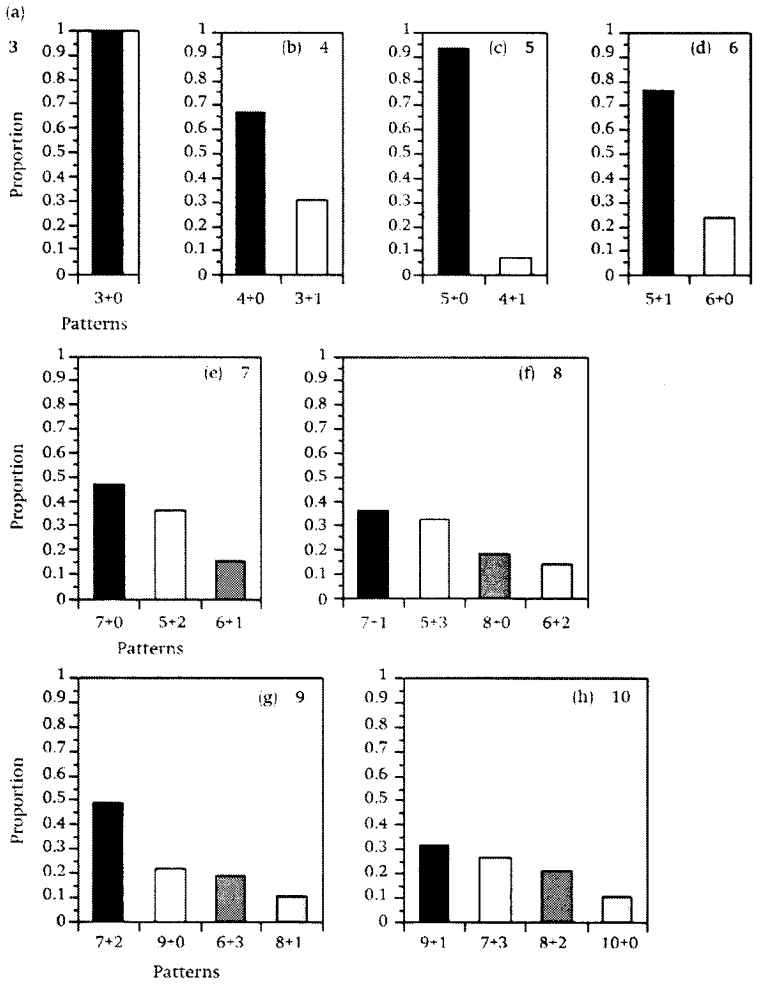


Figure 2: Proportions of languages with from three to ten vowels that exhibit particular common patterns of peripheral and central vowels for languages. Some proportions don't add up to 1 because patterns aren't shown that appear in only a very few languages or that can't be classified as one of these patterns.

Two patterns are common among the seven vowel languages (Figure 17.2e), 7+0 in 36 languages and 5+2 in 28. The most common 7+0 inventory adds a height distinction between mid vowels to the 5+0 pattern, distinguishing /i, e̞, ɛ, a, ɔ, o̞, u/. Languages with five peripheral vowels and two central vowels are much less common than those with five peripheral vowels and just one central vowel, 28 vs. 64. A dozen languages have the third most common pattern among the seven vowel languages, 6+1.

The large proportions of languages with the 5+1 and 5+2 patterns shows that many languages with six or seven vowels have 'added' central vowels to the very popular 5+0 peripheral vowel inventory.[9] The heights and other properties of the central vowels in these languages are unpredictable from what peripheral vowels they have, which suggests that central and peripheral inventories are independent of one another.

The other way to increase inventory size is to add one or two peripheral vowels. Adding central vowels doesn't change the distribution of peripheral vowels, but adding one peripheral mid vowel, in either the front or the back, usually entails adding the other at the same height, as well as shifting the existing mid vowel's height to equalize the intervals between vowels of different heights.

All 26 languages with just three vowels have the 3+0 pattern (Figure 17.22a).[10] 22 of them are missing the two mid vowels that occur in the 5+0 pattern, while just 4 are missing one or both high vowels. The 39 languages with four vowels divide unevenly into 24 with four peripheral qualities and 14 with three peripheral qualities and one central quality – the remaining four-vowel inventory /i, ɜ, a, ʔ/ is unclassifiable (Figure 17.22b).

Turning now to the languages with more than seven vowels, Figures 17.22f–h show that the most common patterns have odd numbers of peripheral vowels: five, seven, or even nine. An odd number of peripheral vowels occurs in 13 of the 22 languages with eight vowels (8 with 7+1, 5 with 5+3), in 28 of the 37 languages with nine vowels (19 with 7+2, 9 with 9+0), and 11 of the 19 languages with ten vowels (6 with 9+1 and 5 with 7+3). More often than not, the peripheral vowels in larger inventories also symmetrically contrast front unrounded with back rounded vowels at all non-low heights.

In summary, the most common vowel patterns contrast front unrounded with back rounded peripheral vowels at all but the lowest height, where only a single vowel is found. Larger vowel inventories differ from smaller ones in two ways: they may have central vowels that are absent in the smaller inventories and/or they may have more contrasts between the high and low peripheral extremes. The most common inventory by far is the 5+0 pattern. The next smallest inventory is more likely to have lost one of the mid vowels, producing the 4+0 pattern in 24 languages, than to lose both and add one central vowel, producing the 3+1 pattern in only 14 languages. It's far more likely that a language will add one central vowel than to add one peripheral vowel (64 languages with 5+1 vs. just 22 with 6+0). However, when two or more vowels are added to 5+0, it is far more likely that a single front:back pair will be added before any other vowels: 7+0 > 5+2, 7+1 > 5+3, 7+2 > 9+0, and the 7+3 occurs in nearly as many languages as 9+1. Once this additional pair of peripheral vowels is added, any more vowels are likely to be central.

17.3.2.4 The explanation

Why are certain total numbers of vowels and particular patterns within each total number preferred over others? Liljencrantz & Lindblom (1972)

and Lindblom (1986) showed that five to seven vowels are preferred over fewer or more vowels because these numbers of vowels divide the vowel space efficiently. Fewer vowels than five are dispreferred because the space can be divided more finely without crowding the vowels so close together that they're likely to be confused, while more vowels than seven are dispreferred because above that number the vowels *are* crowded too closely together. This outcome would be obtained if vowels are required to contrast or disperse sufficiently but not maximally within the limits of the vowel space. Up to a point, height contrasts can multiply among the peripheral vowels without the vowels coming too close together, but central vowels are resorted to when a vowel inventory gets so large enough that a yet finer division of the height continuum among the peripheral vowels pulls adjacent vowels below the threshold for sufficient contrast.

These are all functional explanations. On the one hand, if a vowel inventory is too small, more consonants or longer strings of segments will have to be used to create distinct messages (see also Flemming 2001, 2004). On the other hand, if an inventory is too large or its members are acoustically too close to one another, then distinct messages will be confused with one another.

Other kinds of explanations can also be imagined. For example, languages may prefer to have a back rounded vowel for every front unrounded non-low vowel it has because languages prefer symmetry. This alternative is not implausible because symmetry is not in this instance an abstract, geometric property of a vowel inventory but instead a requirement that a language efficiently use all the possible combinations of distinctive feature values (Ohala 1980; Clements 2003). If a language has a front unrounded vowel of height *n* and it also has a back rounded vowel at that height, then it combines height *n* with both [−back] and [+back] rather than with just one value of this feature. This, too, is plainly a kind of functional explanation, but one concerned with making maximal use of the available resources for contrast between messages rather than with the distinctness of messages.

Sufficient contrast or dispersion is a property of an entire vowel inventory. The sum of the auditory distances between all pairs of vowels is calculated, and then the reciprocal is taken of this sum. The resulting value is larger when the vowels are crowded together in an inventory, so it reflects the energy with which the vowels mutually repel one another, the 'dispersion energy'.

Some vowel qualities, e.g. /i/, occur so often, in inventories of different sizes and compositions, that they appear to be favored *intrinsically* and not just for their auditory distance from other vowels. These vowels may be special because two of their formants are so close in frequency that they merge auditorily into a single, relatively narrow yet intense spectral prominence. Formant frequencies converge when an articulatory change switches the resonating cavities adjacent formants come from (Stevens 1989). The

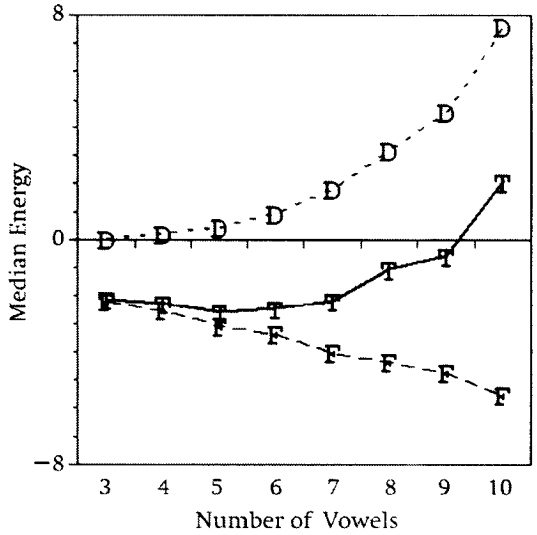


Figure 3: Median focalization energy (F), dispersion energy (D), and total energy (T) for languages with from three to ten vowels.

resulting 'focalization' of acoustic energy makes these vowels more salient than acoustically similar vowels whose formants are farther apart (Schwartz, et al. 1997b). Focalization values are calculated for each vowel as a function of how close adjacent formants are to one another, these values are then summed, and the reciprocal of this sum is taken. The resulting 'focalization energy' is larger for languages with fewer focal vowels. (See Schwartz, et al. 1997b, for the algorithms for calculating these energies.)

The hypothesis tested here is that both dispersion and focalization energies and their sum 'total energy' are larger for less favored inventories. Figure 17.3 shows median focalization, dispersion, and total energy (F, D, and T) for vowel inventories with from three to ten vowels.[11]

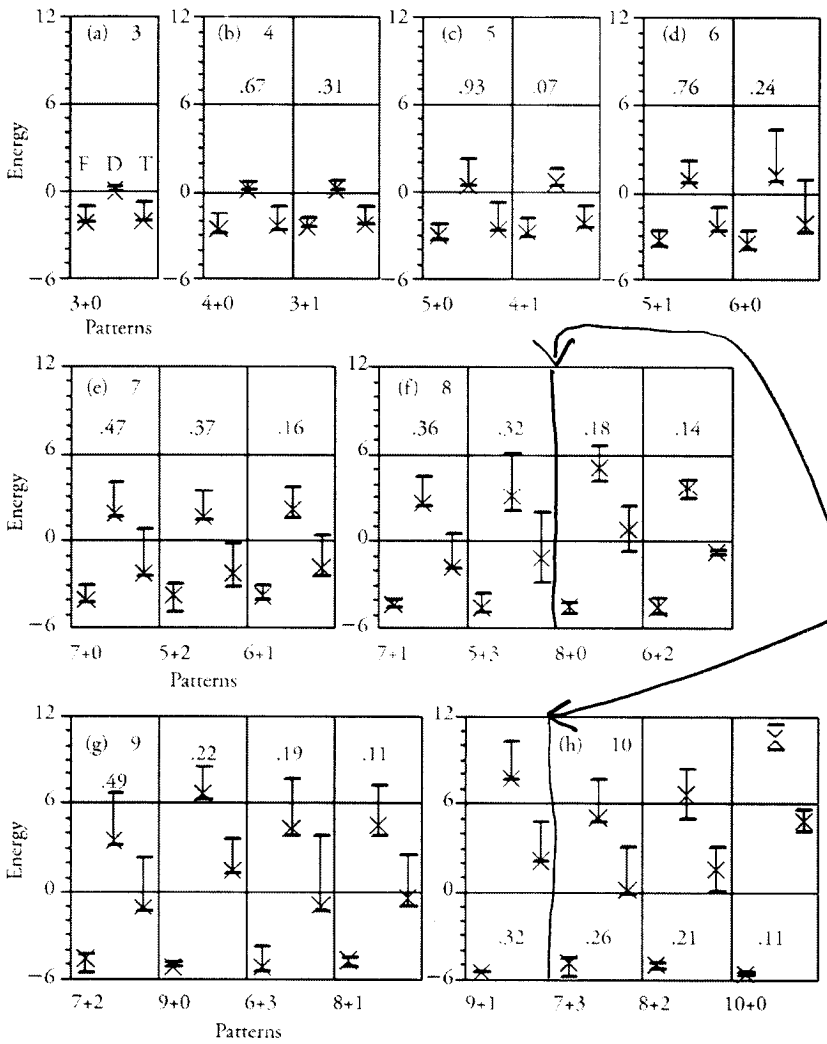
Unsurprisingly, dispersion energy grows with the number of vowels. Inventories with more vowels also include more in which two formants are close together, and focalization energy drops steadily as the number of vowels increases. Up to seven vowels, this steady drop in focalization energy offsets the growth in dispersion energy, and total energy remains relatively unchanged. Indeed, focalization energy drops more than dispersion energy grows between four and five vowels, and total energy is thus somewhat lower in a typical five than four vowel inventory, which may contribute to five vowels being more popular than four. Total energy then grows only modestly from five to seven vowels. However, as the number of vowels in an inventory increases beyond seven, dispersion energy grows much faster than focalization energy drops, and total energy climbs with increasing steepness. This jump in crowdedness probably explains the

markedly lower frequency of languages with eight or more vowels compared to seven or fewer. Total energy also grows less steeply between eight and nine vowels than between seven and eight or nine and ten vowels, which may partly explain why nine vowel inventories are surprisingly popular: they contain more vowels which are made salient by the closeness in frequency of adjacent formants but which are not crowded excessively closely together. (Languages with fewer than five vowels are probably less frequent for a very different reason: they under-use the capacity of the vowel space to distinguish messages reliably from one another.)

The minima and maxima for focalization (F), dispersion (D), and total energy (T) are shown together with the medians in Figure 17.4 for inventories of different sizes and compositions.

For inventories of four to seven vowels (Figures 17.4a–e), the more popular patterns aren't noticeably less energetic. The maximum dispersion and total energies are smaller for the more popular 5+1 pattern than the less popular 6+0 pattern (Figure 17.4d). However, the median for the 6+0 pattern also lies very close to the minimum, which shows that very few languages with the less popular pattern have the higher energy versions of this pattern. Moreover, energy values differ very little between the wildly popular 5+0 pattern and the decidedly unpopular 4+1 pattern (Figure 17.4c), and they differ equally little between the 4+0 vs. 3+1 and 7+0 or 5+2 vs. 6+1 patterns, despite their marked differences in popularity (Figures 17.4b,e). Energies are also not uniformly lower for the more popular patterns in languages with from eight to ten vowels (Figures 17.4f–h).

To see if some relationship might nonetheless be hidden in the data, the proportions with each pattern occurred were correlated with the median focalization, dispersion, and total energies for vowel inventories containing four to ten vowels. If the more popular inventories have lower energies, then all these correlations should be negative. The correlations were significantly negative for dispersion and total energies (dispersion $r(21) = -0.472$, $p = 0.031$; total $r(18) = -0.447$, $p = 0.042$; two-tailed), but, curiously, significantly positive for focalization energy ($r(21) = 0.455$, $p = 0.038$). This correlation turns out positive because focalization energy drops as inventory size increases, and larger inventories are divided into more patterns, each making up a smaller proportion of the total than do the fewer divisions of smaller inventories. The correlations with dispersion or total energy are also influenced by this artifact, but it's hidden in their case because it works in the same direction as the prediction. Accordingly, the correlations were recalculated using only the proportions of the most popular pattern for each inventory size. The results are quite similar: the most popular patterns' proportions correlate negatively with dispersion and total energies (dispersion $r(7) = -0.741$, $p = 0.056$; total $r(7) = -0.752$, $p = 0.051$) and positively with focalization energy ($r(7) = 0.777$, $p = 0.040$), except that the correlations with dispersion and total energy are now only marginally significant. These correlations show that the more popular



there were vertical lines dividing the panels here in the version of the figure originally supplied.

Figure 4: Minimum (bottom whisker), median (X), and maximum (top whisker) values of focalization (F, left), dispersion (D, middle), and total energies (T, right) for inventories of three to ten vowels, broken down by pattern. The arrangement matches that in Figure 17.2. The numbers in each division of a panel are the proportions of languages with that inventory pattern; they correspond to the values that are displayed graphically in Figure 17.2.

patterns are less energetic but also that the energy differences between them and the less popular inventories aren't enormous.

In summary, the popularity of five to seven vowels is well explained by the combined effects of dispersion and focalization energy, but energy differences explain little about why certain patterns are preferred within inventories of a given size.

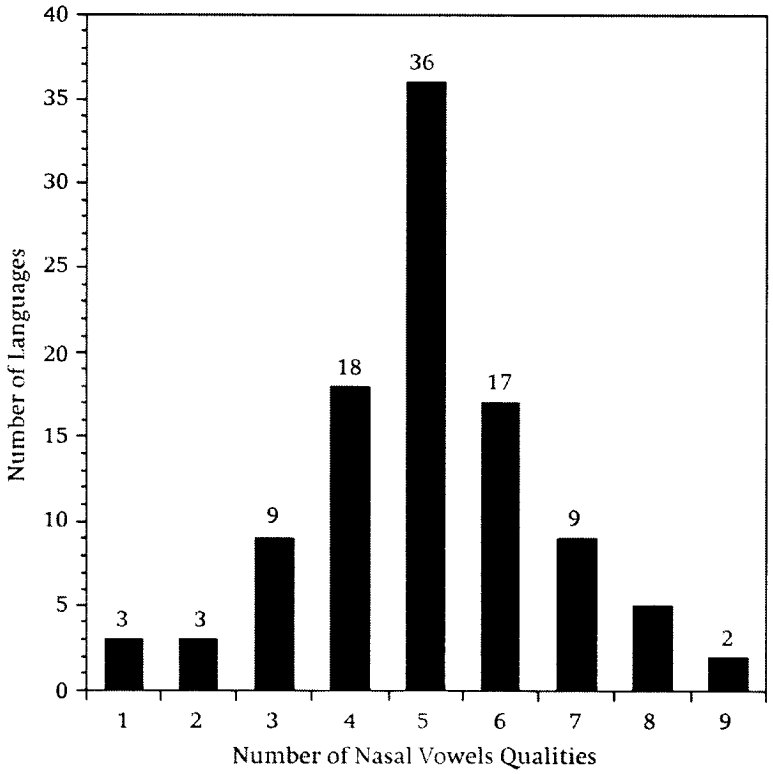


Figure 5: Number of languages with between one and nine nasal vowels.

17.3.2.5 Nasal vowels

Nasalization is the only property other than length that distinguishes vowels of the same quality in more than a very few languages. 102 nasal vowel inventories were extracted from UPSID in exactly the same way as the oral inventories. Nasal inventories are structured much like oral inventories, except they're often smaller. Figure 17.5 shows that languages have fewer nasal than oral vowels: one more language has three or four nasal vowels than has six or seven, 6 languages have fewer than three nasal vowels, and no languages have more than nine (cf. Figure 17.1).

Nasal vowels never occur in an inventory without oral vowels. Though their presence unequivocally implies the presence of oral vowels, is the size and composition of a language's nasal inventory otherwise related to its oral inventory?

Taking up size first, a little over half the languages with nasal vowels, 53 of 102, have fewer nasal than oral vowels in their inventories, and none have more nasal than oral vowels. The languages with fewer nasal than oral vowels have on average 2–3 fewer,[12] and some as many as 6 fewer.

What vowels are missing in the nasal inventories that are found in the corresponding oral inventories? In two languages, Zoque and Cherokee, all the oral vowel qualities are missing from the nasal inventories, which each consist of a single central nasal vowel. Otherwise, one or more mid nasal vowels are missing in 41 languages ("gutless" inventories), one or more high nasal vowels are missing in 20 languages ("headless"), and one or more low nasal vowels are missing in just 6 languages ("footless").[13] Senadi exemplifies the gutless type, with oral /i ɐ̯ ɛ a ɔ o̯ u/ vs. nasal /ĩ ɛ̃ ã ɔ̃ ũ/, Amuzgo is headless, with oral /i e æ a o u/ vs. nasal /ẽ æ̃ ã õ̞ õ/, and Chatino shows what the rare footless type is like, with oral /i e a o u/ vs nasal /ĩ ẽ õ ũ/. Some languages lack nasal counterparts to their oral vowels at more than one these three height divisions: of the 6 footless languages, 2 are also headless, 1 is also gutless, and 1 is also headless and gutless, while 12 of the 20 headless languages are also gutless. In short, nasalization reduces height contrasts, and it does so most often by eliminating mid vowels.

D (#313 in IPA)

Why should it do so? The answer lies in the perceptual consequences of acoustically coupling the nasal to the oral cavity. Coupling adds pairs of poles and zeroes to the poles produced in the oral cavity. The lowest nasal pole (N1) and zero (Z1) occur close to the lowest oral pole (F1) and change both the center of gravity and the bandwidth of this lowest spectral prominence.

N1 is below F1 when the F1 is high and lowers the prominence's center of gravity, but N1 is above F1 when F1 is low, and raises the prominence's center of gravity. Z1 is just above N1. When N1 is below F1, Z1 is likely to coincide with F1 and attenuate it. This attenuation also lowers the center of gravity of the lowest spectral prominence. When both N1 and Z1 are above F1, the center of gravity is instead likely to be raised. Lowering the center of gravity makes the vowel sound higher, raising it makes the vowel sound lower. Headless inventories such as Amuzgo's may be more common than footless inventories such as Chatino's because adding N1 and Z1 more often raises than lowers the lowest spectral prominence's center of gravity.

N1 and Z1 also increase the bandwidth of this lowest spectral prominence, which may make the vowel sound lower. What's probably more important perceptually is that a broader bandwidth makes it harder to detect this prominence's center of gravity and thus to determine the vowel's height. Gutless nasal inventories such as Senadi's may be most common simply because fine distinctions in height between mid vowels or between mid and high or low vowels are made very hard to detect by this bandwidth increase. The perceived centers of gravity differ enough between the remaining high and low vowels that they're preserved. (See also Wright 1986 where it's shown that nasal vowels are perceptually closer to one another than their oral counterparts.)

Perceptual results reported in Kingston (1991), Kingston & Macmillan (1995), and Macmillan, et al. (1999) add to the explanation of gutless inventories' greater frequency. Listeners in these studies identified and discriminated vowels in which vowel height and nasalization were

manipulated independently. They were more likely to identify a vowel as high when it was more nasalized, and more likely to identify a vowel as oral when it was lower. Listeners were also consistently better at discriminating a higher, more nasalized vowel from a lower, less nasalized one than at discriminating a higher, less nasalized vowel from a lower, more nasalized one.[14] Both results show that height and nasalization integrate perceptually, and their integration disfavors intermediate percepts for both height and nasalization.

The perceptual integration of nasalization and height predicts incorrectly that low nasal vowels should often denasalize because a lower vowel is more likely to be identified as oral. Two factors keep low vowels nasalized. First, the soft palate is actually permitted to lower more in lower nasalized vowels (Clumeck 1976; Bell-Berti, Baer, Harris, & Niimi 1979; Al-Bamerni 1983; Henderson 1984) and is actively kept high in higher vowels by contracting the levator palatini and relaxing the palatoglossus (Moll & Shriner 1967; Lubker 1968; Fritzell 1969; Lubker, Fritzell, & Lindqvist 1970; Bell-Berti 1976; Kuehn, Folkins, & Cutting 1982; Henderson 1984). Second, low vowels are longer than higher vowels, apparently because the jaw must lower more (Lehiste 1970; Westbury & Keating 1980), and even light nasalization is easier to detect when the vowel lasts longer (Whalen & Beddor 1989; Hajek 1997). Indeed, nasal vowels of a given height are often longer than the corresponding oral vowels (Whalen & Beddor 1989). Whether a low vowel lasts longer merely because the jaw moves slowly or it is also deliberately prolonged, its greater duration compensates for their height's reducing the perceptibility of nasalization.

Once again, competing factors trade off, delicately: the broadening of the lowest spectral prominence's bandwidth obscures its center of gravity, the integration of vowel height with nasalization discourages mid percepts, while the lower soft palate and greater duration of low vowels ensures they remain nasalized. The next section portrays the consequences for the contents of reduced vowel inventories of having to shorten a vowel. Rather than dispersing vowels in terms of height, shortening compresses them upward.

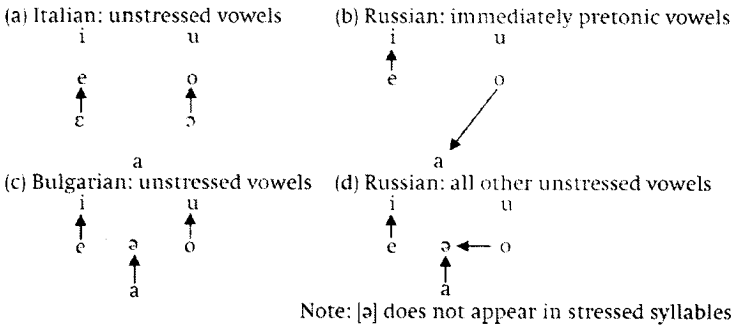
17.3.2.6 Vowel reduction

In many languages, fewer vowels contrast in unstressed than stressed syllables. The proper characterization of unstressed vowel reduction has raised fundamental questions about how the phonetics influences phonology. At least three proposals can be distinguished in the recent literature, Crosswhite's (2001, 2004), Barnes's (2002), and Flemming's (2001, 2004, submitted). The proposals agree that vowel contrasts are reduced in unstressed syllables because these syllables are shorter than stressed syllables, but they disagree as to how.

Crosswhite distinguishes contrast-enhancing reduction, in Italian (1a) and immediately pre-tonic syllables in Standard Russian (1b), from

prominence-reducing reduction, in (Eastern)[15] Bulgarian (1c) and all other unstressed syllables in Standard Russian (1d). Each chart shows the vowels in stressed syllables (except see note in (1d)) and the arrows indicate what they reduce to when they are unstressed (1a,c), or in Russian immediately pretonic (1b) vs. all other positions (1d).

(1)



Contrasts are enhanced by reduction in Italian and in immediately pretonic syllables in Standard Russian because the neutralization of height contrasts involving mid vowels leaves the remaining contrasting vowels farther apart. Contrasts are enhanced in unstressed syllables because their short duration makes it hard to maintain small differences in vowel height, particularly for vowels that aren't at the corners of the vowel space. Crosswhite formalizes this result as a phonological constraint licensing short vowels only at extreme heights.

Prominence is reduced in Bulgarian and in all other unstressed syllables in Standard Russian because the low vowel is raised, in both cases to a mid central unrounded vowel with the quality of [ə]. Raising lowers the vowel's F1 and shortens it – higher vowels are shorter than lower ones (Lehiste 1970). Both these changes reduce the vowel's overall intensity and presumably its prominence; in doing so, they make the vowel more compatible with its prosodically weak position. The shorter duration of unstressed syllables is the effect of reduction when prominence is reduced rather than its cause, as when contrasts are enhanced. Crosswhite formalizes this compatibility requirement in a scale that values higher vowels in unstressed syllables more than lower ones.

Because the high vowels remain unchanged and mid vowels rise in both kinds of reduction, they are distinguished by whether the low vowel remains low – contrast-enhancing – or is raised – prominence-reducing.

Crosswhite's analyses of both kinds are explicitly functional: contrast-enhancing reduction maintains only the vowel height contrasts whose members can be reliably distinguished in the short span of an unstressed syllable, and prominence-reducing reduction ensures that unstressed syllables

aren't so prominent that they're mistaken for stressed ones. This functionalism is, moreover, built explicitly into the phonological formalizations.

Neither Barnes (2002) nor Flemming (2001, 2004, submitted) distinguish between two kinds of reduction, and both treat the shortening of unstressed vowels as reduction's *primum mobile*. Both rely on Lindblom's (1963) finding that the F1 frequencies of Swedish non-high vowels decrease with vowel duration, because speakers undershoot the non-high vowel height targets when there's too little time to reach them. Shortening causes undershoot because speakers don't speed up articulatory movements, particularly of the massive jaw,[16] to reach those targets in the time available. The result is that all the non-high vowels are raised, compressing the vowel space upward from bottom. This outcome resembles Crosswhite's prominence-reducing reduction but both reverses its *explicanda* and *explicandum* and loses its functional motivation. Instead of speakers raising a vowel to lower its F1's frequency, shorten it, and thereby reduce its intensity and prominence, the vowel is raised, more or less automatically, because it's shortened.

Both Barnes and Flemming also argue that there may be no contrast-enhancing reduction. They cite instrumental studies of Italian (Farnetani & Vayra 1991; Albano Leoni, Caputo, Cerrato, Cutugno, Maturi, & Savy 1995) which show that in unstressed syllables the low vowel [a] is realized with a considerably lower F1 frequency, i.e. as a higher vowel, perhaps [ɐ]. For Standard Russian, Barnes shows that the low vowel produced by reduction in immediately pre-tonic syllables does not have a categorically different quality from that produced in other unstressed syllables. Instead, that vowel remains long enough that speakers have the time needed to lower the jaw and tongue and raise the F1's frequency to a value that sounds like [a]. Reduced low vowels in other unstressed syllables are usually shorter and this target is often undershot as a result. Thus, in both Italian and Standard Russian, the phonetic evidence indicates that the low vowel is raised when reduced, as in so-called prominence-reducing reduction, rather than remaining low as would be expected if reduction were contrast-enhancing.

Here, Barnes and Flemming part ways. Barnes argues that the undershooting that occurs when a vowel is shortened in an unstressed syllable can be phonologized, as categorical alternations between the vowel qualities in stressed syllables and the higher vowel qualities heard in unstressed ones. Once phonologization occurs, reduction is no longer governed by the phonetic constraints that originally motivated it, and the vowels that participate in the alternation may freely undergo further sound changes (see the end of Section 17.3.3.2 below for discussion of this claim).

Unlike Barnes, Flemming does build the phonetic motivation for vowel reduction into his phonological account. He uses constraints on what contrasts may occur in a language in place of Crosswhite's licensing or markedness constraints limiting the circumstances in which individual segments may occur. Two kinds of constraints regulate contrasts: the first

requires that contrasting sounds be some minimal distance apart within the phonetic space occupied by the sounds (MINDIST=n) and the second requires that the number of contrasts be maximized (MAXCON). These constraints obviously conflict with one another, as requiring that contrasting sounds be far apart limits the number of possible contrasts, while requiring a large number of contrasts forces contrasting sounds close together. This conflict is resolved by ranking the contrast maximization constraint relative to the minimal distance constraints.

Imagine the range of vowel heights is divided into seven, equally spaced heights: (1) high [i u], (2) lowered high [ɪ ʊ], (3) raised mid [e̝ o̝], (4) mid [e o], (5) lowered mid [ɛ ɔ], (6) raised low [æ ɐ ɐ̞], and (7) low [a, a, ɑ]. In a language with the seven vowels /i, u, e̝, o̝, e/o, a/ found in Italian stressed syllables, the constraint requiring that contrasting vowels differ by at least two steps (MINDIST=2) is ranked immediately above the constraint requiring that the number of contrasts be maximized. MAXCON is a positive requirement rather than a prohibition, and a ✔ is listed in tableau (2) for each contrasting category:

close up
ɛ (303 in IPA) ɔ (306)

(2)

	MINDIST=1	MINDIST=2	MAXCON	MINDIST=3	MINDIST=4
(a) i:a			✔✔!		
(b) i:e:a			✔✔✔!		*(i:e)*(e:a)
☞(c) i:e̝:ɛ:a			✔✔✔✔	*(i:e̝)*(e̝ɛ) *(ɛ:a)	*(i:e)*(e̝ɛ) *(ɛ:a)
(d) i:ɪ:e:ɛ:a		*(i:ɪ)!*(e:ɛ)	✔✔✔✔✔	*(i:ɪ)*(ɪ:e) *(e:ɛ)*(ɛ:a)	*(i:ɪ)*(ɪ:e) *(e:ɛ)*(ɛ:a)

To reduce this inventory to the five vowels found in Italian's unstressed syllables, /i u e̝ o̝ ɐ/, Flemming (2004) adds a constraint prohibiting short low vowels (*SHORTLOW) and ranks it above MINDIST=2 (see (3)):

(3)

	SHORT LOW	MINDIST=2	MAXCON	MINDIST=3
☞ (a) i:e̝:ɐ			✔✔✔!	
(b) i:e:ɛ:a	*(a)!		✔✔✔✔	*(i:e̝)*(e̝:ɛ)*(ɛ:a)
(c) i:e̝:ɛ:ɐ		*(ɛ:ɐ)!	✔✔✔✔	*(i:e̝)*(e̝:ɛ)*(ɛ:ɐ)

This new constraint rules out the candidate with [a] (3b), while MINDIST=2 rules out the candidate which retains [ɛ] (3c), because the raised low vowel [ɐ] is only one step away from it. What's left as the optimal candidate has only three heights and a raised low vowel [ɐ] (3a).

Flemming (submitted) lays out a more explicitly quantitative account of vowel reduction. Rather than requiring that all contrasting vowels be at

least some minimum distance apart, the minimum distance between any pair of vowels now must be maximized. This approach closely resembles that in Section 17.3.2.3, except here an inventory's dispersion energy and goodness is measured by the distance between its closest pair of vowels. Unstressed and stressed vowels remain subject to the same distance requirements, so the minimum distance is the smallest distance found in either inventory. The new account also retains the requirement that the number of contrasts be maximized.

Following Lindblom (1963), the actual formant frequencies of the vowels in the full and reduced inventories are predicted with functions that undershoot their target frequencies as declining exponential functions of the vowel's duration. Since unstressed vowels are shorter than stressed ones, these functions correctly predict that height contrasts neutralize in unstressed vowels and that the unstressed low vowel is raised relative to its stressed counterpart, so long as the functions' other parameters are appropriately set. The parameter settings also ensure that F2 is undershot less than F1, which coincides with observed patterns of vowel reduction. Shortening doesn't affect the extent to which vowels coarticulate with consonants with respect to the articulations that implement the vowels' front-back or rounding contrasts, but it does affect the vowels' height because a consonant's constriction is usually just as close in an unstressed syllable as a stressed syllable, and coarticulated vowels are raised as a consequence.

Flemming's new account invokes the phonetic motivation for and mechanisms of vowel reduction far more directly that his earlier one did. The minimum distance and maximize contrast constraints are now implemented with continuous rather than discrete mathematics, and their ranking or priority is expressed by varying their relative weights. The only explicitly phonological aspect of this account is its outcome: fewer vowels contrast in the reduced inventories. That's an emergent property, not one that's in any way preordained by obedience to some phonological constraint.

17.3.2.7 The phonological consequences of vowel reduction vs. nasalization

Vowel reduction raises vowels by compressing them upward, while nasalization either disperses vowels toward the top and bottom of the vowel space or more rarely lowers them. Their effects differ because the speaker's goals are different.

The goal in pronouncing an unstressed vowel is to produce one that's shorter than a stressed vowel, which leads to the vowel's target being undershot. Shortening is not the goal in a nasal vowel, quite the reverse. Conveying the vowel's height and nasalization instead demands that the vowel last long enough for the spectral modification caused by nasalization to be detected and may even demand that the vowel be prolonged (Whalen & Beddor 1989). Nasalization's need for a longer vowel can even have phonological consequences: contrastive nasalization may only develop on

long or stressed vowels, as in Guaraní (Gregores & Suárez 1967), Copala Trique (Hollenbach 1977), Teke (Hombert 1987), and northern Italian dialects (Hajek 1997), and it may inhibit extreme vowel reduction, as in Brazilian Portuguese (Major 1985).

Mid vowels are eliminated from nasal vowel inventories because nasalization alters and obscures the center of gravity of the vowel's lowest spectral prominence, effects which are exacerbated by the perceptual integration of vowel height and nasalization, which disfavors percepts of intermediate height and nasalization. Low nasal vowels, which integration would otherwise denasalize, remain nasal because they last long enough for listeners to detect whatever nasalization may be present and because speakers actually lower the soft palate more.

17.3.3 Explaining sound change

17.3.3.1 Introduction

Many, perhaps most of the sound changes languages undergo are phonetically motivated. What does this mean? It means that something about how the speaker pronounces the sound that changes, how that sound is transmitted to the listener, or how the listener perceives that sound makes it possible for that sound to be phonologically different at some later time in its history than it was at some earlier time.

In the next section, I describe tonogenesis in Athabaskan as an example of a phonetically motivated sound change. Discussing this sound change is an occasion to evaluate Steriade's (1999b) 'licensing by cue' proposal that contrasts are maintained in those contexts where their phonetic cues are easy to detect and neutralized in contexts where their cues hard to detect – as discussed in Section 17.3.2.5 above, Crosswhite used such a licensing constraint in explain the loss of mid vowels in contrast-enhancing reduction. This example is also evidence that the phonetic motivation for a sound change persists after it's been phonologized, contrary to the claims of Barnes (2002) and Blevins (2004).

17.3.3.2 The phonetics of Athabaskan tonogenesis

In Proto-Athabaskan, glottalic and non-glottalic stops, affricates, nasals, and glides contrasted at the ends as well as the beginnings of stems (Krauss in 2005), but in many present-day members of this family, the stem-final contrast has been replaced by tone in stems ending in stops and affricates (henceforth just 'stops').

The development of tone from an earlier contrast in laryngeal articulations of an adjacent consonant is an extremely common sound change (Hombert, Ohala, & Ewan 1979), particularly in the language families of East and Southeast Asia. It can occur because one of the phonetic correlates of a laryngeal contrast in consonants is differences in the fundamental frequency (F0) of adjacent vowels. These F0 differences become tone contrasts

in the vowels and replace the original laryngeal contrast between the consonants when they lose the other phonetic correlates of the contrast.

Explaining tonogenesis in Athabaskan is complicated by three factors. First, nearly all the tonal Athabaskan languages maintain the contrast between stem-final glottalic and non-glottalic sonorants and between stems ending in glottal stop vs. a vowel. Even so, the same tone appears in stems ending in glottalic sonorants and glottal stop as in stems that once ended in glottalic stops, and the other tone in stems ending in non-glottalic sonorants or a vowel. In these stems, the F0 differences remain synchronically predictable from other properties of the stem-final consonant, and tone doesn't convey the contrast alone.

The second complication is that the tone which developed in stems that ended in glottalic stops only did so when the stem vowel was short. When the vowel was long, the tone appears that otherwise developed in stems that ended in non-glottalic stops. In certain morphological constructions, however, the stem-final stop was spirantized, and the same tone then developed on long vowels as on short ones. This tone also develops on long as well as short vowels in stems ending in glottalic sonorants. Short and long vowels don't contrast before stem-final glottal stop; their modern reflexes uniformly indicate that the vowel is long. Nonetheless, the same tone appears in these stems as in those in which a short vowel preceded a glottalic stop in the protolanguage or those ending in a glottalic sonorant. In short, vowel length doesn't matter if the stem-final consonant was not an oral stop.

Both complications can be accounted for by a difference in the relative timing of laryngeal and oral articulations in stops vs. other manners of articulation (Kingston 1985, in 2005). Their relative timing differs because a stop closure's release differs acoustically from its onset. The brief but intense noise burst that occurs when the stop is released is apparently salient enough that the stop's laryngeal articulation is timed to coincide with it. This timing ensures that different laryngeal articulations modify the burst in characteristic ways that convey their nature to the listener. No comparably salient acoustic event occurs at the closure's onset, or at either the onsets or releases of fricatives and sonorants.[17] Because no similarly salient acoustic event occurs at either the onset or release of the oral constriction in fricatives or sonorants, the timing of laryngeal articulations relative to oral ones is freer. In many languages, however, the laryngeal articulation is timed to coincide with the onset of the oral constriction in these manners of articulation (Kingston 1985, 1990).

Because the laryngeal articulation coincides with the release of the oral constriction in a stop, it is farther from and coarticulates less with a preceding vowel than it would in a fricative or sonorant, at least when the laryngeal articulation coincides with the onset of the constriction in those manners of articulation. Tonogenesis indicates that coarticulation

with a stop was still extensive enough to alter a short vowel's pronunciation if not a long one's. Enough of the vowel would be altered by coarticulating with the nearer laryngeal articulation in a sonorant or fricative to change a long as well as a short vowel.

The timing difference also explains why tone replaced the glottalic:non-glottalic contrast in stem-final stops but merely supplements it in stem-final sonorants. If the stop were not released in some contexts or the release were inaudible, the principal cue to the consonant's identity would become the acoustic effects of its coarticulation with the preceding vowel and not any properties heard during the consonant itself. The absence or inaudibility of the release in a sonorant would be of little consequence for conveying its laryngeal articulation, particularly if that articulation coincides with the onset of the oral constriction. The principal cues to that articulation are already its coarticulatory effects on the preceding vowel, so there's little reason to expect them to change or shift off the consonant.

For the contrast to shift to the vowel from the consonant, the listener has to misinterpret the coarticulatory effects of the consonant's laryngeal articulation as the speaker intending to alter the vowel (Ohala 1981). The listener may be inclined to do so if other evidence that these effects are properties of the consonant is frequently weak or missing.

The third complication is perhaps the most intriguing: some of the present-day daughter languages have high tones in stems that end in a glottalic consonant and low tones elsewhere, while others have low tones in such stems and high tones elsewhere. One of these developments could be original and the other as a reversal, but in Kingston (in 2005) I show that it's actually possible to get both high and low tone directly from different pronunciations of the glottalic consonants.[18] Glottalic consonants are distinguished from non-glottalic ones by a constriction of the glottis that is tight enough to curtail or even cut off air flow through the glottis. The glottis is closed by contracting the interarytenoid and lateral cricoarytenoid muscles while relaxing the posterior cricoarytenoid muscles and the constriction is tightened by the forceful contraction of the thyroarytenoid muscles, which stiffens the inner bodies of the vocal folds and causes the folds to press firmly against one another. If this is all the speaker does, the voice quality of adjacent vowels is creaky and its F0 is low because the vibrating outer covers of the folds remain slack. However, if the speaker also contracts the cricothyroid muscle at the same time, the folds' outer covers are stretched and the voice quality in the adjacent vowel is tense and its F0 high instead. The available evidence suggests that speakers choose to contract the cricothyroid as well as the thyroarytenoid muscles independently of other choices they make about how to pronounce glottalic consonants (Kingston 1982, 1985; Bird 2002; Wright, Hargus, & Davis 2002).

In this account, speakers choose (1) whether to contract the cricothyroid as well as the thyroarytenoid muscles in pronouncing glottalic consonants, a choice which determines whether high or low tone eventually develops on preceding vowels, (2) not to release stem-final stops or to release them inaudibly such that the only the F0 and voice quality of the preceding vowel are reliable cues to the stops' laryngeal articulations, and (3) to time the laryngeal articulations of sonorants and fricatives so that they coincide with the onset of the oral constriction, and thus noticeably alter the F0 and voice quality in long as well as short preceding vowels. None of these choices are obligatory, even if they are more typical than alternatives. Crucially, listeners also mistakenly interpret the coarticulatory effects of the consonant's laryngeal articulation on the preceding vowel's F0 and voice quality as intentional. This mistake is encouraged if speakers fail to release stops or do so inaudibly and if laryngeal articulations are timed to coincide with the onset of the oral constriction in fricatives and sonorants.

17.3.3.3 Licensing by cue

Laryngeal contrasts are kept and lost from consonants in other languages in similar circumstances to Athabaskan. For example, in Lithuanian and Klamath laryngeal contrasts in obstruents are maintained before sonorants and lost elsewhere. Steriade (1999b) proposes that contrasts are maintained in contexts where the cues to their identity are robust and neutralized where those cues are reduced, obscured, or absent. Laryngeal contrasts are maintained before sonorants because cues to those articulations in the consonant's release and in the transition to the following sonorant are robust in that context. They are neutralized before obstruents and word-finally because the release cues are absent in these contexts, and the cues in the transitions from preceding vowels are less robust. The Athabaskan case is quite similar: the glottalic:non-glottalic contrast is kept in stem-initial stops because they reliably precede vowels, and lost stem-finally, where they do not. Stem-final stops may even have been unreleased when the contrast was lost from them, inducing listeners at that time to think the consonants themselves weren't different. Moreover, laryngeal contrasts are probably maintained to this day in stem-final sonorants because the cues are timed to occur early enough that they are robustly signalled during the transition from the preceding vowel.

The licensing by cue account of Athabaskan tonogenesis is, however, incomplete: the glottalic:non-glottalic contrast didn't in fact neutralize in stem-final stops, but instead shifted to a tonal contrast on preceding vowels. How were speakers of Athabaskan languages able to keep morphemes distinct whose stem-final consonants once contrasted in their laryngeal articulations while speakers of Lithuanian or Klamath failed to do so? If

the phonetic correlates *available* to act as cues to a particular laryngeal contrast are the same in all languages where that contrast is found, then Lithuanian and Klamath speakers and listeners had at their disposal more or less the same materials to convey these contrasts, among them differences in F0 and voice quality on preceding vowels, as Athabaskan speakers. Yet they failed to use them. The solution to this conundrum lies in the idea that speakers choose how they are going to pronounce a contrast, and therefore which of the available phonetic materials they're going to use.

Licensing by cue falls short because it conceives the phonetics as something that happens to speakers, rather than also conceiving them as actively manipulating the phonetics to meet their communicative needs. Contrasts are certainly more robustly signaled in some contexts that others, but phonetic materials are available for speakers to use to increase the robustness with which they're signaled in other, ostensibly less favorable contexts. The unanswered question then becomes: why do speakers choose to do so in some languages but not others? The answer to this question will probably turn out to be that speakers make this choice when the contrast is lexically or morphological informative and not otherwise. (For further criticism of licensing by cue see Gerfen 2001; Kingston 2002.)

17.3.3.4 Evolutionary phonology

Tonogenesis in Athabaskan also clearly shows that a sound change's phonetic motivation remains active even after the sound change has been phonologized, contrary to the claims of Barnes (2002) and Blevins (2004). In the tonal Athabaskan languages, the tones that appear in stems ending in glottalic sonorants and glottal stop are always the same as those appearing in stems that ended in glottalic stops in the protolanguage, modulo the effects of vowel length. If it were once possible to constrict the glottis in such a way as to either lower or raise F0, then it should still be possible to do either one, and therefore it should have been possible in the subsequent history of a tonal Athabaskan language for its speakers to adopt the pronunciation of glottal constriction that has the opposite effect on F0 and tone in the preceding vowel. The result would be that stems which originally ended in glottalic stops in the protolanguage would have one tone, while those that end today in glottalic sonorants or glottal stop would have the opposite tone. This has never happened. It hasn't because when the sound change was phonologized, the phonetics of the pronunciation of glottal constriction were, too.[19] Doing so has constrained glottalic sonorants and glottal stop to be pronounced in the same way throughout the subsequent history of each tonal Athabaskan language as its own glottalic stops were when the sound change was actuated.

17.4 Implementation

17.4.1 Introduction

In this concluding section, I discuss two examples of how the phonetics implements phonological representations. In the first example, I take up the question of what it means phonetically for a sound to be phonologically marked vs. unmarked. The discussion shows that the pronunciation of the unmarked member of the contrast is more variable and hypo-articulated than that of the marked member(s). Listeners apparently expect this variability and adjust for it. The adjustments they make suggest that the unmarked sound is phonologically unspecified.

It is often proposed that phonetics manipulates gradients, while phonology instead manipulates categories. This distinction is the central issue in the second example where it is extended to differences in how phonetic and phonological constraints are prioritized. One phonetic constraint isn't categorically ranked above or below another in the way phonological constraints are, but the phonetic constraint with higher priority is weighted more heavily in evaluating a possible output's well-formedness. Weight is inherently gradient rather than categorical.

17.4.2 The phonetics of place and markedness

In heterosyllabic sequences of a coronal stop followed by a non-coronal stop in English, e.g. [t.k] or [d.g], the coronal articulation is typically briefer, it may be substantially reduced, even to the point where the tongue tip and blade don't reach the alveolar ridge, and it is often fully overlapped by the following non-coronal articulation (Nolan 1992; Byrd 1996). For some speakers, coronal stops in this context assimilate completely to the following non-coronal, in some or all tokens (Ellis & Hardcastle 2002). When the order of the places of articulation are reversed, the non-coronal isn't shortened, reduced, nor overlapped nearly as much, nor does it assimilate to the coronal.

This articulatory asymmetry is matched by a corresponding perceptual one. Gaskell & Marslen-Wilson (1996) report the results of a cross-modal priming task, in which, relative to a control stimulus, an assimilated pronunciation of a coronal consonant, e.g. *lea[m] bacon*, sped up recognition that a simultaneous visual probe *lean* was a word just as much as did the unassimilated pronunciation, *lea[n] bacon* (Gaskell & Marslen-Wilson 1996). Gow (2002) reports similar results for heavily overlapped but not fully assimilated coronals. However, an assimilated pronunciation of a non-coronal consonant, e.g. *la[ŋ]e goat*, slowed recognition that the visual probe *lame* was a word significantly compared to the unassimilated pronunciation, *la[m]e goat* (Gaskell & Marslen-Wilson 1996). Monitoring for the phoneme beginning the second word, e.g. the /b/ in *lean bacon*, was also facilitated by an assimilated pronunciation of the preceding coronal,

whether the coronal was assimilated (Gaskell & Marslen-Wilson 1998) or only heavily overlapped (Gow 2003). These phoneme monitoring results suggest that the listener parses the non-coronal place information off the assimilated consonant and attributes it to the following non-coronal.

Both Gaskell & Marslen-Wilson and Gow argue that listeners parse the place information like this because they know that coronal stops are extensively overlapped by and even assimilate to following non-coronals. Their results show that when listeners hear, for example, the non-word *lea[m]* before *bacon*, and this non-word would become the word *lea[n]* if its final non-coronal were replaced by a coronal, they infer that the non-coronal place information belongs to the following consonant and that the intended consonant is coronal. They don't infer another non-coronal when they hear the non-word *la[ŋ]e* before *goat* because they have no comparable experience of non-coronals being extensively overlapped by or assimilating to the place of articulation of the following consonant. This interpretation is supported by Gaskell & Marslen-Wilson's (1996, 1998) findings that an assimilated coronal that isn't homorganic with a following non-coronal, e.g. *lea[m] goat*, neither primes recognition of the visual probe *lean* nor facilitates detection of the initial /g/ in the following word. The inferences are blocked in this 'non-viable' assimilation because the non-coronal place of the [m] cannot be parsed onto the following [g].

Coenen, Zwitserlood, & Bölte (2001) report cross-modal priming experiments run with German listeners in which the procedures and results closely resemble those reported by Gaskell & Marslen-Wilson (1996). Lahiri & Reetz (2002) also report the results of cross-modal priming experiments with German listeners, but they presented the primes in isolation, without any following word whose initial consonant might be an assimilation trigger. In the first experiment, the auditory primes were words ending in either a coronal or non-coronal, e.g. *Bahn* 'railway' or *Lärm* 'noise', and non-words made by replacing the final coronal with a non-coronal or vice versa, *Bahm* vs *Lärn*. Both *Bahn* and *Bahm* primed recognition of the related visual probe *Zug* 'train' but only *Lärm* primed *Krach* 'bang, racket'. Although this result can't be attributed to the listeners' actually parsing the non-coronal place information at the end of *Bahm* onto a following homorganic consonant, they may still separate the labial place information from *Bahm* because *Bahn* is sometimes pronounced *Bahm* in front of a word beginning with a bilabial consonant. This alternative is ruled out by the second experiment, where the manipulated consonants were inter-vocalic rather than final, and therefore in a context where there's never a following consonant to assimilate to. Auditory primes were words with medial coronal or non-coronal consonants, e.g. *Düne* 'dune' or *Schramme* 'a scratch', and corresponding non-words, *Düme* or *Schranne*. Both *Düne* and *Düme* primed recognition of the related visual probe *Sand* 'sand' but only *Schramme* primed *Kratzer* 'a scratch'. These results definitively rule out the inferential parsing account proposed by Gaskell & Marslen-Wilson or Gow.

Lahiri & Reetz interpret their results as evidence that coronal place is not specified phonologically, while the labial and dorsal places are specified. When there is phonetic evidence in the signal for a non-coronal place, as in *Bahm* or *Düme*, this evidence doesn't mismatch the stored forms of the words *Bahn* or *Düne*, because the /n/ in these words isn't specified for place, and these words are activated. Because *Bahm* and *Düme* aren't words, this evidence for non-coronal place also doesn't activate any competing words. Phonetic evidence of coronal place, as in *Lärn* or *Schranne*, however, does mismatch the phonological specification for labial place in the words *Lärm* or *Schramme*, which inhibits their activation.

This interpretation doesn't easily handle the failure of the non-viable assimilation in *lea[m] goat* to prime *lean*. The phonetic evidence for the labial place in the [m] wouldn't mismatch the missing place specification of the /n/ in this string any more than in an isolated word. However, in an earlier cross-modal priming study with German listeners where the auditory primes were followed by another word whose initial consonant could be an assimilation trigger (1995), Lahiri obtained priming for non-viably as well as viably-assimilated coronals, i.e. *Bahm* primed *Zug* even when the following word didn't begin with a labial consonant. This result indicates that viability needs to be reexamined.

The articulatory data shows that the unmarked member of a contrast may vary substantially more in its pronunciation than the marked member (s). The perceptual data indicate that listeners can readily tolerate the phonetic effects of the unmarked member's variation, either because they've had long experience of it or because the unmarked member is actually not specified phonologically, and the variation creates no mismatch between the phonetic evidence and the phonological specification.

17.4.3 Categories and gradients

Phonology is commonly thought to deal in categories, while phonetics deals instead in gradients. Keating (1988c), Pierrehumbert (1990), Cohn (1993a), Zsiga (1995), Holst & Nolan (1995), and Nolan, Holst, & Kühnert (1996) explicitly use the distinction between categories and gradients to *define* phenomena as phonological vs. phonetic.

Recently, Zsiga (2000) has extended this use of the distinction between gradients and categories to separate phonetic from phonological constraints. Using acoustic evidence, she shows that in English the end of the coronal gesture in an [s] is overlapped by the palatal gesture of a following [j] across word boundaries. This evidence agrees with palatographic evidence reported in Zsiga (1995), which showed a shift from coronal to palatal contact at the end of an [s] preceding [j]. Starting after the middle of the fricative, the [s]'s coronal articulation gradually blends with the following [j]'s palatal articulation and produces an articulation midway between these two articulations by the end of the fricative. The acoustics

of Russian speakers' pronunciations, however, show that the two articulations don't overlap in [s# #j] sequences, and that in palatalized [s^j], they overlap completely. Even though the coronal and palatal articulations are simultaneous in the Russian speakers' palatalized [s^j], they aren't blended: both coronal and palatal articulations are produced, not an articulation midway between them.

Zsiga proposes that English and Russian differ in the relative priorities of faithfulness-like phonetic constraints requiring the speaker to achieve particular articulatory targets specified in the phonological representation. For English speakers, the requirement to maintain the coronal constriction specified by /s/ gradually gives way over the last half of the fricative to the requirement to reach the following palatal constriction. The two constrictions blend progressively until a constriction is produced midway between the alveolar ridge and the palate by the end of the fricative. For Russian speakers, however, the requirement to maintain the coronal constriction remains a higher priority all the way to the end of the fricative constriction in [s# #j] sequences, as well as in palatalized [s^j], where the coronal articulation is maintained despite complete overlap with the palatal articulation.[20] Because the coronal constraint's priority doesn't change in Russian even when the coronal articulation is completely overlapped by the palatal articulation, while its priority diminishes gradually as a result of overlap in English, Zsiga argues that these phonetic constraints are weighted continuously with respect to one another and not ranked categorically. Because the priority conflict between phonetic constraints is resolved by continuous weighting and not strict ranking, she also argues that phonetic constraint evaluation is autonomous from and follows phonological constraint evaluation.

This sequential model is quite different from that advocated by Steriade (1999b) or Flemming (2004, submitted) in which phonetic constraints are integrated among and even supplant phonological constraints, and where the phonetic constraints are also strictly, i.e. categorically ranked. Their models do not, as far as I know, try to account for phonetic detail to the extent that Zsiga's proposal does, but there is no formal barrier to their doing so. Future research will determine whether phonological and phonetic constraint evaluation are a single, integrated process, as advocated by Steriade and Flemming or instead sequential, as advocated by Zsiga.

17.5 Summary and concluding remarks

I have tried to show here how distinctive features might be defined, how phonological patterns might be explained, and how phonological representations might be implemented.

The essential problem that has to be solved in defining distinctive features is that their articulations and acoustics vary so enormously that it's

impossible to identify any articulatory or acoustic property that's essential to defining a feature. This variability can be largely eliminated by moving away from the details of particular phonetic realizations, either toward the articulatory plan for the utterance embodied in Articulatory Phonology's gestures or toward the auditory effects of the signal's acoustic properties. Evidence was reviewed that taken together pointed to the second move as the right one.

Explaining phonological patterns is difficult because they are typically determined by more than one phonetic constraint, as well as by phonological constraints, and these constraints may conflict with one another. The eventual explanation is a description of the resolution of this conflict. It is largely because phonetic explanations are complex in this way that I think no bright line can be drawn between the phonetic and phonological components of a grammar. It is interesting in this connection that many of those who advocate such bright lines (e.g. Hale & Reiss 2000b; Blevins 2004) also reject phonological models in which the surface phonological representation corresponding to a particular underlying representation is selected by applying well-formedness constraints in parallel to all possible surface representations, as in Optimality Theory. Replacing serial derivation by parallel evaluation removes the barrier to phonetic constraints being interspersed among and interacting with phonological constraints. (Zsiga's 2000, proposal, as described in Section 17.4.3, is an obvious exception to this generalization.)

The problem in trying to understand phonetic implementation is actually very similar to that arising in attempts to explain phonological patterns in phonetic terms: phonetic constraints not only regulate how a phonological representation can be realized but also determine at least some of its properties. These properties of the phonological representation emerge out of its implementation in much the same way that the distinctive features emerge out of the solution to the variability problem.

Notes

I could not have written this chapter without the feedback and stimulation I received from the students in the seminar I taught on the phonetics-phonology interface at the University of Massachusetts, Amherst in the fall of 2004: Timothy Beechey, Kathryn Flack, Shigeto Kawahara, Anne-Michelle Tessier, and Matthew Wolf. Finally, I must thank Paul de Lacy for inviting me to write this chapter in the first place; his doing so forced me to take up a task that I had been putting off for some time: working out just what I think the interface is (and isn't) between phonetics and phonology. His reactions to earlier versions were also very helpful in getting the chapter into its final form. In short, I owe many of the virtues of the present chapter, such as they are, to others, particularly to my students. As is customary, I keep the faults, of which there are many, for myself.

1 In Articulatory Phonology, this collection of gestures is actually the phonological representation of the utterance. Browman & Goldstein make this move because it is very difficult to translate an utterance's linguistic representation as a sequence of discrete cognitive categories into its physical realization as continuous and overlapping actions having spatial and temporal extents (Fowler, Rubin, Remez, & Turvey 1980). If the phonological and phonetic representations differ to this extent, it's also hard to see how either could constrain the other, yet they do (Browman & Goldstein 1995).

2 Keating's (1990b, 1996) window model of coarticulation achieves a similar result by specifying spatial and temporal ranges for particular articulators' movements. These ranges' limits ensure that the speaker reaches the goal, but their width permits individual articulators' movements to vary in extent depending on context.

3 Lindblom et al.'s (1979) data show that the vocal tract's configuration does differ substantially elsewhere than at the point of constriction when the bite block prevents jaw movement.

4 The vowel space has a fixed acoustic volume because when the articulators move past certain limits, they impede air flow enough that what was a vowel becomes a fricative.

5 A very different and apparently more successful approach to predicting the contents of vowel inventories of particular sizes is presented by de Boer (2000).

6 51 languages in the sample contrast long with short vowels: 23 distinguish more short than long vowel qualities, 14 distinguish more long than short vowel qualities, and 14 distinguish the same number of qualities in long as short vowels.

7 An otherwise central vowel, e.g. /ɯ/ or /ʉ/, was counted as peripheral if a language lacked a more peripheral vowel at that height or backness. A vowel quality is only central if it contrasts with a minimally different peripheral vowel.

8 These patterns and their frequencies closely resemble those in Table I in Schwartz et al. (1997a), who analyzed an earlier, smaller version of UPSID consisting of 317 languages (Maddieson 1984).

9 Vowels were not literally added to an earlier 5+0 inventory at some time in these languages' histories. This is instead a description of how vowel patterns differ or remain the same when additional vowels are present in an inventory.

10 This generalization holds if /ə/ is treated as a front vowel in the Qawasqar inventory /ə, a, o/. It is at least more front than /o/.

11 Because half the values in the range are below the median and half are above, all the values are on average closer to the median than to any other value, and it is less affected by extreme values than the mean.

12 The average deficit in nasal vowels is 2.45, but a language cannot have a fraction of a vowel.

13 Here 'mid' encompasses the range from lower to higher mid, 'high' includes high and lowered high, and 'low' includes low and raised low.

14 Other studies also report that a higher vowel sounds more nasalized than a lower one for a given degree of nasal-oral cavity coupling (House & Stevens 1956; Lubker 1968; Ohala 1975; Abramson, Nye, Henderson, & Marshall 1981; Benguerel & Lafargue 1981; Stevens, Fant, & Hawkins 1987; Maeda 1993, cf. Lintz & Sherman 1961; Massengill & Bryson 1967; Bream 1968; Ali, Gallagher, Goldstein, & Daniloff 1971).

15 Barnes (2002) restricts the reduction pattern shown in (1c) to Eastern Bulgarian; Crosswhite does not.

16 Pettersson & Wood's (1987a,1987b) cineradiographic study of Bulgarian vowels shows that the jaw but not the tongue undershoots its target in unstressed syllables. The tongue remains lower for the non-high vowels [e o a] than for [i u ə], but the jaw is higher, close to its position in [i u ə], apparently enough to lower F1 and make unstressed [e o a] sound like the vowels just above them.

17 Steriade (1993) discusses other phonological consequences of the acoustic difference between a stop's onset and release.

18 In Kingston (2004, see also Kingston & Solnit 1989; Solnit & Kingston 1988), I show that apparent tone reversals of this kind are widespread and also occur when the historical sources of the tones are an earlier contrast between voiced and voiceless obstruents or between aspirated and unaspirated consonants – the latter include sonorants as well as obstruents. In these cases, too, it may be possible to pronounce the consonants such that they either raise or lower F0.

19 Blevins might appeal to structural analogy as the source of this uniformity. Its influence would be exerted through alternations in the verbs, but it's hard to see how it could be extended to the nouns where few if any helpful alternations occur. This is not to say that analogy has played no role in Athabaskan tonogenesis, but its role is limited to the extension of tonogenesis to other morphemes than stems (Kingston 2005).

20 Zsiga notes that Russian speakers may wish to avoid any blending in [s# #j] because they must keep /s/, /ʃ/, and /s^j/ distinct.