studia grammatica 41

URSULA KLEINHENZ (ED.)

Interfaces in Phonology

THEORIA CUM PRAXI

Akademie Verlag

The Phonology-Phonetics Interface

Patricia A. Keating (Phonetics Laboratory, Department of Linguistics, UCLA)

1. Introduction

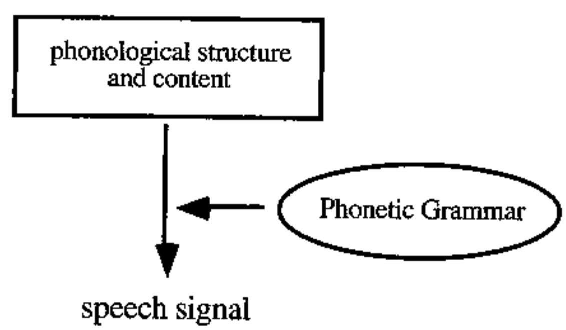
The term "Phonology-Phonetics Interface" can refer to many things. Some topics that come under this rubric which I will not be able to discuss here are listed in (1). A number of recent UCLA dissertations which treat topics of these kinds are included among the references.

(1) Some topics that traditionally fall under "The Phonology-Phonetics Interface":

- motivating the elements of phonological description
  - features, segments, syllables, feet, ...
  - sonority, stress, ...
- explaining typological findings
  - segments, contrasts, inventories
  - synchronic, diachronic processes (after Keating 1988a)

Instead, I will talk about a central question in linguistic phonetics, namely, how relatively abstract phonological structure and content is physically realized — how articulators are moved, how sound is acoustically and therefore auditorily shaped, so that the linguistic structure is made available to a listener. The answer to this question has been conceived of as a component of grammar: a phonetic component that relates the output of the phonology and observable aspects of speech events. This relation has been conceived of as a derivational one, and cast in terms of phonetic rules. These rules are sometimes called rules of ***phonetic implementation.***

(2) Phonetic implementation as the mapping of phonological structure into phonetic output (Following Pierrehumbert 1980).



This conception assumes in turn an answer to another question: Where is the boundary between phonology and phonetics? The basic answer underlying most work in this framework is that *phonology* deals in **discrete symbolic elements**, while *phonetics* deals in **numbers** (on continuous dimensions).

Because phonology deals in symbolic representations, it incorporates certain idealizations from the physical speech signal, including chunking speech in temporal and qualitative dimensions. These chunks can then be represented by symbolic elements, e.g. feature values. Phonology also then confines its operations to operations on and between these symbolic elements. So phonological operations are *categorical*, meaning that category labels are affected.

Because phonetics deals in numbers, it relates these phonological idealizations to speech. It undoes chunking in time to deal in real (or at least notional) time; it undoes chunking in qualities to deal in continuous dimensions. Gradual changes over time along quality dimensions are said to be ***gradient***.

These different properties associated phonological and phonetic representations and processes are summarized in (3). Later I will give examples of phonetic effects that should help make this distinction clearer.

(3) *Phonology*
symbolic representations
allow idealizations of
  temporal chunking (segmentation)
  qualitative categorization (labels)
  timelessness
rules manipulate
  features and feature values
  associations
thus phonological rules can be category changing, produce static changes over whole segments; can be lexical/cyclic

Phonetics
physical representations
continuous in time and space
internal temporal structure
  allows overlap
quantitative values on multiple independent dimensions
rules interpret feature values in time and space, can be gradient

(after Keating 1988b, Cohn 1990)

2. Some Models

Autosegmental/CV Phonology was an important development from the point of view of relating phonology and phonetics. Goldsmith (1976)'s break with strict segmentation was crucial because speech characteristics are not always chunked into neat segments. Strictly segmental SPE phonological representations had made it harder to bridge this gap conceptually.

The traditional class of models of this relationship (which include many speech synthesizers) is summarized under (4) and can be described as providing ***targets***, and ***interpolations*** between targets. Features (or segments) specify "targets" that the human articulators or a computer synthesizer aim at, moving, or "interpolating", from target to target. These can be stated in articulatory or acoustic or auditory domains.

(4) **Targets and Interpolation**

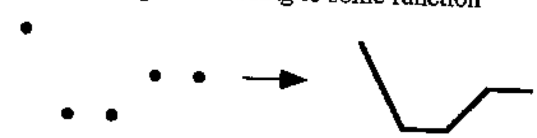
Step 1: Relate a feature to one or more parameters in the domain (e.g. articulatory)

e.g. [Nasal] --> velic opening
e.g. [Voice] --> glottal configuration AND other parameters that facilitate or prevent vocal cord vibration

Step 2: Interpret value of feature as value(s) along parameter(s)

e.g. [+nasal] --> some amount of velic opening over some time interval
e.g. [+voice]--> vocal cords appropriately positioned and tensed for vibration, (for a stop) oral tract walls lax

Step 3: Connect up the targets according to some function



Result: A plan for each articulator to follow, or a predicted acoustic output .

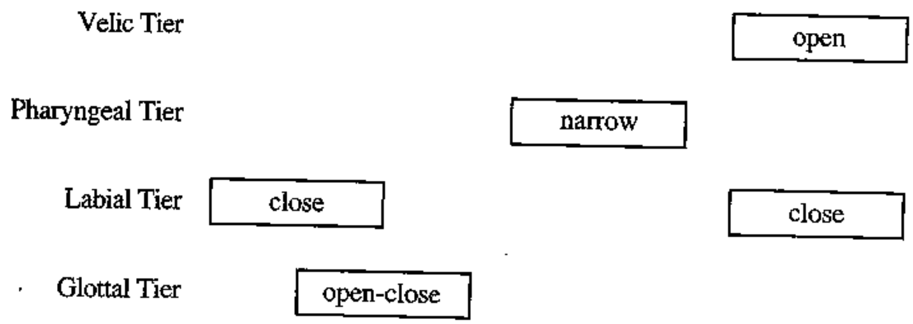
The first two steps result in a sequence of targets along each dimension. The third step connects the targets in continuous space, resulting in a plan for an articulator to follow, or a predicted acoustic contour. The important point is that conceptually, the targets are what realize the feature values, and they are prior to the interpolation function. Two instances of such models are the intonational model of Pierrehumbert 1980 (and subsequent work) and the segmental models in my own work (and subsequent).

A similar way of making this distinction has been proposed by Hertz (e.g. Hertz 1991; Clements and Hertz, ms). The acoustic speech stream is segmented into *phones*, which are quasi-steady-state targets, and *transitions* which are the interpolated intervals between phones.

Targets on each physical dimension are completely autosegmentalized. For example, the different articulations for a single segment can begin and end at different times. This relatively free alignment results in overlaps, where part of an articulation from one segment occurs at the same time as an articulation from a different segment. This overlap is traditionally called co-articulation. It needs to be stressed that this is something we control when we speak — articulations can be more or less overlapped, and languages differ in this respect. Target models in general have tended to focus more on the individual dimensions than on their inter-alignment.

In contrast, the Articulatory Phonology model of Browman & Goldstein (e.g. 1992) focuses more on these inter-alignments, or ***phasings***. This theory also conceives of events on the individual dimensions differently. In a Targets and Interpolation model, we think of the targets as primary and the movements as at best secondary. In contrast, we could think of the movements towards and away from the targets, and their durations, as potentially equally important. In Articulatory Phonology, the phonological primitives are not features, or segments that they can belong to, but abstract articulatory gestures. Within words, gestures are specified for their durations so that the whole time-course is primary. When targets are represented, they are represented as having notional durations. The example frequently given by Browman & Goldstein is schematized in (5), which shows the targets of articulatory gestures in the initial (presumably lexical) representation of the word "palm".

(5) Schematic gestural score for "palm" [pʰɑm] (after Browman & Goldstein 1990)

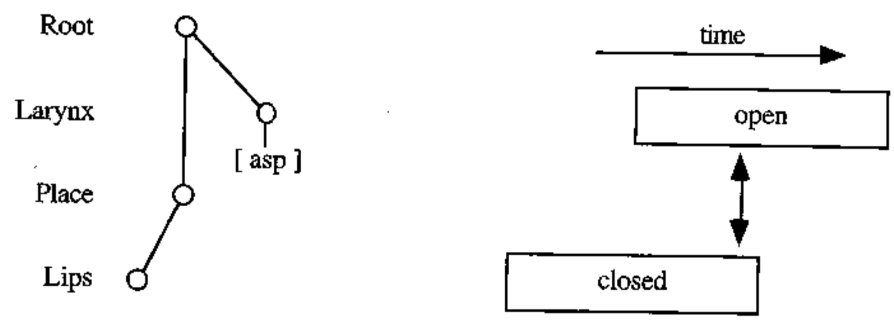


In this schematic, time runs from left to right and articulatory tiers run from top to bottom. The first event is a labial closure (made by the lips); this is closely followed by glottal spreading, indicated as a sequence of opening and closing of the vocal cords — i.e. assuming the velum has been closed, the two together make an aspirated [pʰ] Next comes a constriction in the pharynx (for a low back vowel). The last events, which are indicated as simultaneous, are another labial closure and a velic opening — i.e. nasal [m]. Note that there is no segmentation *per se*. The [ɑ] looks like a segment because it has only one gesture specified; and the [m] looks like a segment because its two gestures are simultaneous. But the [pʰ] looks like a segment only insofar as its two gestures are associated; they are not simultaneous. Note also that this timing relation says that the aspiration of this [pʰ] is lexically specified; it is not derived allophonically.

Articulatory Phonology aims for a closer relation between phonetics and phonology by making phonological representations more like phonetic ones, using the same primitives (the gestures) and putting a fair amount of predictable information into lexical representations. Nonetheless, even this theory needs phonetic implementation just as Targets and Interpolations models do, in the form of postlexical adjustments to the pre-specified gestures and alignments, and a decision mechanism for allocating gestures to articulators (called Task Dynamics). For example, the two gestures of the [m] in (5) are shown as simultaneous. Yet this [m] is in coda position, where a velic opening gesture comes before a labial gesture (Krakow 1989). Therefore presumably some realignment occurs between these gestures from what is represented here.

Compare this approach with Zsiga 1993: she treats Articulatory Phonology not as a combination of phonology and phonetics, but as a model of the phonetic component that is an alternative to Targets & Interpolation models. Zsiga maintains the distinction between categorical phonological and gradient phonetic representations and rules that are central to the Targets and Interpolation framework, but transforms phonological feature values into articulatory gestures, in a kind of Articulatory Phonetics. An example from her dissertation is shown in (6). This is an appealing approach but there are some relations between features and gestures that are more complex than this model is designed to handle, because it assumes that every feature maps to one gesture.

(6) Time-aligned gestural score (right) as a phonetic elaboration of timeless phonological features (left; after Zsiga 1993)



Another model which takes something like traditional phonological features as the input to phonetic implementation, and which takes dynamic gestures rather than targets to be the phonetic primitives, is Fujimura's C/D model (e.g. Fujimura 1994). This model is more concerned with questions of prosodically-conditioned and speaker-conditioned timing and scaling, than with the distinction between target models and non-target models.

Rather than go into the differences among these models, I would like to make a more general point more germane to the interface between phonetics and phonology. A result of work under both kinds of phonetic models is the following: what has been taken to be "phonological" assimilation can be phonetic, by which I mean gradient rather than categorical. Many examples are available; here I will discuss four.

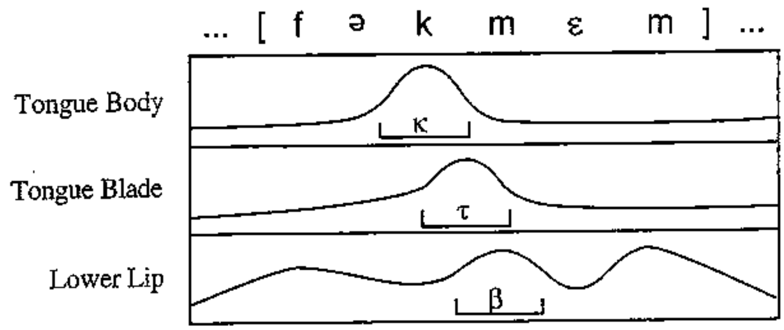
3. Phonetic Assimilation

3.1. When Different Articulations Compete for Some Interval of Time

3.1.1. Overlap in English

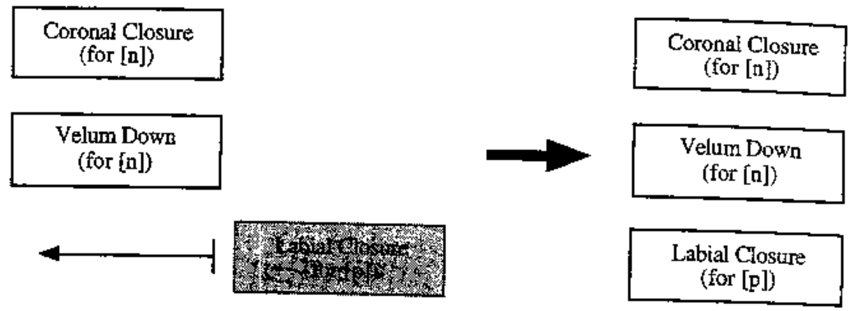
In English, and in many languages, consonants in clusters overlap in time, and the faster we speak the more this is so (discussed at length by Byrd 1994). Browman and Goldstein used this fact to explain the optional deletion of coronal stops between two other consonants. Their now-famous example was the phrase "perfect_memory" in which no [t] is heard. Usually if no /t/ is heard it would be described as having been deleted. It might be expected that as a result of deletion any features of the [t] that might provide a phonetic target are absent. The phonetic data presented by Browman and Goldstein bearing on this phonological deletion analysis are schematized in (7). These are records of the movements of small pellets attached to articulators (listed on the left of picture) during a token of "perfect memory" transcribed without [t]. (Browman and Goldstein use Greek letters for gestures, here kappa, tau, beta for three active aticulators.) This picture shows that there is a clear tongue blade articulations for a [t], as well as for the [k] and [m] also. The key thing is not the presence or absence of a [t] articulation, but its timing relative to the other two stop articulations. The [t] starts just after [k] and ends during [m]. The [t] is completely hidden because of the temporal overlap of the different articulations. It is deleted only in the sense that it is inaudible. Browman and Goldstein suggested that this is the typical situation in deletion.

(7) Coronal place assimilation (after Browman & Goldstein 1990)



They then suggested that the same is true of some place assimilations, like "seven plus" where the /n/ appears as [m]. A schematic of this is shown in (8). The /p/'s labial articulation overlaps with the /n/'s coronal articulation, so that even if the /n/ is not reduced, its lingual articulation is not heard. On the other hand, the nasalization for the /n/, which now also overlaps with the /p/, is not rendered inaudible by this overlap, but instead makes part of the /p/ an [m].

(8) Schematic of n --> m / __ p, as in "seven plus"

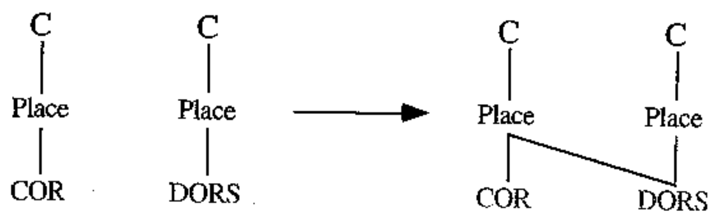


Browman & Goldstein did not actually observe this timing pattern, and since then various people, most recently J. Jun at UCLA (J. Jun 1995), have shown that this kind of assimilation, in which both articulations are intact, is probably pretty rare. Nonetheless, the point Browman & Goldstein made still holds in principle: that the same kind of event, overlap, could lead to apparent deletion in one case and apparent assimilation in another.

Hayes (1992) has discussed how such cases may be treated as phonological spreading: the Place node of the last consonant spreads onto the preceding consonant without delinking any of the preceding consonant's own features, as reproduced in (9). This is phonological overlap, as for a double-articulation or "complex segment".

(9) Phonological overlap (after Hayes 1992):

"late calls": /tk/ --> $\left[\left\{ \begin{matrix} k \\ t \end{matrix} \right\} k \right]$



So the point is not that articulatory overlap cannot be represented phonologically. The point is Browman and Goldstein's suggestion that the degree of overlap depends on rate or casualness of speech, so that the resulting assimilation occurs more-or-less, not all-or-none. It's not simply that the spreading is optional but more likely to occur in fast speech; instead the overlap usually does occur, but to different degrees in different situations, only the most extreme of which result in auditory deletion or assimilation. Phonological spreading is not generally taken to operate gradiently in this way, whereas phonetic implementation must.

3.1.2. Korean Lenis Stop Voicing

In Korean, lenis stops are voiced when they are between vowels, as could be crudely described by the rule in (10). S. Jun's example sentence includes two examples, a word-medial and a cross-word case.

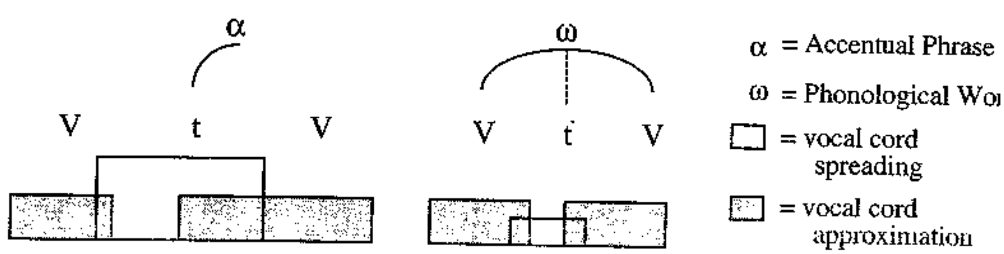
(10) Lenis stop voicing (example from S. Jun 1993) as a phonological rule

Example: [jəmega#darannja], *Was the FRUIT sweet?*

Phonological Rule: [-spread, -constricted] --> [+voice] / V __ V

The rule as written is assimilatory, suggesting that the glottal configuration normally specified by a lenis stop is given up in favor of the glottal configuration of the surrounding sonorants, so that there is one target through all three segments. Instead, Jun shows that the voicing is a function of the relative phonetic duration of the consonant. Short lenis consonants next to long voiced segments voice. Her interpretation of this is as follows: a lenis stop which is voiceless is known to involve some opening of the vocal cords — not as much as for an aspirated stop, but enough to prevent voicing (sometimes enough to make the vowel breathy). This opening requires some minimum amount of time; if there is too little time, then the opening won't be enough to prevent voicing, and the consonant will be voiced. When will there be too little time? When the consonant itself is short, because of its prosodic position, and when the segments on either side of the stop involve glottal approximation, not opening. The glottal targets cannot all be achieved in the available time, and they compete. The voiced segments then win out over the lenis consonant if that consonant is short. Jun's schematic of this situation, in terms of articulatory gestures, is reproduced in (11).

(11) Lenis stop voicing as a phonetic effect



There are three glottal gestures (for V, C, V), in one case initial in the Accentual Phrase, in the other case not. The glottal gesture of the consonant is longer in Accentual Phrase initial position than it is when internal to the Word. In the former case, but not the latter, vocal cord spreading thus leads to a clearly audible gesture. It turns out that reduction of the consonant's glottal gesture is gradient, and as a result, even among the voiced consonants, some will be more voiced than others. So the assimilation is not categorical, in the sense that the glottal articulation for the target consonant is not necessarily the same as for the triggering voiced segments. It is simply that the listener's perception of the result is largely categorical.

3.1.3 The Role of Speech Rate in Such Cases

Both of these cases are gradient in that articulation depends on rate of speech. As rate of speech increases, articulations overlap more. In this voicing example, different articulatory targets on a single articulator come too close together, so one of them gets reduced in degree. In the first example, different articulatory targets on different articulators overlap, so one gets hidden even if it is not reduced. In both examples, an articulation can *sound* as if it has been entirely deleted.

3.2 Assimilation of Unspecified Features

Next, we will consider two examples in which the segment targeted by assimilation is unspecified and the trigger context is specified.

3.2.1 Contextual Nasalization of Vowels

In many languages vowels are described as nasalized before nasal consonants. For example, the English word "bean" has a nasalized vowel [i]. A phonological rule of assimilation is shown in (12).

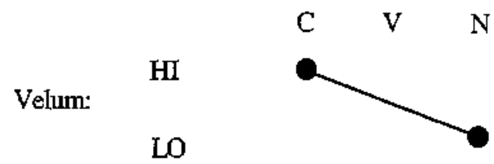
(12) Vowel nasalization as a phonological assimilation

Phonological Rule: C V N

[+nasal]

Cohn (1990, 1993) investigated this in detail and showed that airflow from the nose in such words begins at vowel onset and builds up gradually over the vowel, indicating that the velum opens over the course of the vowel (as X-ray studies have also shown). The nasalization of these vowels looks very different from phonemic vowel nasalization in other languages, in which the vowel is nasal over its whole time interval. That kind of nasalization is seen in English between two nasals, as in "meany". On the other hand, "bean" looks very much like the mirror image of the nasalization of English vowels after nasal consonants, as in "need". If featural spreading is taken to occur only before nasal consonants, as in the typical rule, then the implication is that "bean" and "mean" are alike, but "need" is different. The phonetic implementation will then have to group these otherwise, because phonetically "bean" and "need" are alike and "mean" is different. The phonetic implementation makes more sense if instead we say that the vowel has no target of its own, and its degree and timecourse of nasalization are determined by interpolation from the consonants. This is schematized in (13): an oral consonant specifies a raised velum, the vowel does not care, and a nasal consonant specifies an open velum — so the vowel shows a C-to-N or N-to-C transition.

(13) Phonetic interpretation of vowel nasalization
Phonetic data: velum opening begins at vowel onset



A different view of the facts and their interpretation has been put forward by Bell-Berti, Boyce, and colleagues (Bell-Berti and Krakow 1991, Boyce et al. 1991). They present English examples in which non-nasal segments appear to have their own targets for velum position. Most notably, the vowel /ɑ/ has a lower velum than would be consistent with the schematic in (13) above; and /l/ appears to have a consistently lower velum than /s/ does. /s/ would be expected to have a very high velum position because of the phonetic constraints imposed by the feature value [+strident] (and to a lesser extent [-voice]). A strident fricative requires a high rate of airflow through the vocal tract, and therefore no leakage through the nose. Furthermore, for both /l/ and /ɑ/, the position of the tongue body doubtless influences the velum position. It is certainly well-known that the velum is lower when the tongue is lower, and such effects need to be factored in as part of phonetic implementation.

3.2.2 Vowel Allophony in Marshallese

Another, similar, example that I won't discuss in detail here is the allophonic fronting, backing, and rounding of the vowels of Marshallese by consonants with secondary articulations, described by Choi (1992). In brief, the short vowels of Marshallese contrast in height but not in backness or rounding. Variation in backness and rounding occurs, but as the predictable influence of consonants, which have contrastive lip and tongue body positions. (That is, Marshallese is unusual in using [Back] and [Round] as consonant, but not vowel, features.) Choi showed that the acoustic properties of the Marshallese short vowels are quantitatively predictable from the consonant context, and that no target values associated with backness or rounding need be assumed for them.

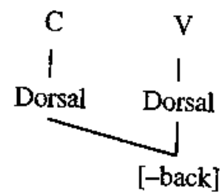
3.2.3 Velar Fronting

In many (if not most) languages, velar consonants are described as fronted before front vowels. For example, in English "key" the [k] is articulated on the hard, rather than the soft, palate. The position of the dorsal occlusion depends gradiently on the vowel. Assimilation to a front vowel can be formulated as a spread of the vowel's value for [Back]. Now if that were a complete account of the facts, this phonological rule would suffice; it would pose no problems in terms of phonetic implementation. The fact that the position of the tongue on the palate depends gradiently on that of the vowel would not stand in the way of a phonological assimilation rule: the rule says the consonant gets the same value as the vowel, and when the vowel is implemented in the front-back dimension, the consonant's implementation will simply follow from that. This is illustrated in (14).

(14) Velar fronting before front vowels as a phonological rule.

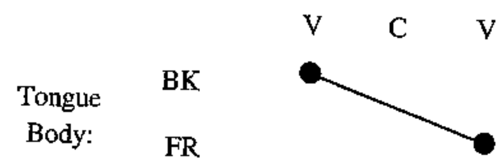
Example: "key" [k̟i]

Phonological Rule:



Such a phonological account of velar fronting is problematic for another reason: the fact that a preceding vowel's frontness also matters to the velar. As summarized in (15), in a vowel-velar-vowel sequence, the tongue moves from vowel to vowel during the velar consonant. Because the tongue forms both the vowels and the velar, the position of the velar articulation itself therefore moves during the consonant. The velar closure typically has no single location but depends entirely on the context. Spreading from the second vowel thus makes a wrong prediction about the realization of the velar. Instead, we say that the velar has no specification or target for backness (beyond the property of being dorsal), and let interpolation arrange its physical backness. (15) illustrates this. The back vowel and front vowel both have their tongue body positions specified, the consonant does not. (See Keating and Lahiri 1993 for further discussion.)

(15) Velar Fronting as phonetic interpolation



3.2.4 Phonetic Assimilation

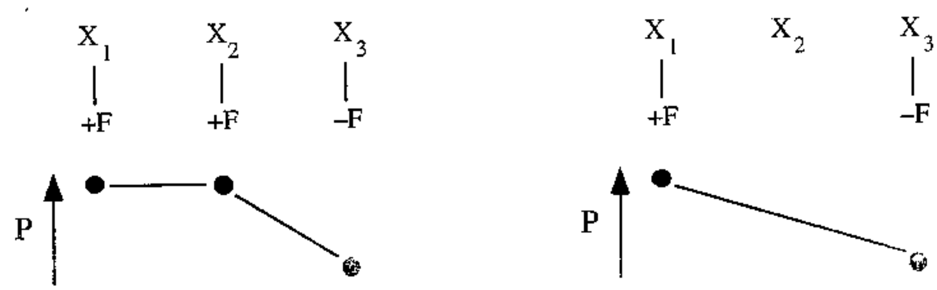
As already noted, the point of these examples is to show that "assimilation" need not be phonological (categorical); it can also be phonetic (gradient). It should be clear from these examples that gradient assimilation, due to overlap, reduction, or interpolation, is much more common than generally appreciated. This does not mean that sub-phonemic assimilation can never be categorical. Feature-spreading (or similar phonological operations of assimilation) in principle can produce categorical allophones, that is, derived feature combinations with targets different from those of the underlying segments. Nonetheless, this seems to be a rare occurrence. When suspected cases are examined carefully, they generally show gradient characteristics.

Consider the case of the many vowel allophones traditionally noted for Russian. Russian consonants have contrastive tongue-body articulations (palatalization/velarization) and these clearly have strong, gradient, effects on adjacent vowels. Most of the surface variation in Russian vowels is surely gradient, not featural. In Keating (1987) I showed that in general, variation in /i/ and /a/ can be accounted for in terms of gradient F2 target undershoot. For each vowel, though, there was one allophone that could not be attributed to target undershoot in this way. The vowel /i/ has a backed allophone target after non-palatalized consonants (this allophone is so well-established that it is considered a separate phoneme in many analyses of Russian); and the vowel /a/ has a fronted allophone target after palatalized consonants. That these are separate targets is seen pre-pausally, when the vowel has plenty of time to reach its usual target, but does not do so, instead reaching and holding the special target. It seems plausible that this is a case of phonetic feature-spreading: these special vowel targets could result from the spreading of [Back] from consonant to vowel. However, it is also possible that some version of the Window model (discussed below) could account for this variation.

3.3 The Role of Phonetic Underspecification

These examples thus illustrate what I call phonetic underspecification. This is a property incorporated into both Targets and Interpolation and Articulatory Phonology models. Phonetic underspecification means that not every segment has to have a specification, or target, for every feature/gesture. The particular consequences of this aspect of representation are different in the different models; I will limit myself here to Targets and Interpolation. Interpolation functions often do not care whether adjacent featural target specifications are from adjacent segments or not; they connect them up through an empty time interval between them, as illustrated in (16).

(16) Role of surface (phonetic) underspecification in such cases



The diagnostic for phonetic underspecification, then, is variability across contexts. If there is no phonetic specification, then what you see will depend entirely on the surrounding specifications, which will trigger interpolation through the unspecified span in a temporally-gradient fashion.

While contrast is an extremely important influence on variability, its influence can be quite indirect. We have already noted that a given feature can be realized along several parameters, so that several parameters can be somewhat influenced by a single feature value. Furthermore, sometimes languages seem to specify aspects of sounds that are not determined by contrast alone. For example, English /l/ seems to be specified as having a tongue-backing ("darkening") component (Sproat and Fujimura 1993), though this cannot be attributed to contrast. For these various reasons, then, there is more phonetic specification than phonological specification. This point has been forcefully made (though in other descriptive terms) by Recasens (e.g. Recasens 1987). He notes that different consonant articulations place different phonetic demands on the tongue body; as a result, for some consonants the tongue seems hardly free to vary at all as a function of context, though it is not contrastively specified for tongue body features. Clearly, then, the theory of phonetic implementation needs to take all such effects into account.

3.4 Contrast and the Window Model of Phonetic (Under)Specification

Thus on this view, there can be a direct relation between phonological contrast and phonetic form. When there is contrast, there is feature specification, and therefore targets. The targets will be the main influence on the contour at that time. But when at some point in time there is no contrast that uses a given parameter, there will generally be no target at that time on that parameter, and the influence of context will obviously be strong. That is, contrast restricts contextual variability while lack of contrast gives rise to contextual variability.

I developed this idea as the "window" model of surface phonetics, summarized in (17) (Keating 1990). In this model, targets are not single points in space. Instead, they are spatial ranges of permitted values. You can think of these for articulation as constraints that say how much it matters how precise an articulation is. Some targets are very narrow ranges or windows; they permit little variation. Other targets are wide ranges or windows; they permit correspondingly more variation. In effect, windows turn phonetic underspecification from an all-or-none proposition to a gradient proposition. They allow an interaction between the constraints imposed by phonological contrast, and phonetic constraints such as mechanical linkages or aerodynamic requirements.

(17) Window model of phonetic (under)specification

Targets are ranges of permitted values, not single spatial points,
- narrow window: a precise target; little variation across contexts
- wide window: not a precise target; much variation across contexts

where window width is a continuous notion (narrower to wider)

In general, targets are narrow when in contrast, wide when not in contrast
Widest possible window = "phonetic underspecification" as in (15)

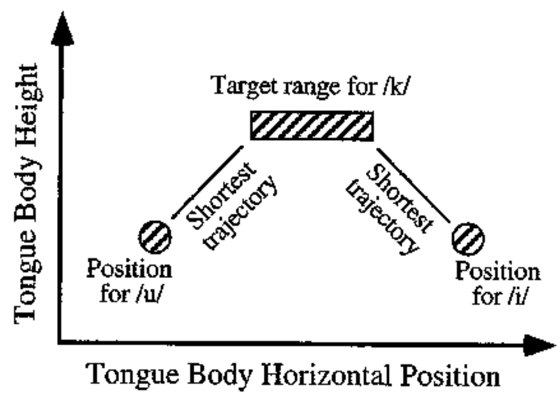
Interpolation through windows
general idea = go as slow as possible, keep smooth trajectory

Looking at some acoustic data on Arabic uvular and pharyngeal fricatives in different vowel contexts has recently suggested to me that this proposal needs to be elaborated. The kind of window described so far is a range that must be reached no matter what — no matter how short the available time, no matter how great the velocity required, this is the absolute minimum that must be achieved. Yet clearly speakers often do more than this minimum when more time is available, as is the case for geminates and at slow speaking rates. It would seem that there is another, smaller, window within the larger window. This is the range that is preferred and which is reached when more time is available, but which is undershot when time is limited.

With target *ranges* rather than *points*, interpolation becomes a more complicated function. The general idea is to go as slow as you can while still making it into the required target range, but I had no precise proposals about this.

Guenther (1995) has implemented a neural net model of articulation that incorporates a windows-like idea. Articulatory targets are spatial ranges of acceptable positions, and these ranges function like my windows. Large target ranges give rise to shorter movements and more contextual variation. (18) schematizes one of his examples, which is velar position after /u/ vs. /i/, but which extends easily to velar position between /u/ and /i/. Guenther plots target ranges in two dimensions of tongue body position, Height vs. Backness. The velar stop /k/ has a limited range of possible heights but a wide range of possible backnesses, while the vowels in this illustration have limited heights and backnesses. The model makes the shortest connection between targets in this 2-dimensional space, so gets back velars after /u/ and front velars after /i/ and back-to-front movement between /u/ and /i/.

(18) Implementation by Guenther (1995) in a neural net model of articulation: tongue position for /k/ in /uk/ vs. /ik/



3.5 A Window-Style Model of Hyperarticulation

Guenther also develops his target ranges as an implementation of Lindblom's dimension ranging from hyper- (or *over*) articulation to hypo- (or *under*) articulation, under (19).

(19) Window-style model of hyperarticulation
- shrinks targets
  - slow speech (this is one attested strategy)
  - careful speech
  - prominence
  - other prosodic conditions
- reduces coarticulation (Lindblom 1990)

A small window is a kind of hyperarticulation because it requires more careful speech to reach the small target and it limits coarticulation. So slower speech, and more careful speech, would be modeled as a shrinking of target sizes, and faster/less careful speech as an expansion of targets. He also follows up on a result of deJong, Beckman, and Edwards (1993) that phrasal prominence results in a decrease in contextual variation and thus involves hyperarticulation. This can be modeled straightforwardly as some decrease in target range of the head of a prosodic domain. That is, depending on prosodic factors, we can shrink or expand a window. Thus depending on prosodic position, the degree of contrast to be observed in some case is smaller or larger. I believe that this approach (or something along these lines) will be very fruitful in accounting for a variety of prosodically conditioned variations: cases in which position means a narrower window (a strong specification) and other cases in which position means a very wide window (a weak specification, or near-neutralization).

4. Concluding Discussion

Finally, I would like to place this discussion into the context of the distinction I have sometimes heard, that "phonology = formal, phonetics = functional". Such statements often are meant to imply that phonology and phonetics are two competing perspectives on the same phenomena, one of which must be better than the other. While there certainly are some individuals who believe this, the view outlined here is quite a different one. Phonology and phonetics are accounts of different phenomena — different levels in the grammar which carry out different computations. This is true in principle even when it is an unanswered empirical question whether any particular phenomenon belongs at one level or the other. Phonetics and phonology, formal and functional are orthogonal dimensions of inquiry and all combinations of them can be pursued.

Formal phonology: Probably most readers of this volume consider themselves to be in this category. Note that although almost all formal phonological theories incorporate at least some phonetic elements, such as phonetically-based distinctive features, they generally do not aim to account for (what are taken to be) phonetic facts.

Formal phonetics: This is a possible characterization (though not typical terminology) of the approach discussed here. The phonological part of the grammar produces symbolic representations which are passed to the phonetic component of the (formal) grammar. One consequence of the "Laboratory Phonology" movement has been to put this kind of work forward. The Conference in Laboratory Phonology is a conference which as of 1996 will be held in even-numbered years, alternating between the U.S. and abroad, and attended by phoneticians and phonologists; it is somewhat eclectic, but usually focused primarily on phonologically-informed phonetics. The proceedings are published (Kingston & Beckman 1990, Docherty & Ladd 1992, Keating 1994, Connell & Arvaniti in press). It has had a real impact on phonetics in that respect, perhaps a lesser one in phonology, but growing.

Functional phonology: By this is usually meant phonological theories or descriptions which are not formal in character, e.g. appealing to processing or learning constraints. These functional considerations may, but need not, be phonetic, i.e. pertaining to speech production or perception.

Functional phonetics: Presumably this includes much of traditional phonetics, focused on understanding the basic mechanisms of speech production and perception considered apart from linguistic structure. There is what might be called a "functional phonetics/

functional phonology interface" which is concerned largely to explain typological generalizations about inventories.

However, there is no necessary restriction to one approach at a time, or to interface only formal-to-formal, functional-to-functional. For example, at UCLA phonologists are trying to base formal Optimality Theoretic constraints on functional phonetic considerations (e.g. Jun 1995, Hayes 1995, Silverman 1995, Kaun 1995, Flemming 1995, Kirchner 1995, Steriade in prep.). These studies highlight the importance for phonology of the properties of the speech medium.

Acknowledgment

My great thanks to Rob Hagiwara for help with ms. preparation, and to colleagues at UCLA and an anonymous reviewer for useful comments.

5. References

Bell-Berti, F. and R. A. Krakow (1991). Anticipatory velar lowering: a coproduction account. *J. Acoust. Soc. Am.* 90: 112-123.

Boyce, S. E., R. A. Krakow and F. Bell-Berti (1991). Phonological underspecification and speech motor organization, *Phonology 8*: 219-236.

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

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

Browman, C and L. Goldstein (1992). Articulatory phonology: an overview. *Phonetica* 49:155-80.

Byrd, D. (1994), *Articulatory Timing in English Consonant Sequences*, UCLA Ph.D. dissertation distributed as *UCLA Working Papers in Phonetics 86*.

Choi, J.D. (1992), *Phonetic Underspecification and Target Interpolation: An Acoustic Study of Marshallese Vowel Allophony,* UCLA Ph.D. dissertation distributed as *UCLA Working Papers in Phonetics 82*.

Clements, G. N and S. Hertz (ms), An Integrated Representational System for phonology and acoustic phonetics, with a case study of English vocalic nuclei. Unpublished ms (1995).

Cohn, A. (1990), *Phonetic and Phonological Rules of Nasalization*, UCLA Ph.D. dissertation distributed as *UCLA Working Papers in Phonetics 76*.

Cohn, A. (1993). Nasalisation in English: phonology or phonetics, *Phonology 10*:43-82.

Connell, B. and A. Arvaniti, eds. (in press). *Papers in Laboratory Phonology IV* Cambridge University Press.

deJong, K., M. Beckman, and J. Edwards (1993), The interplay between prosodic structure and coarticulation, *Language and Speech 36*, 197-212.

Docherty, G. J. and D. R. Ladd, eds. (1992). *Papers in Laboratory Phonology II*, Cambridge University Press.

Flemming, E. (1995). *Perceptual Features in Phonology.* UCLA Ph.D. dissertation.

Fujimura, O. (1994). C/D Model: A Computational Model of Phonetic Implementation. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, Volume 17, pp. 1-20.

Goldsmith, John. 1976. *Autosegmental Phonology.* MIT Ph.D. dissertation.

Guenther, F. H. (1995), Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psych. Rev.*, in press.

Hayes, B. (1992), Comments on Nolan's paper, in G. Docherty and D. R. Ladd (eds), *Papers in Laboratory Phonology II,* Cambridge University Press, pp. 280-286.

Hayes, B. (1995). A phonetically-driven, Optimality-Theoretic account of post-nasal voicing. Talk presented at MIT, April 4, 1995.

Hertz, S. R. (1991). Streams, phones, and transitions: toward a new phonological and phonetic model of formant timing. *J. Phon.* 19(1): 91-109.

Jun, J. (1995), *A Gestural Approach to Assimilation in Consonant Clusters*, , UCLA Ph.D. dissertation.

Jun, S. (1993), Asymmetrical prosodic effects on the laryngeal gesture in Korean, Paper presented at Fourth Conference in Laboratory Phonology, to appear in proceedings edited by B. Connell and A. Arvaniti (in press).

Kaun, A. (1995). *The Typology of Rounding Harmony: an Optimality-Theoretic Approach.* UCLA Ph.D. dissertation.

Keating, P. A. (1987). Acoustic study of Russian vowel allophones. Poster presented at Spring meeting of the Acoust. Soc. Am. in Indianapolis. Abstract in *J. Acoust. Soc. Am.* 81 (Suppl.1): S67.

Keating, P. A. (1988a), The phonology-phonetics interface, in F.J. Newmeyer (ed.) *The Cambridge Survey, Vol.I: Linguistic Theory*, Cambridge University Press, pp. 281-302.

Keating, P. A. (1988b), Underspecification in phonetics, *Phonology* 5::275-297.

Keating, P. A . (1990), The window model of coarticulation: Articulatory evidence, in J. Kingston and M. Beckman (eds.) *Papers in Laboratory Phonology I*, Cambridge University Press, pp. 451-475.

Keating, P. A., ed. (1994). *Papers in Laboratory Phonology III.* Cambridge University Press.

Keating, P. A. and A. Lahiri (1993). Fronted velars, palatalized velars, and palatals. *Phonetica* 50:73-101.

Kingston, J. and M. E. Beckman, eds. (1990). *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech.* Cambridge University Press.

Kirchner, R. (1995). *Lenition in Optimality-Theoretic Phonology.* UCLA dissertation, to appear.

Krakow, R. A. (1989). *The Articulatory Organization of Syllables: a Kinematic Analysis of Labial and Velar Gestures.* Yale Ph.D. dissertation.

Lindblom, B. (1990), Explaining phonetic variation: A sketch of the H & H theory, in W. J. Hardcastle and Alain Marchal (eds.) *Speech Production and Speech Modelling*, Kluwer Academic Publishers, pp. 403-439.

Pierrehumbert, J.B. (1980), *The Phonology and Phonetics of English Intonation*, MIT Ph.D. dissertation.

Recasens, D. (1987). An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *J. Phonetics* 15: 299-312.

Silverman, D. (1995). *Phasing and Recoverability.* UCLA Ph.D. dissertation.

Steriade, D. (in prep.). Neutralization and the expression of contrast. UCLA ms.

Sproat, R. and O. Fujimura (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *J. Phonetics* 21: 291-311.

Zsiga (1993), *Features, Gestures, and the Temporal Aspects of Phonological Organization*, Yale Ph.D. dissertation.