Journal of Phonetics (1990) **18**, 321–334

Phonetic representations in a generative grammar

Patricia A. Keating

Phonetics Laboratory, Linguistics Department, UCLA, Los Angeles, CA 90024-1543, U.S.A.

Within the context of a grammar containing both phonological and phonetic rules, three kinds of representation all qualify as "phonetic". One is the categorical output of the phonology. The others are two forms of output of the quantitative phonetics: articulatory and acoustic. These representations are discussed and exemplified primarily with respect to laryngeal features for consonants.

1. Introduction

In this paper I will discuss the general notion of "phonetic representation" as it plays a role in a generative grammar. There are of course other quite useful notions of phonetic representation, such as impressionistic auditory transcription, but these are not my topic here.

From the generative phonetician's point of view, the phonological and phonetic components of grammar, taken together, have as their job making lexical representations pronounceable. In gross terms, they build the structures needed to determine appropriate features values and thence to supply phonetic information. Broadly speaking, phonetic representation is a level of representation derived by rules of grammar, containing phonetic information. Phoneticians usually use the term "phonetic representation" to refer to the output of the phonetic rules—something like what is observed and measured. However, they also sometimes use the term as phonologists do, to refer to the output of the phonological rules, which contains sub-phonemic detail. Both uses have their justification, and this ambiguity will be a theme of this paper.

What does it mean to say that a representation is "pronounceable"? One version of this idea is that by "pronounceable" we mean language-independent, i.e. pronounceable by someone other than a native speaker. As Anderson (1985, p. 263) notes:

> Virtually all views of sound structure have assumed that utterances can be represented in (at least) two distinct ways, each with its own importance. One of these is a phonemic (or phonological) representation, which indicates the properties of the form that differentiate it from other forms in the same language, and another is a phonetic representation, which indicates in an objective, language-independent way how the form is pronounced.

0095-4470/90/030321 + 14 $03.00/0

That is, the phonetic representation is in terms of elements with a universally valid interpretation, so that no language-specific information is omitted. (Of course this phonetic representation need not be seen as the result of application of rules of grammar, but that is the position I assume here.)

Anderson (1985, p. 266) makes clear that the phonetic representation is more abstract than a physical record of an utterance:

> A phonetic transcription can be said to have linguistic significance insofar as it indicates all and only those physical properties of an utterance that are potentially under linguistic control, in the sense that their distribution can constitute a component of the difference between one language and another. . . . There is thus a principled basis for positing a level of representation of utterances which is neither a complete physical record nor confined to the distinctive features of the language in question. This is exactly a phonetic representation in the traditional sense;

Again, those phonetic properties specific to a given language must be indicated in the phonetic representation, whereas language-independent properties are not. A different, and rather extreme, version of this position is given by Kenstowicz & Kisseberth (1979, p. 140), that potentially the phonetic representation is the same as the phonological representation:

> The view of phonology espoused in this book is that all utterances (and morphemes composing them) have an underlying representation and a phonetic representation. These representations are linked by one or more phonological rules (*or no rules if the two representations happen to be identical*) that express the predictable features of pronunciation found in the phonetic representation of each morpheme [emphasis added].

This statement makes sense only if there is a fairly substantial body of interpretive phonetic rules that can relate underlying representations directly and language-independently to physical utterances. I have argued against this view elsewhere (Keating, 1985). In this regard, it is interesting to consider Kenstowicz & Kisseberth's examples (1979, p. 171) of redundant phonetic information that need not be represented in grammars:

> That is, a [k] will tend to be somewhat fronted before [i] and a vowel will tend to be somewhat nasalized before a nasal consonant. These universal tendencies are, as such, not part of the phonologies of individual languages, but rather belong to the realm of universal phonetics.

Is it surprising that languages may differ in precisely these respects? The first of these supposed universals is discussed by Keating & Cohn (1988), and the second, most recently, by Huffman (1989) and Cohn (1990). The general point to be made, however, is to highlight the changes in thinking about phonetic representation that have occurred since the standard theory represented by Kenstowicz & Kisseberth. Languages can differ in slight enough ways that the language-specific phonetic rules must deal in fine differences, particularly differences in timing. Effects such as velar fronting and vowel nasalization thus may properly belong to the phonetics rather than the phonology, but not because they are universal. Instead, it is because only phonetic rules deal with numerical values.

Less-detailed phonetic representations which cannot be interpreted by universal conventions are also possible, and indeed are also necessary. While these do not indicate every aspect of pronunciation, they do serve another important goal for phonetic representation: they delimit phonetic variants which count as "the same" at a categorical level. As Anderson (1985, pp. 8–9) discusses, linguistic representations allow us to express the equivalence of items which are in fact different:

> Whenever we study any of these sorts of systematic relations between sound forms, we seek to determine the range and conditions of variation in what counts as 'the same' linguistic element. . . . But whenever we say that (by virtue of rule R), *x* and *y* count as linguistically 'the same thing', the temptation is virtually irresistable to ask what that 'thing' is. When we propose an answer to this sort of question, we are no longer describing the rule relating *x* and *y*, but are rather proposing a *representation* of what they have in common [emphasis in original].

Phonological representations express equivalence at the level of overall contrast within a language. Phonetic representations can express equivalence at the level of contrast in a given context. As an example, consider the stop voicing contrast in English. Phonologically, English stops may be said to contrast as "voiced" *vs.* "voiceless"; all the realizations of "voiceless" stops count as "the same" at this level of abstractness. For example, they pattern along with the voiced and voiceless fricatives, respectively. Syllable-initially, however, English stops contrast as unaspirated *vs.* aspirated. At this level of derivation only aspirated voiceless stops count as the same thing as voiceless fricatives. The further fact that systematically different degrees of aspirated stops can be physically distinguished (e.g. for different places of articulation, in different vowel contexts, under different degrees of stress) does not matter at this intermediate level of abstractness. Degree of aspiration (above and beyond its presence *vs.* absence) is never contrastive in and of itself; it always follows from something else. We can call this the "phonetic category" effect: there are certain phonetic types that are used for phonological contrasts and that can also be provided by phonological rules of particular languages, whereas other phonetic differences arise only in phonetic implementation.

Phonetic categories are part of the output of the phonology into the phonetics. Details of timing are provided by phonetic rules. By "phonetic" representation, do we mean inputs to phonetic rules, or outputs from them? Clearly both levels of representation are possible, and either might be called "phonetic". My view is that *both* levels of representation are "phonetic", and thus my title is "phonetic representations". In fact, I will discuss *three* representations that I consider phonetic.

The first representation, like the "systematic phonetic transcription" of Chomsky & Halle (1968), is the output of the phonology, but differs from that earlier proposal in the ways already outlined above. In the same sense, this representation corresponds roughly to the "phonetic representation" in my paper on stop consonant voicing (Keating, 1984), but differs from it in both concept and form. For clarity I will refer to this representation as a *categorical* phonetic representation, because it idealizes across members of a category. I take the categorical phonetic representation to be the output of the phonology (and therefore the same thing as a "derived phonological representation"), though this is *perhaps* a somewhat more elaborated output than normally assumed by phonologists. My current view of this representation derives most directly from work by Kingston (1985), Inouye (1988), Huffman (1989) and Steriade (1989), but also from work on the general notion of sub-segmental structure by Stevens (1980, 1986), Seneff & Zue (1988) and (indirectly, since they do not have segments) by Browman & Goldstein (1986).

I define the categorical phonetic representation as *clusters of feature values aligned with elements of internal segment structure*. The features are basically the usual ones adopted by phonologists, motivated by the usual phonological considerations. They may be binary, or possibly unary. The elements of internal segment structure are very limited and will be discussed below. I see this representation as, ideally, neutral with respect to articulation and acoustics/perception, a representation that can be input to articulatory planning by a speaker, or the result of perceptual processing by a listener.

The other two "phonetic" representations correspond to the idealized "physical" representations I have discussed in earlier work, where by physical representation I mean one which refers to continuous time and space. This notion is inspired by, if not the same as, Pierrehumbert (1980). Clearly a physical representation must be domain-specific, and thus there are two of them. One is particularly, the output of articulatory rules. The other is acoustic, the output of acoustic rules. I will call these the *articulatory parametric representation* and the *acoustic parametric representation*, respectively. The term "parameters" refers to the fact that these representations are in terms of physical parameters, not phonological features. The parameters are related to, but not the same as, the features; nor are they necessarily the same as the "parameters" used in e.g. formant synthesizers. (For views of the relation of parameters to features see Ladefoged & Lindau, 1986, Keating, 1988*a*, and Stevens & Keyser, 1989.) A stronger model is that the features are defined so that, taken together with acoustic theory, they are identical with single articulatory and acoustic parameters. This is so not only for Chomsky & Halle (1968), but also for Stevens. The parametric representations consist of continuous values over the many simultaneous physical parameters.

The relations among these levels of representation, a subset of the grammar, can be summarized as in (1). Both of the parameteric representations are derivable from the categorical representations, but it may be useful to think of their relations in terms of speakers and listeners. Speakers would use the categorical representation to arrive at an articulatory one, while listeners would use an acoustic representation to arrive at the categorical one.

(1) Phonology
|
Categorical phonetic representation
/ \
Articulatory parametric representation   Acoustic parametric representation

At this point it may be useful to ask about the relation of the many phonetic representations to the traditional phonetic object, the allophone. If by allophonic we simply mean subphonemic, then allophones may occur at any of these levels. Or we could restrict the "allophone" to the (featural) categorical phonetic representation.

This happens to correspond to my own usage, but I do not know which is more common, nor do I see anything at stake beyond terminological clarity.

2. Categorical phonetic representation

2.1. *Current model*

Since it is the output of the phonology, the form of the categorical phonetic representation will be given by phonological rules. I adopt some standard assumptions about initial phonological representations, and about the effects of phonological rules on those representations. Thus I assume some kind of skeleton with timing units to which features values are linked; a feature value may be linked to more than one timing unit, and may become multiply linked through spreading rules. I assume that the features are organized into some kind of hierarchical structure. And I assume that segments may be underspecified for feature values. However, these assumptions play no special role in what follows; I state them here only as general considerations. Furthermore, categorical representations, whether phonological or phonetic, need to specify very little information about the internal temporal structure of segments. Basically, segment length, in the sense of short *vs.* long segments (one *vs.* two timing slots) must be indicated, and in some segment types, a sequence of feature values (in a particular order) may occur within the segment.

Several recent proposals have addressed the kinds of sequences of feature values which may occur within segments, e.g. Sagey (1986), and most recently Steriade (1989). Steriade's proposals draw on earlier work by Kingston (1985) on glottal features and are similar to proposals by Huffman (1989) for phonetic implementation of Nasal and by Inouye (1988) on indicating release *vs.* unrelease. Like the others, Steriade draws on the distinction between constriction and release to motivate differences between stops and non-stops, and gives these phonetic events a formal status in derived representations. However, Steriade's proposals go further; she allows sequences of feature values within a segment only when each value is linked to a separate structural element. This greatly constrains possible segment types. For our purposes here, suffice to note that Steriade's proposal is more general than the others, and provides simpler representations for the cases to be discussed here, and so is the one I adopt in this paper.

Under Steriade's account, a non-continuant segment will project two aperture nodes, in sequence. The first is the closure and has a stop aperture. This node thus represents an interval of time. The second is the release of the closure, and may have either a fricative or an approximant aperture. This node is a brief interval, more like an idealized moment in time (Huffman's "landmarks"). Particular features may align with one or the other or both of these structural elements. As noted above, the feature values in the output of the phonology are still categorical.

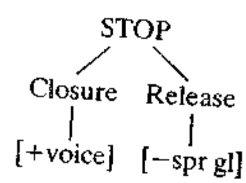
In this paper I apply this system of representation to account for some segment types and phenomena involving (non-tonal) laryngeal features. I adopt the features for voicing, aspiration and glottalization that have become fairly standard among phonologists since Halle & Stevens (1971): Voice, Spread Glottis and Constricted Glottis.[1] The feature Voice describes events during closure, and so will be aligned

[1] This is a hybrid feature set. Voice is not among the Halle & Stevens features; their companion features Stiff Vocal Cords and Slack Vocal Cords have not been widely adopted, and are not included here. See Keating (1988*b*) for discussion.

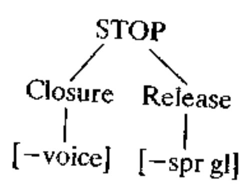
with the closure node. The value [+voice] then refers to vocal cord vibration, and low-frequency periodicity, during the closure. On the other hand, the features Spread Glottis and Constricted Glottis, which describe the position of the vocal cords, are not by their definitions necessarily aligned with either closure or release. (Alignment with release makes both the glottal feature, and the release itself, more salient, as Kingston (1985) has discussed, but alignment with closure is also possible, as we will see.) As an example of how this works, I will first show how to characterize the three most common voicing categories in stop consonants, the "Voice Onset Time" categories, using Voice, Spread Glottis and the closure/release distinction.

The feature Voice aligns with the closure to distinguish phonetically voiced from voiceless closure intervals. For presence or absence of aspiration, the feature Spread Glottis goes on the release: the value [+spread glottis] at the release refers to an open position of the vocal cords which will result in aspiration, while the value [−spread glottis] at the release refers to a close position which will result in no aspiration. These can be represented as in (2)–(4), where STOP serves as an abbreviation for all other features and structure characterizing stop consonants:

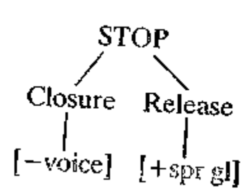
(2) Voiced



(3) Voiceless unaspirated



(4) Voiceless aspirated



2.2. *Reanalysis of earlier work*

The form of phonetic representations given above appears very different from what I proposed in my paper on the same topic (Keating, 1984), representation of stop consonant voicing. In this section I re-examine aspects of that proposal in terms of the current one.

The 1984 paper, like the present proposal, distinguished three levels of representation. One, phonological representation, is not at issue here. Briefly, it expresses contrasts within a language by means of features which may be defined relatively (more or less of some property) rather than absolutely (a threshold value of a property). In the example treated in that paper, the feature Voice is defined relatively along a physical scale from fully voiced to fully aspirated (i.e. is defined as more voiced *vs.* less voiced). Actual amounts of voicing would depend on the context. This is related in spirit to Pierrehumbert's (1980) use of two rather abstract tonal categories, defined only as a relative difference in pitch. One of the main points of the paper was that phonological features for voicing which are phonetically absolute fail to characterize all of the relevant cases. The next level, a feature-based representation of major phonetic categories, is the one at issue here. The 1984 paper also assumed a further, quantitative, level of representation, called the physical representation, as the locus of language-specific phonetic details. However, the properties of this level of representation were left vague.

In the example of stop consonant voicing, three phonetic categories were defined along the physical voicing scale. The three categories are, in traditional terms, {voiced}, {voiceless unaspirated}, and {voiceless aspirated}, though these names are just mnemonics, and formally the categories are simply arrayed along the voicing continuum. They can be distinguished as [1voice], [2voice], and [3voice], summarized in (5). Within and across languages, [+voice] can correspond to either [1voice] or [2voice], while [−voice] can correspond to either [2voice] or [3voice].

(5)

Traditional name		Voicing continuum
Voiced	=	1voice
Voiceless unaspirated		2voice
Voiceless aspirated		3voice

These phonetic categories function basically as unanalyzed primitives in the system and thus at first glance do not resemble anything like feature values aligned with segment structures. However, the information conveyed by the category *names* is quite compatible with such a view. The terms voiced and voiceless describe the state of the stop closure re vocal cord vibration and low-frequency periodicity. The terms unaspirated and aspirated describe the state of the release re glottal friction. The category names thus encode sequences of feature values tied to intervals or landmarks within the stop consonant. In this way the three phonetic categories in (5) can readily be equated with representations as in (2)–(4) above.

This way of distinguishing voiceless unaspirated from aspirated categories contrasts with that of Browman & Goldstein (1986). Whereas the distinction between (3) and (4) above is in the value of Spread Glottis aligned with the release, Browman & Goldstein essentially distinguish alignment of glottal spreading with the closure from alignment with the release.[2] They put glottal spreading during the closure to indicate lack of aspiration at the release. However, their approach faces empirical difficulties which mine does not.

Voiceless unaspirates vary in their realization as they can be produced with or without glottal opening gestures (and also with a glottal constriction gesture). As reviewed in my 1984 paper, and briefly mentioned by Goldstein & Browman (1986), English /b d g/ and /p t k/, in their voiceless unaspirated versions, have

[2] Obviously I translate into my own terms here, since Browman & Goldstein's representations contain neither features in the usual sense, nor closures or releases of segments, nor segments themselves. However, this should not affect the point intended.

both been observed to be produced with, and without, glottal spreading gestures. Thus word-initial /b d g/ may be produced either with or without an opening gesture yet in either case be voiceless unaspirated. Flege (1982) found that four out of the 10 speakers in his study produced initial /b d g/ as voiceless stops with no spreading during closure, while two out of the 10 speakers unexpectedly produced them with spreading during closure. (The other four produced voiced, not voiceless, /b d g/.) For /p t k/, Lisker, Abramson, Cooper & Schvey (1969) observed that voiceless unaspirated (word-medial, ambisyllabic) tokens were usually (84% of tokens in their study) produced with spreading during closure, but that 16% of tokens were not. Finally, it is well-known that syllable-final /p t k/ can be produced with glottal constriction, not spreading, during closure (Westbury & Niimi, 1979; Manuel & Vatikiotis-Bateson, 1988). (Goldstein & Browman ignore these counterexamples, explicitly dismissing the first two.) Thus it is not possible to define voiceless unaspirates by a glottal opening gesture during closure, nor to define English /p t k/ by a glottal spreading gesture.

The key observation in all these cases of voiceless unaspirates is that, whether or not there is a glottal gesture during closure, the vocal cords are together by the release. In the case of [+constricted glottis], the vocal cords are closer together than in the neutral case, but in both cases, they are not spread. That is, these different variants of voiceless unaspirates share a common characterization as [−spread glottis] at the release. This is what the representation in (3) above captures. The advantage of this representation is that it makes no claims about—and therefore makes no *wrong* claims about—vocal cord position during the closure.

So far we have aligned values for Spread Glottis only with the release, but clearly they may associate instead with closures in phonetic representation. The representation in (3) leaves the closure interval unspecified for Spread Glottis. Whether there actually is spreading during closure for this segment type is left open to variation, e.g. language variation or speaker variation. However, the choice of position for the vocal cords during closure is generally influenced by the value of the feature Voice, as will be discussed below.

As noted earlier, the idea that releases play a key role in phonetic form, especially with respect to laryngeal gestures, has been proposed and exploited by Kingston. However, it can be seen that the form of representation under the present proposal makes use of this idea in a different way. Here, there is no *necessary* connection between glottal gestures and stop release. Rather, it simply follows that because stops have releases while other segments do not, the glottal spreading in a stop may align with the release while in other segments it cannot. Because stops have constriction intervals like other segment types, they may also have glottal gestures during that interval, just as, e.g., fricatives do.

An additional point worth noting concerns the use of features for both vibration and glottal state within one and the same consonant. The feature Voice describes actual vocal cord vibration, while Spread Glottis and Constricted Glottis describe glottal states. This dual characterization offers a solution to earlier debates over which of these aspects of description is relevant to the phonology. Chomsky & Halle (1968), Halle & Stevens (1971) and Hayes (1984) have taken the glottal position to be crucial. Most other practice has assumed vibration to be primary. Assuming that in fact both kinds of characterizations are needed, then this is an advantage for the present proposal.

2.3. *Extensions*

Further improvements over the earlier analysis result from a greater coverage of segment types. The greater coverage is due to the fact that the structure of representations proposed here is completely general, not developed specifically for stops.

2.3.1. *Continuants*

My 1984 system was based only on stops, and did not make clear why continuants should not participate. Consider, for example, fricatives. Generally speaking, fricatives are not as varied as stops in their use of glottal featues. Across the languages in Maddieson (1984), there are no aspirated fricatives, and few glottalized fricatives (e.g. Korean, Yuchi). Fricatives usually contrast simply as voiced–voiceless. Within languages, it is rare for fricatives to parallel the glottal contrasts of stops (e.g. Amharic, with voiced, voiceless, and ejective stops and fricatives). In other languages with spread-constricted glottis contrasts on stops, the fricatives are either plain voiceless only (E. Armenian) or contrast only in voicing (e.g. Georgian, Lak).

On Steriade's account, this follows from the fact that continuants, including fricatives, do not have a release node separate from the constriction node. (While continuants may have a physical "release" in the sense of offset of constriction, they do not have a phonological release *node*.) Thus any [+spread glottis] feature value necessarily aligns with the constriction node, and so cannot indicate aspiration.

This structural difference between stops and continuants also relates to the difference in timing of the glottal spreading gesture for syllable-initial fricatives *vs.* stops noted by Browman & Goldstein (1986) and Goldstein (1989)—the gesture peak comes in the middle of a fricative but at the end (release) of a stop.

2.3.2. *Voiced aspirates*

Again, my 1984 system offered no account of these. Dixit (1989) summarizes his own and other's work on production of Hindi voiced aspirates as showing a moderate (half of voiceless aspirate) amplitude of glottal opening, beginning well into the closure and peaking after the release. As a result of the smaller glottal opening gesture, the aspiration of voiced aspirates can be voiced. However, closures are argued to be virtually identical for voiced aspirates and unaspirates. Only characteristics at the release are different. Under the present proposal we have [+voice] during closure and [+spread glottis] at release. Thus voiced aspirates are treated as the fourth possible combination of values Voice and Spread Glottis, which was impossible in the 1984 model. The phonetic component will have the further task of specifying a smaller glottal opening and the late alignment of the glottal gesture relative to the oral release, both presumably attributable to the requirement of closure voicing.

In discussing phonetically descriptive names for the voiced aspirate category, Dixit suggests terms which describe their sequence of events, such as "voiced murmured" or even "voiced breathy voiced". These terms refer to the closure and the release portions of the stop. (Dixit indicates that the aspiration of voiced aspirates is always voiced, suggesting that the release is redundantly [+voice]. However, K. Davis (personal communication 1990) finds in her acoustic corpus of

Hindi stops that this aspiration is rarely voiced, and furthermore, she finds that it is not necessary for perception of voiced aspirates. Thus this point remains unclear to me.)

In most languages, the feature value [+spread glottis] is associated with voicelessness: a segment which is [+spread glottis] is redundantly [−voice]. In English, the segment /h/ is specified only as [+spread glottis]; it is redundantly [−voice] unless it voices by assimilation to context. Languages with voiced aspirates, or with breathy voicing, are exceptions to this generalization.

2.3.3. *Preaspirates*

Again, these were not related to the more usual voicing categories under the 1984 system. However, they require a rather different analysis from the single-segment cases seen so far. It is clear that preaspirates are not just the mirror image of voiceless (post)aspirates, i.e. with [+spread glottis] aligned with the closure of a stop. Instead, as Kingston (1985) discusses, they derive from a geminate structure. I know of no cases of single-segment preaspirates.

2.3.4. *Constricted glottis segments*

Most of the discussion of glottal configuration so far has focused on the feature Spread Glottis, and especially on the value [+spread glottis], which when aligned with a stop release indicates aspiration. If no glottal spreading is involved, there are two other possibilities. One is a neutral glottal opening: [−spread glottis], [−constricted glottis]. The other is a constricted, or narrowed, glottis, or glottalization: [+constricted glottis]. Like [+spread glottis], [+constricted glottis] usually aligns with the release in derived representation, since these values are more salient when realized there. However, often it is realized through both the closure and the release, making the phonetic behavior of Constricted Glottis somewhat different from Spread Glottis. Constricted Glottis can doubly link in a way that Spread Glottis does not.

However, [+constricted glottis] can also align only with a closure. English uses a glottal constriction to make syllable-final voiceless stops completely voiceless (Westbury & Niimi, 1979; Manuel & Vatikiotis-Bateson, 1988). Since these stops are also often unreleased, clearly the feature is aligned with the stop closure. Indeed, the glottalization is seen at the end of the preceding vowel.

In English, this glottalization is one part of a more general use of the glottal features to enhance voicelessness in stops. In syllable-initial position, glottal spreading at the release (aspiration) is obligatory; ambisyllabically, spreading in closure is optional but common (as outlined above); and syllable-finally, glottalization is optional but common.

2.3.5. *Unreleased stops*

Unreleased stops (typically, in word or syllable-final position) by definition have no (audible) release node and so are just like continuants in their options for aligning glottal features. If an unreleased stop has the feature value [+spread glottis], it must be realized as glottal spreading during the closure, either voiceless or (breathy) voiced, and cannot be realized as aspiration. The same holds for [+constricted glottis]. It would be unlikely (though not impossible) for a language to contrast two different kinds of unreleased vocieless stops, since spread, constricted and neutral glottal positions would all sound much the same (without other enhancements). Similarly, different glottal positions are unlikely to contrast during a voiced unreleased stop; voicing during stop closure is heard only through the vocal tract walls, making it difficult to distinguish different kinds of voicing.

In languages with released final stops, a full range of glottal contrasts is possible and attested (e.g. E. Armenian, Indic languages), but in languages with unreleased final stops, only voicing contrasts are reported (e.g. English). Languages with glottal but not voicing contrasts have neutralized final unreleased stops (e.g. Cantonese, Korean). The presence *vs.* absence of a release node in derived representation allows us to characterize these differences appropriately.

3. Parametric representations

Representations derived from the (categorical) phonetic representation are defined not in terms of discrete features, but in terms of continuous articulatory and acoustic parameters. My own general framework for deriving such representations, called the targets and interpolation model, is similar in spirit to many speech synthesis models, and has been described in earlier papers (see also Huffman 1989, Cohn 1990, for summaries and elaborations). Briefly, after default feature vaues (if any) have been supplied, phonological features are mapped to parameters in a many-to-many relation which is probably to some extent language-specific. The feature values are mapped to targets, which extend in time along these parameters. Targets are not necessarily precise values, but may be ranges, with further requirements added in the form of windows, or limits, of possible values. Targets are connected up by interpolation functions selected from a set of possible functions, which may be language-specific and/or parameter-specific. This model is intended to apply in both articulatory and acoustic domains.

Of course, the targets and interpolation model is not the only possible way to derive parametric representations from a categorical representation. The task dynamic approach adopted by Browman & Goldstein (1990, and references cited there) is one alternative in deriving the articulatory parametric representation.

3.1. *Articulatory*

In the cases at hand, there are three features to be implemented: Voice, Spread Glottis and Constricted Glottis. One parameter is vocal cord opening. The spatial value of this parameter is controlled primarily by the Glottis features taken together, but in a way compatible with the value for Voice. Thus, [+spread glottis] is realized by a glottal opening gesture, but if the value for Voice at that point is [+voice], then the glottal gesture will be smaller than if the value for Voice is [−voice]. The timing (point of onset, duration) of a glottal gesture will depend in part on any other nearby glottal features, so that all of these specifications can be satisfied.

A feature value [+constricted glottis] will be realized as a glottal closing gesture, but again the exact gesture will depend on the concomitant value for Voice. In combination with [−voice], glottal constriction will be interpreted as a complete constriction, or glottal stop; in combination with [+voice], it will be interpreted as a partial constriction.

The feature Voice thus influences the parameter of vocal cord opening. In many segment types, proper vocal cord position, plus a flow of air from the lungs, is enough to effect vocal cord vibration. For stop consonants, however, a flow of air from the lungs in the face of oral closure is not guaranteed. Therefore in utterance-initial position, further measures are needed to continue flow and thus give enough of a pressure drop across the larynx to effect vibration. To a lesser extent, the same holds of utterance-final position, and of voiced geminates. Some of the articulatory maneuvers that can be used are documented in Westbury (1983) and include oral cavity and pharyngeal expansion, larynx lowering, laxing of the vocal tract walls, and increased respiratory effort. Slight velum opening would be another possibility. Thus a value [+voice] may be implemented along any or all of these articulatory parameters to give actual vocal cord vibration in a given phonetic context.

The hypothesis that a single feature may call on more than one parameter (see Ladefoged & Lindau, 1986) is intended here to have a cross-linguistic implication, namely, that languages may differ in how they realize a given feature value. Such a difference would be related to saliency: the more parameters are used for a given feature, the more robust and salient that feature's values will be (see Stevens & Keyser, 1989). Thus in some languages, or for some speakers (such as Westbury), [+voice] is rigorously implemented as a full interval of vocal cord vibration through the use of several parameters. In contrast, for other speakers of, e.g., English, fewer such extra parameters, or even none, may be marshalled, resulting in briefer intervals of vocal cord vibration, i.e., partial devoicing. In time, such partial devoicing may come to be interpreted as the realization of a [−voice] value. In this way, an apparent neutralization between [+voice] and [−voice] consonants might still involve physical phonetic differences, but not enough for them to be auditorily categorized as different.

In the case of [+voice], multiple parameters can be controlled so as to guarantee a robust and salient instance of the defining property of the feature value, namely, vocal cord vibration. This basic idea can be extended in a less direct sense. The feature Constricted Glottis may also marshal various parameters in its implementation, not to make the glottis "more constricted", but instead to make clear to a listener that the glottis was indeed constricted. These added parameters would include vertical larynx movement, resulting in a glottalic airstream mechanism, or the articulatory parameters described by Dart (1987) for Korean.

Each parameter would receive a spatial value (or range of values) for a given moment or interval of time. Interpolation between adjacent values follows language-specific and parameter-specific functions.

3.2. *Acoustic representation*

The acoustic speech signal consists of changing values along many simultaneous acoustic parameters. Maxima and minima of these parameters, or sudden changes in values, may be readily interpreted by listeners as localized feature values. That is, the listener may derive features values from perceived parameter values. In terms of the grammar, we derive the parameter values from the feature values. However, I will phrase this section in terms of the listener rather than the grammar.

An interval of low-frequency periodicity, or voicing, will be interpreted as a feature value [+voice] occurring during that interval. An interval without such periodicity will be interpreted as a feature value [−voice] occurring during that interval, if the interval is long enough to outlast any masking effects of previous voicing. (That is, it is not enough for some acoustic event to occur—the event must be perceivable to a listener in the face of masking and other auditory processing effects. A too-short interval of voicelessness will not be noticed by the listener.)

The presence of a feature value [+spread glottis] during the closure interval of a stop consonant perhaps may be detected in the quality of voicing offset in a preceding segment. The presence of this feature value at release will be detected through a long Voice Onset Time and/or aspiration. Laryngealized voicing indicates the presence of a [+constricted glottis] value; amplitude characteristics at and after a consonant offset do so for various glottalized obstruents. In English, a word-initial glottal stop can result in laryngealization of both adjacent vowels, and a syllable-final glottalized stop may be detected in the quality of voicing offset before the stop, and in the (lack of) transitions into the stop.

The physical characteristics of the signal corresponding to various feature values may thus be found in different locations in different cases. Interpretation of these acoustic events in terms of phonological structure and lexical items clearly must depend on knowledge of the language in question, as many researchers have shown.

4. Conclusion

My point of departure in this paper was the principle that what someone has to know about a language to behave like a native speaker is the grammar of that language; since languages differ in phonetic detail, some account of those differences must be provided by the grammar. Two more-specific points also played a role: first, that the phonetic component of the grammar is concerned with quantitative operations, and second, that the phonological component of the grammar is concerned with operations on symbols. One of these operations is to derive aperture nodes for segments, with two such nodes (closure and release) projected for stops. Given all of these ingredients, we find ourselves with two levels of representation which are candidates for the label "phonetic", one the input to the phonetic component and the other its output. I conclude that these levels are both phonetic levels, but different according to the elements that compose the representations. Phonetic representations are linguistic forms indicating derived aspects of the sound structure of utterances, but they may do this in more or less detail, with varying degrees of approximation to "pronounceability".

Earlier versions of portions of this paper were presented at the 1990 University of Arizona Phonology Workshop and at an MIT Linguistics Department colloquium; members of those audiences, especially Juliette Blevins, made helpful comments. Bruce Hayes, Marie Huffman and Donca Steriade gave useful suggestions on a draft of the paper. All errors remain my own.

References

Anderson, S. (1985) *Phonology in the twentieth century*. Chicago: University of Chicago Press.

Browman, C. & Goldstein, L. (1986) Towards an articulatory phonology, *Phonology Yearbook*, **3**, 219–252.

Browman, C. & Goldstein, L. (1990) Gestural specification using dynamically-defined articulatory structures, *Journal of Phonetics*, **18**, 299–320.

Chomsky, N. & Halle, M. (1968) *The sound pattern of English*. New York: Harper and Row.

Cohn, A. (1990) Phonetic and phonological rules of nasalization. UCLA dissertation; also *UCLA Working Papers in Phonetics*, **76**, 1–224.

Dart, S. (1987) An aerodynamic study of Korean stop consonants: measurements and modeling, *Journal of the Acoustical Society of America*, **81**, 138–147.

Dixit, P. (1989) Glottal gestures in Hindi plosives, *Journal of Phonetics*, **17**, 213–237.

Flege, J. (1982) Laryngeal timing and phonation onset in utterance-initial English stops, *Journal of Phonetics*, **10**, 177–192.

Goldstein, L. & Browman, C. (1986) Representation of voicing contrasts using articulatory gestures, *Journal of Phonetics*, **14**, 339–342.

Goldstein, L. (1989) On articulatory binding: comments on Kingston's paper. In *Papers in laboratory phonology I: between the grammar and the physics of speech* (J. Kingston & M. Beckman, editors), pp. 445–450. Cambridge: Cambridge University Press.

Halle, M. & Stevens, K. (1971) A note on laryngeal features. Quarterly Progress Report No. 101. Cambridge, MA: Research Laboratory of Electronics, MIT.

Hayes, B. (1984) The phonetics and phonology of Russian voicing assimilation. In *Language sound structure* (M. Aronoff & R. Oehrle, editors), pp. 318–328. Cambridge, MA: MIT Press.

Huffman, M. (1989). Implementation of Nasal: timing and articulatory landmarks. UCLA dissertation, also *UCLA Working Papers in Phonetics*, **75**, 1–149.

Inouye, S. B. (1988) Flap as a contour segment, *UCLA Working Papers in Phonetics*, **72**, 40–82.

Keating, P. (1984) Phonetic and phonological representation of stop consonant voicing, *Language*, **60**, 286–319.

Keating, P. (1985) Universal phonetics and the organization of grammars. In *Phonetics linguistics: essays in honor of Peter Ladefoged* (V. Fromkin, editor), pp. 115–132. Orlando: Academic Press.

Keating, P. (1988*a*) The phonology-phonetics interface. In *Linguistics: the Cambridge survey*, Volume 1: *Grammatical theory* (F. Newmeyer, editor), pp. 281–302. Cambridge: Cambridge University Press.

Keating, P. (1988*b*) A survey of phonological features. Distributed by IULC.

Keating, P. & Cohn, A. (1988) Cross-language effects of vowels on consonant onsets, *Journal of the Acoustical Society of America*, **84**, Suppl. 1, S84.

Kenstowicz, M. & Kisseberth, C. (1979) *Generative phonology: description and theory*. New York: Academic Press.

Kingston, J. (1985) The phonetics and phonology of the timing of oral and glottal events. UC Berkeley dissertation.

Ladefoged, P. & Lindau, M. (1986) Variability of feature specifications. In *Symposium on invariance and variability of speech processes* (J. Perkell & D. Klatt, editors). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lisker, L., Abramson, A., Cooper, F. & Schvey, M. (1969) Transillumination of the larynx in running speech, *Journal of the Acoustical Society of America*, **45**, 1544–1546.

Maddieson, I. (1984) *Patterns of sounds*. Cambridge: Cambridge University Press.

Manuel, S. & Vatikiotis-Bateson, E. (1988) Oral and glottal gestures and acoustics of underlying /t/ in English, *Journal of the Acoustical Society of America*, **84**, Suppl. 1, S84.

Pierrehumbert, J. B. (1980) The phonology and phonetics of English intonation. MIT dissertation. Also distributed by IULC, 1987.

Sagey, E. (1986) The representation of features and relations in nonlinear phonology. MIT dissertation.

Seneff, S. & Zue, V. (1988) Transcription and alignment of the TIMIT database. Manuscript distributed by the National Bureau of Standards.

Steriade, D. (1989) Affricates and release. Paper presented at the conference on features, MIT, October.

Stevens, K. N. (1980) Acoustic correlates of some phonetic categories, *Journal of the Acoustical Society of America*, **68**, 836–842.

Stevens, K. N. (1986) Models of phonetic recognition II: an approach to feature-based recognition. In *Proceedings of the Montreal symposium on speech recognition*, pp. 67–68.

Stevens, K. N. & Keyser, S. J. (1989) Primary features and their enhancement in consonants, *Language*, **65**, 81–106.

Westbury, J. R. (1983) Enlargement of the supraglottal cavity and its relation to stop consonant voicing, *Journal of the Acoustical Society of America*, **73**, 1322–1336.

Westbury, J. R. & Niimi, S. (1979) *An effect of phonetic environment* [illegible] *voicing control* [illegible] *during stop consonants*. In *Speech communication papers presented at the 97th meeting of the Acoustical Society of America* (J. Wolf & D. Klatt, editors). New York: Acoustical Society of America.