PHONETIC AND PHONOLOGICAL REPRESENTATION OF STOP CONSONANT VOICING

PATRICIA A. KEATING

University of California, Los Angeles

This paper argues for a more structured view of the relation between the phonological feature [voice] and its specific phonetic implementations. Under the theory of universal phonetics proposed here, the implementation of [voice] is sharply constrained: the opposition is defined relatively, as more or less voicing, along a dimension consisting of exactly three discrete, ordered categories, which can be shown to have clear articulatory and acoustic bases. While the phonological feature allows certain rule equivalences across languages to be expressed, the phonetic categories describe possible contrasts within languages, and express markedness relations.*

1. INTRODUCTION. It is common practice to represent sounds in different languages which are phonetically similar, but not absolutely identical, with the same symbol or set of feature specifications. Sometimes this practice is simply a convenience, for ease in transcription or typesetting. At other times, the sounds in question are said to be different only at the phonetic level; at the phonemic level, they may indeed have the same feature specifications, and so are thought of as the 'same' sound. For example, the distinction in the *SPE* model (Chomsky & Halle 1968) between systematic phonemic vs. phonetic levels makes such a view possible. However, I will argue here that the version of the model proposed in *SPE* can be improved, allowing a more adequate treatment of a variety of facts. I will discuss how surface phonetic variation, within and across languages, can be derived in a synchronic grammar from the interaction of three relatively simple systems: the possible phonological features and their values, their possible phonetic category mappings, and phonetic detail rules accounting for variation within these phonetic categories. More generally, this paper contributes to an important goal of linguistic phonetics: that of relating discrete and timeless phonological units to physical reality, with its continuous articulatory and acoustic manifestations.

1.1. THE *SPE* MODEL represents lexical items as matrices of binary-valued phonetic features; each row is a feature, and each column a segment. Phonological rules may change the values of features, or may add or delete segments, but may not change the inventory of features which form the rows of the matrices. By contrast, phonetic rules convert the binary values into quantitative values along continuous phonetic scales; these rules specify the value along each scale at which the given language divides the scale into its phonetic

* This research was supported in part by a postdoctoral fellowship at MIT from NIH, by a research grant from UCLA, and by an NSF grant to Peter Ladefoged. I would like to thank Elan Dresher, Louis Goldstein, Bruce Hayes, Peter Ladefoged, Ian Maddieson, John Ohala, Patti Jo Price, Stephanie Shattuck-Hufnagel, and John Westbury for comments and/or advice; Don Brusasco, Marie Huffman, Ellen Jackson, and Wendy Linker for data collection; and Philip Lieberman and Sheila Blumstein for the initial idea. While not all of them agree with me, they have all helped in some way.

categories. Languages differ phonetically in part in just this way: two languages can differ in how nasalized a nasalized vowel must be, or how retroflexed a retroflexed stop must be. These rules, like the binary phonological rules, are part of the 'phonological' component of the grammar, and no claim is made that quantitative rules are necessarily ordered after the binary rules.

The output of this phonological component is the phonetic transcription, containing all facts about pronunciation that are determined by the grammar. The phonetic scales available as rows of the phonetic matrix are universally fixed, but the quantitative values found along those scales will differ across languages. A further universal phonetic component—not technically part of the grammar—will convert these scalar values into a representation of articulations that are continuous in time. Given the universal phonetic scales, this translation or realization is assumed to be automatic. Various universal phonetic principles will produce phonetic patterns that will not need to be specified by the grammar of any one language.

Phonological feature representation serves to describe natural classes in rule application and to differentiate underlying forms. Phonetic feature (scalar) representation provides more detail—serving to describe any possible systematic differences between sounds that could distinguish two languages, and might therefore be supposed to be part of the grammar of any one language. Following Ladefoged 1978, we might suppose that the appropriate level of detail to be represented is that required to distinguish a native from a foreign accent; but note that phonetic feature representation is still removed from the physical level. In the *SPE* model, it is the last form of discrete segmental representation prior to a continuous-in-time physical representation, the exact form of which remains to be established.

1.2. THE CASE OF CONSONANT VOICING. This paper will explore how the above framework can be used to describe variation in stop consonant voicing within and across languages with two phonemic classes ('voiced' and 'voiceless'). The discussion will be limited to this single case for two reasons. First, as will be seen, a wide variety of data can be brought to bear on the issues, and these data are available mainly for voicing. Second, there has traditionally been confusion over the terms 'voiced' and 'voiceless', and the phonetic symbols defined in terms of them; these are typically used to mean different things in different languages. That is, the problem is clearly defined for voicing, and the relevant data are at hand.

The problem is this: the symbols for voiced and voiceless stops *b d g* and *p t k* are used for a variety of physical events. Strictly speaking, [b d g] are supposed to be reserved for stops with voicing during their closures, and [p t k] for voiceless unaspirated stops. The symbols defined in this way can be used in Polish, French, and many other languages quite straightforwardly. In contrast, their use in English and other languages is more complicated: [b d g] occur mainly in medial position, sometimes in initial position; [p t k] occur after [s], and sometimes medially. But /b d g/ are used as broad-transcription symbols for both [b d g] and voiceless [b̥ d̥ g̥], which occur mainly in initial

position. The latter set has been called voiceless lenis, or voiceless lax, or even 'voiced as in English'. However, such phonetic detail is often ignored in descriptions of English, e.g. in Ladefoged 1982. More often discussed is the distinction between unaspirated [p t k] and aspirated [p^h t^h k^h]. However, Heffner 1969 uses [b̥ d̥ g̥] for the medial unaspirated stops (as in *capper*), and in general it is unclear how [p t k] and [b̥ d̥ g̥] are thought to differ. The former may be said to be 'fortes' or tense, and the latter 'lenes' or lax; but definitions of the feature concerned are usually vague or contradictory.

There has also been difficulty in formulating an appropriate theoretical framework. In all recent rigorous attempts to define phonological features in phonetic terms, the common practice described above has been abandoned. In the system of Jakobson et al. 1952, Polish would be described as having a voicing contrast, but English would not: it would have distinctive tenseness. Neither *SPE* nor Halle & Stevens 1971 (H&S) has a 'voicing' feature in the usual sense; in these papers, Polish and English would use different phonological features to distinguish /p t k/ from /b d g/. This is because these feature systems use physical features describing specific articulatory states, both to represent phonetic categories and to serve as the basis for phonological representations. Recall that differences in scalar values express cross-language phonetic differences in the *SPE* system. Thus, if Polish and English differ simply in the amount of aspiration required for a stop to count as voiceless, they can be described as having the same phonological contrast. But if they use different phonetic CATEGORIES or articulatory mechanisms, then their phonological descriptions will necessarily be different, since the phonetic categories are used as the phonological categories.

Consider how the features in H&S describe voiced and voiceless stops of different kinds. They propose four features to characterize the state of the larynx at the moment of stop release: [±stiff vocal cords], [±slack vocal cords], [±spread vocal cords], and [±constricted vocal cords]. The combination [−spread, +constricted] means a glottalized stop; [+spread, −constricted], an aspirated stop, and [−spread, −constricted], a plain stop. The combination [+stiff, −slack] means a voiceless stop; [−stiff, +slack], a voiced stop. Therefore a [−stiff, +slack, −spread, −constricted] stop is voiced and unaspirated; a [+stiff, −slack, −spread, −constricted] stop is voiceless and unaspirated; a [+stiff, −slack, +spread, −constricted] stop is voiceless and unaspirated; and a [−stiff, +slack, +spread, −constricted] stop is a voiced aspirate. Other combinations describe voiced implosives, voiced laryngealized stops, voiceless glottalized or ejective stops, and the 'moderately aspirated' Korean stop that contrasts with the more aspirated variant already specified above. The main point of interest to us is H&S's description of lax or voiceless [b], occurring in initial position in Danish and English, as [−stiff, −slack, −spread, −constricted], distinct from [p]. These features were originally designed to relate pitch to voicing differences with a single laryngeal mechanism; they have been criticized on that basis by, e.g., Anderson 1978.

The above features are descriptively inadequate for voicing independent of its relation to pitch. For example, H&S posit that the difference between the

initial 'voiceless lax' /b d g/ of some English speakers and the initial fully voiced (often called prevoiced) stops of other English speakers is one in degree of vocal-cord slackness. In fact, the glottal configuration and state can be identical for these two categories. The determining factor for closure voicing is the pressure in the oral cavity: if oral pressure is kept low, either through cavity expansion or leak, then closure voicing will occur (Westbury 1983). By contrast, the difference between English medial voiceless unaspirated and (fully) voiced stops (e.g. *rapid* vs. *rabid*) is posited by H&S to be one of vocal-cord stiffness. In fact, the difference is more likely to be that the medial voiceless stops involve a glottal spreading gesture (Lisker et al. 1969); by the time of release, the vocal cords will be back together, and may well be slack just as for the voiced stops. Thus the glottal state at the moment of consonant release (the basis of the H&S features) can be identical for English initial 'voiceless lax' and initial prevoiced—as well as for medial 'voiceless unaspirated' and medial voiced-through stops, requiring the addition of features for supraglottal pressure and occurrence of gestures at times other than release.

While such feature additions may save the system from these particular criticisms, this would be the wrong direction to take in enumerating phonetic features. A set of features predicated on phonetic accuracy will require ever more additional features as new articulatory mechanisms are discovered. The proliferation of features is the price paid for using the same set of features for phonological as well as low-level phonetic representation—an otherwise appealing constraint on the relation of phonology to phonetics. But at the same time, no claim is made that these largely redundant features can structure linguistic contrasts. Thus the H&S features distinguish voiced laryngealized stops from true implosives, whereas languages never do (Greenberg 1970, Ladefoged 1982); similarly, H&S needlessly distinguish voiceless unaspirated from voiceless lax stops. The point, then, is that H&S (and *SPE*) don't simply have the wrong features in these instances; they will ALWAYS have TOO MANY features because they want to describe exactly how individual sounds are articulated. While we want the phonological features to have some phonetic basis, we also want to distinguish possible contrasts from possible differences. Our goal must be to find some feature framework in which the phonetic basis of phonological features is not explicit phonetic detail.

1.3. A DIRECTION. It's clear that, in the *SPE* type of feature system, Polish and English /b d g p t k/ can look very little alike, and can not be referred to by the same symbols in a phonological representation. My starting point here will be to consider how the alternative result might be allowed, by relaxing the constraint that phonological feature representations incorporate specific articulatory information. Instead of concentrating on accurate phonetic descriptions of individual segments, let us consider how each level of representation can characterize some aspect of sound systems.

Suppose we try to modify the *SPE* features so that, within the general model of the relation between phonology and phonetics, we have only as many phonetic categories given by the phonetic features as there are contrasting phonetic types in languages. Let us call these the (major) PHONETIC CATEGORIES, and

think of them as defining the phonetic symbols which we normally use in transcription; i.e., the categories are imposed along the phonetic scales which underlie the phonological feature contrasts. Smaller-scale possible differences will not be represented in this transcription, as in fact they typically are not. Suppose further that the phonological representation will not directly reflect phonetic facts, but rather will organize natural classes for phonological rules. A phonological feature, then, corresponds to the use of terms like 'voicing' for various kinds of contrasts. Will such revisions in the feature system do any more than justify the traditional, non-rigorous, use of terminology? What can it mean to say that two languages have the same contrast, like voicing, when the corresponding physical events are somewhat different?

2.1. A PROPOSAL. Here I will explore the consequences of a proposal by Lieberman 1970, 1977 that [±voice] be used as a binary phonological feature which can be implemented differently in different languages along the continuous dimension of V[oice] O[nset] T[ime]. VOT was originally proposed by Lisker & Abramson 1964 as the time between the onset of voicing and the release of a stop consonant; when voicing onset follows the release by any appreciable amount of time, aspiration occurs during that time interval. Thus the VOT dimension relates aspiration to voicing phonetically. However, Lisker & Abramson 1964, 1971 provide no suggestion concerning the role of this phonetic dimension in phonologies, or in phonological feature representation.

As given by Lieberman, this proposal simply amounts to a claim within the *SPE* framework that the physical scale appropriate to a voicing feature is the VOT scale, and that plus/minus values of the voicing feature will have different quantitative VOT values in different languages. However, this specification does not deal with the question of the possible number of contrasts available to languages—since, presumably, any arbitrary set of VOT values can be specified under Lieberman's proposal. In addition, VOT is not always the most useful way to characterize voicing across environments. Therefore I will modify the proposal so that the binary phonological feature values will be implemented as CATEGORIES chosen from a fixed and universally specified set. This set, which will be motivated below, consists of three categories: fully voiced, voiceless unaspirated, and voiceless aspirated stop consonants. These correspond directly to the standard division of the VOT continuum into lead, short-lag, and long-lag values for stops in initial position; however, they should be viewed as more abstract categories which include a number of acoustic correlates and articulatory mechanisms. To keep the phonological and phonetic representations distinct, I will follow the convention that [±voice] (or non-binary [voice]) refers only to the classificatory feature and its values, while {voiced}, {vl.unasp.}, and {vl.asp.} refer to the major phonetic categories. As in the *SPE* model, these categories will be further realized as articulatory and acoustic parameters represented continuously in time. To some extent, these mappings will be part of the definition of the phonetic categories, and therefore universal; e.g., {voiced} will involve vocal-cord vibration and low-frequency periodicity during consonant closure. To some extent, however, they will be language-specific, in ways to be discussed below.

Given that the phonetic categories encode possible contrasts, we can return to the question of Eng. vs. Pol. /b d g p t k/. Both languages may be said to contrast [+voice] and [−voice] stops, but the phonetic categories which implement the phonological contrast will differ. Thus, in some cases, Polish [+voice] stops will be {voiced} while the English ones will be {vl.unasp.}, and the Polish [−voice] stops will be {vl.unasp.} while the English ones will be {vl.asp.} The details of these differences will be the topic of §4. For now, it suffices to note that the framework allows us to say that the stops of the two languages are always the same phonologically, though they may be different phonetically.

To summarize, three kinds of representation are being proposed. One is phonological: just as many features and feature values as are needed to distinguish the natural classes in a given language. Another, dealing with major phonetic categories, is 'modified systematic phonetic': just as many phonetic categories as are needed to distinguish categories contrasting in any language. The last is pseudo-physical: continuous in time and encompassing as many parameters as necessary for phonetic description—possibly autosegmental in character (cf. Goldsmith 1976).

Before the implications of this proposal can be discussed, it should be made explicit that a limited set of cases is being considered. Basically, I am dealing only with languages which contrast no more than two phonetic categories in any one context. It will be assumed, lacking evidence to the contrary, that all such contrasts should be described as [±voice]. I will exclude languages like Thai and Hindi, with more than two contrastive categories, largely because it is unclear whether such languages should be analysed as having a single non-binary feature [voice], or more than one binary feature. It is immaterial to the present arguments how these cases are treated; however, their analysis would certainly be relevant to a more complete account of voicing contrasts and their possible phonetic implementations.

In fact, extensive data would be required to bear on this question. Consider Thai, which in initial position contrasts voiced, voiceless unaspirated, and voiceless aspirated stops (at all except the velar place of articulation; Haas 1956). In final position, only unreleased, largely voiceless, stops are allowed (Abramson 1972). At first glance, this limitation appears to be phonological evidence that we are dealing with a single three-valued feature, since we hesitate to say that two independent features happen to be neutralized (in the sense of having defective distribution) at the same time and in the same way. But when further data are considered, it is evident that more than voicing is involved, since the only coronal obstruent found in final position is [t°]. Thus not only voicing, but also place and manner features, are limited in final position; thus the evidence for a single, ternary [voice] feature in Thai is at best equivocal.

2.2. THE PHONOLOGICAL FEATURE. If we limit ourselves, then, to cases like Polish and English, what does it mean to say that there is a level of representation in the grammar at which various phonetic sorts of /b d g/ are all [+voice]? What kind of evidence could be relevant in supporting this claim? I want to

argue that evidence can be found in a PREDICTION that the system makes about the relation between phonological rules and phonetic categories: namely, that the feature outputs of phonological rules should have different phonetic category implementations in different languages. Since phonological rules apply to binary feature values, before phonetic category implementation, they cannot anticipate the phonetic values onto which those binary values are mapped; i.e., rules cannot look ahead to future steps in a derivation. Thus a rule which changes a binary feature value cannot know whether a resulting [+voice] value will be implemented as {voiced} or {vl.unasp.}. A curious consequence of this lack of look-ahead is that it should be possible for such a rule to occur with EITHER sort of phonetic category implementation. That is, the occurrence of a phonological rule in a language should not depend on, or be correlated with, the phonetic details of the language. One could well object to this seemingly implausible prediction, since we tend to think of the occurrence of a rule as being intimately tied to its phonetic effects. But note that the issue does not even arise in a system where phonological and phonetic features are identical, as in *SPE*, since a change in a binary-valued feature automatically entails a particular phonetic change.

However, my prediction is borne out by various data. It appears that a distinction between phonological and phonetic category levels of representation offers an important advantage in describing phonological rules. In a system like *SPE*, which equates phonological with phonetic representation, rules that occur across languages will look different in each language, depending on the phonetics. In a system like the one proposed here, which distinguishes the two levels, these rules will look the same regardless of the phonetics. Thus, if rules which affect voicing recur consistently across languages, but differ in their phonetic categories, then we have evidence in favor of distinguishing phonological from phonetic representation. The generalization that certain rules occur across languages will be missed if phonological rules apply to phonetic features which are different across languages, but it will be expressed if such rules apply to phonological features that are similar across languages. Consider now three relevant cases of rules which refer to 'voicing' in a similar way across phonetically different languages, and which thus support the existence of a [voice] feature.

(a) Chen 1970 considers vowel duration before word-final 'voiced' and 'voiceless' stops in several languages; and he shows that, in all of them, vowels are longer before 'voiced' than before 'voiceless' stops.[1] This apparent generalization would have to be abandoned if phonetic differences were incorporated into phonological feature representation, as required in standard theories, since

[1] Chen's study shows at least this, although two caveats are in order. First, some languages do not show this effect (Flege 1979, Keating 1979). Second, Chen's claim that different languages show different degrees of lengthening raises some problems; e.g., it is known that there is less lengthening in disyllables than in monosyllables, yet Chen's samples are not matched for number of syllables. Some of Chen's 'language' effects are surely sample effects. However, Mack 1982 compares French and English with carefully matched word lists, and does find a difference in lengthening; so Chen's claim may still prove true.

some of these languages can be described as having 'tense' vs. 'lax' stops (or whatever feature is used), while others have 'voiced' vs. 'voiceless'. That is, if vowels are longer before phonetically 'voiced' stops and before phonetically 'lax' stops, we can conclude that, in general, they are longer before phonologically [+voice] stops. Further supporting data can be found in Mack, which uses precisely matched word lists for Eng. vs. Fr. /b d g p t k/ word-finally in monosyllables. French, which contrasts {voiced} stops (with fully voiced closure) with {vl.unasp.} stops, can be described as having a traditional 'voicing' contrast, unlike English. Mack found that, in English, vowels before [−voice] stops were 53% as long as vowels before [+voice] stops; in French, the ratio was 74%. Both the English sets of vowels were longer than the corresponding French sets (154 and 285 msec. for English vs. 133 and 182 msec. for French).

It might be thought that vowel duration in general is phonetically governed—with, let us say, vowel duration being proportional to the phonetic 'voicedness' of the following stop: the longest vowels might occur before the most voiced stops, and the shortest vowels before the most aspirated stops. Three arguments can be made against this account. First, Mack's data above indicate that such a prediction is false. Presumably the ranking for degree of phonetic voicing in her data (from least voicing to most) is Eng. [−voice] < Fr. [−voice] < Eng. [+voice] < Fr. [+voice], yet this does not correspond at all to the vowel duration ranking, Fr. [−voice] < Engl. [−voice] < Fr. [+voice] < Eng. [+voice]. Degree of lengthening does not appear to depend on degree of phonetic voicing: thus English has less closure voicing in its 'voiced' final stops than many other languages, given its tendency partially to devoice final [+voice] stops, yet more vowel lengthening. Hyman (1975:171) actually describes English as having more vowel lengthening BECAUSE it has less consonant voicing.

Consider further languages with three or four categories: Does vowel duration correspond to degree of phonetic voicing across those categories? Maddieson 1977 looked at vowel durations before voiced and voiceless aspirated and unaspirated stops in Assamese, Bengali, Hindi, Marathi, and Eastern Armenian. The predominant, but not uniform, result was for vowels to be longer before voiced than before voiceless stops, and longer before aspirated than before unaspirated stops. That is, taking labials as an example, the effect of consonants on vowel duration can be ranked as:

$$p < p^h < b < b^h$$

The effect of voicing is pairwise larger than that of aspiration. Presumably because Maddieson used a carrier phrase (but possibly because of measurement criteria), all his duration values are much smaller than most of Mack's; hence direct comparison between his three/four-category languages and her two-category languages is difficult. But once again we see that {vl.asp.} stops, the least phonetically voiced of all, do not co-occur with the shortest vowels.

Finally, consider English vowels before flaps. Phonetically, flaps may be voiced or voiceless; but Fox & Terbeek 1977 show that the duration of a preceding vowel, for speakers who make a distinction at all, depends on the underlying phonological value of [voice], and is not correlated with surface pho-

netic voicing. That is, vowel length does not depend mechanically on phonetic detail. These types of data all argue that vowel duration is not conditioned directly by the degree of voicing during a stop consonant; whatever phonetic correlations may be found, a more compelling correlation is that between vowel duration and the value of the phonological feature [±voice]. This correlation can be expressed only if some phonological feature is independent of phonetic categories.

(b) Cluster voicing assimilation is another common phonological rule which appears to apply generally across phonetic categories. Thus Polish has regressive voicing assimilation (Mikoś 1977), and a [voice] contrast of {voiced} vs. {vl.unasp.} stops; Danish, however, has progressive 'voicing' assimilation, but an aspiration contrast in initial position (Fischer-Jørgensen 1954).

(c) Evidence on fundamental frequency after stop consonant release indicates that phonological [voice] values are more important than phonetic voicing in determining pitch patterns. Hombert et al. 1979 note that English {vl.asp.} [−voice] stops and French {vl.unasp.} [−voice] stops perturb the F_0 of a following vowel by about the same magnitude. Current work by M. Caisse (at Berkeley) makes a similar point: {vl.unasp.} stops in initial position differ in F_0 across languages depending on whether they are [+voice] or [−voice] (my paraphrase). The F_0 differences must, of course, have some articulatory cause, and in that sense there must be two kinds of voiceless unaspirated stops; Hombert et al. suggest a difference in larynx height. Another possibility is a difference in extent of glottal opening, combined with a difference in oral occlusion durations. In any case, the distribution of these stops appears to depend on their phonological function, and can be simply stated in terms of phonological feature values.

Thus the distinction between phonological and phonetic features appears not only plausible but necessary, if rules such as those discussed above are to be properly defined across languages. Other rules that come to mind as depending on phonetic content of segments can best be described under the phonetic implementation rules, e.g. the fact that certain phonetic segments alternate as implementations of a [+voice] specification. Statements about segment frequencies in phonological inventories (e.g. lack of /g/ relative to /k/) also come to mind as depending crucially on segment content. In §3 below, it will be seen that one function of the phonetic level of representation is precisely to allow for this kind of statement about markedness, outside the set of synchronic rules contained in the grammar. It is conceivable that rules can be found which must refer to the phonetic implementation of a phonological segment or contrast; but the fact remains that, for at least some rules, the proposed phonological representation elucidates cross-language generalizations.

2.3. PHONETIC CATEGORIES. Having seen the need for phonological feature values which are phonetically somewhat abstract, we turn now to motivating the phonetic categories that implement these values. Careful consideration is required, since positing such categories along the phonetic voicing dimension is one of the main modifications of the *SPE* framework being proposed here.

These categories will be defined along a phonetic voicing dimension which, in initial position, coincides with the VOT dimension. Four steps are necessary. First, we must define VOT and the voicing dimension. Second, we must show that languages are limited to three contrasting categories—i.e., that three is the right number of major phonetic categories to posit. Third, we want to see that the three categories are the SAME three in various languages—in particular, the three posited here as provided by the phonetic theory. Fourth, we will examine the functional, extralinguistic basis of these categories.

2.31. DESCRIPTION OF VOICING AND VOT. As the time interval between the release of a stop consonant occlusion and the onset of vocal-fold vibration, VOT is meant as a cover term for various laryngeal and supralaryngeal events associated with this timing relation. The acoustic manifestations of the so-called VOT dimension are diverse. In practice, VOT is measured from acoustic displays as the time between the release burst and the first quasi-periodicity in the acoustic signal; this time interval is referred to in a narrow sense as 'acoustic VOT'. Graphically, VOT is usually represented as a continuum of time values. Stop release is the arbitrary reference point in time, 0 msec; and Voice Onset is measured relative to that point. Voice Onset occurring coincident with stop release is thus called 0 msec VOT. Voice Onset occurring before stop release is assigned a negative VOT value, and is said to lead the release. Voice Onset occurring after stop release is assigned a positive VOT value, and is said to lag the release. Positive VOT values to about 20–35 msec (depending on the place of articulation) are called 'short lag'; higher values are called 'long lag'. In general, stops traditionally described as being voiceless unaspirated (or devoiced) have short-lag VOT's, while stops traditionally described as being voiceless aspirated have long-lag VOT's.[2]

What then of other positions in the word, and their relation to the VOT dimension? Recall that {voiced} stops in initial position are characterized by voicing lead, i.e. voicing during stop closure. This voicing may or may not begin at the moment of closure, and may or may not extend throughout the entire closure; but at least some glottal vibration and low-frequency periodicity occur during closure. In other positions as well, {voiced} stops are characterized by voicing during stop closure. If the stop follows a sonorant, then the voicing of the sonorant and the stop closure will typically be continuous. No measurement can be made of the VOT, since voicing is already on. However, a measurement of closure duration, from the onset of this voiced closure to stop release, is most equivalent to a VOT measurement for {voiced} stops following pause. If a {voiced} stop is in final position, or before another stop, it may not be released; in that case, the only possible measurement would be the amount of voicing during the closure. In all these cases, regardless of which endpoints

[2] A lag VOT value reflects the time it takes the vocal cords to start vibrating; a long-lag VOT value indicates that the cords were quite separated at release, and took some time to get back together. Aspiration is friction noise generated at the still-open glottis by the flow of air through the vocal tract after stop release. Therefore, as long as the vocal cords are apart, air can flow between them and generate aspiration. It is in this sense that VOT is a measure of aspiration.

can be ascertained, the stop closure crucially contains some low-frequency vibration.

Similarly, in all positions a {vl.asp.} stop has a measurable amount of aspiration after the release. In medial positions before a sonorant, the measurement of VOT proceeds just as for initial position. While there may be one or two pitch periods of voicing at the beginning of closure, carrying over from a preceding voiced segment, closure voicing does not extend beyond that small amount. The only difficulty arises when the {vl.asp.} stop is not released into a sonorant, and so no VOT measurement is possible (since voicing does not begin again). In these cases the duration of the aspiration, which is quite audible, can often be measured from acoustic displays. Closure duration for {vl.asp.} stops is typically fairly short.

A {vl.unasp.} stop is one with no more than a few pitch periods of voicing during its closure, and a short interval of voicelessness following its release (if any). This category differs from {voiced} in its relative lack of closure voicing, and from {vl.asp.} in its short-lag VOT values following release. If there is no release, the two {voiceless} categories cannot be distinguished, so all unreleased stops without closure voicing are {vl.unasp.} Data from Polish listeners (Keating 1979) suggest that a stop closure must be about one-half voiced for a Polish medial /d/ to be heard.[3] Such phonological judgments do not, of course, directly indicate perception of the phonetic categories. However, they do indicate that a substantial range of degrees of closure voicing are not distinguished, and are categorized as [−voice]. We have already seen that fundamental frequencies after {vl.unasp.} stops vary; closure durations also are non-uniform. In sum, this category appears to show the most acoustic variation of the three—except that when VOT can be measured, it is confined to the narrow short-lag region of the VOT continuum.

2.32. NUMBER OF PHONETIC CATEGORIES. Lisker & Abramson 1964 point out that no language appears to contrast more than three categories along the VOT dimension. If more than three categories are contrasted at all, at least two will have similar VOT values; and these will differ along some other dimension. Certainly a survey of the traditional literature on a number of languages supports the view that three general categories are sufficient for descriptions of contrasts, and even for most cases of allophonic variation. The 51 languages surveyed by Keating et al. 1983 all use at least some kind of voiceless unaspirated stops in virtually every position (at least, according to the sources). As categories contrasting with them, fully voiced and voiceless aspirated stops are about equally common. Voiced aspirated, prenasalized, 'tense', and implosive categories are also found; but in each such case, the VOT value is the same as for one of the three basic categories. The languages which contrast no more than two VOT categories in any one environment include English, German, Spanish, Polish, French, Tagalog, Dutch, Swedish, Mandarin, Can-

[3] Other unpublished data of Leigh Lisker and Patti Jo Price suggest a similar situation for English. However, since {vl. unasp.} stops typically have one or two pitch periods of voicing at the beginning of closure, and since {voiced} stops are often not voiced throughout the closure and release, it is not obvious what is meant by listening for 'closure voicing'.

tonese, Russian, and Kirghiz. Languages using all three, at least in some environment, include Thai and Eastern Armenian. Languages using these three plus some other include Hindi and other languages of India, which add voiced aspirates. In sum, cross-language distributional evidence supports Lisker & Abramson's idealization of three basic voicing categories for contrastive and allophonic use in initial position. And in fact, they are also sufficient elsewhere, since no greater number of contrasts is found in any other position.

2.33. THE PARTICULAR PHONETIC CATEGORIES. Lisker & Abramson 1964 observe that, while differences are found across languages in the exact VOT values used, rough agreement exists on the areas of the VOT continuum. These areas correspond to the traditional phonetic categories of prevoiced, voiceless unaspirated, and voiceless aspirated stops, taken here to be components of the feature system. Following this observation, it is generally assumed that some universal framework provides these categories, and that the cross-language variability is secondary in importance. Let us consider this assumption in some detail.

If no language requires us to posit more than three categories along the phonetic voicing dimension, can we go further and say that the same three categories are used in different languages? It could be that languages divide up the continuum differently, even if it is always into three or fewer categories. To be sure, the identity of categories across languages is not crucial to the formal system being proposed; but such identity would offer a further constraint on possible phonetic implementations, and would be relevant as regards the contribution of a universal phonetic theory to feature systems. Here I will review the evidence that there are really three discrete categories—rather than fuzzy areas of the continuum—found across a variety of languages.

As a baseline for the division of the voicing dimension into three categories, let us refer to languages with three or more contrastive phonetic categories. Lisker & Abramson 1964 give data on VOT in initial position for a number of such languages. Thus in Thai, which has three contrasting categories in initial and medial position for labials and apicals, the {voiced} stops have lead VOT values, up to about −40 msec; the {vl.unasp.} stops have short lag values, from 0 to +10, +20, or +30 msec VOT, depending on place of articulation; and the {vl.asp.} stops have higher VOT values. Donald 1978 shows that Thai listeners have discrimination peaks at about −20 and +20 msec VOT, corresponding to these distributions.

The Thai data show an obvious gap between the lead and short-lag VOT values, i.e. a lack of low negative values. This gap can also be seen in data presented below in §3, and in fact seems to be the general rule for languages, regardless of the number of voicing categories. Whatever its cause, it has the effect of clearly separating {voiced} from {vl.unasp.} stops acoustically, and makes the {voiced} category largely coincide across languages. Furthermore, the {vl.unasp.} and {vl.asp.} categories, as they occur in Thai, appear quite typical of languages that contrast such categories. Inspection of the VOT values for {vl.unasp.} stops in a variety of languages indicates that they typically lie within a narrow area of the VOT continuum, and are essentially normally dis-

tributed within that area. (This normal shape is more apparent in graphs plotted to a finer scale than 10 msec intervals, as is done below and in most spectrogram-based studies.) Although slight differences occur across languages in these distributions, in general we can say that labial {vl.unasp.} stops have VOT values up to about +20 msec, apicals up to about +30 msec, and velars up to about +40 msec. The distributions appear to be constrained on the low VOT side by the 0 value and the gap, and on the high VOT side by the long-lag values with which they contrast. Crucially, the division between short and long lag is quite similar across languages, and the corresponding perceptual boundary is also similar across languages (for a summary of the literature, see Keating 1979).

What about languages contrasting two categories, {voiced} and {vl.unasp.}? Because of the gap in VOT values, we expect the {voiced} values to look like those of Thai etc.; but there is no reason for the {vl.unasp.} values to be constrained on the high VOT side, and we might expect to see an acoustic contrast of lead vs. general-lag VOT. Usually this does not happen; {vl.unasp.} values in such cases typically resemble those that are constrained by a {vl.asp.} category. In medial and final positions, {vl.unasp.} stops are less clearly distinguished from {voiced} stops, since the amount of voicing during closure varies continuously.

To the extent that there are three fixed phonetic categories whose values are constrained in this way, the case of stop consonant voicing provides a counterexample to the ideal of 'maximal dispersion' (Liljencrants & Lindblom 1972), by which languages keep their contrasts maximally distinct in the phonetic space. On that hypothesis, the most favored contrast should be {voiced} vs. {vl.asp.}, with extreme VOT values; but such a contrast is at best rare in languages (Flege 1979). The most common category across all environments and languages is {vl.unasp.}; it is nearly universal, both alone and in contrast with one or both of the other categories, and its acoustic values are highly constrained.

However, some cases of phonological categories have phonetic values beyond these observed category limits. Thus [+voice] stops which are basically {voiced} may have a few {vl.unasp.} tokens; and [+voice] stops which are basically {vl.unasp.} may have some {voiced} tokens. Furthermore, {vl.unasp.} stops, as in the Polish data to be presented below, may display a tail of values into the higher VOT range, resulting from high vowel contexts, or from extra emphasis, or for no apparent reason other than spreading over the phonetic space, as Pol. /k/ does. Still, such spreading need not occur: English speakers do not generally prevoice more often in more careful speech.[4] And Spanish

[4] In Lisker & Abramson 1967, speakers were shown to produce more prevoicing in minimal-pair readings, which were assumed to be a more careful style. However, the opposite was found by Flege & Massey 1980, who also found that speakers prevoiced the same amount as a session progressed. (They had thought that, as speakers relaxed, their more casual speech would have less prevoicing.) In support of this latter interpretation of the relation of prevoicing to carefulness of speech, Malsheen 1980 shows that mothers do not prevoice more in speaking to their children, although they do aspirate more strongly; and Chen et al. 1980 show that speakers do not prevoice more in speaking 'clearly' to deaf listeners.

speakers do not spread their {vl.unasp.} /p t k/ into other phonetic categories even in contexts in which VOT is not the only contrastive dimension. Dent 1976 hypothesizes that since, in running speech, most instances of Sp. /b d g/ are spirantized, while /p t k/ are stops (i.e., the voicing contrast co-occurs with a manner contrast) the {vl.unasp.} /p t k/ category should therefore be free to expand into the {voiced} (or, we might add, the {vl.asp.}) category. However, her acoustic measurements of VOT for [− voice] stops, in contrasting and non-contrasting contexts, showed no difference: the Spanish stops did not spread into other VOT regions. Thus languages may have phonetic spreading, but do not always do so.

The discussion of the three phonetic categories thus far has been limited to the question of acoustic similarities and constraints across languages. There may also be articulatory similarities. Although cross-language physiological data are limited, it appears that glottal gestures for {vl.asp.} stops, for example, are quite similar in English and Swedish (e.g. Löfqvist 1980, Yoshioka et al. 1981). And despite various acoustic differences, Weismer 1980 reports a constant voiceless (open glottis) interval for [− voice] stops and fricatives, even across place of articulation. Word-initial clusters of /s/ plus a [− voice] stop also appear to have the same constant opening (cf. Yoshioka et al.), accounting for the fact that such stops are {vl.unasp.}. The glottal gesture may be quite similar across segments, contexts, and languages because it is a ballistic movement, not controlled for duration or extent. The same gesture can be used for both aspirated and unaspirated stops, so long as the timing of the consonant closure is manipulated. If it is, then [− voice] stops of either category will have a constant glottal definition; however, the references cited indicate that some {vl.unasp.} [− voice] stops have a very small glottal opening, or none. Thus the goal of the speaker may be the state of no vibration, as much as the state of an open glottis. That the goal may be statable in terms of vibration is suggested by English speakers who prevoice their [+ voice] initial stops. These speakers use an extra articulatory gesture to achieve the same output, {voiced}, as is found in medial position, where {voiced} stops require no extra gestures (cf. Westbury). However, Hayes 1983 presents evidence, from Russian voicing assimilation, that speakers' goals can be glottal states rather than vibration. Over-all, then, the issue of what is being controlled to achieve what goals remains problematic; for present purposes, we can note only that various possibilities exist, and that the phonetic categories must be defined in all relevant domains. It is likely that, in some ways, the three phonetic categories have a functional motivation in terms of articulatory mechanics and control.

Evidence also exists for a perceptual basis for three fixed phonetic categories. Boundaries between these phonetic categories can be elicited from listeners in languages where they do not represent a linguistic contrast: the boundaries are present, but cannot have been induced by phonological experience. Both Abramson & Lisker 1972 and Williams 1974 found extra discrimination peaks for Spanish labials at about +25 msec VOT, between {vl.unasp.} and {vl.asp.} categories, besides the (linguistic) peak at about −10 msec VOT. English listeners can also show such a non-linguistic peak, between {voiced} and {vl.unasp.} stops, given the right experimental procedures (Pisoni 1977, Carney

et al. 1977). And Kikuyu listeners, for whom VOT differences correlate with a contrast between {vl.unasp.} and prenasalized apical stops, show two discrimination peaks for labial stops, at −15 and +20 msec VOT (Streeter 1976). Thus we see that not only do languages agree on roughly where the perceptual boundaries fall, but listeners whose linguistic experience does not include those boundaries also put them in the same place.

2.34. FUNCTIONAL BASIS. The above evidence suggests some extra-phonological basis for the two category boundaries; the auditory system seems to impose a discontinuity on the perception of VOT that is exploited by linguistic categorization. Evidence that the discontinuity is also extra-phonetic comes from studies of animal perception of VOT. Perception of VOT has been studied in rhesus monkeys (Waters & Wilson 1976) and chinchillas (Kuhl & Miller 1975, 1978, Kuhl 1978). The methodology of Kuhl & Miller allows a more direct comparison with adult humans, although only in the lag VOT region.[5] The result is that the animals show the same boundaries as adult English speakers on the same stimuli, indicating common non-linguistic properties of the auditory system.

Further research has been directed at identifying such psycho-acoustic factors, starting from a hypothesis by Hirsh 1959, 1975 about limits on the ability of the auditory system to resolve and to sequence acoustic events separated in time. Suppose that voicing onset and stop release are two separate events that must be perceived in the right sequence in order for a voicing judgment to be made; if they are too close in time (say, 20 msec), they cannot be ordered accurately; they will be perceived as being simultaneous, and the perception will be {vl.unasp.} Following a suggestion to this effect by Stevens & Klatt 1974, the work of Hirsh was applied by Miller et al. 1976 and by Pisoni 1977 to the perception of non-speech stimuli designed to be analogous to VOT stimuli. In discrimination tasks, most listeners showed peaks at the analog values of about +20 and −20 msec VOT. Some debate has ensued as to whether these results are enough to account for all VOT perception. More recently, Soli 1983 has shown that spectral effects, such as F_1 onset frequency, are crucial to the discontinuous perception of VOT. Such discontinuities along the VOT (and possibly other relevant dimensions) indicate that some perceptual basis exists for the distinctiveness of the three phonetic categories. Whatever that basis may turn out to be, it would constitute a functional explanation for the

[5] The most interesting part of their research was the use of three VOT continua—one at each of the labial, alveolar, and velar places of articulation. This allowed them to test the effect of place of articulation on VOT boundaries as known from adult humans. The chinchillas' boundaries did in fact vary according to place of articulation like those of humans, from about +25 msec for labials to about +42 msec for velars. However, their identification functions were less steep. Such sharpening of category boundaries with linguistic experience is not unique to the human/animal comparison: the same kind of result was shown by an experiment in which Czech and American listeners divided a continuum of Czech words, varying in their vowel duration, between pairs of phonemic short and long vowels (Keating 1978). The cross-over points for the Czechs (who were performing a linguistic task) and the Americans (who were performing a non-linguistic task) were the same, but the boundaries were much steeper for the Czechs.

consistency of the categories across languages. While the phonetic categories have a formal representation in the grammar, they also have a non-linguistic perceptual motivation, plus whatever articulatory bases may be identified.

3. APPLYING THE SYSTEM TO DATA. Given these three major phonetic categories, how are they used in languages as implementations of phonological feature values? Here I will consider phonological to phonetic category-mapping, across allophonic variation, within each of three languages: Polish with a rather simple system, and then English and German with more complex systems (§3.1). It will then be seen that the descriptive framework motivated by these language-specific considerations is also sufficient for cross-language comparisons (§3.2). This will necessitate a fairly detailed examination of a certain amount of acoustic data.

3.1. CONTEXTUAL ALLOPHONES IN THREE LANGUAGES. First consider voicing in Polish, where the [voice] contrast is extremely straightforward. The language contrasts /b d g/ with /p t k/ in initial and medial positions. Word-final stops before pause are neutralized to [p t k]; but before a sonorant-initial word, this neutralization is optional. Before an obstruent-initial word, a final stop is subject to regressive voicing assimilation (cf. Mikoś).

Figure 1 (cf. Keating et al. 1981) shows acoustic measurements of VOT in post-pausal initial position. A list of 42 disyllabic words beginning with all phonologically legal sequences of a stop consonant [b d g p t k] followed by a vowel [i e ẽ a ã o u ɨ] was read ten times each by five monolingual Polish speakers in Łódź, Poland. Palatalized allophones of /t d/ before [i] were not included. The VOT distributions for the voiceless stops show a normal distribution in the short-lag region, but are skewed with some long lag values because of high vowel contexts. (High vowels generally cause higher VOT values, since pressure in the oral cavity behind the constriction is vented more slowly.) Nonetheless, it is striking how little overlap appears across vowel contexts.

Figure 2 (cf. Keating 1979) shows acoustic measurements for medial post-stress position for [t d]—including lag VOT values for items without closure voicing and closure duration for items with closure voicing, since voicing continues largely uninterrupted throughout the V[d]V sequence. Medial [+voice] stops have somewhat shorter closure, and therefore less measured voicing than the initial stops; but the medial [−voice] stops have VOT values which are very similar to the initial ones. In Polish, therefore, stops in initial and medial positions are closely similar. [+Voice] stops have voicing during closure, and sometimes through the burst. [−Voice] stops do not, with voicing always beginning after the burst. That is, Polish appears to be an uncomplicated, uncontroversial case of a [voice] contrast in which the surface [+voice] members are (fully) {voiced}, and the surface [−voice] members are {vl.unasp.}

Compare Polish now with English phonetic implementation. It is well known that, in initial position, English [+voice] stops are {voiced} or {vl.unasp.}, and that the [−voice] stops are {vl.asp.}. This result was replicated in a way directly comparable to the Polish study described above. Figure 3 shows VOT measurements for English post-pausal initial stops before 12 vowels, from a list of

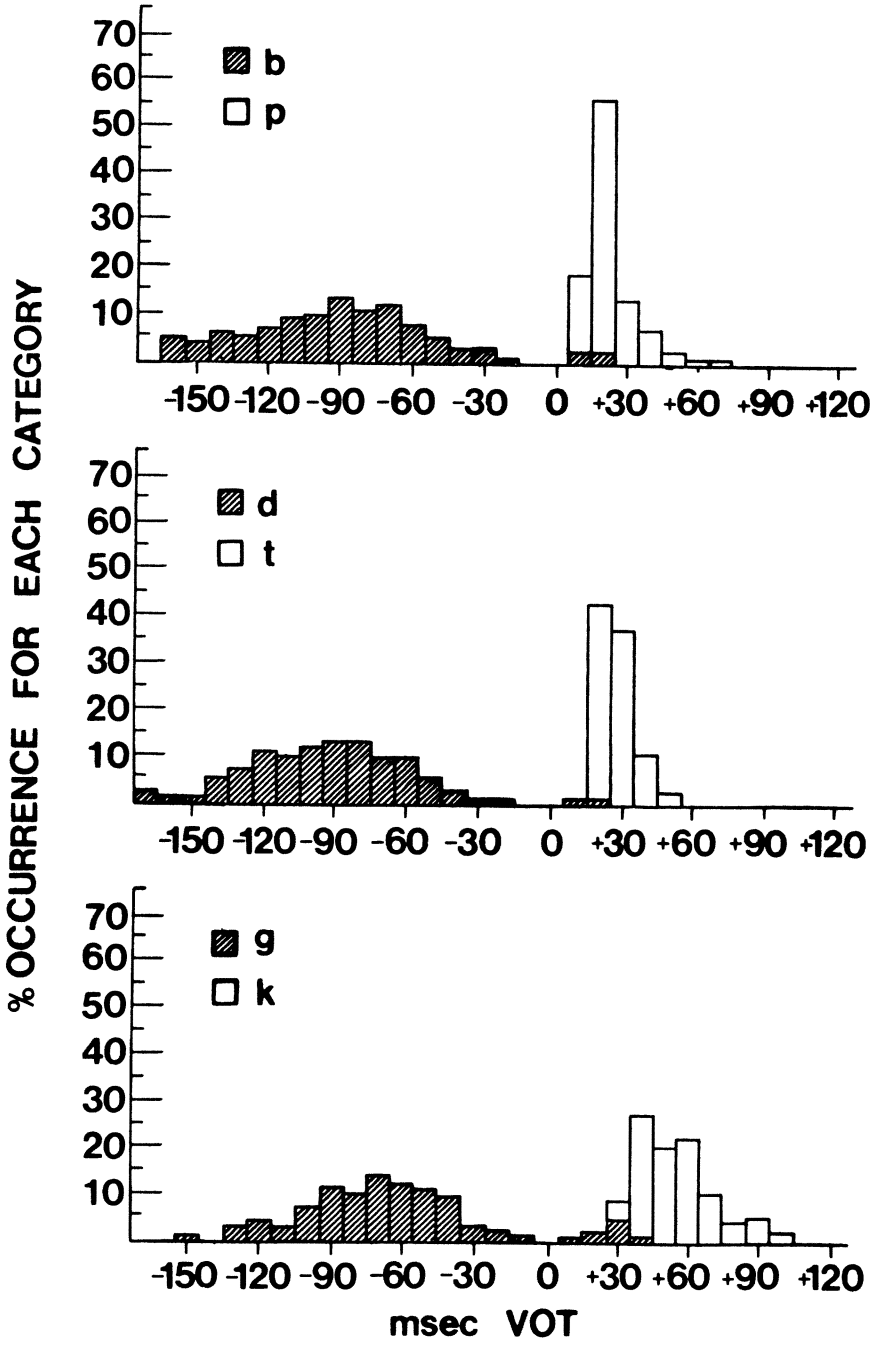

FIGURE 1. VOT values for Polish utterance-initial stops at three places of articulation before eight vowels. Measurements are for five speakers reading 42 words, ten times each.

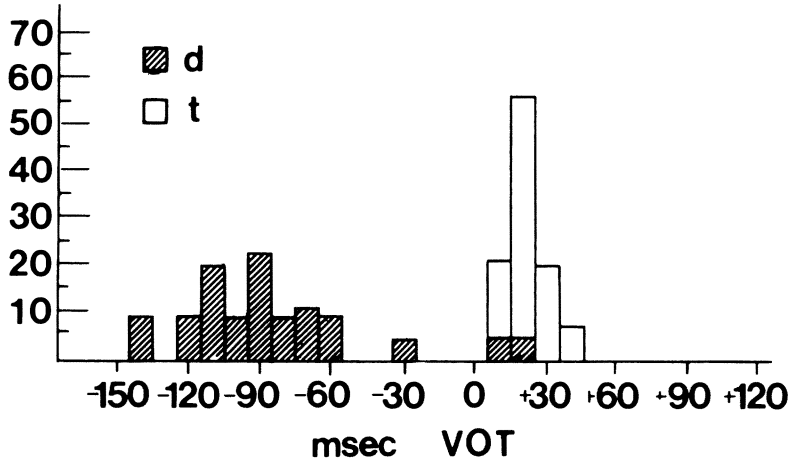

FIGURE 2. VOT values for Polish medial /t/ in *data*, *tata*, and *rata*, read once or more by 24 speakers; and voiced closure durations or lag VOT values for Polish /d/ in *rada*, read once or more by 22 speakers.

79 disyllabic words read four times by a monolingual American English speaker, and twice by a second speaker (cf. Keating et al. 1981). It can be seen that English divides up the VOT continuum differently from Polish (Fig. 1)—with some lead values, but mainly short lag vs. long lag.

Next, compare these initial English values with medial post-stress values, which have generally been noted to differ (Lisker & Abramson 1967, Flege & Brown 1982). Although /t d/ were used in the Polish data, in American English they are generally flapped medially, so the English data presented are for /p b/. Six speakers read a list of words containing medial /p/ or /b/ before a reduced low vowel, resulting in 25 /p/ measurements and 24 /b/ measurements. The lag VOT values for /p/ are displayed in Figure 4 (as the Polish data were shown in Fig. 2). While all the /b/'s had a substantial amount of closure voicing, the figure shows lag VOT values for stops which had any voicelessness during their closures, and closure durations for the others. The two languages differ in that the voiced closure durations are shorter in English than in Polish. Over-all, however, the values are remarkably similar in this context.

We see then that English has more variation in its phonetic implementation of [±voice] than Polish does. Such variation has been noted before; e.g., Kahn 1976 contains a valuable study of the phonetic variants of Eng. /t/, including [t^h], [t], [ɾ], and [$t^?$]. General syllable-based rules for deriving these allophones are proposed, largely also extendable to /p k/. However, no corresponding analysis of voiced stops is given. It is often assumed in the phonological literature that only voiceless stops are subject to such rules—that (ignoring flapping) /b d g/ are always 'voiced', while /p t k/ are voiceless but [±aspirated]. In fact, though, as we have already seen in comparing initial stops with stops in one medial context, /b d g/ also vary phonetically, being sometimes voiced throughout their closures, and sometimes voiceless unaspirated.

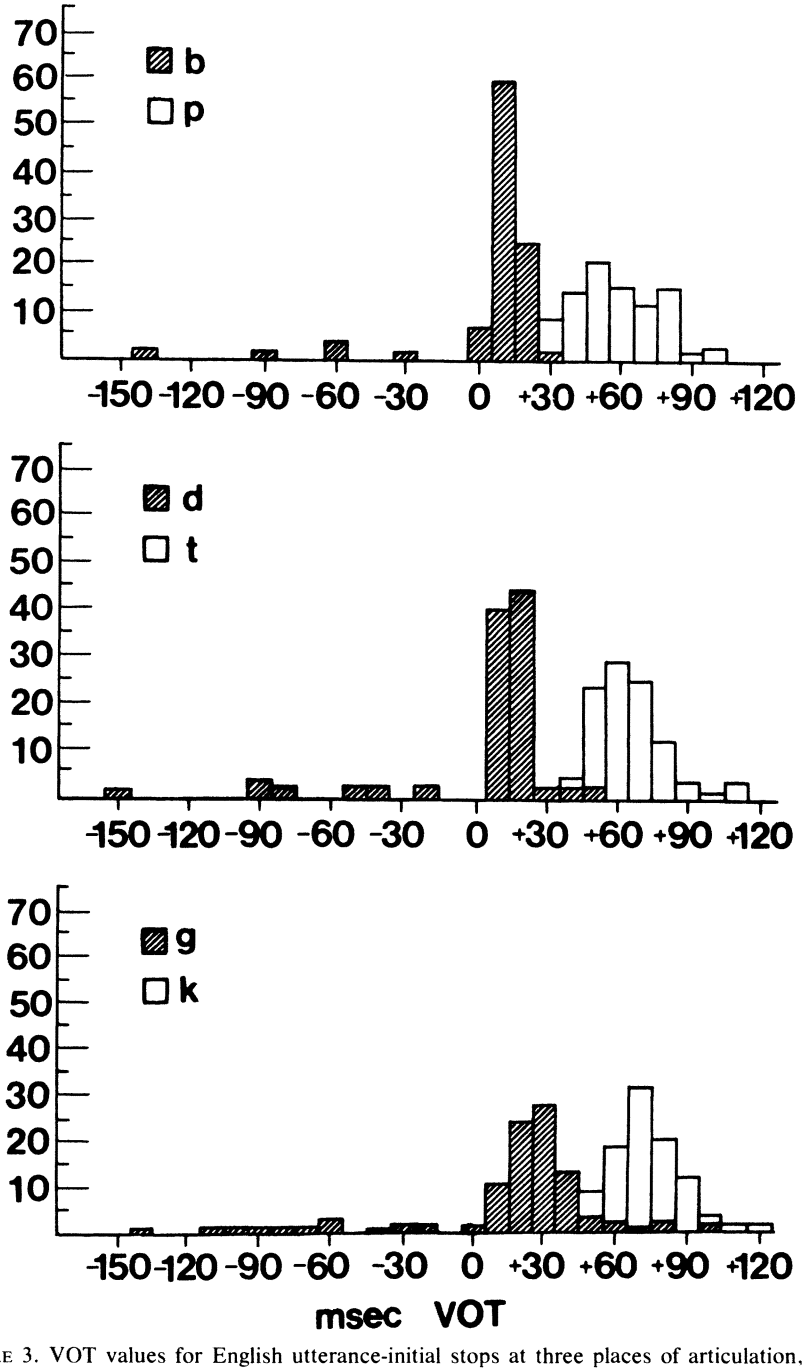

FIGURE 3. VOT values for English utterance-initial stops at three places of articulation, before twelve vowels. Measurements are for two speakers reading 72 words—one, four times each, and the other, two times each.

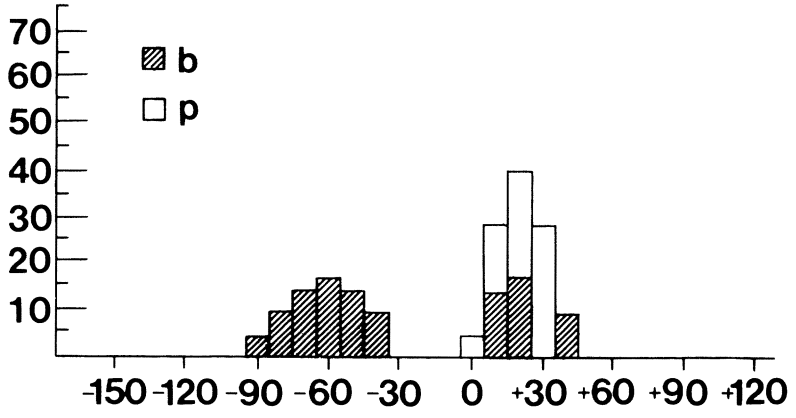

FIGURE 4. VOT values for English medial /p/, and voiced closure durations or lag VOT values for English medial /b/; from a list of words read once by six speakers.

In addition to variation across contexts, English has also been shown to display substantial differences among speakers. Previous studies (Lisker & Abramson 1964, 1967, Smith & Westbury 1975, Flege 1982) have shown that some English speakers produce some or all of their initial voiced stops with prevoicing rather than short-lag VOT, although other conditioning factors are also involved.[6] Furthermore, Flege 1982 shows that speakers' glottal gestures, at least to the extent that they can be determined with an electroglottograph, differ in the production of English initial /b/. Some speakers open the glottis, while others do not; however, no single speaker has two glottal timing patterns for initial /b/. This is true even though individual speakers sometimes prevoice, and sometimes do not; i.e., the observed glottal patterns are quite consistent across tokens, while the occurrence of prevoicing is not. For a given speaker, the acoustic variation must result from additional articulatory mechanisms.

To summarize to this point, Polish implementation is quite simple: [+voice] as {voiced}, and [−voice] as {vl.unasp.} English shows more variation, both across positions and speakers: [+voice] is implemented as {voiced} and as {vl.unasp.}, [−voice] as {vl.unasp.} and as {vl.asp.} However, to be more certain of the variation that occurs in English, more data from a single group of speakers have been obtained.

Six Americans, four female and two male, read a list of 215 words, each containing one of the six stops of English before a low vowel. Stops occurred either initially or intervocalically; the following vowel had primary stress, secondary stress (full vowel), or was reduced. (The medial unstressed /b p/ tokens are the ones already presented.) There were six words for all but one of the 36 combinations of stop, position, and stress. For each recorded token, up to three measurements were made from a computer-implemented oscillographic

[6] Such as place of articulation, vowel context, and speaker's sex.

display. For post-pausal stops, VOT was measured as described before. For post-vocalic stops, the duration of voiced and voiceless closure intervals was measured, as well as lag VOT. The sum of the two closure measurements gives the total closure duration. Figure 5 shows a schematic waveform for each

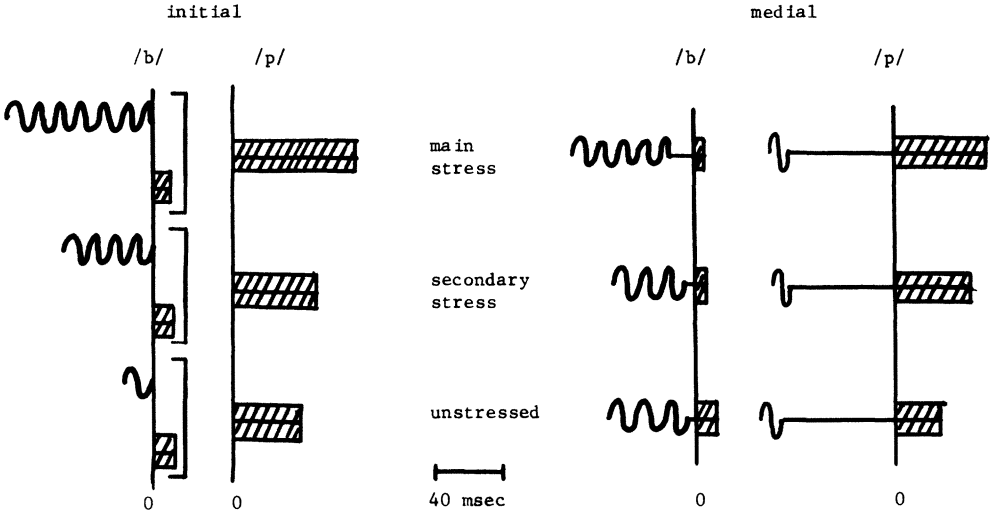

FIGURE 5. Phonetic implementations of English /b p/ in initial and medial positions, for three stress environments. The vertical lines, labeled with zero, represent the moment of stop release. Each schematic shows mean values for three measurements: duration of closure voicing (wavy lines), duration of closure voicelessness (straight lines), and duration of voicing lag (striped bars). For initial /b/, prevoiced and short lag values are graphed separately.

context, for the labials only; each section of the waveform represents the mean across the six speakers for that measurement.

One general pattern across the three places of articulation is that initial [−voice] stops usually have long-lag VOT values, confirming Kahn's results; and initial [+voice] stops usually have short-lag VOT values, regardless of stress. The degree of aspiration of the [−voice] stops is somewhat sensitive to stress level—although, for initial /t/, the VOT values are the same before vowels with secondary stress and no stress.[7] The [+voice] stops are occasionally prevoiced; the occurrence and duration of prevoicing vary with place of articulation and stress. Another pattern is that medial [+voice] stops generally have at least some voicing during closures. Medial [−voice] stops have

[7] The result that VOT value for initial [−voice] stops does not depend on the stress of the following vowel would seem to refute the claim in the literature (e.g. Hoard 1971, Kahn 1976) that aspiration correlates with stress. However, that claim might still be correct. The high VOT values obtained for initial [−voice] stops in stressless syllables result from reduction and devoicing of the entire stressless syllable. That is, the high VOT values are not, strictly speaking, the same as aspiration of the initial consonant. This reduced syllable devoicing does not necessarily require a special mechanism. Lack of stress could cause shortening of the syllable, while a minimum time interval could be required to bring the vocal cords into position for vibration; the syllable would fall entirely within this interval and so would be voiceless.

long-lag VOT values before main and secondary stress. Before reduced vowels, the values are much lower, falling within the range of short rather than long lag. Medial /t d/ before reduced vowels are, of course, typically flapped.

We see, then, that fairly consistent stop consonant variation really occurs across contexts in English. Is this phonetic variation just an idiosyncrasy of English, or is it more general? This pattern was investigated further by looking at a subset of these environments for one speaker of standard German, another language with an initial aspiration contrast. This speaker read a set of words containing labial or alveolar stops before low and mid vowels—initially with main or secondary stress, and medially before main stress or a reduced vowel. Results of spectrographic analysis are summarized in Figure 6. Initial [− voice]

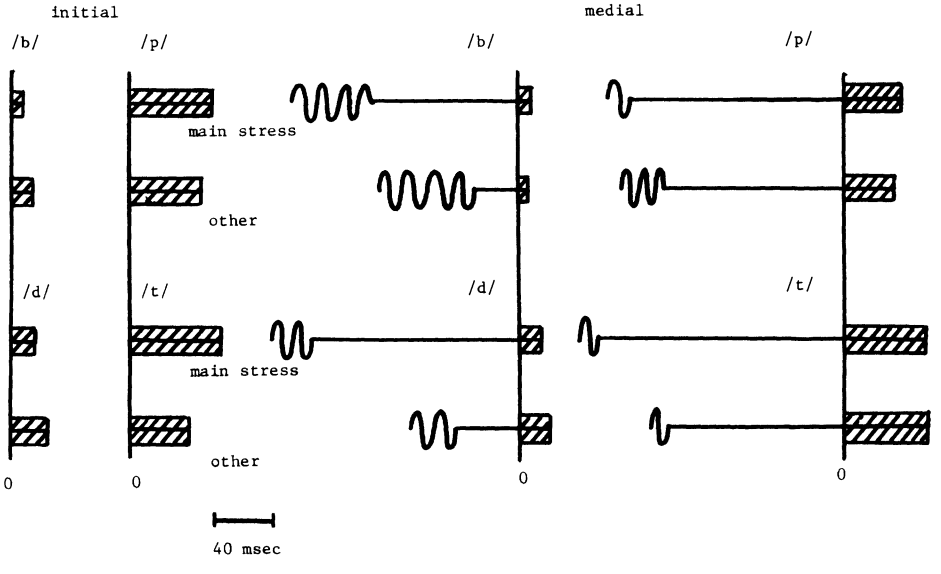

FIGURE 6. Phonetic implementation of German /b p d t/ in initial and medial positions for various stress conditions. The 'other' stress for initial stops is largely secondary stress, while the 'other' stress for medial stops is largely non-stress. The form of the graph is like that of Fig. 5.

stops are all somewhat aspirated. Medial [− voice] stops are somewhat less aspirated; but surprisingly, even with this relatively slight degree of aspiration, they still contrast with voiceless unaspirated [+ voice] stops.[8] The [+ voice] stops are typically {vl.unasp.} both initially and medially. The exception to this generalization is medial /b/ before a stressless vowel, which is often voiced; i.e., /b/ and /d/ differ with respect to medial voicing. Basically, however, the German speaker shows less variation in use of major categories, and more variation in the degree of voicelessness associated with the different environments, than the English speakers did. Not only English, then, shows positional

[8] William Moulton informs me that German speakers are taught always to aspirate [− voice] stops in Standard German, and that therefore the failure of this speaker to de-aspirate before stressless vowels, as in English and other languages, may result from an artificial style of speech.

variation; but the variation in the two languages is not the same. Each language must have its own implementation rules.

From descriptions in the literature, it seems likely that such languages as Norwegian (Vanvik 1972), Mandarin (Dow 1972), and Kirghiz (Hebert & Poppe 1963)—all having aspirated initial [− voice] stops—show some variation of this sort. For example, the [+ voice] stops of Kirghiz are described as being voiceless unaspirated initially, medially adjacent to another 'voiced' stop, and finally after a long or heavy-stressed vowel, but voiced elsewhere. The [− voice] stops are described as being voiceless aspirated initially, and voiceless unaspirated elsewhere. By this account, some neutralization occurs in clusters and word-finally after certain vowels; otherwise two contrasts are possible: an initial one of aspirated vs. unaspirated stops, and a medial/final one of unaspirated vs. voiced stops. While more thorough acoustic studies are needed of such cases, the general pattern appears to be that languages like Polish, without contrastive aspiration, show little allophonic variation; but languages like German and English, with contrastive aspiration after pause, show a great deal of positional variation (cf. Keating et al. 1983). That is, there may be some general tendency against medial aspirated stops which influences the phonetic implementation of [− voice] stops. At the same time, there are other cross-language differences, and the implementation rules of aspiration languages will be complex.

3.2. Cross-language description. By looking at considerable acoustic data, we have seen three degrees of complication in the phonetic implementation of a [± voice] contrast. Polish, like other non-aspiration languages, is the least complex case: [+ voice] is always {voiced}, and [− voice] is always {vl.unasp.} English, for speakers who prevoice, and German—as spoken by at least one person—are somewhat more complex, since two implementations are used in each case. For this kind of English, [+ voice] is {voiced}, and [− voice] varies between {vl.unasp.} and {vl.asp.} For German, however, [+ voice] varies between {voiced} and {vl.unasp.}, while [− voice] is {vl.asp.} The most complex case is English for speakers who do not prevoice: [+ voice] varies between {voiced} and {vl.unasp.}, and [− voice] varies between {vl.unasp.} and {vl.asp.}, sometimes in a correlated fashion.

While there is clearly no general pattern across languages, it is also true that phonetic implementation of [voice] is constrained somewhat; thus, not surprisingly, {voiced} stops are never [− voice], and {vl.asp.} stops are never [+ voice]. Regardless of which phonetic pair is contrasted in a given case, the one implementing [+ voice] is always phonetically more voiced than the one implementing [− voice]. So English speakers who do not prevoice often implement their phonological [± voice] contrast as {voiced} vs. {vl.unasp.} in certain contexts, or as {vl.unasp.} vs. {vl.asp.} Whichever pair is chosen, the more voiced one implements [+ voice], and the less voiced one implements [− voice]. This system is illustrated in Figure 7 for non-alveolars in two representative contexts. In either case, the {vl.unasp.} category is used, sometimes as [+ voice] and sometimes as [− voice].

Comparing English with Polish implementation, it appears that Fig. 7 can be taken to represent two different languages, as well as two different environ-

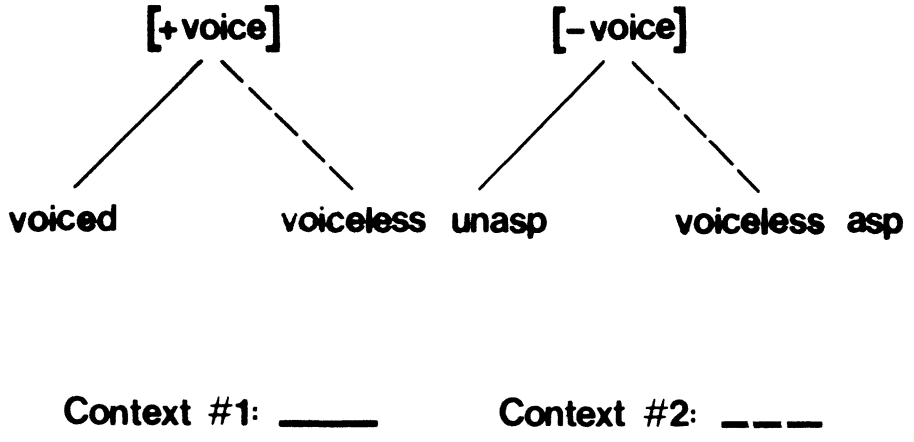

FIGURE 7. Schematic of phonetic categories used to implement the [voice] feature in different contexts.

ments. The phonological identity of the {vl.unasp.} category will be different in various languages, depending on the phonetic contrast used. It will be phonologically [+voice] if it is the left member of the [voice] contrast, and it will be phonologically [−voice] if it is the right member of the [voice] contrast. Thus the [voice] feature's plus and minus values can be phonetically interpreted only in a relative sense, as 'more' and 'less' voicing, not 'with' and 'without'. Such a relative definition is also required by the somewhat abstract nature of the phonetic voicing dimension, whose physical characteristics will vary with position and context.

Depending on the phonetic implementation used, the phonetic {vl.unasp.} category may be unused, may implement [+voice] stops, or may implement [−voice] stops; i.e., it acts as a 'swing' category for phonological implementation. If it is known that a class of stops in a language has short-lag VOT's, it cannot be said whether they are [+voice] or [−voice]. However, this does not mean that no phonetic differences exist between [+voice] and [−voice] short-lag stops, as we have already seen. It simply means that the possible phonetic differences are never employed contrastively.

3.3. PHONETIC DETAIL RULES. The rules of phonetic category implementation are thus language-specific, and draw on the universal set of phonetic categories. Consider now the rules that map the three phonetic categories into their concrete physical realizations. Are these rules language-specific or universal? Superficial examination would suggest that the former must be the case. For example, compare the VOT distributions for short-lag stops in Polish and English (Figs. 1 and 3). It can be seen that the distributions for Polish are about 5 msec VOT higher than those for English, although both sets are clearly short-lag VOT. We may have to state that Polish short-lag stops are slightly more aspirated than English ones by means of slightly different quantitative rules for the two languages.

However, such a rule would not be necessary if the observed variation could

be derived by a general principle. One such principle would be 'polarization' of two adjacent categories along the voicing dimension. According to this principle, within the limits of the implementation chosen—i.e. the phonetic categories—there is maximal separation of the distributions of values. In effect, this principle says that the 'dispersion' theory of Liljencrants & Lindblom operates not over a continuous phonetic space or dimension, but within the discrete categories of the phonology.

In this account, the contrasts in Polish and in English are heightened through polarization, with a differential effect on voiceless unaspirated stops. Data on Spanish vs. English VOT can be interpreted in the same way. Ladefoged & Kim 1965 compare Eng. /b/ with Sp. /b p/, looking for VOT differences. They find roughly that Eng. /b/ covers a less extreme range of VOT values than the two Spanish categories combined—i.e., it includes lead values (but not as voiced) and lag values (but not as voiceless). They propose a fourth phonetic category corresponding to the Eng. /b/ distribution. However, these data can be handled with three categories and the polarization principle. Under this account, Eng. /b/ has a bicategory distribution as lead and short-lag VOT values, while Sp. /b/ and /p/ are each realized by a single VOT category. The English /b/ lag VOT values are then polarized away from the higher-lag values which implement Eng. /p/. The Sp. /b/ distribution is polarized away from Sp. /p/, toward long lead values, while the Sp. /p/ distribution is polarized in the opposite direction.

However, some of the data presented here and elsewhere on languages with initial contrasts of {vl.unasp.} vs. {vl.asp.} stops are less clearly explained by polarization. The English and German data presented above, with the similar data on Swedish in Keating et al. 1983, show that, even for similar data sets, some differences in VOT are found for the same phonetic category implementations. Furthermore, English basically has more aspirated [−voice] stops—even though, with greater likelihood of [+voice] {voiced} stops, there is less pressure for dispersion. The German medial pre-stress contrast is so marginal in terms of VOT that polarization does not seem to be at work there, either.

Data from more languages are clearly required to verify this hypothesis about distribution dispersion. While polarization may turn out to be a wrong hypothesis, the principle involved is attractive: surface differences across languages are accounted for—not by an elaborate set of phonological distinctions or language-specific phonetic rules, but by the interaction of phonetic mapping constraints and a universal phonetic principle.

A similar case arises in the treatment of stress and positional allophones, as discussed above: To what extent do stress and position, as such, universally affect quantitative detail and even phonetic category implementation? The issue here is whether rules of category choice are the only phonetic rules for voicing in the grammar. If differences in the way a given category appears are found across languages with identical phonological systems, then the grammars of those languages will have to contain rather specific quantitative rules. Obviously, a fair degree of uniformity across similarly specified contrasts is predicted. However, it must be noted that such comparisons of distributions re-

quire data sets that are large enough to be reliable, and that have similar vowels following the consonants in question—since vowel identity can affect VOT. It is possible that at least some cross-language differences observable in the literature result from failure to satisfy one or both of these requirements. In any event, universal principles must be entertained as hypotheses before we resort to language-specific quantitative rules, since they might otherwise be missed.

4. NEUTRALIZATION AND MARKEDNESS. The framework that has been proposed distinguishes a phonological from a phonetic level in order adequately to characterize cross-language and cross-context differences and similarities. Does this proliferation of formal levels, and the concomitant abstractness of the phonetic categories, play any other role in the grammar? I will now suggest that the formal, rather abstract phonetic level proposed here provides an additional advantage: that it is the correct level for statements about markedness phenomena, and thus provides a new approach to an old problem.

Markedness phenomena have generally been treated as phonological—as a property of phonemes and/or phonological features (Trubetzkoy 1939, Jakobson 1962, Chomsky & Halle 1968, Kean 1975). Thus 'voiceless' stops are found in more languages, and in more environments, than 'voiced' stops; and voiceless stops are taken as the unmarked category value. Languages without a voicing contrast will have voiceless, rather than voiced, stops. And when neutralization is not contextual, then, as Trubetzkoy pointed out, voiceless rather than voiced stops are found in positions of neutralization (where neutralization refers either to phonological rules or defective distribution). The same sort of preference holds for unaspirated over aspirated stops, and those unaspirated stops are typically voiceless; but discussions of markedness rarely relate voicing to aspiration.

A simple physical generalization underlies the apparent phonological one: physically, most environments favor {vl.unasp.} stops over other categories. Westbury & Keating 1980 use a computer implementation of an electrical analog model of vocal-tract aerodynamics to demonstrate that, in absolute initial and final positions, the most typical vocal-tract settings result in this category. (In inter-sonorant position, they result in voiced stops, which are not unambiguously preferred over voiceless stops; cf. Houlihan 1982.) It is claimed that non-contextual neutralization reflects this phonetic preference, thereby resulting in the physically unmarked category. However, the final devoiced stops found in neutralization are not physically identical to the voiceless stops with which they are neutralized. Westbury & Keating suggest that, in Polish, the underlying [+voice] stops have more closure voicing; and Dinnsen 1982 finds that final neutralized stops in Catalan differ in closure duration. These differences apparently are not audible; but they do indicate that, for speakers, underlying phonological feature values must be available to the physical detail rules. I will return to this point below; for now, note that since neutralization is not physically complete, it can be captured only by a somewhat abstract phonetic category system that equates, as voiceless unaspirated, two slightly different kinds of stops.

With this account, a re-analysis of part of Trubetzkoy's theory of markedness and neutralization is possible. In Trubetzkoy's theory ([1939] 1969:146), if two phonemes differ in a single feature, the one which is articulatorily simpler is 'unmarked', and has the minus value of the feature in question. However, this phonetic basis for markedness could be ignored by a particular language—in which case the plus value of the feature would appear to be unmarked, since the phoneme with that value would occur in positions of neutralization (147). (His theory included other types of neutralization, but this type is probably the best known.) Nonetheless, Trubetzkoy tried to maintain the rather interesting correlation between the phonetically and phonologically unmarked feature values, going so far as to use the correlation as a criterion for deciding some 'doubtful cases'. For example, 'in a language in which voiced lenis consonants form a neutralizable opposition with voiceless fortis consonants, and in which the archiphoneme in the positions of neutralization is represented by a voiceless fortis consonant, the correlation of voice is present', since voicelessness but not fortisness is phonetically unmarked. Similarly, in a *t/d* opposition, if the *t* is phonologically unmarked, then the relevant feature must be voice, since *t* is voiceless; but if the *d* is phonologically unmarked, then the relevant feature must be tension, since *d* is lax ([− tense]) (76–7). That is, Trubetzkoy's tenseness feature bore the burden of maintaining the correlation between phonetic and phonological markedness in cases involving voicing. Otherwise, tenseness did little work: it was left phonetically vague (and has eluded phoneticians since), and was largely redundant with voicing. Generalizations about voicing had to be repeated for tenseness (80), much as with Chen's vowel-length data. Because Trubetzkoy thought of neutralization as a phonological phenomenon, the situation with voicing encouraged the positing of this second feature to act as a mirror image of Voice, with a minus value (Lax) which would be phonetically voiced in certain cases.

The hypothesis that phonetically natural categories will often be unmarked, and be more likely to occur, is similar to the hypothesis formulated here. The difference is that my proposal does not necessitate a pseudo-phonetic feature to represent the fact that phonological categories work differently in different languages. Instead, an additional level of representation permits markedness to be stated regardless of which phonological feature value the phonetically unmarked voiceless unaspirated category is implementing. A language will show neutralization to that phonetic category, regardless of which phonological value is entailed. Thus the multiplication of phonological features like Tension, to the extent that it results from assumptions about markedness and neutralization, can be seen to be an unnecessary complication arising from Trubetzkoy's theory.

Furthermore, the separation of levels of representation allows statements about inventory preferences across languages clearly to distinguish phonological from phonetic preferences. Languages may in fact prefer certain phonological feature contrasts over others; but within those contrasts, they appear also to prefer particular phonetic categories. Thus, while it may be said that languages will not have /g/ without /k/, it is probably the case that the preferred

phonetic category is {vl.unasp.}, although for labials the preferred category may be {voiced} (cf. Maddieson 1981).

5.1. COMPARISON WITH OTHER MODELS. Let us return now to the question of how the proposals made here differ from, and improve upon, previous ones. The present model is essentially an extension of that in *SPE*. Like *SPE*, it has a phonological component with feature-changing rules and feature-implementing rules, which provide non-binary values with some language-specific detail. However, the modified systematic phonetic level consists of phonetic categories along a phonetic dimension; the inventory of categories imposes constraints on possible contrasts that a simple scale does not. The phonetic categories necessitate that less phonetic detail be provided at both the phonetic and phonological levels of representation; but in this way they allow generalizations at each level to be expressed.

Because this phonetic category representation is more abstract than *SPE*'s systematic phonetic level, its phonetic transcription level is also more abstract. Various subtle phonetic differences that would distinguish one dialect or language from another are not represented, with the result that Ladefoged's goal of describing a native accent precisely is abandoned at this level. The phonetic category level is designed to be the one level to which phoneticians, including Halle & Stevens, refer constantly: that of possible contrasts. Phoneticians have had to use the surface phonemic level surreptitiously when they talk about constructs such as dispersion of categories in phonetic space, or freedom of allophonic variation. But with the phonetic category level, one can refer directly to those categories that a language uses in its phonetic realizations. The phonetic category level is not limited to surface phonemic contrasts; at the same time, it may turn out to meet some needs which the surface phonemic level has fulfilled.

The alternative to such an intermediate level would appear to be a system like the one proposed by Pierrehumbert 1980 for intonation, in which phonological values are mapped directly into continuous (quantitative) values. While this proposal is intriguing, Pierrehumbert herself suggests that segmental features would have an additional representation corresponding to a phonetic transcription. In the case of [voice], a mapping directly into quantitative values would fail to express the severe limitations on possible phonetic contrasts. Presumably, an attempt to make those constraints explicit would amount to a system of the sort proposed here.

Another aspect of Pierrehumbert's proposals, however, is supported by my system. She argues that the phonetic rules which derive pitch contours must not replace the phonological values (H or L) of the tonal system, but simply give them further specification with phonetic values. This is because certain rules of Pierrehumbert's system must refer simultaneously to a tone's phonetic value and to its phonological category. Such a corepresentation of phonological and phonetic values is required for the proper treatment of two phenomena mentioned above: final (non-neutralizing) devoicing (e.g. in Polish), and fundamental frequency control. In both cases, we saw slight physical differences

between members of the 'same' phonetic category, {vl.unasp.} That such differences should not be expressed by having two phonetic categories follows from the fact that no language uses these kinds of voiceless unaspirated stops contrastively: where Polish does use two, in final position, no contrast is perceived by native speakers. The differences correlate with the phonological feature value for [voice]. Therefore phonetic implementation cannot REPLACE phonological values WITH phonetic category values. Rather, it must add phonetic values without removing phonological values, so that both kinds will be available at the point of physical (quantitative) implementation. At that point, the co-occurrence of, e.g., [+voice] and {vl.unasp.} will result in a lowered fundamental frequency at Voice Onset, or a larger amount of closure voicing in a final stop. In this sense, neutralizing rules will not be formally identical to the usual kind of feature-changing rule (cf. Port et al. 1982). The phonological value of [voice] must also be available for the duration specification of a preceding vowel, as we have seen.

Consider the sort of grammatical organization which is implied by these proposals. Underlying phonological feature values ([±voice]) may be transformed by phonological rules into opposite values. Then the language-specific phonetic implementation rules select major phonetic categories corresponding to the phonological values, according to context (including adjacent segments) and possibly following general trends. The two representations, however, are present simultaneously. In the case of [voice], it has been convenient to think of the VOT dimension as being represented in the implementation rules. More properly, however, these rules will convert representations as [±voice] to the three basic phonetic voicing categories, for which VOT values are an approximation. These phonetic categories must then be converted to values along physical scales, in ways which are sensitive not only to phonological value, but also to context, place of articulation etc. The scales can be thought of as acoustic dimensions such as VOT and closure duration (with more than one scale specified for a given set of categories); or they can be thought of as articulatory specifications which control actual speech production mechanisms. Whether this conversion can be accomplished entirely by universal rules of phonetics, or requires additional language-specific rules, remains an open question—although limited evidence suggests that at least some aspects of conversion can be universally predicted.

Contrast this framework with the type of theory advanced by Ladefoged (e.g. Ladefoged & Bhaskararao 1983). To the extent that any clear relation exists between phonological segments and phonetic specifications, it is essentially that a phonological segment can have any number of phonetic realizations. Segments differ phonetically across languages in ways that may not be contrastive in any one language. No constraints are posited or envisioned. By contrast, the present hypothesis is that interesting and meaningful constraints exist on phonetic variation. I agree that a number of physical implementations are possible for a given phonological segment type and for a given phonological contrast. The phonetic category level has been added to the grammar to provide a limited, but descriptively adequate, mapping between the phonological and

physical representations of sounds. In this way, two goals that Ladefoged eschews can be met: first, to relate phonological and phonetic entities explicitly, and second, to place limits on the relation.

5.2. SUMMARY AND CONCLUSIONS. I have tried to show that certain cross-language phonetic differences can best be expressed as differences in the realization or implementation of phonological feature contrasts as phonetic categories. A phonological feature, [±voice], has been proposed to account for rule equivalence across languages with phonetically different contrasts, justifying the use of the same symbols for different sounds. This system is motivated by the allophonic variation found across contexts within a single language such as English. Three phonetic categories—{voiced}, {vl.unasp.} and {vl.asp.}—have been proposed to express the maximum number of contrasts found along the voicing dimension, and to describe the markedness relation among those categories. Evidence has been presented that the boundaries defining the three categories derive from physiological—mainly auditory—constraints. It is also proposed that subtle differences between languages in the exact position where the categories lie, between the category boundaries, can be accounted for by principles such as polarization, which separates categories within the limits imposed by the boundaries.

My analysis accords with a view that languages do not differ without limit; rather, variation is constrained at different levels in the grammar. In the course of this account, several empirical claims have been made that must be tested further. One is that languages with different phonetic implementations will nonetheless share the same phonological rules. Another is that contrastive voicing categories are restricted to certain parts of the phonetic voicing dimension, respecting the category boundaries imposed by universal phonetics. Still another claim is that certain phonetic categories may be in a sense 'ambiguous', receiving minutely different physical realizations, and inducing different contextual effects on the basis of their phonological category. More generally, it is posited that a level of representation between that of phonological features and the output of phonetic detail rules—a level which to some extent abstracts away from phonetic detail—is useful and necessary.

We have seen that, while universal phonetics provides a constrained set of phonetic voicing categories, various languages use all possible combinations in their implementations of [±voice]. The choice of implementation rules must be specified for each context in each language, since there seems to be no way to predict categories across environments—even though there are only three main possibilities to choose from. Further, we have seen that the need for language-specific quantitative rules is an open question, depending on the viability of hypotheses such as the polarization principle. The answer to this question must await more data on slight differences between languages with the same implementation of a contrast, so that principles such as polarization can be proposed and tested. To the extent that such principles are insufficient to account for observed variation, even low-level phonetic rules will be seen to be language-specific. In that case, the role of universal phonetics will largely

be to constrain the form and substance of such rules, in part by establishing the phonetic category boundaries that would limit possible cross-language differences. Such constraints will obviously contribute to the learnability of the phonological and phonetic systems.

Rules of implementation and of detail must eventually be provided for [voice] in other kinds of segments (as well as for other phonological features), if this model is to be supported further. The success of the enterprise will depend on finding evidence for a limited number of major phonetic categories available for each feature. The extent of psycho-acoustic investigations of phonetic dimensions suggests that no great progress will be made in the near future. Nonetheless, hypotheses such as those of Stevens & Blumstein 1981 on the acoustic properties underlying various phonological features, coupled with studies of the distribution of major categories, could prove rewarding. Promising starting points for further research include dental vs. alveolar places of articulation (Lahiri & Blumstein 1981) and different kinds of flaps (Price 1981).

It is assumed by phonologists that the phonological component of the grammar is highly structured—and that its formal properties, as well as its substance, are of linguistic interest. It might be thought that, in contrast, the organization of phonetics is relatively trivial, comprising a physical scale for each phonological feature—and, for each language, a value along the scale at which it is divided into phonological categories. I have tried to show that such a simple system will not work for [voice] contrasts. The phonetic rules developed to account for the data presented, and the phonetic level of representation derived by them, have interesting properties, which suggest that the phonetic component of the grammar is more structured and richer than has been supposed. It is to be hoped that further work along these lines will clarify the place of phonetics in a grammar of linguistic competence.

REFERENCES

ABRAMSON, ARTHUR S. 1972. Word-final stops in Thai. Tai phonetics and phonology, ed. by Jimmy G. Harris & Richard B. Noss, 1–7. Bangkok: Central Institute of English Language.

——, and LEIGH LISKER. 1972. Voice-timing perception in Spanish word-initial stops. J. Phon. 1.1–8.

ANDERSON, STEPHEN R. 1978. Tone features. Tone: A linguistic survey, ed. by Victoria A. Fromkin, 133–76. New York: Academic Press.

CARNEY, ARLENE EARLEY; GREGORY P. WIDIN; and NEAL F. VIEMEISTER. 1977. Noncategorical perception of stop consonants differing in VOT. JASA 62.961–70.

CHEN, F.; V. ZUE; M. PICHENY; N. DURLACH; and L. BRAIDA. 1980. Speaking clearly: Acoustic characteristics and intelligibility of stop consonants. Paper presented at 99th meeting of the Acoustical Society of America.

CHEN, MATTHEW. 1970. Vowel length variation as a function of the voicing of the consonant environment. Phonetica 22.129–59.

CHOMSKY, NOAM, and MORRIS HALLE. 1968. The sound pattern of English. New York: Harper & Row.

DENT, LAUREL. 1976. Voice Onset Time of spontaneously spoken Spanish voiceless stops. Paper presented at 91st meeting of the Acoustical Society of America.

DINNSEN, DANIEL. 1982. Abstract phonetic implementation rules and word-final devoicing in Catalan. Paper presented at LSA Annual Meeting, San Diego.

DONALD, SUSAN LEA. 1978. The perception of voicing contrasts in Thai and English. Dissertation, University of Connecticut at Storrs.

DOW, FRANCIS. 1972. An outline of Mandarin phonetics. Canberra: Faculty of Asian Studies, in association with the Australian National University Press.

FISCHER-JØRGENSEN, ELI. 1954. Acoustic analysis of stop consonants. Miscellanea Phonetica 2.42–59.

FLEGE, JAMES E. 1979. Phonetic interference in second language acquisition. Indiana University dissertation.

——. 1982. Laryngeal timing and phonation onset in utterance-initial English stops. J. Phon. 10.177–92.

——, and WILLIAM BROWN. 1982. Effects of utterance position on English speech timing. Phonetica 39.337–57.

——, and KAREN MASSEY. 1980. English prevoicing: Random or controlled? Paper presented at LSA Summer Meeting.

FOX, ROBERT A., and DALE TERBEEK. 1977. Dental flaps, vowel duration and rule ordering in American English. J. Phon. 5.27–34.

GOLDSMITH, JOHN. 1976. Autosegmental phonology. MIT dissertation. [Distributed by Indiana University Linguistics Club.]

GREENBERG, JOSEPH H. 1970. Some generalizations concerning glottalic consonants, especially implosives. IJAL 36.123–45.

HAAS, MARY R. 1956. The Thai system of writing. Washington, DC: American Council of Learned Societies.

HALLE, MORRIS, and KENNETH N. STEVENS. 1971. A note on laryngeal features. Quarterly Progress Report, Research Laboratory of Electronics, MIT, 101.198–213.

HAYES, BRUCE. 1983. The phonetics and phonology of Russian voicing assimilation. To appear in a Festschrift for Morris Halle, ed. by Mark Aronoff & Richard Oehrle. Cambridge, MA: MIT Press.

HEBERT, R., and N. POPPE. 1963. Kirghiz manual. (Indiana University publications, Uralic and Altaic series, 33.) Bloomington: Indiana University.

HEFFNER, R.-M. S. 1969. General phonetics. Madison: University of Wisconsin Press.

HIRSH, IRA. 1959. Auditory perception of temporal order. JASA 31.759–67.

——. 1975. Temporal aspects of hearing. The nervous system, III: Human communication and its disorders, ed. by D. B. Tower, 157–62. New York: Raven Press.

HOARD, JAMES. 1971. Aspiration, tenseness, and syllabication in English. Lg. 47.133–40.

HOMBERT, JEAN-MARIE; JOHN J. OHALA; and WILLIAM G. EWAN. 1979. Phonetic explanations for the development of tones. Lg. 55.37–58.

HOULIHAN, KATHLEEN. 1982. Is intervocalic voicing a natural rule? Paper presented at LSA Annual Meeting, San Diego.

HYMAN, LARRY M. 1975. Phonology: Theory and analysis. New York: Holt, Rinehart & Winston.

JAKOBSON, ROMAN. 1962. Selected writings, I: Phonology. The Hague: Mouton.

——; GUNNAR FANT; and MORRIS HALLE. 1952. Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge, MA: MIT Press.

KAHN, DANIEL. 1976. Syllable-based generalizations in English phonology. MIT dissertation. [Distributed by Indiana University Linguistics Club.]

KEAN, MARY-LOUISE. 1975. The theory of markedness in generative grammar. MIT dissertation. [Distributed by Indiana University Linguistics Club.]

KEATING, PATRICIA A. 1978. Perception of vowel length in Czech and English. Paper presented at LSA Annual Meeting.

——. 1979. A phonetic study of a voicing contrast in Polish. Brown University dissertation.

——; WENDY LINKER; and MARIE HUFFMAN. 1983. Patterns in allophone distribution for voiced and voiceless stops. J. Phon. 11.277–90.

——; MICHAEL J. MIKOŚ; and WILLIAM F. GANONG III. 1981. A cross-language study of range of Voice Onset Time in the perception of initial stop voicing. JASA 70.1261–71.

KUHL, PATRICIA K. 1978. Predispositions for the perception of speech-sound categories: A species-specific phenomenon? Communicative and cognitive abilities: Early be-

havioral assessment, ed. by Fred D. Minifie & Lyle L. Lloyd, 229–56. Baltimore: University Park Press.

——, and JAMES D. MILLER. 1975. Speech perception by the chinchilla: Voiced–voiceless distinction in alveolar plosive consonants. Science 190.69–72.

——, ——. 1978. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. JASA 63.905–17.

LADEFOGED, PETER. 1978. Phonetic differences within and between languages. UCLA Working Papers in Phonetics 41.32–40.

——. 1982. A course in phonetics. 2nd edn. New York: Harcourt Brace Jovanovich.

——, and CHIN W. KIM. 1965. Comments on 'Voicing in initial stops'. UCLA, MS.

——, and PERI BHASKARARAO. 1983. Non-quantal aspects of consonant production: A study of retroflex consonants. J. Phon. 11.291–302.

LAHIRI, ADITI, and SHEILA E. BLUMSTEIN. 1981. A reconsideration of acoustic invariance for place of articulation in stop consonants: Evidence from cross-language studies. Paper presented at the 102nd meeting of the Acoustical Society of America.

LIEBERMAN, PHILIP. 1970. Towards a unified phonetic theory. LI 1.307–22.

——. 1977. Speech physiology and acoustic phonetics. New York: Macmillan.

LILJENCRANTS, JOHAN, and BJÖRN LINDBLOM. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. Lg. 48.839–62.

LISKER, LEIGH, and ARTHUR S. ABRAMSON. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20.384–422.

——, ——. 1967. Some effects of context on Voice Onset Time in English stops. Language and Speech 10.1–28.

——, ——. 1971. Distinctive features and laryngeal control. Lg. 47.767–85.

——; ——; FRANKLIN S. COOPER; and MALCOLM H. SCHVEY. 1969. Transillumination of the larynx in running speech. JASA 45.1544–6.

LÖFQVIST, ANDERS. 1980. Interarticulator programming in stop production. J. Phon. 8.475–90.

MACK, MOLLY. 1982. Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. JASA 71.173–8.

MADDIESON, IAN. 1977. Further studies on vowel length before aspirated consonants. UCLA Working Papers in Phonetics 38.82–90.

——. 1981. UPSID: Data and Index. (UCLA Working Papers in Phonetics, 53.) Los Angeles.

MALSHEEN, BATHSEBA J. 1980. Two hypotheses for phonetic clarification in the speech of mothers to children. Child phonology, II: Perception, ed. by Grace H. Yeni-Komshian et al., 173–84. New York: Academic Press.

MIKOŚ, MICHAEL J. 1977. Problems in Polish phonology. Brown University dissertation.

MILLER, J. D.; C. C. WIER; R. PASTORE; W. J. KELLY; and R. J. DOOLING. 1976. Discrimination and labeling of noise–buzz sequences with varying noise-lead times: An example of categorical perception. JASA 60.410–17.

PIERREHUMBERT, JANET BRECKENRIDGE. 1980. The phonology and phonetics of English intonation. MIT dissertation.

PISONI, DAVID B. 1977. Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops. JASA 61.1352–61.

PORT, ROBERT; F. MITLEB; and M. O'DELL. 1982. Neutralization of obstruent voicing in German is incomplete. Paper presented at the 102nd meeting of the Acoustical Society of America.

PRICE, PATTI JO. 1981. A cross-linguistic study of flaps in Japanese and in American English. University of Pennsylvania dissertation.

SMITH, BRUCE, and JOHN R. WESTBURY. 1975. Temporal control of voicing during occlusion. Paper presented at the 89th meeting of the Acoustical Society of America.

SOLI, SIGFRID. 1983. The role of spectral cues in discrimination of Voice Onset Time differences. JASA 73.2150–65.

STEVENS, KENNETH N., and SHEILA E. BLUMSTEIN. 1981. The search for invariant acoustic correlates of phonetic features. Perspectives on the study of speech, ed. by Peter D. Eimas & Joanne L. Miller, 1–38. Hillsdale, NJ: Erlbaum.

STEVENS, KENNETH N., and DENNIS H. KLATT. 1974. Role of formant transitions in the voiced–voiceless distinction for stops. JASA 55.653–9.

STREETER, LYNN A. 1976. Kikuyu labial and apical stop discrimination. J. Phon. 4.43–9.

TRUBETZKOY, NIKOLAI S. 1939. Grundzüge der Phonologie. Prague: Cercle Linguistique. [Translated by Christiane A. M. Baltaxe: Principles of phonology. Berkeley & Los Angeles: University of California Press, 1969.]

VANVIK, A. 1972. A phonetic-phonemic analysis of Standard Eastern Norwegian, I. Norwegian Journal of Linguistics 26.1–64.

WATERS, R., and W. WILSON. 1976. Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. Perc. Psychophys. 19.285–9.

WEISMER, GARY. 1980. Control of the voicing distinction for intervocalic stops and fricatives: Some data and theoretical considerations. J. Phon. 8.427–38.

WESTBURY, JOHN R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. JASA 73.1322–36.

——, and PATRICIA A. KEATING. 1980. A model of stop consonant voicing and a theory of markedness. Paper presented at the Linguistic Society of America. [Revised and expanded version to appear.]

WILLIAMS, LEE. 1974. Speech perception and production as a function of exposure to a second language. Harvard University dissertation.

YOSHIOKA, HIROHIDE; ANDERS LÖFQVIST; and HAJIME HIROSE. 1981. Laryngeal adjustments in the production of consonant clusters and geminates in American English. JASA 70.1615–23.

[Received 2 November 1982;
revision received 24 August 1983;
accepted 6 September 1983.]