Linking linguistic contrasts to reality: The case of VOT*

Peter Ladefoged and Taehong Cho

In the early days of acoustic phonetics Eli Fischer-Jørgensen (1954) described a number of important properties of stop consonants, including the fact that the length of the aspiration in voiceless stops (Voice Onset Time, or VOT, as it was later termed) varied with place of articulation. Cho and Ladefoged (1999) summarize several factors underlying this universal tendency, showing how it may depend on aerodynamic circumstances, the mass and mobility of different articulators, temporal adjustment between the closure duration and VOT, and perceptual concerns. These factors are given different weights in different languages, resulting in variations across languages in the way contrasts in VOT are manifested.

This paper is based on the data reported in Cho and Ladefoged (1999). It will discuss the extent to which the phonological inventory of a language affects the VOT, and conclude by discussing how a phonological description of a single language can make explicit statements about the physical manifestation of VOT in that language, noting how it differs or is the same as that in other languages. We will not consider within languages variations in VOT due to place or articulation or other factors. It is sufficient for our purposes to consider VOT only in voiceless velar stops in a number of different languages.

Previous discussions of cross-linguistic variations in VOT (e.g. Lisker & Abramson 1964, Keating 1984) have been hampered by not being able to refer to a large body of data that had been collected in accordance with the same protocol. When comparing the phonetic properties of different languages it is important to collect data from a number of speakers of each language, so that any individual bias is discounted, and to analyze the data in the same way, so that differences in recording techniques and measurement procedures do not affect the results. We were fortunate to have access to a body of data that fulfilled these prerequisites.

The data we used consisted of measurements of VOT in 18 languages. For the last 8 years Peter Ladefoged and Ian Maddieson have been the Principal Investigators of a National Science Foundation sponsored project to study the phonetic structures of endangered languages. (This project is continuing for a further three year period under the sole direction of Ian Maddieson.) Endangered languages are no different from any other languages in their possible phonetic structures. These languages are endangered in that they are losing speakers, usually for socio-economic reasons that have no linguistic biases. Consequently a sample of the world's languages consisting solely of languages spoken by communities that are disappearing could theoretically be representative of the world's languages as a whole. Of course, as Maddieson (1997a) and others have pointed out, there are enormous problems in getting a valid sample of languages that is truly representative of the 6,703 (Grimes 1999) current languages in the world (not to mention the further problems of extending this sample to cover all human languages that ever have been or ever could be). The 18 languages that are used in the present work fail completely as a

* This paper appeared in N. Gronnum and J. Rischel (eds.), Travaux Du Cercle Linguistique De Copenhague, vol. XXXI (To Honour Eli Fischer-Forgensen), pp.212-225. C.A. Reitzel, Copenhagen, 2001.

representative sample of possible languages. But they are diverse enough to be at least indicative of the range of VOT that can be found.

The 18 languages in the data set represent 12 different language families as shown in Table 1. Most of the languages are spoken by a comparatively small number of speakers, but Navajo and Apache are fairly widely spoken. Navajo is not an endangered language, but it was investigated in the same way as the other languages. Jalapa Mazatec is also not dying rapidly. It is spoken by nearly all the inhabitants of Jalapa de Diaz in Mexico, including the children. It is endangered in the sense that it is changing rapidly due to the influence of Spanish. Many distinctions are no longer made by younger speakers. Scottish Gaelic may be spoken by 70,000 people, as we have been told, but it is clearly an endangered language, spoken by very few young people. Eastern and Western Aleut are closely related, but they have a number of phonological differences.

Table 1. Languages in the data set.

LANGUAGE	FAMILY	LOCATION
Aleut (Eastern)	Eskimo-Aleut	Alaska, U.S.A.
Aleut (Western)	Eskimo-Aleut	Alaska, U.S.A.
Apache	Athabaskan	Arizona, U.S.A.
Banawá	Arawan	Northern Brazil
Bowiri	Niger-Congo	Ghana
Chickasaw	Muskogean	Oklahoma, U.S.A.
Dahalo	Cushitic	Kenya
Defaka	Niger-Congo	Nigeria
Gaelic	Indo-European	Scotland, U.K.
Hupa	Athabaskan	California, U.S.A
Jalapa Mazatec	Otomanguean	Mexico
Khonoma Angami	Tibeto-Burman	Nagaland, India
Montana Salish	Salishan	Montana, U.S.A.
Navajo	Athabaskan	New Mexico, U.S.A.
Tlingit	Athabaskan	Alaska, U.S.A.
Tsou	Austronesian	Taiwan
Wari’	Chapacuran	Northern Brazil
Yapese	Austronesian	Western Pacific

The recordings of these 18 languages were made in the field in a standardized way by one or other of the two Principal Investigators, Peter Ladefoged and Ian Maddieson, with the exception of the Hupa data, which were recorded by Matthew Gordon, at that time a graduate student in the UCLA Phonetics Lab. Material illustrating the full range of segmental contrasts in each language was recorded, but in this paper we will refer only to the data on voiceless unaspirated and voiceless aspirated velar stops. These stops were always recorded in initial position in citation forms of contrasting words before a non-high vowel. It is arguable that this is not the most appropriate data in that it does not reflect natural utterances in the languages. We thought, however, that it was preferable to ensure unity of style across languages, even at the expense of naturalness.

Several speakers of each language were recorded using high quality equipment. All the speakers were adult native speakers who used the language in their daily life. As noted above, most of the languages investigated are moribund (the children no longer speak them), but all our speakers were completely fluent. The recordings were all analyzed by Research Associates in the UCLA Phonetics Lab in the same way (for details see Cho and Ladefoged, 1999). The differences between languages that emerged are almost certainly not artifacts of the slight differences in the circumstances in which the recordings were made, nor are they due to variations in the measurement techniques of the graduate students who worked on this project, all of whom were closely supervised and trained in the same way. We can conclude that the data reflect real differences between languages.

Figure 1 shows the mean VOT of the velar stops in these 18 languages. If a language contrasts aspirated and unaspirated velar stops, both values are shown, so that there are 25 columns of mean values. There is a very wide range of values, going from20 ms for the Khonoma Angami voiceless unaspirated stops to 154 ms for the Navajo aspirated stops. The lower of these figures is not unexpected; voiceless velar stops are known to have some voicing lag, so it is no surprise that the lowest value recorded in the data set is significantly greater than zero. But, to those unfamiliar with Navajo, the value of 154 ms for aspirated velar stops may seem excessive — almost as if it were an artifact. It is not. As listeners to the language can attest, aspiration is a very salient feature of Navajo speech.

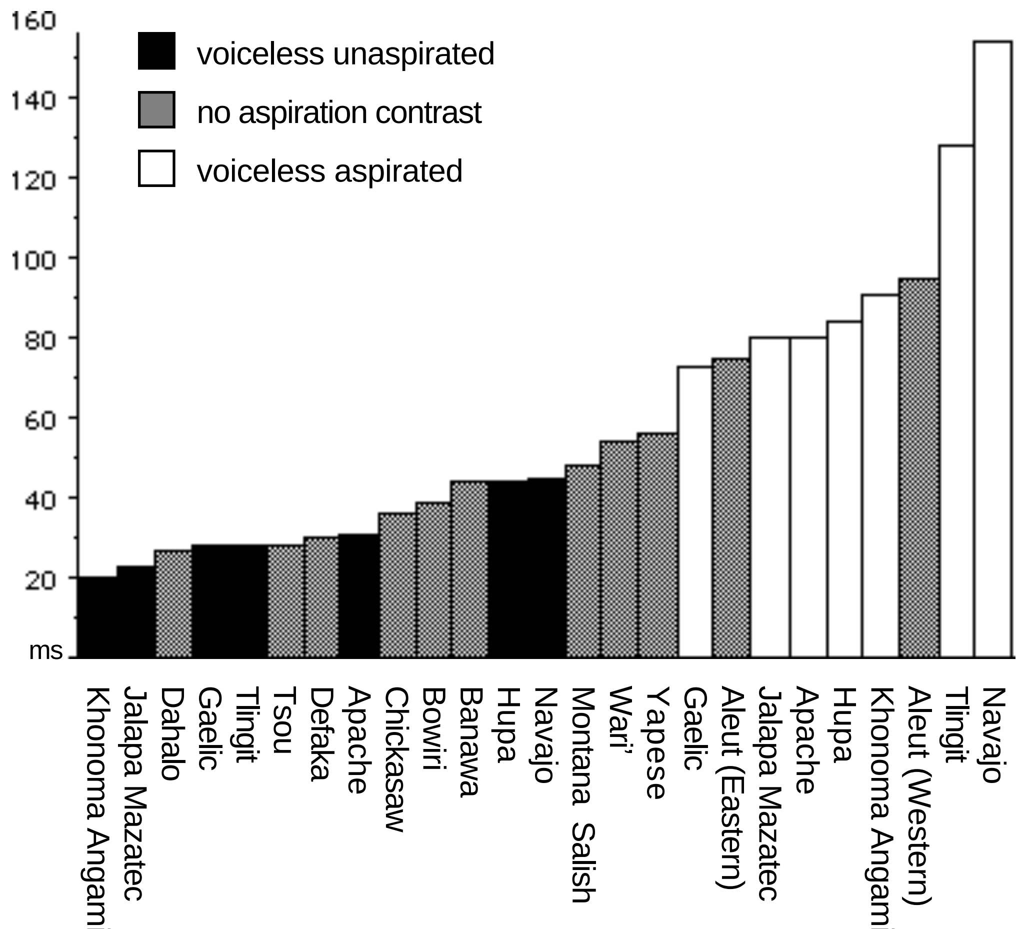

Figure 1. Mean VOTs (ms) for voiceless velar stops in 18 languages. (Values for voiced velar stops are not shown.)

Before discussing the data in detail, we must consider the set of oppositions that occurs among these velar stops in each language. The fact that in some languages there is only a single velar stop, while in others there are two, and in yet others three, might be expected to influence the VOT chosen by each language. If there is no need to make a perceptual distinction between two sounds, then one might expect a language to use the simplest articulatory gesture, what Docherty (1992) has called the low-cost option. This simplest gesture — whatever it is — is presumably the same for all human beings. So one might expect all languages that have only one velar stop to use the same gesture. When there is more than one voiceless velar stop, then each of them must be kept distinct and something other than the low-cost option will have to be used. It is important, therefore, to consider the phonological oppositions. The contrasts within each language are shown in Table 2. We have omitted labialized velars, considering them to be irrelevant to the present discussion of VOT. But we have included information on ejectives, as these sounds are perceptually similar to velar stops that are distinguished simply by VOT. None of the languages in the sample has distinctions due to variations in phonation type, such as creaky voice or breathy voice. The symbols used are those chosen by the authors of the individual studies.

Table 2. The contrasting velar stops (not including labialized stops) that occur in each of the 18 languages.

LANGUAGE	CONTRASTING VELAR STOPS	SOURCE
Aleut (Eastern)	**k**	Cho et al. 1997
Aleut (Western)	**k**	Cho et al. 1997
Apache	**k, k^h, k’**	Gordon et al. 2000
Banawá	**k**	Ladefoged et al. 1997
Bowiri	**k, g**	Maddieson p.c.
Chickasaw	**k**	Gordon et. al. 1997
Dahalo	**k, g**	Maddieson et al. 1993
Defaka	**k, g**	Shryock, et al. 1996
Gaelic	**k, k^h**	Ladefoged et al.1997
Hupa	**k, k^h, k’**	Gordon, 1996
Jalapa Mazatec	**k, k^h**	Silverman et al. 1995
Khonoma Angami	**k, g**	Blankenship et al. 1993
Montana Salish	**k, k’**	Flemming et al. 1994
Navajo	**k, k^h, k’**	McDonough & Ladefoged 1993
Tlingit	**k, k^h, k’**	Maddieson et al. 1996
Tsou	**k**	Wright & Ladefoged 1997
Wari’	**k**	MacEachern et al. 1997
Yapese	**k, k’, g**	Maddieson 1997b

There are six languages in the set that have only one velar stop. In none of them is this stop voiced, so, quite understandably from a phonological point of view, the authors of the original studies have represented each of these stops by the symbol **k.** But from a phonetic point of view, these stops vary considerably. Some might well be considered phonetically aspirated stops, and the others unaspirated. Both forms of Aleut have stops that are among the most aspirated in the set. The other four languages that have only a single velar stop, Banawá, Chickasaw, Tsou and Wari’, vary in their choice of VOT, Tsou being among the lower group and Wari’ among the higher. None of these six languages needs to make a perceptual distinction between two sounds, and they might all be expected to use the simplest articulatory gesture. But the data show that they do not necessarily choose the same simplest, low-cost, articulation.

There are only four languages, Bowiri, Dahalo, Defaka and Khonoma Angami, that have what those with a European bias might consider to be the typical velar contrasts, **k** vs **g**. A fifth, Yapese, has this contrast plus an ejective. We do not have adequate data on the voiced stops in these five languages (except that we know that they all have some voicing during the stop closure), so we cannot say which of them are more like French, with a contrast between a fully voiced stop and a voiceless unaspirated stop, and which more like English, with a contrast between a partly voiced stop and an aspirated stop.

In our pursuit of the distinction between voiceless unaspirated and aspirated stops, we can consider all 11 languages that have a single voiceless velar stop that has, from a phonetic point of view, to be called voiceless unaspirated or aspirated. When describing these languages we do not need to make the distinction for phonological reasons, but we must, within the usual techniques of phonetic description, say that these stops fall into the one phonetic category or the other. The VOT's for these 11 languages are represented by the gray columns in Figure 1. Where should we draw the line for the phonetic boundary between voiceless unaspirated stops and aspirated stops? If we look at the data in Figure 1 it might appear as if we could draw it after Banawá. But if we remove the distraction of the data from languages that contrast voiceless unaspirated and aspirated stops, and look at languages that have just one voiceless stop, as in Figure 2, the answer is not so obvious. There is a steady increase in VOT, and no obvious way of dividing the data into two phonetic categories until we get to Eastern and Western Aleut, both of which clearly have aspirated stops. The difference between Banawá and Montana Salish, the dividing line suggested by the data in Figure 1, is smaller than that between Banawá and Bowiri, and that between Montana Salish and Wari'. There is a smooth increase in VOT, and we can make only an arbitrary decision about which languages have voiceless unaspirated stops, and which have aspirated stops.

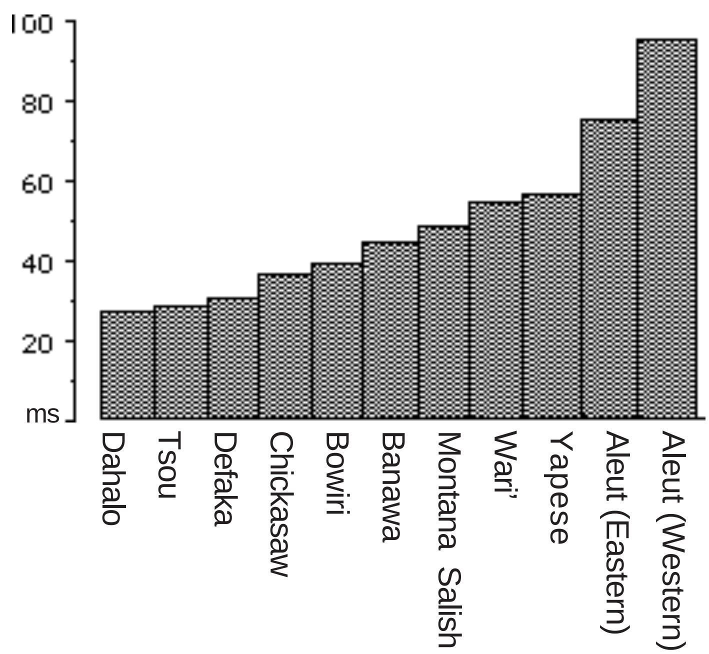

Figure 2. The 11 languages in the sample that do not distinguish voiceless unaspirated and aspirated stops.

Now consider the 7 languages in the sample that contrast voiceless unaspirated and aspirated stops, shown in Figure 3. We noted earlier that when there is no need to make a perceptual distinction we might expect that languages will choose the low-cost option, and make the simplest articulation. This expectation turned out to be wrong. Here, where there is a need to make a perceptual distinction between similar sounds, one might expect that languages would maximize the perceptual difference between them. But, as in the previous case, this expectation is not met. Languages do not behave in this way. Some languages make a large difference, others do not. Hupa has a difference of only 40 ms between voiceless unaspirated and aspirated stops. This is less than half Tlingit's 100 ms.

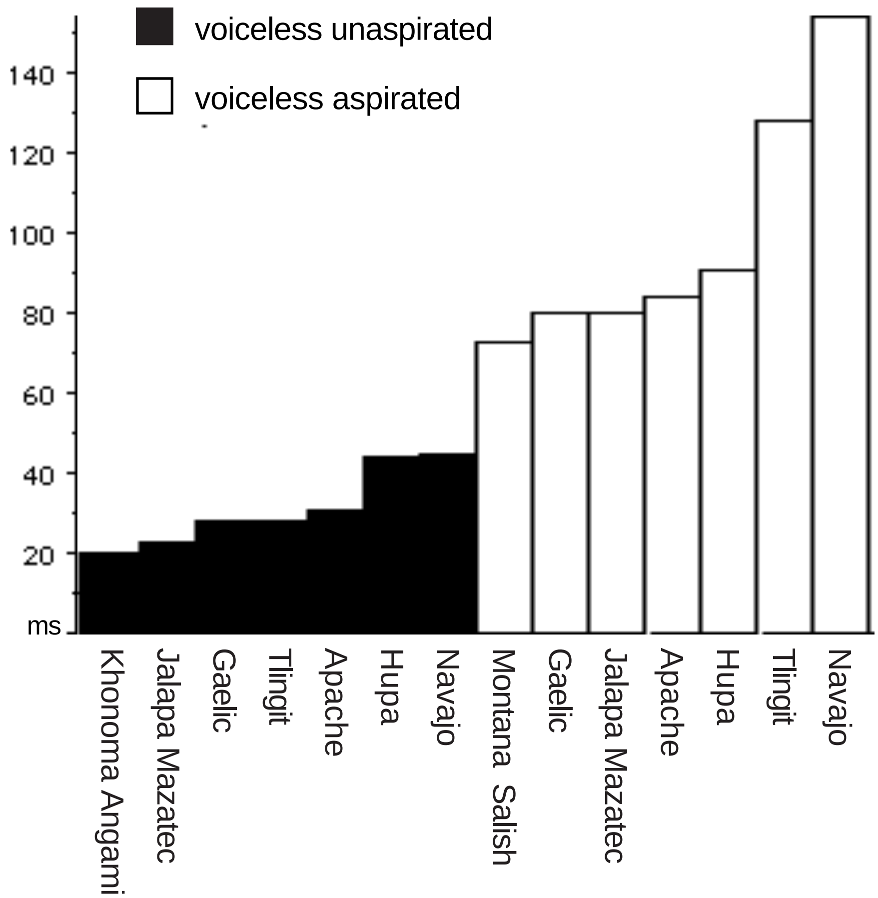

Figure 3. The 7 languages in the sample that distinguish voiceless unaspirated and aspirated stops.

There are six languages that have ejectives. Apache, Hupa, Montana Salish, Navajo and Tlingit contrast **k, k', k^h**. Yapese contrasts **k, k', g** and Montana Salish has just **k, k'**. These languages are interesting in that they may use VOT as a helping feature (Stevens, Keyser & Kawasaki,1986) to further the distinction between ejectives and other stops. There is a tendency for this to happen, as can be seen from the data in Figure 4, in which the shading has been arranged so as to make it easier to compare languages. In every language except Hupa the ejectives are clearly distinguished from the other stops by VOT.

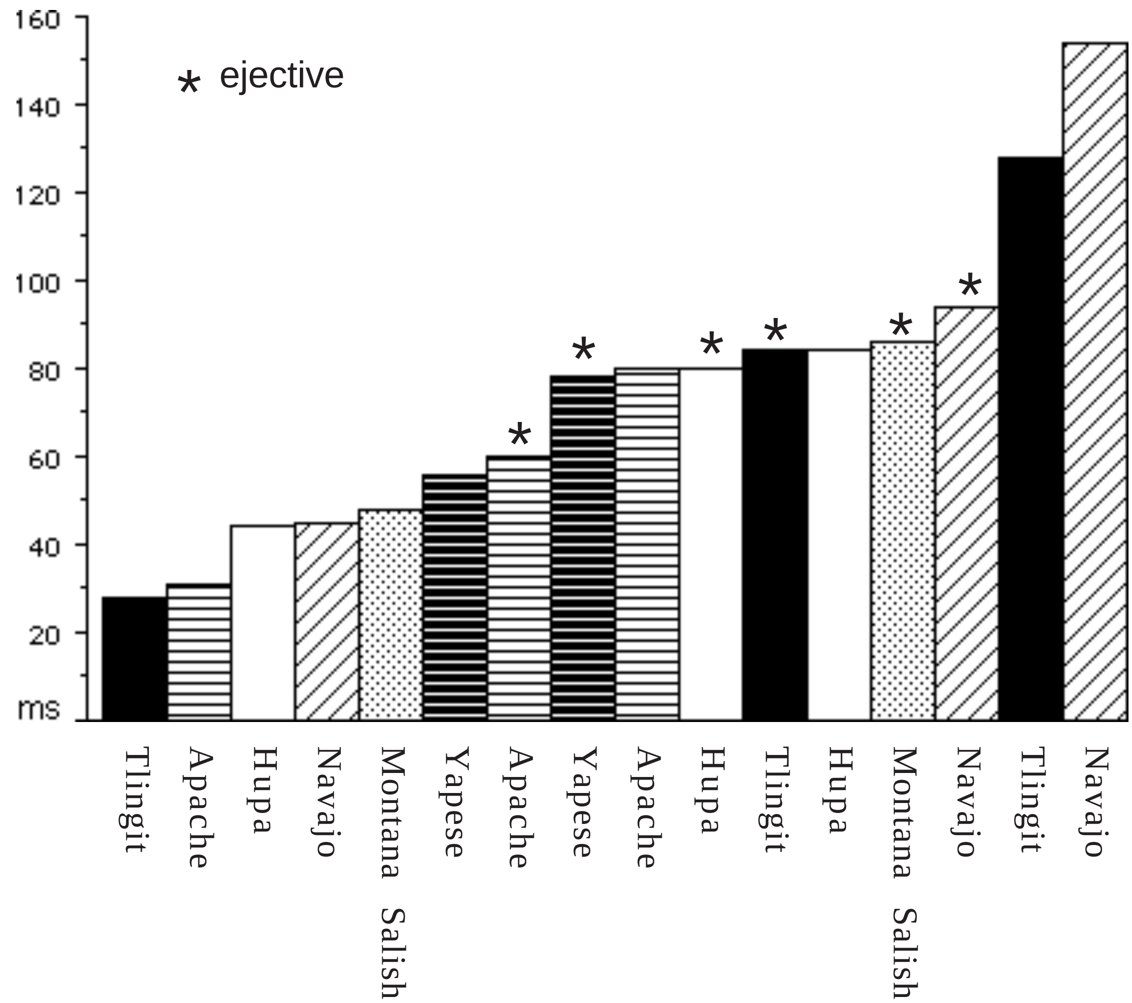

Figure 4. The 6 languages in the sample that have ejectives. Each language has been given a different shading so as to make it easy to compare the ejectives with the other stops in the language.

We will now assess the data as a whole, and consider how phonological statements about each language can be related to observable VOT differences. We presume that phonological descriptions will need to distinguish at least three possibilities [voiced], [voiceless unaspirated] and [aspirated]. We have considered only the [voiceless unaspirated] and [aspirated] categories in this paper, partly because we do not have data on voiced stops, and partly because we do not know of any claims that different degrees of negative VOT are phonologically contrastive. However, languages vary in the amount of voicing that can occur in a phonologically voiced stop, and any mechanism that we propose for the realization of phonologically voiceless stops in terms of physical phonetic variables should apply in a similar way to phonologically voiced stops.

How should the three possibilities {voiced, voiceless unaspirated, aspirated} be realized? Because we want a variable that will be constant within a language across different places of articulation, we propose making phonetic specifications not in terms of the directly observable acoustic measure, VOT, but in terms of an underlying physiological measure. We suggest that there is a phonetic parameter, which we will call Articulatory

VOT, definable in terms of the difference in time between the initiation of the articulatory gesture responsible for the release of a closure and the initiation of the laryngeal gesture responsible for vocal fold vibration. We think it likely that speakers aim for a certain timing difference between articulatory and glottal gestures irrespective of the articulatory gesture involved. This is the low-cost option suggested by Docherty (1992). Differences in VOT within a language are usually the inevitable consequence of the physiological movements and the aerodynamic forces that occur at different places of articulation and in different syntagmatic contexts. Cho and Ladefoged (1999), however, found a few cases in which a single VOT target cannot account for all the observed variations in VOT within a language. Sometimes there may be variations in VOT ascribable to aerodynamic causes (e.g. the variations due to place of articulation in unaspirated stops) that a language may choose to use in other circumstances (e.g. as an aid to the perception of places of articulation of aspirated stops in which different aerodynamic forces occur and would, unless prevented, have produced a different acoustic effect).

Even if we specify VOT in terms of an underlying physiological parameter, our data show that there is a great deal of between language variation. Moreover, it is impossible to predict the differences between languages from knowledge of the phonological contrasts within a language. It is not the case that if a language lacks a contrast between **k** and $\mathbf{k^h}$ it will have the simplest possible VOT, with a value between the modal value for **k** and that for $\mathbf{k^h}$. Nor is it the case that if a language does have a contrast between **k** and $\mathbf{k^h}$ will it make that contrast with a larger than usual VOT for $\mathbf{k^h}$ and a smaller one for **k**, so as to make sure that the difference is easy to hear. Nor does the VOT in ejectives have any simple relation to the VOT of other phonological contrasts. We propose, as in Cho and Ladefoged (1999), that each language chooses a modal VOT value for each of the categories [voiced], [voiceless unaspirated] and [voiceless aspirated] that are specified in the phonology. The statement of these values is the link between phonology and measurable phonetic parameters. A phonological description of a language that does not include statements of this kind is incomplete.

Many thanks are owed to the members of UCLA Phonetics Lab and all the speakers who participated in the project studying ‘Phonetic structures of endangered languages’. This work was supported by NSF grant SBR 951118 to Peter Ladefoged and Ian Maddieson.

References

Blankenship, Barbara, Peter Ladefoged, P. Bhaskararao and N. Chase. 1993. Phonetic structures of Khonoma Angami. Linguistics of the Tibeto-Burman area 16.69-88.

Cho, Taehong, and Peter Ladefoged. 1999. Variations and universals in VOT: evidence from 18 languages. Journal of Phonetics 27. 207-229.

Cho, Taehong, Alice Taff, Moses Dirks and Peter Ladefoged. 1997. Some phonetic structures of Aleut. UCLA Working Papers in Phonetics 95.41-67.

Docherty, Gerry. 1992. The timing of voicing in British English obstruents. Berlin: NewYork.

Fischer-Jørgensen, Eli. 1954 Acoustic analysis of stop consonants. Miscellanea Phonetica 2.42-59.

Flemming, Edward, Peter Ladefoged and Sarah Thomason. 1994. Phonetic structures of Montana Salish. UCLA Working Papers in Phonetics 87.1-34.

Gordon, Matt. 1996. The phonetic structures of Hupa. UCLA Working Papers in Phonetics 93.164-187.

Gordon, Matt, Pamela Munro and Peter Ladefoged. 1997. The phonetic structures of Chickasaw. UCLA Working Papers in Phonetics 95.41-67.

Gordon, Matt, Brian Potter, Willem de Reuse and Peter Ladefoged. 2000. Phonetic structures of Apache. UCLA Working Papers in Phonetics 97. forthcoming.

Grimes, Barbara E. (ed.) 1999. Ethnologue: Languages of the world. Thirteenth edition. Dallas: Summer Institute of Linguistics.

Keating, Patricia A. 1984. Phonetic and phonological representation of stop consonant voicing. Language 60.286-319.

Ladefoged, Peter, Jenny Ladefoged and Daniel Everett. 1997. Phonetic structures of Banawá, an endangered language. Phonetica 54.94-111.

Lisker, Leigh and Arthur Abramson. 1964. Cross-language study of voicing in initial stops: Acoustical measurements. Word 20.384-422.

MacEachern, Margaret, Barbara Kern and Peter Ladefoged. 1997. Wari' phonetic structures. Journal of Amazonian Languages 1.5-25.

McDonough, Joyce and Peter Ladefoged. 1993. Navajo Stops. UCLA Working Papers in Phonetics 84.151-164.

Maddieson, Ian. 1997a. Phonetic universals. Handbook of phonetic sciences, ed. by John Laver and William Hardcastle, 619-639. Oxford: Blackwells Publishers.

Maddieson, Ian. 1997b. The many glottalized consonants in Yapese – how, why and where? Paper presented at the Third Austronesian Formal Linguistics Association Conference, UCLA.

Maddieson, Ian, Bonny Sands, Peter Ladefoged and Sinisa Spajic. 1993). The phonetic structures of Dahalo. Afrikanistische Arbeitspapiere 36.5-53.

Maddieson, Ian, Nicola Bessel and Caroline Smith. 1996. A preliminary phonetic report on the sounds of Tlingit. UCLA Working Papers in Phonetics, 93, 125-148.

Shryock, Aaron, Peter Ladefoged and Kay Williamson. 1996. Phonetic structures of Defaka. The Journal of West African Languages. 26.3-27.

Silverman, Daniel, Barbara Blankenship, Paul Kirk and Peter Ladefoged. 1995. Phonetic structures in Jalapa Mazatec. Anthropological Linguistics 37.70-88.

Stevens, Kenneth. N., S. Jay Keyser and Hiroko Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. Invariance and variability in speech processes, ed. by Joseph Perkell and Dennis Klatt, 426-463. New Jersey: Lawrence Erlbaum Associates.

Wright, Richard, and Peter Ladefoged. 1997. A phonetic study of Tsou. Bulletin of the Institute of History and Philology, Academia Sinica 68.987-1028.