Journal of Phonetics (1999) **27**, 207–229
Article No. jpho.1999.0094
Available online at http://www.idealibrary.com on IDEAL®

Variation and universals in VOT: evidence from 18 languages

Taehong Cho and Peter Ladefoged

Phonetics Lab, Department of Linguistics, UCLA, Los Angeles, CA 90095, U.S.A.

Received 2nd December 1998, and accepted 26th July 1999

Voice onset time (VOT) is known to vary with place of articulation. For any given place of articulation there are differences from one language to another. Using data from multiple speakers of 18 languages, all of which were recorded and analyzed in the same way, we show that most, but not all, of the within language place of articulation variation can be described by universally applicable phonetic rules (although the physiological bases for these rules are not entirely clear). The between language variation is also largely (but not entirely) predictable by assuming that languages choose one of the three possibilities for the degree of aspiration of voiceless stops. Some languages, however, have VOTs that are markedly different from the generally observed values. The phonetic output of a grammar has to contain language specific components to account for these results.

1. Introduction

When a pattern recurs in hundreds of languages it may seem inevitable. For example, many phoneticians have noticed that vowels are usually longer before voiced than before voiceless stops (Halle & Stevens, 1967; Chen, 1970; Lisker, 1974; Maddieson & Gandour, 1977; Maddieson, 1997*a*, *inter alia*). It is also a common observation that high vowels in stressed monosyllables are shorter than low vowels in comparable syllables (Lindblom, 1967; Lehiste, 1970; Lisker, 1974; Westbury & Keating, 1980; Maddieson, 1997*a*). But neither of these patterns is inevitable. A language that at one time had a contrast between long and short vowels could lose this possibility and keep just the long high vowels and the short low vowels. A language of this kind might be slightly more difficult to learn, but it would not be impossible.

There are, however, other kinds of phonetic events that have inevitable consequences. Whenever the tongue goes from a raised position in the front of the mouth to a low position in the back, the frequency of the first formant will go up and that of the second formant will go down. Similarly, if there is no compensatory adjustment, stretching the vibrating vocal folds will always raise the pitch of a voiced sound. Again, other things being equal, whenever a contraction of the internal intercostal muscles occurs to produce a stressed syllable, then the syllable will have a higher pitch and an increase in loudness.

Address correspondence to T. Cho, Phonetics Laboratory, Department of Linguistics, UCLA, 405 Hilgard Ave, Los Angeles, CA 90095, U.S.A. E-mail: taehong@humnet.ucla.edu.

0095–4470/99/020228 + 23 $30.00/0

In discussing phonetic universals we should keep these two kinds of phonetic events distinct. It is physically impossible to move the tongue from a high front to a low back position without raising F1 and lowering F2. It is perfectly possible to reverse the usual vowel length differences between high and low vowels, although the resulting gestures may be more difficult to make. In this paper, we will discuss differences among aspirated and unaspirated stop consonants as reflected by variations in voice onset timing (VOT). We will mainly be concerned with variations in VOT due to place of articulation, and will consider which, if any, of these variations are inevitable consequences of some physiological adjustment, and which are simply the most favored (perhaps the easiest) articulatory gestures.

It is well known that VOT varies to some extent with place of articulation. The principal findings are that: (1) the further back the closure, the longer the VOT (Fischer–Jørgensen, 1954; Peterson & Lehiste, 1960); (2) the more extended the contact area, the longer the VOT (Stevens, Keyser & Kawasaki, 1986); and (3) the faster the movement of the articulator, the shorter the VOT (Hardcastle, 1973). These patterns have been known for many years. They can be observed in Lisker and Abramson's (1964) classic cross-linguistic study of VOT—although they themselves did not go into details concerning variations of VOT conditioned by place by articulation. Tables I and II show that, in their data, velar stops always have a longer VOT. Furthermore, in both aspirated and unaspirated stops, VOT is shortest before bilabial stops and intermediate before alveolar stops, with the exception of the unaspirated stops in Tamil and the aspirated stops in Cantonese and Eastern Armenian.

In early forms of generative phonology, such patterns were considered to be attributable to low level (automatic) phonetic implementation rules, constrained by physiological (biomechanical) factors, and thus not a necessary part of the grammar of any one language. This is the view expressed by Chomsky & Halle (1968) in the *Sound Pattern of English* (SPE). In SPE, for any given language, once binary features have been converted into scalar featural values, the physical output is completely determined by universal phonetic implementation rules.

TABLE I. Summary of VOT (ms) in unaspirated stops reported by Lisker & Abramson (1964)

	Dutch	Puerto Rican Spanish	Hungarian	Cantonese	Eastern Armenian	Korean	Tamil
/p/	10	4	2	9	3	18	12
/t/	15	9	16	14	15	25	8
/k/	25	29	29	34	30	47	24

TABLE II. Summary of VOT (ms) in aspirated stops reported by Lisker & Abramson (1964)

	Cantonese	English	Eastern Armenian	Korean
/p^h/	77	58	78	91
/t^h/	75	70	59	94
/k^h/	87	80	98	126

It has also been known for many years that the SPE view is not correct, and that there are language specific phonetic rules which must be part of the grammar of each language (Pierrehumbert, 1980, 1990; Keating, 1984, 1985, 1990; Fourakis & Port 1986; Cohn, 1993, among others). In particular, Keating (1985) convincingly shows that three assumed phonetic universals—intrinsic vowel duration, extrinsic vowel duration, and voicing timing—are not automatic results of speech physiology. They are not universal attributes of sounds, but are at least in part determined by language specific rules. Docherty (1992) reaches a similar conclusion with respect to VOT in British English.

There have been several recent reports of variations in VOT, the most important being those of Cooper (1991*a*, 1991*b*), Docherty (1992), and Jessen (1998). These studies present data on VOT in many different contexts, but in each case the comparable data are limited to a single language (Cooper on American English, Docherty on British English, and Jessen on German). The present study is limited in a different way. It presents carefully matched data from a large number of languages, but the data set involves only a single phonetic context.

1.2. *Explanations of VOT patterns*

There have been several explanations in the literature for the general voice onset differences found in the studies reported above. These explanations depend on a number of factors, including laws of aerodynamics, articulatory movement velocity, and differences in the mass of the articulators. In addition, there is an alternative analysis that suggests there is a temporal adjustment between stop closure duration and VOT (Weismer, 1980; Maddieson, 1997*a*). There might also be a perceptual advantage to place related VOT differences (Jessen, 1998). All these explanations attempt to account for what Docherty (1992) calls the low cost procedure for producing stops—one which is favored, but might not be used in a particular language. They also apply only if all other things are equal, which they seldom are. It is precisely because things are seldom equal that it is difficult to pull apart that bit of the variance that might be attributable to physical causes.

1.2.1. *General law of aerodynamics*

Many phoneticians (e.g., Hardcastle, 1973; Maddieson, 1997*a*) have suggested that one of the factors which contribute to VOT differences is the relative size of the supraglottal cavity behind the point of constriction. There are two ways of considering this. Firstly, the cavity behind the velar stop has a smaller volume than that behind the alveolar or bilabial stops. Secondly, the cavity in front of the velar stop has a larger volume than that in front of the alveolar or bilabial stops.

In order to produce voicing there must be a difference in air pressure across the vocal folds (van den Berg, 1958). If the air in the oral cavity is at a pressure similar to that below the vocal folds, there will be no airflow between them and they will not vibrate. From the first point of view, the notion that the cavity behind the velar stop has a smaller volume than that behind the alveolar or bilabial stops, it may follow that the velar stop has a greater pressure behind it at the beginning of the release phase. The air in the lungs and the vocal tract has to be considered as a single volume, which is smaller in a velar stop than it is in an alveolar or bilabial stop. During an utterance, the air is compressed by the action of the respiratory muscles. If the volume being compressed is small, a given reduction in size will produce a greater increase in pressure. As a result the air pressure in the vocal tract may be higher for a velar stop. If this is so, it will take a longer time for the pressure behind the closure to fall and allow an adequate transglottal pressure for the initiation of the vocal fold vibration.

The second point of view considers the fact that there is a larger body of air in front of the velar stop. This body of contained air will act like a mass that has to be moved before the compressed air behind the velar closure can be released into the open air. Irrespective of whether there is or there is not a higher air pressure behind velar closures, the drop in the pressure of the air in the vocal tract will be slower for velars, again resulting in more time to attain the crucial transglottal pressure difference required for voicing.

1.2.2. *Movement of articulators*

In addition to noting the effect of differences in the relative size of the supraglottal cavities, Hardcastle (1973) postulates that the voice onset difference can be due in part to the fact that the tip of the tongue and the lips move faster than the back of the tongue. This notion is supported by a cineradiographic study of VC and CV articulatory velocities by Kuehn & Moll (1976), who report that the articulatory movement is fastest for the tongue tip, intermediate for the lower lip, and slowest for the tongue body. This may be partly due to the relative masses of the articulators involved; the tongue tip is smaller and lighter than the lips or the body of the tongue. It is also due to the fact that jaw movements affect lip and tongue movements in different ways; tongue dorsum movement is least affected by jaw movement, while lower lip movement is accelerated by jaw movement.

In line with this latter point, Maddieson (1997*a*) suggests that one of the reasons for the difference in VOT between English stops /p/ and /k/ is the distance from the pivot point of the jaw rotation. A schematized representation of the effect jaw rotation is shown in Fig. 1. As illustrated in the figure, because the pivot of jaw rotation is further from the lip than from the tongue body, the movement of the lower lip will be greater than that of the tongue body for a given angular motion of the jaw (see also Vatikotis-Bateson & Ostry, 1995). When the articulator is the lower lip, the compressed air behind

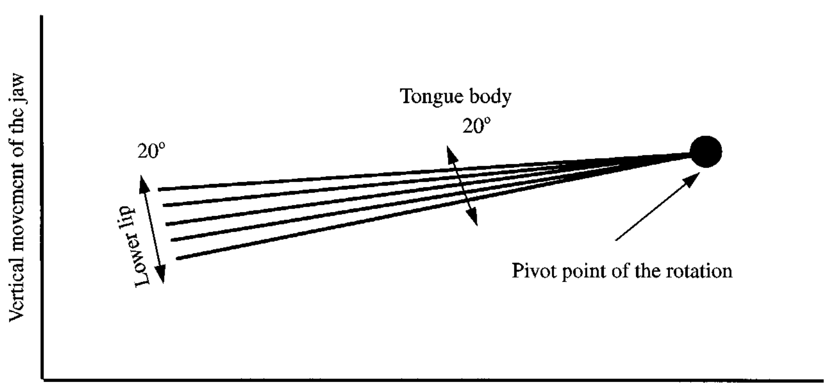


Figure 1. Schematic representation of the effect of jaw rotation. A 20° shift in jaw angle separated the lips apart more than the tongue back and velum.

the constriction escapes at a faster rate, resulting in a shorter time before building up an appropriate transglottal pressure for the initiation of voicing.

As Maddieson (1997*a*) notes, this explanation does not account fully for the place-related difference in VOTs between bilabial and alveolar stops. Recall that Kuehn & Moll (1976) report that the tongue tip moves faster than the lower lip. If the articulatory velocity is the primary physiological factor for the voice onset difference, we would expect that the VOT would be shorter for apical alveolar stops than for either bilabials or velars, which is not the general finding. This suggests there may be some other factors accounting for the place-related voice onset difference.

1.2.3. *The extent of articulatory contact*

The VOT variations can also be partly accounted for in terms of the extent of the contact area between the articulators. As velar stops are produced with a constriction between the rounded upper body of the tongue (the dorsum) and the similarly rounded soft plate, the contact area is more extended than that in bilabial and alveolar stops. There is a similar difference in contact length between laminal and apical stops which almost always accompanies dental *vs.* alveolar stop contrasts (Ladefoged & Maddieson, 1996). In general, stops with a more extended articulatory contact have a longer VOT.

Stevens (1999) provides an aerodynamic explanation for these differences. His main point is that the rate of change in intraoral pressure following the release depends on the rate of increase in cross-sectional area at the constriction. This is significantly different for different places of articulation, primarily due to the differences in the extent of articulatory contact. When there is a long narrow constriction the Bernoullli effect causes the articulators forming the constriction to be sucked together. Because the velar stop has extensive contact between the tongue body and the palate, there is a larger Bernoulli force so that the change in cross-sectional area is relatively slow compared with that for the bilabial or alveolar stops. Consequently, the decrease in intraoral pressure after the closure is gradual for the velar and rapid for the bilabial. Stevens' aerodynamic data show that the volume velocity of airflow at both the constriction and the glottis increases roughly in proportion to the rate of the decrease in intraoral pressure for the first 50 ms immediately following the release of the closure. Schematized curves of airflow and intraoral pressure at the release of voiceless stops appear in Fig. 2.

The timing of the vocal folds vibration is determined by the two inter-related aerodynamic factors shown in the figure: (1) the rate of decrease in intraoral pressure and (2) the rate of increase in volume velocity of the airflow. In his discussion of these relationships Stevens was not concerned with differences in laminality. He noted that alveolar stops can be considered to be produced at an intermediate rate of change of both the intraoral pressure and the volume velocity of airflow, assuming that the alveolar contact area is longer than the bilabial, but shorter than the velar.

1.2.4. *Change of glottal opening area*

In addition to the factors described above, Stevens (1999) ascribes differences in VOT among voiceless aspirated stops to the different degrees of glottal opening area that accompany the different places of articulation. For the aspirated stops, the glottis is already open well before the release to allow for aspiration. After the release, this glottal opening must be reduced to reach approximately 0.12 cm^2 in order to initiate vocal fold vibration. Stevens suggests that the glottal opening area after the release will decrease

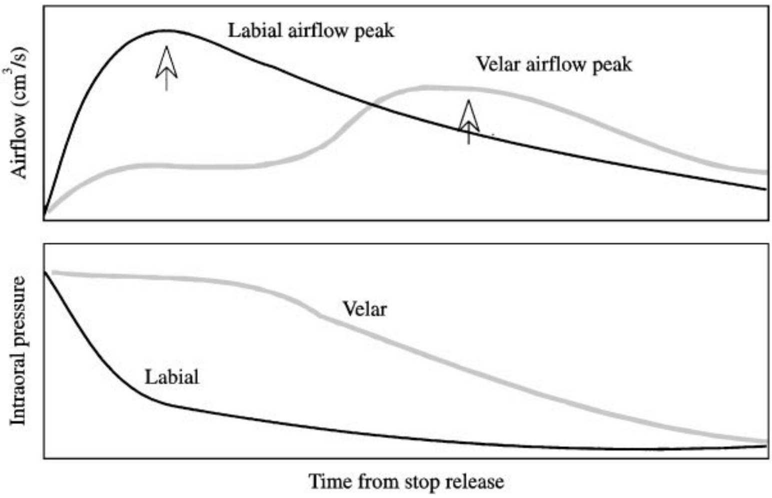


Figure 2. Schematized curves of airflow and intraoral pressure at the release of voiceless stops, based on data in Stevens (1999).

less rapidly for the velar than for the alveolar or for the labial stop because the intraoral pressure for the velar stop drops more slowly. A build-up of intraoral pressure induces an outward displacement (i.e., abduction forces) on the glottal folds (as well as on the walls of the vocal tract). In addition, during the closure interval the stiffness of the walls of both the glottis and the vocal tract is increased, presumably to counteract the increased intraoral pressure. The decrease of intraoral pressure immediately following the release causes an inward force of the walls of the glottal folds, coupled with a corresponding relaxation of the stiffness. However, the stiffness is still preserved to some degree, thus inhibiting the vibration of vocal folds immediately following the release. On the basis of these assumptions, Stevens posits that the glottal area decreases somewhat more rapidly following the release of bilabial or alveolar stops than the velar stops, since the decrease in intraoral pressure following the release of the bilabial or the alveolar stop is more rapid, and there is a more rapid formation of the adduction forces along with a more rapid relaxation of the stiffness. Thus, the voice onset occurs somewhat earlier for a labial or alveolar than for a velar voiceless aspirated stop.

1.2.5. *Temporal adjustment between stop closure duration and VOT*

The stop closure duration for bilabial stops is, in general, longer than that of either alveolar or velar stops, which may be due to different degrees of air pressure in the cavity behind the constriction (Maddieson, 1997*a*). We already noted that a smaller cavity behind the constriction will cause a more rapid build-up of the intraoral air pressure, reaching equity with subglottal air pressure in a relatively shorter time. Based upon this aerodynamic principle and the results of an experiment by Ohala & Riordan (1979), Maddieson (1997*a*) posits that "if the consonant gesture is timed in some way that directly relates to the time of the pressure peak, then broadly speaking, the further back in the oral cavity a stop closure is formed, the shorter its acoustic closure duration will be" (p. 630). This provides an inverse relationship between the closure duration and the

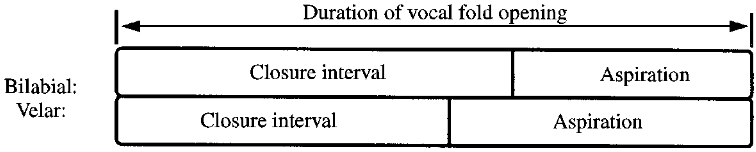


Figure 3. Schematic representation of place differences in aspirated stops from constant vocal fold abduction plus different closure duration. (From Maddieson, 1997*a*, p. 622).

observed VOT variation. Weismer (1980) reports that for word initial English /p/ and /k/, the interval from the onset of the stop closure to the voice onset is the same. Based upon this result and other evidence cited by Weismer, Maddieson (1997*a*) suggests another possible alternative account of the place-dependent VOT: "There is an abduction–adduction cycle of the vocal cords for voiceless stops which is longer in duration than the closure and has a constant time course, anchored to the onset of closure (p. 621)." In other words, the duration of the vocal fold opening is considered to be fixed, and when the closure duration is relatively longer, the following VOT becomes relatively shorter (and *vice versa*). Fig. 3 is a schematic representation from Maddieson (1997*a*, p. 622) showing this relationship. Umeda (1977) and Lisker & Abramson (1964) also discuss the same type of durational relationship between closure and aspiration.

1.2.6. *Summary of reported causes of VOT variations due to place of articulation*

In summary, the literature indicates that the following physiological/aerodynamic characteristics account, to some extent, for the variations of VOT associated with a difference in the place of articulation.

(1) *The volume of the cavity behind the point of constriction.* The relatively smaller volume of the supralaryngeal cavity in velar stops causes a greater pressure, which will take longer to fall and allow an adequate transglottal pressure for the initiation of the vocal folds vibration.

(2) *The volume of the cavity in front of the point of constriction.* The relatively greater mass of the contained air in front of velar stops causes a greater obstruction to the release of the pressure behind the velar stop, so that this pressure will take longer to fall, resulting in a greater delay in producing an adequate transglottal pressure.

(3) *Movement of articulators.* A faster articulatory velocity (e.g., the movement of the lower lip as compared to the tongue dorsum) allows a more rapid decrease in the pressure behind the closure and thus a shorter time before building up an appropriate transglottal pressure.

(4) *Extent of articulatory contact area.* The more extended contact area in laminal dental and velar stops results in a slower release because of the Bernoulli effect pulling the articulators together. Because the articulators come apart more slowly there is a longer time before an appropriate transglottal pressure is produced.

(5) *Change of glottal opening area* (*for voiceless aspirated stops*). The glottal opening area after the release will decrease less rapidly for the velar than for the alveolar or labial because the intraoral pressure drops more slowly for the velar.

(6) *Temporal adjustment between closure duration and VOT*. There is a trade-off between the closure duration and the VOT so that there is a fixed duration of vocal fold opening.

Characteristics (1)–(4) hold better for unaspirated or slightly aspirated stops. They are based on a general principle of aerodynamics: objects such as the vocal folds will vibrate only when there is a sufficient pressure difference across them, and sufficient flow between them. This principle holds, however, only if the vocal folds are adducted so that they are in a suitable position to vibrate. In the case of aspirated stops this does not occur for a considerable period after the release. Place effects on the transglottal pressure occur in the first few milliseconds after release. Even for velar stops the tongue body is expected to have lowered 4–5 mm by 50 ms after the release (Maddieson, 1997*a*). It is therefore unlikely that in any aspirated stop the supraglottal pressure will be high enough to affect the voicing initiation more than 50 ms after the release when the vocal folds are sufficiently adducted.

On the other hand, (5) will hold for aspirated stops, and (6) will hold for both unaspirated and aspirated stops. The characteristics in (5) explain, though indirectly, why the vibrations of the vocal folds are suppressed even after an adequate transglottal pressure is attained. Recall that the stiffness in walls of both vocal folds and vocal tract are maintained to some degree following the release, which presumably inhibits the vocal fold vibration (Stevens, 1999). The explanation in (6) also seems to account better for the variation of the aspirated stops. It depends on notions of speech timing rather than any aspect of the aerodynamic mechanism varying with different places of articulation.

2. VOT variations in 18 languages

To discuss the factors underlying variations in VOT we need a body of data from a number of widely different languages, all of which have been collected and analyzed in the same way. Without such controls it is possible that any observed differences between languages may be due to the procedures used to obtain the data. This paper is based on data from a project investigating the phonetic structures of endangered languages. There is nothing special about endangered languages from the phonetic point of view. They are losing speakers and may not be spoken a hundred years from now, but that is for socio-linguistic reasons and not because they share some phonetic properties. Their linguistic characteristics are as diverse as any other group of languages. What makes the languages special for the purposes of this paper is that all the data were collected within one project, using the same protocol. If there had not been a project of this magnitude to which we could refer, we would not have had such a well-matched data set. So, although we hope our findings have some universal relevance, we must note that our conclusions are limited by the data that has been collected. The most notable lack in the data is information about stop closure duration. As all the measures that are available are based on acoustic records of stops in utterance initial position, we cannot discuss possible interactions between stop closure and VOT.

In the UCLA endangered languages project, all the recordings of endangered languages were made in the field in a standardized way by one or the 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. The recorded material always included lists of words illustrating the segmental contrasts in each language in various contexts. In this paper, we are concerned only with recordings of voiceless unaspirated and voiceless aspirated stops which were always recorded in initial position in contrasting words before a non-high vowel. Contrasts before high vowels such as **i** were also recorded, but will not be considered here as the vowels differed from language to language and represented different impedances to the outgoing air. Contrasts that were elicited in sentences rather than as citation forms will not be considered, as they occurred in different contexts that might have had an effect on the VOT.

In all the cases several speakers of each language were recorded (the exact number for each language is given in the tables below), all of them being adult native speakers who used the language in their daily life. The recordings were made on high-quality equipment as described by Ladefoged (1997). The signal-noise ratio was better than 40 dB, and the frequency response was within $\pm$ 3 dB from 50 to 10,000 Hz. (Later recordings made on DAT recorders were substantially better.) The recordings were analyzed in the UCLA Phonetics lab by Graduate Research Assistants working under the direct supervision of the Principal Investigators. In all the cases, the analytical procedure provided for observation of the waveform and spectrogram of each utterance, usually by means of the Kay CSL system. VOT was measured as the interval between cursors placed at the onset of release (the final release, if there was more than one) and the onset of the first complete vibration of the vocal folds as indicated on the waveform.

Greater control could have been taken in that no special instructions were given to subjects concerning rate of speech, which was usually that of a typical fieldwork elicitation session in which citation forms are being repeated. The lack of such controls is unlikely to have biased our results. We do not have the individual measurements for each speaker in each language; so we cannot use a rigorous procedure such as analysis of variance to assess differences between languages. Nevertheless, we believe that the overall differences between languages that emerged are not artifacts of the different circumstances in which the recordings were made. It seems to us, for example, that the longer VOTs recorded in Navajo are a characteristic of that language and not due to a procedural artifact. We should also note that most of the languages investigated are moribund, but all our speakers were fluent in their native language.

In this paper, we will consider only the mean VOT as reported in the papers published on each language. For further details on the speakers and more detailed statistical analyses individual papers should be consulted.

2.1. *Languages investigated*

The 18 languages considered in this paper are listed in Table III. The home locations of our principal speakers are shown in Fig. 4. These languages are in no way a random sample of the world's languages; several of them come from the same language families. There are, however, 12 different language families represented. Some of the languages are very closely related. Eastern Aleut, spoken on the Pribilof Islands in the Bering Sea, and Western Aleut, spoken in the nearby Aleutian Islands, might be considered simply as different dialects; they do, however, have substantial phonological and other differences.

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 languages in the endangered languages project. Jalapa Mazatec is also not dying rapidly. It is spoken by nearly all the inhabitants of Jalapa de Diaz, 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

TABLE III. Languages, speakers and contrasting stops (not including affricates). The number of speakers for each language is very approximate. Stops enclosed in parentheses are not discussed, either because they have a negative VOT or because the sources do not provide appropriate data

Language	Family	Location	Total spkrs	SS	Stops (not discussed)	Source
Aleut (Eastern)	Eskimo-Aleut	Alaska, U.S.A.	400	11	t, k, q	Cho *et al.* (1997)
Aleut (Western)	Eskimo-Aleut	Alaska, U.S.A.	70	12	t, k, q	Cho *et al.* (1997)
Apache	Athabaskan	Arizona, U.S.A.	11,000	8	p, t, k, t^{h}, k^{h}, t’, k’ (ʔ)	Potter *et al.* (to appear)
Banawá	Arawan	Northern Brazil	75	5	t, k (b, d, ɟ)	Ladefoged *et al.* (1997)
Bowiri	Niger-Congo	Ghana	3500	10	p, t, k (b, d, g, kp, gb)	Maddieson p.c.
Chickasaw	Muskogean	Oklahoma, U.S.A.	12,000	14	p, t, k (b, ʔ)	Gordon *et al.* (1997)
Dahalo	Cushitic	Kenya	400	3	p, t̪, t̠, k (b, d̪, d̠, g, g^{w}, k^{w}, ʔ)	Maddieson *et al.* (1993)
Defaka	Niger-Congo	Nigeria	200	12	p, t, k (ɓ, b, d, g, kp, gb)	Shryock *et al.* (1995)
Gaelic	Indo-European	Scotland, U.K.	70,000	11	p, t, k, p^{h}, t^{h}, k^{h}, p^{j}, t^{j}, k^{j}, p^{hj}, t^{hj}, k^{hj}	Ladefoged *et al.* (1997)
Hupa	Athabaskan	CA, U.S.A	90	3	p, t, k^{j}, q, t^{h}, k^{hj}, t’, k^{j}’, q’ (ʔ)	Gordon (1996)
Jalapa Mazatec	Otomanguean	Mexico	7000	6	t, k, t^{h}, k^{h} (ʔ)	Silverman *et al.* (1995)
Khonoma Angami	Tibeto-Burman	Nagaland, India	4000	6	p, t, k (b, d, g, g^{w}, k^{w}, k^{h}w)	Blankenship *et al.* (1993)
Montana Salish	Salishan	Montana, U.S.A.	70	5	p, t, k, q, t’, k’, q’ (k^{w}, q^{w}, ʔ, p’, k^{w}’, q^{w}’)	Flemming *et al.* (1994)
Navajo	Athabaskan	New Mexico, U.S.A.	130,000	7	t, k, t^{h}, k^{h}, t’, k’ (k^{w}, k^{hw}, ʔ)	McDonough & Ladefoged (1993)

Tiwi	Australian	Australia	1500	5	ʈ, t, ʈ, [t̠] (p, k)	Anderson & Maddieson (1994)
Tlingit	Na-Dene	Alaska, U.S.A.	1500	4	t, k, q, t^h, k^h, q^h, t’, k’, q’ (k^w, q^w, k^{hw}, q^{hw}, $k^{w’}$, $q^{w’}$, ʔ)	Maddieson *et al.* (1996)
Tsou	Austronesian	Taiwan	3000	13	p, t, k (ɓ, ɗ, ʔ)	Wright & Ladefoged (1997)
Wari’	Chapacuran	Northern Brazil	1200	10	p, t, k, k^w	MacEachern *et al.* (1997)
Yapese	Austronesian	Western Pacific	8000	3	p, t, k, p’, t’, k’, (b, d, g, ʔ)	Maddieson (1997*b*)

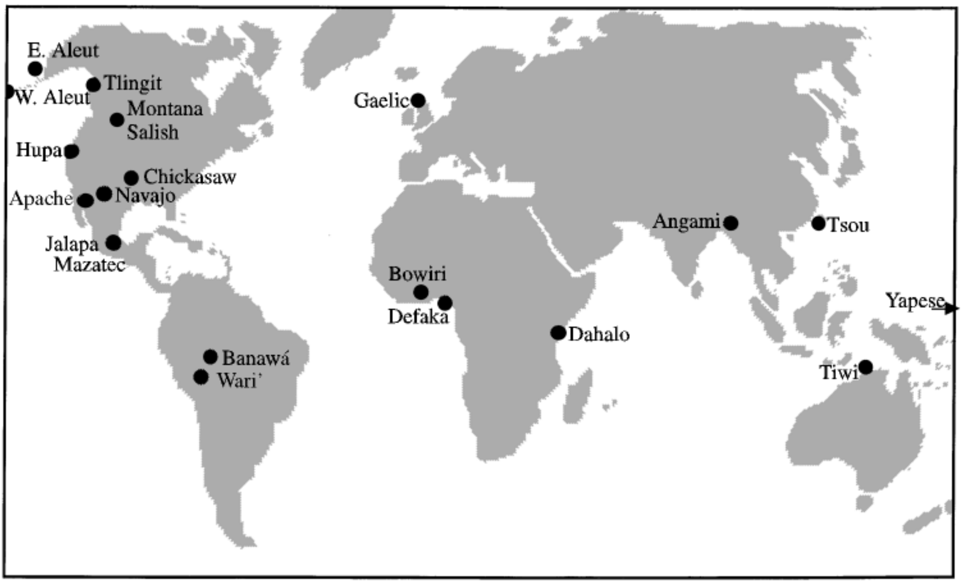


Figure 4. The locations of the 18 languages investigated.

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.

The column headed "Ss" shows the number of subjects used in providing the data reported here. The next column shows the full set of stops (excluding affricates) for each language.

2.2. *VOT data*

The mean VOT (ms) of the stops in the 18 languages is shown in Table IV. When a language contrasts unaspirated and aspirated stops, the latter are shown in a second line for that language. Wari' has plain and labialized velar stops, the latter being shown as the second entry in the velar column.

The first point to note is that, with the exception of Dahalo, velar stops have the longest VOTs in all of the 13 languages that do not have contrasts between velar and uvular stops; and in the remaining five languages either velars or uvulars have the longest VOT. Even Dahalo follows this trend if we disregard the alveolar stops. At first glance, disregarding the unusual pattern of Dahalo VOT values may seem problematic. But Maddieson *et al.* (1994, p. 29) note "In post-hoc analyses [of a one-factor ANOVA] the alveolar stops were distinct from all the others at at least the 0.01 level of significance.… The noise pattern associated with the alveolars is … striking." It seems that Dahalo speakers have simply chosen an unusually slow articulatory velocity for the apical alveolar stop in comparison with the velars, and as a result have a longer VOT for their alveolar stops. If we disregard the Dahalo alveolar stops because of their affrication, then it is true that velar or uvular stops always have the longest VOTs.

An interesting point that emerges from the data in Table IV is the similarity of VOT differences between velar and coronal aspirated and unaspirated stops. The mean for the

TABLE IV. Mean VOT (ms) of the stops in 18 languages studied in the UCLA endangered languages project (as of May 1999)

Language	Bilabial	Dental	Alveolar	Retroflex	Velar	Uvular
Aleut (Eastern)			59		75	78
Aleut (Western)			76		95	92
Apache	13		15		31	
Apache (aspirated)			58		80	
Banawá		22			44	
Bowiri	17		18		39	
Chickasaw	13		22		36	
Dahalo	20	15	42		27	
Defaka	18		20		30	
Gaelic	13	22			28	
Gaelic (aspirated)	64	65			73	
Hupa	11	16			44	27
Hupa (aspirated)		82			84	
Jalapa Mazatec			11		23	
Jalapa Mazatec (aspirated)			63		80	
Khonoma Angami	10	9			20	
Khonoma Angami (aspirated)	83	55			91	
Montana Salish	22		24		48	55
Navajo	12		6		45	
Navajo (aspirated)			130		154	
Tlingit			18		28	30
Tlingit (aspirated)		120			128	128
Tsou	11		17		28	
Wari’	19	26			50–58	
Yapese	20	22			56	

difference between the unaspirated velar and coronal stops is 18.9 ms and that for the corresponding place difference between aspirated stops is 16.7 ms. Even Navajo, which has aspirated stops with an exceptionally long VOT, has longer VOTs for velar aspirated stops than for alveolar aspirated stops. From a physiological or aerodynamic point of view, there must be two different explanations for this similarity. Any appeal to the aerodynamic conditions shortly after the release can apply to unaspirated stops, but not to aspirated stops (certainly not the Navajo stops); and any explanation that considers the special characteristics of aspirated stops cannot apply to unaspirated stops. At this point we must note that it may be perceptually advantageous to make place differences in VOT the same across aspirated and unaspirated stops. If there is a difference in the VOT of velar and coronal stops as a result of a low-cost articulatory strategy (Docherty, 1992) that works in the production of unaspirated stops, this difference in VOT may become part of a perceptual cue distinguishing these places of articulation. Once this happens it may be deliberately used in aspirated stops as a perceptual aid for place distinctions even when there is some articulatory cost to using it.

Figs 5 and 6 show that the range of VOTs associated with dental stops overlaps with that of alveolar stops. The volume of air behind the closure is much the same in the laminal dentals and the apical alveolars. Accordingly, the laminal dentals (which have

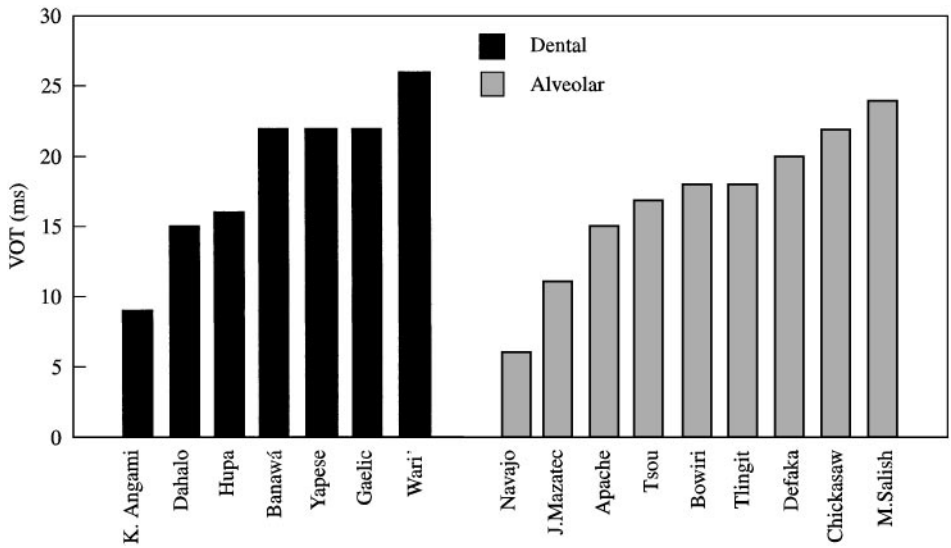


Figure 5. VOTs (ms) for the unaspirated coronal stops. (The Dahalo alveolar stops, which have anomalous VOTs, have been omitted).

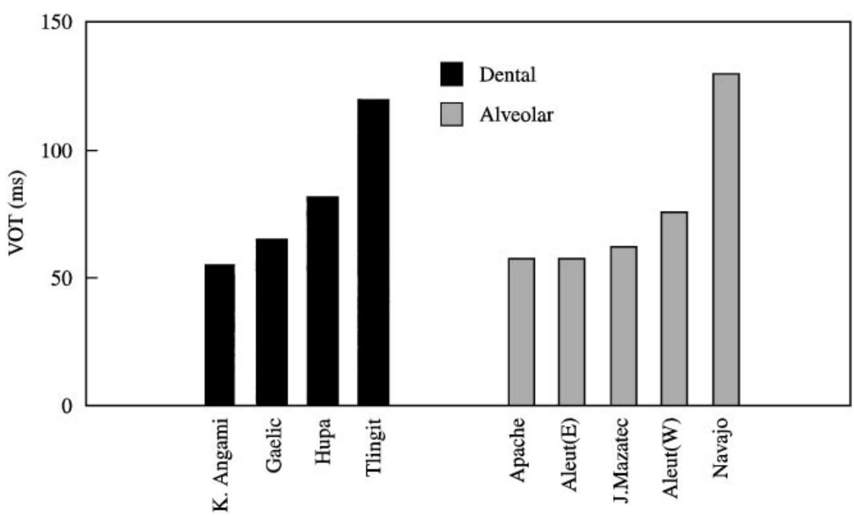


Figure 6. VOTs (ms) for the aspirated coronal stops.

a more extended contact area) might have been expected to have a slower release, and hence a significantly longer VOT. But it seems that the length of the contact is not an important source of differences in VOT for the coronal stops in these languages.

The differences between bilabial stops and coronal stops are also not significant. Many of the languages investigated do not have bilabial stops, and accordingly we are left with only 13 languages to compare as shown in Fig. 7. The mean VOT of the unaspirated bilabial stops is 15.3 ms and that for the coronal stops is 19.9 ms. A one-tailed paired

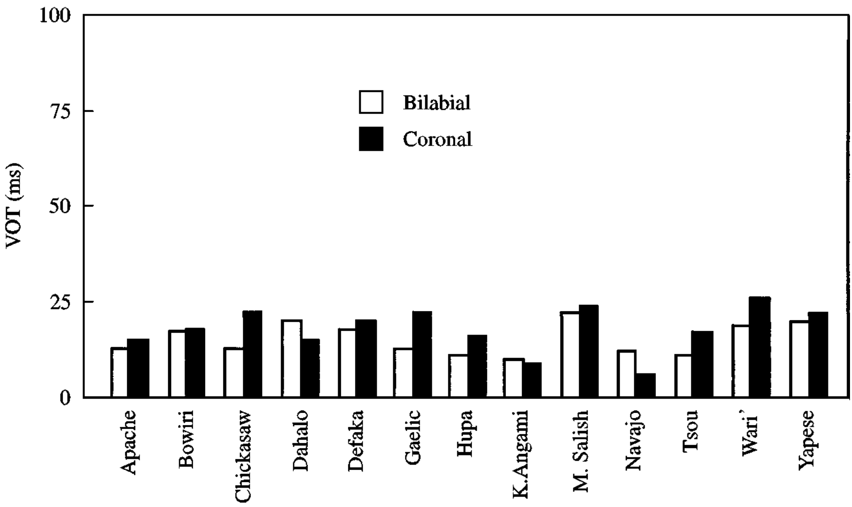


Figure 7. VOT (ms) for bilabial stops contrasted with dental or alveolar stops. In languages with both dental and alveolar stops, the one with the shorter VOT is shown.

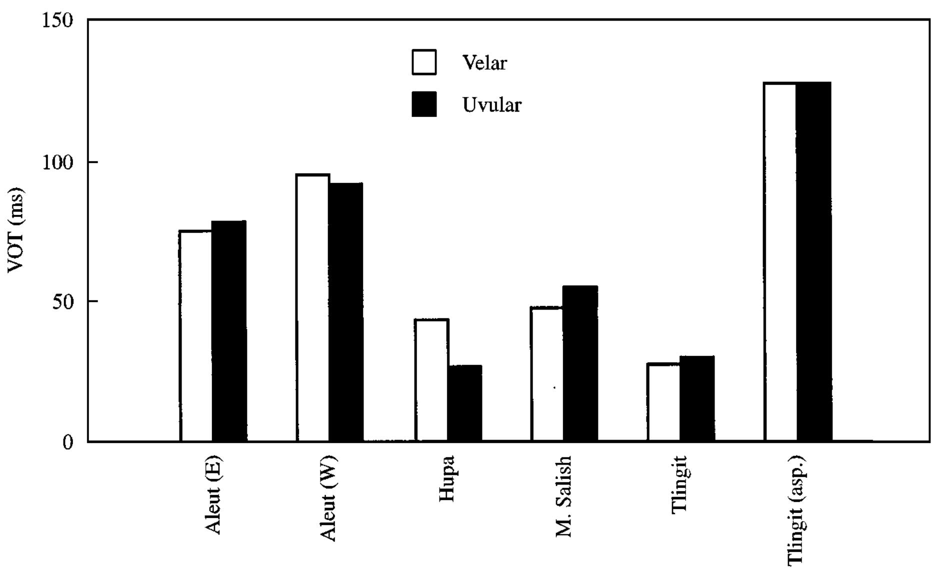


Figure 8. VOTs (ms) for velar and uvular stops.

t-test shows that there is no significant difference between these means. It has been reported that in many other languages the difference in VOT between bilabial and alveolar stops is not significant, or shows substantial overlapping (e.g., Abramson & Lisker, 1971; Lisker & Abramson, 1964).

The differences in VOT between velars and uvulars in languages that contrast these two types of sounds are shown in Fig. 8. There is little consistency in these data. In accounting for the inconsistent variations between velar and uvular stops, we suggest that although the volume of the cavity behind the constriction is smaller for uvulars than for velars, the uvular stop might be produced by a constriction with relatively shorter contact. The first of these two factors might result in a shorter VOT for velars, and the second in a shorter VOT for uvulars. The trade off between these two factors apparently varies from language to language.

2.2.1. *Ejectives*

The contrast between ejectives and other stops does not depend on VOT. Ejectives are produced with a glottalic airstream mechanism in which the air in the pharyngeal cavity is compressed by the upward movement of the closed glottis (Ladefoged, 1993). The same articulators are involved in the production of ejectives as in plosives, but at the time of the articulatory release the vocal folds are pressed tightly together and above their usual position, rather than being potentially in a position such that voicing might occur given an appropriate transglottal pressure drop. Ejectives do not have a VOT in the usual sense of this term, but for the purposes of this paper we may consider the interval between the release of the articulation and the release of the glottal closure to be the VOT. It is, in all these languages, the voice onset time after the stop release. The data for the VOTs in languages with ejectives are interesting, in that the aerodynamic conditions associated with these sounds are different from the conditions that occur in plosives. Table V shows a summary of the mean VOTs for the six languages in our data set that have ejectives. Among these languages, only Montana Salish and Yapese have regularly contrasting bilabial ejectives.

The mean VOTs for velar ejectives are longer than those for the alveolar ejectives in three of these languages, Apache, Salish, and Yapese. However, in the other three, Hupa, Navajo, and Tlingit, the mean VOTs for velars are shorter than those for alveolars. VOTs for the other places of articulation also show irregular patterns. In the two languages that have bilabial ejectives, Montana Salish has a comparatively short VOT and Yapese a comparatively long one. In Tlingit, VOT is longest for uvular ejectives, but in Montana Salish and Hupa uvulars are intermediate between the other stops.

There seems to be no physiological mechanism that accounts for the effect of place of articulation on VOT in the production of ejectives. The irregularities found in these six languages may be partly due to differences among the languages in the degree of the upward movement of the closed glottis. But it seems equally probable that the timing between the oral release and the glottal release is simply a language specific matter. Supporting this notion is the fact that there are major differences in the data for Apache

TABLE V. Voice onset time (ms) for ejectives in six languages

Language	Bilabial	Alveolar	Velar	Uvular
Apache		46	60	
Hupa		93	80	89
Montana Salish	81	65	86	81
Navajo		108	94	
Tlingit		95	84	117
Yapese	60	64	78	

and Navajo, although they are closely related languages belonging to the same language group, southern Athabaskan. Mean VOTs for alveolar and velar ejectives in Navajo are about 234% and 156% of those in Apache. The lengthy pause that follows the release of an ejective stop is a salient aspect of Navajo.

2.2.2. *Unaspirated vs. aspirated stops*

Languages differ in the values of VOT that they choose as the basic value for an unaspirated or an aspirated stop. Let us consider for simplicity just the velar stops in these 18 languages. Fig. 9 shows the complete set of values for both aspirated and unaspirated velar stops, a total of 25 mean values. It would be possible to draw an arbitrary line at, say, 50 ms, and suggest that this separates aspirated from unaspirated stops. But it is not at all clear that there are just two phonetic categories from which languages can choose. The data do not lend themselves to a statistical clumping procedure, but it would certainly be plausible to say that there are four phonetic categories, one around 30 ms representing unaspirated stops, another around 50 ms for slightly aspirated stops, a third for aspirated stops at around 90 ms, and a fourth for the highly aspirated stops of Tlingit and Navajo.

There does not seem to be any phonological reason why there might be four groups as suggested. They do not reflect differences dependent on the number of contrasts in voicing that each language has. Banawá, for example, has only a single velar stop, with no contrast in voicing; the mean VOT for this stop is 44 ms, placing it in the second group. But both Western and Eastern Aleut also have only one velar stop; their mean values are 78 and 95 ms, making them fully aspirated stops. Similarly, it does not matter whether a language contrasts voiceless unapirated stops with aspirated stops. Both

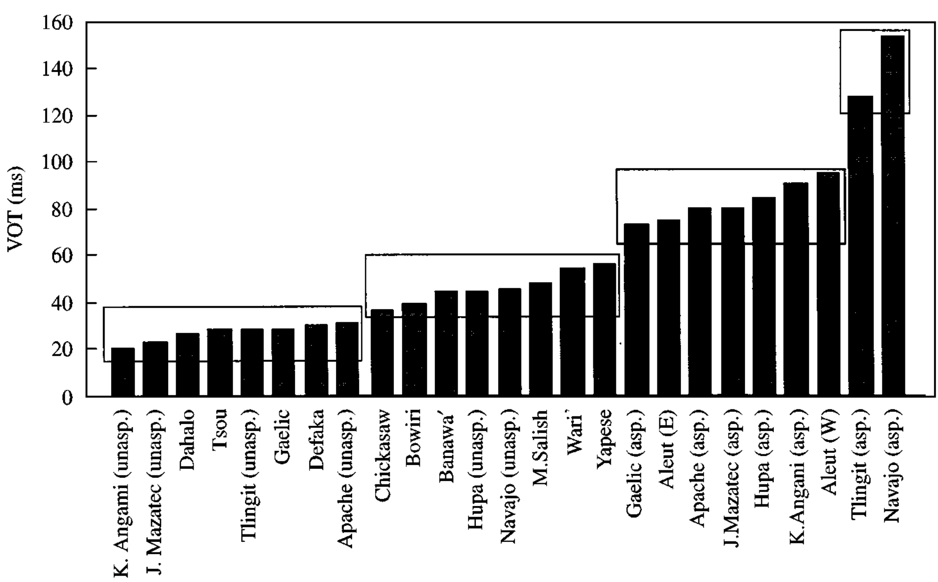


Figure 9. Mean VOTs (ms) for velar stops across languages. The rectangles enclose four regions, representing what might be called unaspirated stops, slightly aspirated stops, aspirated stops and highly aspirated stops.

TABLE VI. VOT (ms) of stops is languages with more than two stop categories

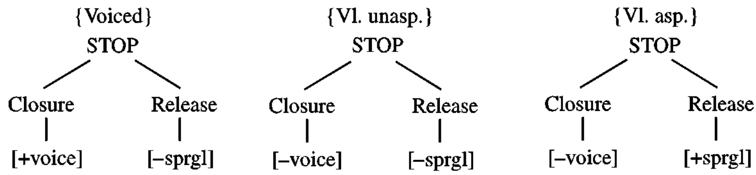
Languages	Unaspirated k	Glottalized k'	Aspirated k^h
Apache	31	60	80
Hupa	44	80	84
Navajo	45	94	154
Tlingit	28	84	128

Angami and Hupa make these contrasts. But the Angami voiceless unaspirated stops have much shorter VOTs than their Hupa counterparts and so they appear in different groups in Fig. 9. The Angami aspirated stops are in the same group as their Hupa counterparts, but have slightly longer VOTs.

Table VI shows mean values for languages that have ejectives as well as a contrast between aspirated and unaspirated stops. In every case, the VOT values for velar ejectives fall in between the VOT values of unaspirated and aspirated stops. Our conjecture is that languages with more than two types of stops tend to enhance the contrastiveness among stops by dispersing VOT values along the VOT continuum. However, some languages may not follow this tendency. The difference in VOT between velar ejectives and aspirated velar stops in Hupa is very small, and unlikely to be significant (we do not have the data to test this statistically). In this language other features, such as the characteristics of the burst, carry the contrast without any enhancement from the VOT. Languages employ different strategies for contrasting these three-way distinctions among stops.

2.3. *Discussion*

In trying to account for what was known about VOT in different languages at the time, Keating (1984, p. 289) proposed a model in which there are "only as many phonetic categories given by the phonetic features as there are contrasting phonetic types in languages." As necessary evidence, she showed that in order to achieve not only phonological generalization but also the contrasting phonetic differences between languages such as English and Polish stops /p, t, k, b, d, g/, there are two different levels of representations in the grammar. At the first level, various phonetic kinds of /b, d, g/ are defined by the feature [+voice] in both languages. At the second level, the phonetic features further distinguish stops in English from those in Polish by the use of three phonetic categories {voiced}, {voiceless unaspirated} and {voiceless aspirated}. In Polish, as in other languages without aspiration such as French, the phonological features [+voice] and [−voice] are realized as {voiced} and {voiceless unaspirated}, respectively, whereas the phonological features [+voice] in English is usually realized as {voiced}, but can be sometime realized as {voiceless unaspirated} (e.g., word-initially); similarly English [−voice] can be either {voiceless unaspirated} or {voiceless aspirated}, depending on the context (cf. Docherty, 1992). Keating notes that the implementation of the phonologically identical feature Voice is different in different languages, but the categories are chosen from a "fixed and universally specified set" which allows only three discrete phonetic categories {voice}, {voiceless unaspirated}, and {voiceless aspirated} without "fuzzy areas of a continuum". In Keating (1990), these three discrete phonetic categories are represented under Aperture Theory (cf. Steriade, 1989, 1993):



Keating's approach has many similarities with that of Ladefoged and Maddieson (1996; see also Ladefoged, 1997), and that proposed here. We differ from Keating in much the same way as Docherty (1992). We consider what might appear to be phonetic categories as at best modal values within the continua formed by the physical scales—the parameters—that define each feature.

We want to be able to characterize contrasts within languages (phonological differences) as well as phonetic differences between languages. We suggest that there is a phonological feature, 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. This is a somewhat different definition of VOT than the traditional phonetic definition that has been used throughout this paper, in which VOT is considered to be the interval between the release of an articulatory gesture, usually (always in this paper) a stop, and the beginning of vocal fold vibration.

If, for phonological purposes, we redefine VOT as the interval between the gestures involved, then the values of this feature cannot be determined by direct observation. They become largely unmeasurable without invoking some of the notions of articulatory phonology as described by Browman and Goldstein (1990, 1992). Articulatory phonology regards gestures as being realized by a task dynamic model (Saltzman, 1986; Saltzman & Munhall, 1989; see also Hawkins, 1992) that would, when fully worked out, take care of the physiological and aerodynamic influences on voicing lag that we have been discussing. The data we have been discussing seem fully compatible with this possibility. This should not, however, be taken as an endorsement of all the notions of Browman and Goldstein's Articulatory phonology. In this paper our major concern is just the description of the phonetic facts about VOT, noting how some of them can be considered to be due to physiological and aerodynamic causes whereas others require language specific specification.

In general, speakers do not deliberately produce different values of the feature VOT for different places of articulation. From the data we have presented it appears that, for some languages, we might be able to account for differences due to places of articulation—if we only knew enough about the exact articulatory movements involved. It is 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). The observed VOT is just the inevitable consequence of the physiological movements and the aerodynamic forces. Given enough knowledge about the gestures involved, the differences due to place of articulation may be as determined as the formant frequency changes that occur with particular vocalic gestures.

There is, however, plenty of evidence that languages differ in the targets that they choose. Our data show that even if we could measure VOT in terms of the difference in time between the initiation of the articulatory gesture and the initiation of the laryngeal gesture, there are still large differences between languages. All the measured VOTs in Fig. 9 are for virtually the same articulatory gesture, and should therefore reflect comparable intervals between the initiation of the gestures. Nevertheless, they show unpredictable variations between languages. Does this mean that there are more than three modal values of VOT, [voiced], [voiceless unaspirated], and [aspirated]?

The answer is probably no. The strongest evidence in favor of there being only three values is that no languages have more than three contrasts. Languages that have four contrasting homorganic stops, such as Hindi, or six such as Owerri Igbo (Ladefoged & Maddieson, 1996) are irrelevant to this discussion. They have only three VOT distinctions, and use some other action of the larynx, specified by one or more other features, to make these additional contrasts. So in a phonological description we need not consider more than three values of VOT. But phonology is concerned with only one language at a time. From the point of view of a phonetic theory that will allow us to specify all the ways in which one language may differ from another, we need a more detailed specification of VOT. There is a continuum of possible VOTs from which languages may choose. The relation between the phonological units and the physical output in a language is illustrated in Fig. 10.

We propose that lexical specifications in a language are made in terms of possible modal values of phonological features such as, for the feature VOT, [voiced] *vs.* [voiceless unaspirated] *vs.* [aspirated]. The language-specific phonetic rules then assign target values for timing between the initiation of the articulatory gesture and the initiation of the laryngeal gesture. (In the current articulatory phonology, such temporal specification is made in the gestural score.)

These two processes, the choosing of an appropriate modal value and the assignment of a target for this value, are conducted by the grammar specific to the language. In many

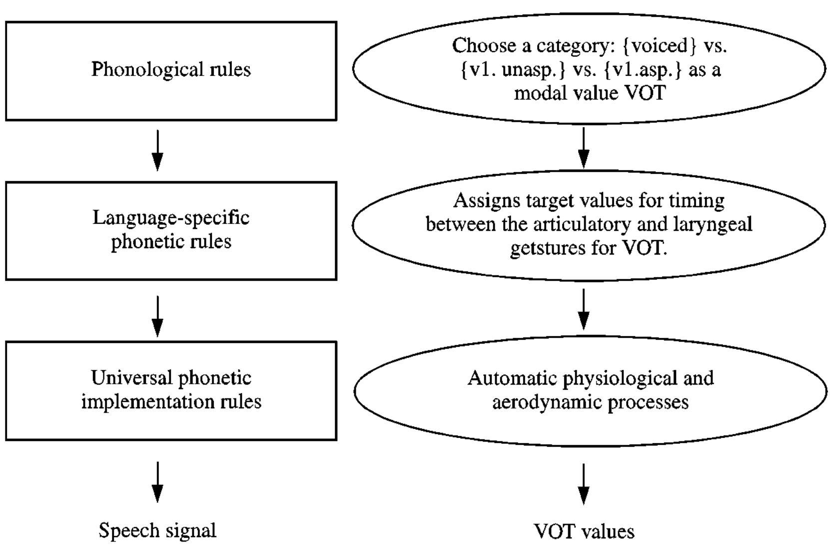


Figure 10. Multiple processes from phonology to speech signal. The model adopted here is based on Keating (1985, 1990) and Cohn (1993).

cases these actions would account for all the observed differences between this language and others, as well as for contrastive differences such as those between [voiceless unaspirated] and [aspirated] stops within the language and for the allophonic differences due to the place of articulation. It might, however, be necessary for the grammar of a particular language to specify more than one target for a given modal value. There are cases in which, even if we knew everything about the articulations involved, we would not be able to predict the differences in VOT associated with the place of articulation. We would need extra statements within the grammar of the language. Knowing all about the articulations will be insufficient, for example, if the VOT differences were deliberately introduced as perceptual cues to the place of articulation in aspirated stops. A language might have voiceless unaspirated stops for which a single target value of VOT would be sufficient. We listed in the introduction six reasons for variations in VOT of which the first four provided ample support for physiological and aerodynamic differences being sufficient to account for place differences. But, we were unable to provide equally convincing reasons for VOT variations among aspirated stops. Specific values for each place of articulation might be required in the grammar for aspirated stops.

This means that the grammar of the language would be supplying context restricted values for features. The value [aspirated] would correspond to one target when it is in the context [velar] and another when it is in the context [labial]. There is nothing new in this notion (Ladefoged, 1992). Feature definitions are often context restricted. For example, when describing English vowels, [high] will have one target value when it is in the context [front] and another when it is in the context [back], irrespective of whether one specifies the targets in terms of formant frequencies or height of the tongue. Similarly, what one means by [alveolar] is different for a stop and a lateral. Accordingly, it should be no surprise that the target for [aspirated] in the case of velar stops might be different from that for [aspirated] for bilabial stops. These possibilities are permitted by the model we have outlined.

After all these language-specific factors have been taken into account, the values assigned for the timing of the targets will still be abstract (as is the case for the comparable gestural score in articulatory phonology). These abstract values are converted to real timing values by universal implementation rules. These rules enable the task dynamic system to use the physiological and aerodynamic constraints to take care of the observed differences due to place. In this way (if we only knew enough) we could account for all the variations in VOT.

This paper owes much to comments by Michael Jessen and Gerry Docherty. Many thanks are also due 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

Abramson, A. & Lisker, L. (1971) Voice timing in Korean stops. In *Proceedings of the Seventh International Congress of Phonetic Sciences*, 439–446.

van den Berg, J. (1958) Myoelastic theory of voice production, *Journal of Speech and Hearing Research*, **1,** 277–244.

Blankenship, B., Ladefoged, P., Bhaskararao, P. & Chase, N. (1993) Phonetic structures of Khonoma Angami, *Linguistics of the Tibeto-Burman Area*, **16**(2), 69–88.

Browman, C. P. & Goldstein, L. (1990) Tiers in articulatory phonology, with some implications for casual speech. In *Papers in laboratory phonology I*: *between the grammar and physics of speech* (J. Kingston & M. Beckman, editors), pp. 341–378. Cambridge: Cambridge University Press.

Browman, C. P. & Goldstein, L. (1992) Articulatory phonology: an overview, *Phonetica*, **49,** 155–180.

Chen, M. (1970) Vowel length variation as a function of the voicing of consonant environment, *Phonetica*, **22,** 129–159.

Cho, T., Taff, A., Dirks, M. & Ladefoged, P. (1997) Phonetic structures of Aleut, *UCLA Working Papers in Phonetics*, **95,** 68–90.

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

Cohn, A. (1993) Nasalization in English: phonology or phonetics. *Phonology*, **10,** 43–81.

Cooper, A. M. (1991*a*) An articulatory account of aspiration in English. PhD dissertation, Yale University.

Copper, A. M. (1991*b*) Laryngeal and oral gestures in English /p, t, k/. In *Proceedings of the International Congress of Phonetic Sciences* 12 (Aix), **2,** 50–53.

Docherty, G. (1992) *The timing of voicing in British English obstruents*. Berlin; New York: Foris.

Everett, D. L. & Kern, B. (in press) *Wari*' (*The Pacaas Novos languages of Western Brazil*). London: Routledge, Descriptive Grammar Series.

Fourakis, M. & Port, R. (1986) Stop Epenthesis in English. *Journal of Phonetics*, **14**(2), 197–221.

Fischer-Jørgensen, E. (1954) Acoustic analysis of stop consonants. *Miscellanea Phonetica*, **2,** 42–59.

Flemming, E., Ladefoged, P. & Thomason, S. (1994) Phonetic structures of Montana Salish. *UCLA Working Papers in Phonetics*, **87,** 1–34.

Gordon, M. (1996) The phonetic structures of Hupa. *UCLA Working Papers in Phonetics*, **93,** 164–187.

Gordon, M., Munro, P. & Ladefoged, P. (1997) The phonetic structures of Chickasaw. *UCLA Working Papers in Phonetics*, **95,** 41–67.

Halle, M. & Stevens, K. N. (1967) On the mechanism of glottal vibration for vowels and consonants. *MIT Quarterly Progress Report*, **85,** 267–271.

Hardcastle, W. J. (1973) Some observations on the Tense-Lax distinction in initial stops in Korean. *Journal of Phonetics*, **1,** 263–271.

Hawkins, S. (1992) An introduction to task dynamics. In *Papers in laboratory phonology II*: *gestures*, *segments*, *prosody* (G. J. Docherty & D. R. Ladd, editors), pp. 9–25. Cambridge: Cambridge University Press.

Jessen, M. (1998) *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: John Benjamins.

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

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

Keating, P. A. (1990) Phonetic representations in a generative grammar. *Journal of Phonetics*, **18,** 321–334.

Kuehn, D. P. & Moll, K. (1976) A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, **4,** 303–320.

Ladefoged, P. (1992) The many interfaces between phonetics and phonology. In *Phonologica* 1988 (W.U. Dressler, H. C. Luschützky, O. E. Pfeiffer & J. R. Rennison, editors), pp. 165–179. Cambridge: Cambridge University Press.

Ladefoged, P. (1993) *A course in phonetics*. 3rd edition. For Worth: Harcourt Brace College Publishers.

Ladefoged, P. (1997) Linguistic phonetic descriptions. In *The handbook of the phonetic sciences* (W. Hardcastle & J. Laver, editors), pp. 589–618. Oxford: Blackwells.

Ladefoged, P. & Maddieson, I. (1996) *Sounds of the world's languages*. Oxford: Blackwells.

Ladefoged, P., Ladefoged, J. & Everett, D. in press. Phonetic structures of Banawá, an endangered language. *Phonetica*.

Lehiste, I. (1970) *Suprasegmentals*. Cambridge: MIT Press.

Lindblom, B. (1967) Vowel duration and a model of lip mandible coordination. *Speech Transmission Laboratory*, *Quarterly Progress and Status Report*, **4**/1967, 1–29. Stockholm: Royal Institute of Technology.

Lisker, L. (1974) On 'explaining' vowel duration variation. *Glossa*, **8**(2), 233–246.

Lisker, L. & Abramson, A. (1964) Cross-language study of voicing in initial stops: acoustical measurements. *Word*, **20,** 384–422.

MacEachern, M. R., Kern, B. & Ladefoged, P. (1997) Wari' phonetic structures. *Journal of Amazonian Languages*, **1,** 5–25.

McDonough, J. & Ladefoged, P. (1993) Navajo Stops. *UCLA Working Papers in Phonetics*, **84,** 151–164.

Maddieson, I. (1997*a*) Phonetic Universals. In *The handbook of phonetic sciences* (J. Laver & W. J. Hardcastle, editors), pp. 619–639. Oxford: Blackwells.

Maddieson, I. (1997*b*) The many glottalized consonants in Yapese — how, why and where?. *Paper presented in the Third Austronesian Formal Linguistics Association Conference*, UCLA.

Maddieson, I. & Gandour, J. (1977) Vowel length before aspirated consonants. *Indiana Linguistics*, **38,** 6–11.

Maddieson, I., Sands, B., Ladefoged, P. & Spajić, S. (1993). The phonetic structures of Dahalo. *Afrikanistische Arbeitspapiere*, **36,** 5–53.

Maddieson, I. & Anderson, V. B. (1994) Phonetic structures of Iaai. *UCLA Working Papers in Phonetics*, **87,** 163–182.
Maddieson, I., Bessel, N. & Smith, C. (1996) A preliminary phonetic report on the sounds of Tlingit. *UCLA Working Papers in Phonetics*, **93,** 125–148.
Ohala, J. J. & Riordan, C. (1979) Passive vocal tract enlargement during voiced stops. In *Speech communication papers* (J.J. Wolf & D. H. Klatt, editors), pp. 89–92. New York: Acoustical Society of America.
Peterson, G. E. & Lehiste, I. (1960) Duration of syllable nuclei in English, *Journal of the Acoustical Society of America*, **32,** 693–703.
Pierrehumbert, J. B. (1980) *The phonology and phonetics of English intonation*. MIT dissertation.
Pierrehumbert, J. B. (1990) Phonological and phonetic representation. *Journal of Phonetics*, **18,** 375–394.
Potter, B., Gordon, M., Dawson, J., de Reuse, W. & Ladefoged, P. (to appear) Phonetic structures of Western Apache. *UCLA Working Papers in Phonetics*.
Saltzman, E. (1986) Task dynamic coordination of the speech articulators: a preliminary model. In *Generation and modulation of action Patterns* (*Experimental Brain Research Series* 15), (H. Heuer & C. Fromm, editors), pp. 129–144. New York: Springer.
Saltzman, E. & Munhall, K. G. (1989) A dynamical approach to gestural patterning in speech production. *Ecological Psychology* **1,** 333–82.
Shalev, M., Ladefoged, P. & Bhaskararao, P. (1994) Phonetics of Toda. *PILC Journal of Dravidian Studies*, **4,** 19–56.
Shryock, A., Ladefoged, P. & Williamson, K. (1995) Phonetic structure of Defaka. *UCLA Working Papers in Phonetics*, **91,** 89–109.
Silverman, D., Blankenship, B., Kirk, P. & Ladefoged, P. (1995). Phonetic structures in Jalapa Mazatec. *Anthropological Linguistics*, **37**(1)**,** 70–88.
Steriade, D. (1989) Affricates and release. *Paper presented at the Conference on Features*, MIT. October.
Steriade, D. (1993) Closure, release, and nasal contours. In *Phonetics and phonology* 5 (M. Huffman & R. Krakow, editors), pp. 401–470. New York: Academic Press.
Stevens, K. N. (1999) *Acoustic phonetics*. Cambridge: MIT Press.
Stevens, K. N., Keyser, S. J. & Kawasaki, H. (1986) Toward a phonetic and phonological theory of redundant features. In *Invariance and variability in speech processes* (J. S. Perkell & D. H. Klatt, editors), pp. 426–463.New Jersey: Lawrence Erlbaum Associates.
Umeda, N. (1977) Consonant duration in American English. *Journal of the Acoustical Society of America*, **61,** 846–858.
Vatikotis-Bateson, E. & Ostry, D. J. (1995) An analysis of the dimensionality of jaw motion in speech. *Journal of Phonetics*, **23,** 101–117.
Weismer, G. (1980) Control of the voicing distinction for intervocalic stops and frictives: Some data and theoretical considerations. *Journal of Phonetics*, **8,** 427–438.
Westbury, J. & Keating, P. (1980) Central representation of vowel duration. *Journal of the Acoustical Society of America*, **67** (Suppl. 1), S37(A).
Wright, R. & Ladefoged, P. (1997). A phonetic study of Tsou. *Bulletin of the Institute of History and Philology*, Academia Sinica.